

国家科学技术学术著作出版基金资助出版

近红外光谱分析

基础与应用

主 编：严衍禄

副主编：赵龙莲、韩东海、杨曙明

 中国轻工业出版社

图书在版编目 (CIP) 数据

近红外光谱分析基础与应用/严衍禄主编. —北京: 中国轻工业出版社, 2005. 1
ISBN 7-5019-4441-5

I. 近… II. 严… III. 红外分光光度法-应用-产品质量-质量检验 IV. 0657. 33

中国版本图书馆 CIP 数据核字 (2004) 第 065407 号

责任编辑: 王 淳 责任终审: 孟寿萱 封面设计: 刘 鹏
版式设计: 丁 夕 责任校对: 李 靖 责任监印: 吴京一

出版发行: 中国轻工业出版社 (北京东长安街 6 号, 邮编: 100740)

印 刷:

经 销: 各地新华书店

版 次: 2005 年 1 月第 1 版 2005 年 1 月第 1 次印刷

开 本: 787×1092 1/16 印张: 37

字 数: 800 千字

书 号: ISBN 7-5019-4441-5/TQ·278 定价: 70.00 元

读者服务部邮购热线电话: 010—65241695 85111729 传真: 85111730

发行电话: 010—88390721 88390722

网 址: <http://www.chlip.com.cn>

Email: club@chlip.com.cn

如发现图书残缺请直接与我社读者服务部联系调换

30129K5X101ZBW

撰写者名单

1. 近红外光谱分析的物理基础 吉海彦（中国农业大学信息与电气工程学院）
 2. 近红外光谱分析的化学信息基础 严衍禄（中国农业大学信息与电气工程学院）
 3. 近红外光谱分析的数学基础 张录达（中国农业大学理学院）
 4. 近红外光谱分析仪器 张晔晖（中国农业大学信息与电气工程学院）
 5. 近红外光谱分析的信号处理 闵顺耕（中国农业大学理学院）
 6. 近红外光谱分析测试技术 赵龙莲（中国农业大学信息与电气工程学院）
 7. 近红外光谱分析的数学模型及其建立 严衍禄（中国农业大学信息与电气工程学院）、
张录达（中国农业大学理学院）
 8. 近红外光谱分析数学模型的优化 李军会 严衍禄（中国农业大学信息与电气工程学院）
 9. 近红外光谱分析数学模型的维护 赵龙莲（中国农业大学信息与电气工程学院）
杨曙明（国家饲料质量监督检验中心）
 10. 近红外光谱定性判别分析 闵顺耕（中国农业大学理学院） 李军会（中国农业大学信息与电气工程学院）
 11. 异常样品的分析 杨曙明（国家饲料质量监督检验中心）
 12. 近红外光谱分析的误差源 杨曙明（国家饲料质量监督检验中心）
 13. 近红外光谱分析的光谱库与网络技术 李军会（中国农业大学信息与电气工程学院）
 14. 近红外光谱技术在作物品质分析中的应用 魏良明（中国农业大学国家玉米改良中心）
王文真（中国农业科学研究院品种资源所） 严衍禄（中国农业大学信息与电气工程学院）
戴景瑞（中国农业大学国家玉米改良中心）
 15. 近红外技术在面粉加工中的应用 郑家丰（北京市粮食科学研究院）
 16. 近红外光谱技术在饲料分析上的应用 丁丽敏（中国农业大学动物科技学院）
任鹏（中国农业科学研究畜牧所）
 17. 近红外光谱技术在食品分析中的应用 韩东海（中国农业大学食品学院） 丁丽敏
（中国农业大学动物科技学院）
 18. 近红外光谱技术在石油化工工业分析中的应用 袁洪福 陆婉珍（石油化工科学研究院）
 19. 近红外光谱分析在制药工业中的应用 相秉仁（中国药科大学）
 20. 近红外光谱分析在烟草与纺织行业中的应用 张建平（上海烟草集团公司） 赵丽丽
（中国农业大学食品学院）
 21. 近红外光谱分析在生物学中的应用 陈斌 周熙成 朱德兵（航天医学工程研究所）
 22. 近红外光谱的矿物学应用 郭立鹤（矿产资源研究所）
 23. 近红外遥感在农业上的应用 王纪华 赵春江 刘良云 黄文江（北京农业信息技术
研究中心）
 24. 近红外技术在过程分析中的应用 劳彩莲（中国农业大学信息与电气工程学院）
- 附录一 杨曙明（国家饲料质量监督检验中心）
- 附录二 杨曙明（国家饲料质量监督检验中心）

现代近红外光谱分析经过半个多世纪的发展，无论在理论上、技术上和应用上都已经趋向成熟。近红外分析作为一种快速、无损、多组分同时分析的技术，在国际许多领域中被越来越广泛地应用。我国改革开放以来由于经济持续快速发展，国民经济结构的调整，一些产业由轻视产品质量发展为重视质量，该技术作为一种有效的品质分析技术在我国日益受到重视。近年来国外近红外分析仪器在我国的销售量迅速增加，国内对近红外分析技术的研发、仪器生产以及近红外分析仪器的用户也以极快的速度增长，因此对该项技术的了解也更为迫切。

本书是继《现代近红外光谱分析技术》（陆婉珍院士等编著，2000年中国石化出版社出版）一书出版后较全面介绍该项技术的第二本专著。本书的出版对推动我国近红外分析技术的发展将是有益的。本书包括3篇（理论篇、技术篇、应用篇）24章，各近红外技术应用领域的专家参加了本书的撰写，并且得到了陆婉珍院士、戴景瑞院士的帮助。本书各章具有相对的独立性，读者不一定需要系统阅读本书，可根据各自的需要再阅读相应的章节。为了便于读者了解和应用近红外光谱分析仪器，在附录中介绍了当前国内外企业生产或研究开发的一些相关仪器资料。

本书的出版得到了国家科学技术学术著作出版基金委员会的资助，也得到了一批国家级科研项目的帮助，本书的部分内容取自这些科研项目的研究成果*。

由于近红外技术与应用发展迅猛，本书的编著需要相当的时间与一定的过程，因此本书内容难以跟上该技术的发展，缺点和错误是难免的，敬请读者指正。

本书由中国轻工业出版社精心编辑出版，诚挚感谢帮助本书出版的各单位和个人！

严衍禄

2004年7月于北京，中国农业大学

* 1. 国家高技术研究发展计划（863计划）课题——稻麦品质遥感监测与预报技术研究（2002AA243011）；

2. 国家高技术研究发展计划（863计划）课题——农畜产品品质快速无损检测技术（2002AA248051-2）；

3. “十五”国家重大科技项目——科学仪器研制与开发课题“科学仪器通用软件平台的研究与开发”（2001BA210A05）；

4. “十五”国家重大科技专项——食品安全关键技术课题“食品添加剂、饲料添加剂，违禁化学品检验技术”（2001BA804A21）；

5. 科技部农业科技成果转化项目——近红外农产品品质分析及应用系统的产业化开发（02EFN216900721）。

绪论	1
----	---

基 础 篇

1. 近红外光谱分析的物理基础	8
1.1 近红外光谱分析的信号特征和应用特征	8
1.2 近红外光谱分析的信息负载——吸收光谱分析	10
1.3 近红外光谱分析的信息负载——漫反射光谱分析	13
1.4 近红外光谱测定的物理学基础	19
2. 近红外光谱分析的化学信息基础	29
2.1 近红外光谱的信息源	29
2.2 近红外光谱的化学信息基础	31
2.3 近红外光谱分析信息提取的难点	39
3. 近红外光谱分析技术的数学基础	44
3.1 近红外光谱分析的统计学基础	44
3.2 近红外光谱定量分析基础模型	48
3.3 近红外光谱定量分析模型的评价	51
3.4 优化定量分析模型的逐步回归法	56
3.5 主成分回归	62
3.6 近红外光谱定量分析中的几种常用算法	67
3.7 聚类分析	73
3.8 贝叶斯判别分析	78
4. 近红外光谱分析仪器	82
4.1 概述	82
4.2 近红外光谱仪器的基本组成与结构	84
4.3 近红外光谱仪器的主要类型	87
4.4 滤光片型近红外光谱仪器	88
4.5 色散型近红外光谱仪器	90
4.6 傅里叶变换型近红外光谱仪器	92
4.7 声光可调滤光型近红外光谱仪器	93
4.8 近红外光谱仪器的主要性能指标	94
5. 近红外光谱分析的信号处理	98
5.1 近红外光谱中噪声和背景的扣除方法	98
5.2 傅里叶变换滤波	103
5.3 小波变换滤波	106
5.4 光谱基线的校正与数据规范化处理	109

6. 近红外光谱分析测试技术	113
6.1 仪器的选择	114
6.2 工作方式的选择	115
6.3 样品的收集与制备	121
6.4 仪器状态、扫描参数的设定	123
6.5 样品状态、装样条件对测量结果的影响	129
6.6 校正各随机因素对近红外测定的影响	137
7. 近红外光谱分析的数学模型及其建立	141
7.1 概述	141
7.2 近红外光谱分析数据的前处理	146
7.3 近红外光谱分析数据的简单关联模型	151
7.4 近红外光谱分析数学模型的结构与预测性能	157
7.5 建立近红外光谱分析数学模型一般过程与资源共享	162
8. 近红外光谱分析数学模型的优化	169
8.1 概述	169
8.2 评价近红外数学模型优劣的标准	172
8.3 优化近红外数学模型的措施与方法	174
8.4 近红外光谱分析数学模型优化技术的发展趋势	188
9. 近红外光谱分析数学模型的维护	190
9.1 数学模型的适配性检验	191
9.2 模型的修正	193
9.3 模型转移	195
9.4 模型转移的数学实现	199
9.5 不同类型近红外光谱仪器间的模型转移	213
9.6 模型转移中的其他问题	217
10. 近红外光谱定性判别分析	222
10.1 概述	222
10.2 近红外光谱定性判别分析的一般程序	222
10.3 近红外定性判别分析的方法——有监督模式识别算法	225
10.4 近红外分类学研究——无监督模式识别方法	234
10.5 定性判别分析的结果评价	235
11. 异常样品的分析	236
11.1 近红外光谱分析中异常样品的概述	236
11.2 双线性模型中的极端值检测	240
12. 近红外光谱分析的误差源	251
12.1 来源于仪器的误差源	251
12.2 来源于样品的误差源	259
12.3 来源于操作过程的误差源	266

13. 近红外光谱分析的光谱库与网络技术	272
13.1 近红外光谱分析的光谱库技术	272
13.2 近红外光谱分析的网络技术	276
13.3 小结	285

应 用 篇

14. 近红外光谱技术在作物品质分析中的应用	286
14.1 概述	286
14.2 近红外光谱分析在我国作物品质分析中的应用概况	287
14.3 小麦、大麦的近红外光谱分析	290
14.4 玉米的近红外光谱分析	296
14.5 稻米的近红外光谱分析	300
14.6 油料作物的近红外光谱分析	302
15. 近红外技术在面粉加工业中的应用	307
15.1 原料小麦质量检测	307
15.2 制粉过程中的质量控制	311
15.3 面粉和副产品的检测	314
15.4 小结	316
16. 近红外光谱技术在饲料分析上的应用	318
16.1 近红外分析在饲料分析中的理论特性及其优势	319
16.2 影响饲料近红外测定准确性的因素	329
16.3 近红外定标方程在饲料检测上的应用	331
16.4 进行饲料近红外分析的仪器类型	332
16.5 近红外在饲料工业中的应用	333
16.6 近红外光谱技术在饲料工业中应用的发展趋势和展望	346
17. 近红外光谱技术在食品分析中的应用	349
17.1 成分的定量分析	350
17.2 食品品质分析	355
17.3 结构分析	362
17.4 奶制品质量分析	364
17.5 饮料质量分析	379
17.6 烘烤食品质量分析	383
18. 近红外光谱分析在石油化工工业分析上的应用	395
18.1 概述	395
18.2 近红外光谱分析技术对石化工业分析的适用性	396
18.3 方法原理与特点	398
18.4 汽油分析	401
18.5 航煤分析	414
18.6 柴油分析	416
18.7 重质油料分析	421

18.8	化工过程中的应用	429
18.9	在线近红外光谱过程分析	436
18.10	小结	440
19.	近红外光谱分析在制药工业中的应用	444
19.1	药物制剂分析	444
19.2	体内药物、生物分子分析	451
19.3	制药过程控制及实时分析	452
19.4	中药分析	455
19.5	药品的无损分析	460
19.6	其他应用	462
19.7	小结	462
20.	近红外光谱分析在烟草与纺织行业中的应用	465
20.1	近红外光谱分析在烟草行业中的应用	465
20.2	近红外光谱分析在纺织工业中的应用	476
21.	近红外光谱分析技术在生物医学中的应用	482
21.1	概述	482
21.2	血氧的测定	482
21.3	血糖的测定	488
21.4	组织及体内外成分分析	490
21.5	光 CT	493
22.	近红外光谱的矿物学应用	495
22.1	用近红外光谱研究基础矿物学问题	495
22.2	近红外光谱的地学应用	502
23.	近红外遥感在农业上的应用	509
23.1	近红外遥感的数据获取	509
23.2	近红外在作物水分监测上的应用	514
23.3	近红外在作物营养监测上的应用	518
23.4	近红外在作物病虫害监测上的应用	523
23.5	近红外在作物长势监测上的应用	528
23.6	近红外在作物估产上的应用	529
23.7	近红外在作物品质监测预报上的应用	530
24.	近红外技术在过程分析中的应用	534
24.1	过程分析技术	534
24.2	在线近红外分析系统	538
24.3	NIR 在线过程分析应用	549
24.4	小结	562
附录一	近红外光谱分析常用名词术语解释	565
附录二	近红外光谱仪器的日常维护	574

绪 论

近红外 (Near Infrared, NIR) 光是指波长介于可见区与中红外区之间的电磁波, 其波长范围约为 $0.8 \sim 2.5 \mu\text{m}$, 波数范围约为 $12500 \sim 4000 \text{cm}^{-1}$ 。近红外光谱分析是指利用近红外谱区包含的物质信息, 主要用于有机物质定性和定量分析的一种分析技术。近红外光谱分析兼备了可见区光谱分析信号容易获取与红外区光谱分析信息量丰富两方面的优点, 加上该谱区自身具有的谱带重叠、吸收强度较低、需要依靠化学计量学方法提取信息等特点, 使近红外光谱分析成为一类新型的分析技术。在近几届匹茨堡会议的方法学分类中, “近红外” 已经独立于 “红外及拉曼”, 成为二十余类分析技术中的一类独立技术, 并且连续成为会议热点。在我国, 众多的研究与生产领域中广泛运用近红外光谱技术的趋势已经逐步形成。鉴于现代近红外光谱分析技术的特点, 出版本书, 向读者较全面地介绍该技术的理论基础、软硬件技术以及在各行各业的应用是一件很有意义的工作。

一、近红外光谱分析技术发展的回顾

比中红外谱区更早, 近红外谱区于 1800 年被天文学家 William Herschel 所发现^[1]。近红外谱区的应用也比中红外谱区更早。在经典定量、定性分析方面, 到 19 世纪前期, 近红外谱区的分析已得到相当发展, 但是当时的理论与技术水平还无法将近红外谱区的 (强度弱、谱带复杂、相互重叠的) 信息充分提取出来。19 世纪中叶, 有机结构分析和研究的重点逐渐转向了中红外谱区, 并被广泛应用, 积累了丰富的知识。而此期间光谱学研究已有的传统模式不完全适用于近红外光谱分析, 影响了近红外光谱分析的应用。以致 1960 年 Wheeler 称近红外谱区为 “被遗忘的谱区”^[2]。

现代近红外光谱分析是从农业分析开始的。20 世纪 50 年代, 美国的 Norris 等人开始研究用可见光透射与反射技术测定鸡蛋、蔬菜和水果的品质^[3,4], 但因可见区的信息量小, 研究受到限制。由于近红外谱区包含的信息比紫外、可见区丰富的多, 他们从农业分析领域中开始了用近红外谱区分析农产品的工作。Norris 等注意到水分对近红外光的强烈吸收, 因此应用近红外谱区测定农产品中的水分含量。为了排除样品中的脂肪、蛋白质等其他成分干扰水分的测定, 他们采用了类似光谱分析中的多组分测定技术: 同时测定几个波长处样品的透射率或反射率, 用统计学方法, 可以分别计算出样品中的水分、蛋白质、脂肪等含量。为了克服固体样品透射率低的困难, 他们采用了漫反射技术; 为了克服样品状态对测定带来的影响, 他们用反射吸光度之差作为光谱参数, 来计算样品中各成分的含量^[5]; 并发展了一套检验方法, 用以决定某种计算方法的取舍。若计算测定值与样品的真实值相符合就采用该方法, 反之则舍去。这些方法与传统的 (研究光谱、确定谱峰归属, 再进行定性、定量分析的) 光谱分析方法有很大不同。因此, 当时这一领域的工作没有受到分析界的足够重视。

按照 Norris 的工作, 1970 年美国的一家公司首先研制了近红外品质分析仪器。该机

使用了六个波长的窄带干涉滤光片，主要用于分析农产品中水分、蛋白质等的含量，并于 1973 年 10 月登记了美国专利（3776642 号）Grain Analysis Computer。由于这类仪器只要事先经过校正，即使不熟悉光谱的人员也能迅速得到分析结果，能够满足粮库、港口对粮食品质分析的要求，因此许多有关粮食加工、贮藏的单位和公司采用了此项技术。到 20 世纪 80 年代的中期已经有数以千计近红外光谱分析仪进入应用单位，并发表了近两千篇应用论文。但当时大量应用近红外光谱分析的人员不是传统的光谱分析工作者，而是普通的品质检验人员；大量应用论文不是发表在传统仪器分析和光谱分析的刊物上，而是发表在一些应用领域的专业刊物上；大量近红外品质分析仪主要是用滤光片作为单色器件，生产厂家大多不是著名的光谱仪器或分析仪器公司，而是另外一些近红外专用仪器生产厂家。这种状况也影响了近红外光谱分析技术自身的发展。

20 世纪 80 年代，仪器分析特别是应用光谱分析领域的有关专家开始注视近红外光谱分析技术的发展。如著名的应用光谱杂志（Applied Spectroscopy）自 80 年代以来刊载了许多有关论文，以及一系列评论，其中包括有机物的近红外光谱分析^[6]、近红外定性定量分析^[7]、近红外药物分析^[8]以及近红外反射光谱分析进展^[9]等。由于光谱界、分析界与近红外应用领域人员的合作，近红外光谱分析进一步采用了化学计量学（Chemometrics）中多元校正方法，以及现代光学、计算机数据处理技术，使近红外光谱分析技术发展为现代近红外光谱分析技术。1987 年在挪威召开了第 1 届近红外光谱分析国际会议；此后，近红外光谱应用的领域越来越广泛。一些国际著名的光谱仪器公司纷纷推出了近红外光谱分析仪器。20 世纪 90 年代国际分析界逐步形成了近红外光谱分析的热潮。近年来近红外光谱分析一直是匹兹堡（PITTCON）会议的热点^[33~35]。在 2000 年的 PITTCON 会议上，近红外光谱分析技术被认为是该次会议中所有光谱法中最受重视的一类方法，在该次会议上直接和近红外光谱技术有关的分会达到 11 个之多^[33]。2001 年第 10 届国际近红外光谱会的主题为“用近红外光谱改变世界”（Changing The World with NIR），该次会议上发表的论文包括食品、精细农业、环境、化学、高聚物、药物、纺织、石油化学、生物医学、生命科学、化妆品以及医学等十几个领域的近红外运用以及有关近红外光谱技术的研究^[37]。

目前，国际上已有两种专门的近红外分析杂志：“Journal of Near Infrared Spectroscopy”与“NIR News”，已有多种近红外光谱分析的专著^[10,11,12,38]出版。我国陆婉珍院士等近红外光谱分析专家于 2000 年出版了国内第一本全面介绍现代近红外光谱分析技术的专著^[28]。

在我国近红外光谱技术的应用研究也是从农业领域开始的。由于 20 世纪 60 年代禁运等原因，我国粮食与农业系统在 1977 年前后才开始由国外引入近红外光谱分析仪，但由于建立数学模型的困难，到 20 世纪 80 年代才发表了我国第一批有关研究论文^[13,14]。我国农业仪器分析工作者在 20 世纪 80 年代初就关注此项技术^[15,16]。1983 年前后开始在大型通用傅里叶变换红外光谱仪上进行近红外光谱分析技术在农业分析的理论研究和应用研究^[17]。我国石油科学与药物科学等领域有许多研究报道^[18,19]（详见本书第 18、19 章）。我国光谱学界也很重视此项技术^[20]，在光谱学的专著中包含了近红外光谱分析的章节^[21,22]。在全国光谱学术报告会上发表了一批近红外光谱应用的论文^[24]。在著名的 PITTCON 会议上也发表了一些近红外光谱应用的论文^[25,26]。20 世纪 90 年代我国开始了近红

外光谱仪器的研制^[23,27]。但建立定量分析数学模型的困难一直是影响该技术在我国广泛应用的重要原因。目前我国一些高产值的产业已经率先广泛运用近红外光谱分析技术,如我国的石油化学、烟草、乳品、酒类与农业等行业,近年来都已经分别投资了数百万甚至上千万美元,进口了一大批高精度近红外光谱仪,其中石油化学与农业领域还应用了一批国产仪器,它们在保证和提高产品质量方面起了重要作用。我国石油化工与药物生产领域开始了近红外光谱在线分析的应用^[39,40]。

二、近红外光谱分析技术简介

近红外光谱分析在应用领域以及方法、技术等方面具有一系列特点,而所有的这些特点都来源于该谱区特殊的信息源。

近红外光谱分析的信息与信号:近红外光谱的信息源是分子内部原子间振动的倍频与合频。该谱区信号的频率比中红外谱区高,介于中红外谱区和可见谱区之间;因此近红外的光谱类似于可见光,容易获取与处理。近红外光谱信息的特点类似于振动光谱的中红外谱区,信息量丰富,本区包含大量含氢基团的结构信息,但信息强度比中红外谱区低、谱峰宽,同一基团的倍频与合频信息常可在近红外谱区的多个波段取得。在近红外谱区同一波段有样品多种信息的叠加。近红外光谱信息和信号的特点贯穿于近红外光谱分析技术的全过程,本书基础篇的第1章和第2章较详细介绍了有关内容。

近红外光谱分析的技术难点:样品的近红外光谱是近红外光谱分析信息的载体,近红外光谱分析是从近红外光谱(分析的信息源)中提取信息的过程。为了实现近红外光谱分析,必须解决一系列由于近红外光谱的特点造成的分析技术难点。主要是:近红外光谱区的吸收强度较弱、光谱的信噪比低;测定不经过预处理的样品的光谱易受样品状态、测量条件的影响而造成光谱的波动性;以及由于近红外光谱分析样品可以不经预处理,因此样品状态、测量条件、成分的组成等造成光谱中背景复杂、谱峰重叠等几个困难。这些困难的核心是样品近红外光谱中的有效信息率低。对复杂样品进行近红外光谱分析是从复杂、重叠、变动的光谱中提取微弱信息,其困难犹如“沙里淘金”。近红外光谱分析必须依靠算法和计算机软硬件技术,因此,近红外光谱分析是一种现代高技术分析。

近红外的光谱与分析的方法都具有统计学的性质,在分析过程中需要不断掌握、处理与确定近红外分析的一些统计变动量的尺度(简称为“度”),这一系列对“度”的掌握目前主要靠操作人员的经验,也需要理论指导,这些工作是近红外光谱分析研究的重点与难点。与传统光谱分析(紫外、可见、中红外光谱分析等)的关系,可以说近红外光谱分析在光学系统方面有较多的继承,在软件(包括样品资源、操作技术、算法、数学模型、数据库、网络技术等众多的领域)方面需要有较多的发展^[29,34]。掌握近红外光谱分析技术需要较广泛的自然科学背景和技术背景,为此,本书在基础篇介绍近红外光谱技术涉及的数学、物理、化学、仪器等基础,其中第4章介绍了近红外光谱分析仪器的有关内容,第3章和第5章介绍了近红外光谱分析有关数学背景和信号处理的背景;在技术篇中详细讨论了有关定性、定量分析的技术,其中第6章介绍了近红外光谱分析的测试技术,第7、8、9章详细讨论了近红外光谱分析的难点:数学模型的建立、优化和维护;第10~13章分别讨论了近红外光谱的定性分析、样品误差源及网络技术等有关内容。

近红外光谱分析的应用特点：近红外光谱的信息和信号特点决定了其应用的特点：近红外光谱几乎可以用于所有与含氢基团有关的样品的化学性质与物理性质分析，可以广泛地应用于定性分析和定量分析领域，被称为“具有解决全球农业分析的潜力”；由于吸收强度低，近红外光谱分析可以不需对样品作任何化学或物理的预处理，可取得样品内部深处的物质信息，因此可用于对复杂样品（如生物样品）进行非破坏性测定、原位分析、在线分析和活体分析；近红外光谱分析速度快，可以实现瞬间分析；近红外谱区光子的能量比可见光还低，在近红外谱区分析是一种安全分析；近红外分析的过程不消耗试剂、不产生污染，属于“绿色分析”技术；虽然近红外光谱分析的检测极限不如中红外光谱低，但对于常量分析，近红外光谱分析确实是一种“多快好省”的、较为理想的现代分析技术，应用领域极为广泛。本书的应用篇的第 14~20 章中介绍了近红外光谱分析在农业、石化等 8 个行业的应用，第 21~24 章还介绍了该技术在生物医学、矿物学、过程控制和遥感技术中的应用。

三、当前需要注意发展的几种近红外技术

现代近红外光谱分析经过约半个世纪的发展，在理论上和技术上已经趋于成熟；但应当指出，近红外方法虽然具“应用广泛”、“操作简便”的优点，但对于还没有建立数学模型，或难以建立数学模型的一些样品，目前还不能用近红外法测定。近红外化学计量方法是通过数学模型来预测未知样品，由于种种复杂背景的变化、测量方法、测量条件的影响，近红外法不能保证预测的每一个样品达到相同高的准确度，测定结果只能用统计方法给出测定结果的置信度，只能达到一定的准确率。对于广大的近红外光谱分析技术用户，建立优秀的数学模型有时有一定困难，影响了该技术的广泛应用。当前近红外技术需要借助于光学技术、计算机技术（硬件技术、算法、软件技术、网络技术等）与电子技术等当代技术的发展，继续改进软硬件技术，特别应侧重在应用领域的开拓，解决应用中的各种问题。

（一）需要注意研究降低近红外光谱分析的技术难度

近红外光谱分析建模过程需要掌握一些参数的度，对一般使用人员有相当的困难，为了使该技术成为一项“大众化”的分析手段^[33]，应当注意研究降低近红外分析技术的使用难度，特别要依靠软件技术和网络技术的支撑，建立、优化、维护、使用数学模型；发展免建模型的专用仪器以及运用网络技术的近红外光谱仪器，发展新的算法，发展智能化软件。本书的技术篇对此做了较详细的介绍。

（二）需要注意研究发展近红外图像分析技术

传统的光谱分析只能得到分析对象空间各点的平均光谱，从而得到对象空间平均的结构与组成，因此通常只能用于均匀物质的分析，但是物体特别是生物体的各种复杂功能主要还决定于空间各个层次的结构，例如细胞与组织的功能决定于空间微米和亚微米尺度的空间结构；又如，遥感技术要得到地面各点的结构特征，就不能只依靠平均光谱。近年来发展的图像分析可以得到空间分布的信息，但只是形象的信息，不能得到结构和组成的信

息。需要把光谱技术和图像技术结合在一起，得到光谱图像即物体空间各个点的光谱，运用化学计量学方法通过对空间各个点的光谱分析，就能得到对象各点的结构和组成信息。由于近红外与中红外谱区信息量丰富，这些谱区的光谱图像技术必将对许多领域，特别是生物学与遥感技术的研究产生重大影响^[31,32]。品质分析从实验室的湿化学分析发展到近红外无损分析，使分析效率提高了一个到几个数量分析级。如果进一步在田间作物收割前用近红外品质分析遥感技术直接分析作物的品质，还可以使分析效率再提高几个数量分析级。从光谱图像中提取空间信息目前还缺少有效的算法，为此还需要发展空间化学计量学。本书第 23 章介绍了近红外遥感技术。

（三）需要注意研究发展近红外过程分析技术

产品质量的保证必须从生产的全过程实现质量的监控，国际产业界对过程控制仪器（PCI, Process Analyzer Instrument）给予了高度的重视^[41]。传统的化学分析方法对原料与产品质量分析费时、费事，难以进行现场分析或在线分析；传统过程控制主要依靠压力、温度、流量等间接参数进行生产监控，而对生产线上的原料、中间产品及最终产品的品质直接快速分析比较困难。由于近红外光谱分析技术可以不需对样品进行前处理，具有同时对多组分进行瞬间分析的能力，因此特别适合于过程分析，直接分析生产线上有关物料的内在品质，并迅速反馈给生产人员进行质量控制或进行自动控制，近红外光谱分析技术是当代过程控制的主导技术。本书第 24 章介绍了近红外过程分析技术。

（四）需要注意研究发展近红外光谱分析技术与其他分析技术的综合应用

近红外技术与其他分析技术一样，有其长处，也有一定局限。例如，因为近红外谱区的吸收较低，近红外法的检测低限不如中红外定量分析，约为 $10^{-3} \sim 10^{-4}$ ，不适宜做含量过低的样品、微量样品的显微分析。因为近红外光谱复杂谱峰重叠，不宜进行官能团定性分析。因此应注意发展近红外光谱分析技术的综合应用。例如，为了实现对复杂物的综合性状品质的分析，需要运用多种仪器分析手段进行分析，并对各种分析结果取长补短，进行综合分析处理。对此可以研究运用军事科学中近来发展的“数据融合”（Data Fusion）技术：运用现代计算机技术与数学的方法，对多种仪器分析方法得到的信息在一定的准则下加以综合、分析、过滤、相关与合成，并以此辅助进行复杂物综合性状品质分析。

四、对当前我国发展近红外光谱分析技术的几点看法

中国经济的持续高速发展以及我国商品融入国际市场的进程，对我国产品的品质分析提出了迫切的要求。据我国政府部门测算，仅农产品和食品质量、农残的快速检测仪器，全国急需数十万台^[37]。根据国际分析界的经验，近红外光谱分析技术作为一种快速、无损、无公害的多组分同时分析的现代技术，在农产品和食品质量分析方面是一种首选技术；我国政府部门对发展近红外光谱分析技术的高度重视，近三个五年计划内与近红外光谱分析技术相关的国家级项目达到了数十个，我国的近红外技术具有应用领域广泛、市场容量大的特点，当前是我国发展近红外光谱分析技术难得的机遇。但由于近红外光谱分析

技术自身的困难, 目前我国的近红外商品仪器中国内产品还不多, 特别是高档近红外光谱分析仪器, 大部分是国外相关公司的产品; 此外, 许多近红外用户的使用水平相对较低, 我国尚未建立近红外光谱分析仪器技术与应用技术的标准, 因此目前迅速发展中国的近红外光谱分析技术还具有相当的困难。

在我国推广应用近红外光谱分析技术, 必须提高仪器的性能、降低仪器的使用难度、增加批量、降低价格。可以采用“建立基础、发展专用、兼顾通用、完善服务”的方针。建立基础指集中力量研发近红外基础软硬件, 包括基本光学结构模块与基础数据库、应用软件网络系统; 当前还应尽快建立我国近红外光谱分析仪器与应用技术的标准。在此基础上与应用结合, 设计专用的附件、配件开发专用数学模型, 研制与化解出各种各样性能优秀、使用方便、价格较低、应用面宽的、免建数学模型的专用仪器; 适当发展实验室通用仪器, 建立与完善我国近红外服务支持系统, 尽可能减少使用人员的困难。有关工作是一项系统工程, 有待于近红外光谱分析技术的专业人员与应用人员在政府职能部门的支持下通力合作才能完成。

本书的作者们希望该书通过基础篇、技术篇、应用篇这 3 篇的有关介绍, 为发展和应用中国近红外光谱分析技术起到一些作用。

参 考 文 献

- [1] Herchel W. Philos. Trans, 1800, 90: 225
- [2] Wheeler O H. J Chem Educ, 1960, 37: 234
- [3] Norris K H, Rowan J D. Food Technol, 1957, 11: 374
- [4] Birth G S, Norris K H. Food Technol, 1958, 12: 592
- [5] Norris K H, Hart J R. In Proceedings. International Symposium on Humidity and Moisture, 1963
- [6] Weter J G. Appl Spectroscopy Riv, 1985, 21 (1, 2): 1~43
- [7] Stark E, Luchter K. Ibid, 1986, 22 (4): 335~399
- [8] Ciurczak E W. Ibid, 1987, 23 (1, 2): 147~163
- [9] Martin K A. Applied Spectroscopy Reviews 1992, 27 (4): 325~383
- [10] Osborne B G, Fearn T. Near-Infrared Spectroscopy in Food Analysis. Longman Scientific and Technical. New York, 1986
- [11] Williams P C, Norris K H. Near-Infrared Technology in the Agricultural. Minneapolis, 1987
- [12] Donald A Burns, Emil W Ciurzak. Handbook of Near- infrared analysis. MARCEL DEKKER, INC, 1992
- [13] 吴秀琴. 应用 51A 型近红外 S 测定小麦种子中赖氨酸的含量. 农业测试分析, 1985, 2 (2)
- [14] 王文真, 蒋自强, 林夕. 近红外光谱分析及其在我国的应用. 国外科学仪器, 1989, 3
- [15] 严国光, 严衍禄. 仪器分析原理及其在农业中的应用. 北京: 科学出版社, 1982
- [16] 王文真, 张玉良. 近红外分析的发展和应用. 国外农业科技, 1983, 6
- [17] 北京农业大学 FT 近红外 DRSA 研究组. 北京农业大学学报, 1991, 16 (增): 1~4
- [18] 袁洪福, 龙义成, 陆婉珍. 近红外光谱仪的研制. 分析化学, 1999, 27 (5): 608
- [19] 刘国林, 陈国广, 相秉仁等. 近红外光谱技术在元胡止痛散定量分析中的初步应用研究. 中国现代应用药学杂志, 2000, 17 (5): 383~385
- [20] 范世福. 近红外光谱分析的复兴. 分析仪器, 1993 (3): 1~8
- [21] 严衍禄. 近红外光谱技术及应用. 实用傅里叶变换红外光谱学. 北京: 中国环境科学出版

社, 1991

- [22] 严衍禄, 吉海彦. 傅里叶变换近红外光谱技术及应用. “近代傅里叶变换红外光谱技术及应用”上卷第八章. 北京: 科学技术文献出版社, 1994
- [23] 袁洪福, 褚小立, 陆婉珍等. 一种新型 CCD 在线近红外光谱分析仪的研制. 分析化学, 2003
- [24] 赵龙莲, 闵顺耕等. 傅里叶变换近红外光谱法测定烟草中 9 种品质参数. 光谱学与光谱分析, 1998, 18 (4): 89~90
- [25] YanLu Yan, CaiLan Lao. Quantitative calibration Transferof Gasoline in FTNIR Analysis. Oral Presentation Abstracts No. 980 at PITTCON' 98, 1998
- [26] Min ShunGeng, Zhou Xueqiu, Zhao Longlian, et al. Can aLower S/N Dispersion NIR Instrment Accurately Determine the Nitrogenin Tobacco. Oral Presentation Abstracts No. 1241 at PITTCON' 98, 1998
- [27] 吉海彦, 严衍禄. 在国产近红外光谱仪实验样机上用偏小二乘法定量分析大麦成分. 分析化学, 1998, 26 (5): 607~611
- [28] 陆婉珍, 袁洪福等. 现代近红外光谱分析技术. 北京: 中国石油化工出版社, 2000
- [29] 严衍禄, 赵龙莲等. 现代近红外光谱分析的信息处理技术. 光谱学与光谱分析, 2000, 20 (6): 777~780
- [30] 李军会, 赵龙莲等. 农业近红外光谱分析技术软件及网络系统的研制. 现代仪器, 2000 (6)
- [31] Pina Colarusso, et al, Infrared Spectroscopic Imaging: from Planetary to collular Systems. Applied Spectroscopy, 1998, 52 (3)
- [32] 肖松山, 范世福, 李昀等. 光谱成像技术进展. 现代仪器, 2003 (5): 5
- [33] 金钦汉. 从 2000 年匹茨堡会议看分析化学和分析仪器发展的一些新动向. 现代科学仪器, 2000 (3): 14~16
- [34] 金钦汉. 从 1999 年匹茨堡会议看分析仪器的跨世纪发展动向. 分析仪器, 1999 (2): 1~5
- [35] 陈焕文, 金钦汉. 2003 年匹茨堡会议简介. 现代科学仪器, 2003 (2): 41~45
- [36] 吴波尔, 张渝英. 大力发展我国科学仪器事业. 现代科学仪器, 2002 (1): 8~16
- [37] “10th International Conference on Near-Infrared Spectroscopy” Abstracts and Programs
- [38] 尾崎幸洋, 河田聪 稿. 近赤外分光法. 日本分光学会测定法シリーズ32, 学会出版センター, 1995, 170~190
- [39] 徐广通, 陆婉珍, 袁洪福等. CCD 近红外光谱仪在柴油生产控制分析中的应用. 石油炼制与化工, 1999, 30 (9): 57
- [40] 蒲登鑫, 王文茂, 李军会等. 近红外在线质量监控技术在中药葛根素生产中的应用. 现代仪器, 2003 (5): 27
- [41] 全球过程分析仪器市场统计报告. 生命科学仪器, 2003 (3/4): 17

基 础 篇

近红外光谱分析是从复杂、重叠、变动的背景中提取弱信息。与常规分析相比，该项技术涉及更多的自然科学基础，需要较多地依靠化学计量学算法与软件技术。在本篇中介绍近红外光谱分析相关的物理、数学、化学信息学基础以及有关近红外光谱分析仪器的基础。

1. 近红外光谱分析的物理基础

1.1 近红外光谱分析的信号特征和应用特征

1.1.1 近红外光谱的信号特征

在近红外光谱分析中，装载分析信息的信号是近红外光，因此近红外光谱分析的信号特征决定于近红外辐射的物理特征，并决定了近红外光谱仪器应用的光源、检测器与光学材料。

近红外光谱分析的信号有下列特征：

近红外光的产生：近红外光的波长范围介于 $0.8\sim 2.5\mu\text{m}$ ，由黑体辐射定律可知，在 3000 度左右的色温下在近红外谱区的辐射效率很高，因此近红外谱区的光源比中红外谱区的光源要方便得多，一般用卤素灯。

近红外光的检测：近红外谱区光子的能量比可见谱区低，其能量范围为 $1.65\sim 0.5\text{eV}$ ，该能量值低于碱金属外层电子的逸出功，因此检测器不能用光电管或光电倍增管等光电材料。常用的测光材料是半导体材料，如 Si、PbS 以及 InSb、InAs 等，而 InSb 和 InAs 的导电机制是少数载流子，近红外谱区的检测效率很高，但温度引起的热噪声很大，因此这类检测器应该在液氮下工作，或者进行半导体制冷以保持较低的恒定温度。近红外谱区的信号检测也可以用真空热电偶等热电器件，但检测效率很低。

近红外谱区的光学材料：适用于近红外谱区的光学材料比中红外谱区多得多，参见表 1-1。特别是一些透红外的玻璃，在近红外区有良好的光学性能与物理性能，价格又比较便宜，这就使近红外区仪器的造价下降。近红外区的光导纤维已实现了商品化，使得近红外仪器应用更加方便，便于实时分析。

1.1.2 近红外光谱的应用特征

1.1.2.1 近红外光谱的应用范围

近红外光谱主要是有机分子的倍频与合频吸收光谱，与中红外光谱一样，本谱区也能

表 1-1

近红外谱区一些光学材料的透过特征

材 料	透过范围/ $\mu\text{m}(\text{cm}^{-1})$	折射率 [*]	溶解度/(g/100g 水)
熔融石英	0.18~3.3(15555~3030)	1.46	0
石英	0.2~3.3(50000~3030)	1.54	0
火石玻璃	0.35~2.2(28572~4545)	1.66	0
CaF	0.12~12(83333~833)	1.43	0.0016
NaCl	0.3~17(33333~588)	1.54	35
AgCl	0.4~25(25000~400)	1.98	0
KBr	0.3~29(33333~345)	1.56	54
KRS-5 ^{**}	1~40(10000~250)	2.63	0.05

* 折射率测定光波长为 589nm。 ** 成分为 TlBr 42%；TlI 58%。

够得到分子的结构、组成、状态的信息，而且从近红外反射光谱还能得到样品的密度、粒度、高分子物的聚合度及纤维的直径等物质的物理状态信息。因为近红外谱区的吸收主要是分子或原子振动基频在 2000cm^{-1} 以上的倍频、合频吸收，所以有机物近红外光谱主要包括 C—H、N—H、O—H 等含氢基团的倍频与合频吸收带。这些含氢基团的吸收频率特征性强，受分子内外环境的影响小，而且在近红外谱区比中红外谱区的样品光谱特性更稳定，这些是在近红外谱区做复杂天然物质品质分析的前提。

1.1.2.2 近红外光谱分析的样品状态

有机物近红外光谱谱带重叠与吸收强度弱的特点给分析带来了困难：用于定量分析其检测限比中红外区的定量分析差 1~2 个数量级，用于结构分析因信息量小及谱峰重叠严重，也不如中红外区。但是吸收强度弱也给分析带来方便。例如：农业样品不需要经过化学等预处理直接进行分析，可以用长光径的样品池分析样品，光径的误差对分析的影响较小，黏度较大的样品也可以在近红外区进行流动分析。

1.1.2.3 近红外光谱分析的信息处理

倍频、合频吸收带比基频吸收带宽得多，这使得多组分样品的近红外光谱中不同组分的谱带、同一组分中不同基团的谱带以及同一基团不同形式的倍频、合频谱带发生严重的重叠，从而使近红外光谱的图谱解析异常困难。这也是近红外光谱在光谱分析中长期未受到重视的原因。复杂样品近红外光谱分析的信息提取主要依靠化学计量学方法与计算机数据处理才能实现。

1.1.2.4 近红外漫反射光谱分析

因为物体对光的散射率随波长的减少而增大，近红外谱区光的波长比中红外区短，其散射的效率比中红外区高，因此近红外谱区适合做漫反射光谱或散射光谱分析，可以得到较高的信噪比和较宽的线性范围。近红外分析可以测各种类型的样品，除了可检测一般气体、液体、固体样品外，还可以用近红外漫反射光谱检测粉末样品、纤维样品、糊状样品、乳类肉类样品等。利用近红外光谱的信息量丰富、图谱的稳定性高、取得图谱比较容易、漫反射光谱分析不需要对样品做任何化学处理的优点，如果再加上傅里叶光谱仪的信息多通道传输，收集一个样品的近红外漫反射光谱相当容易。这几种优点的综合，结合现代数学方法和计算机技术，可以克服从复杂背景中提取微弱信息的困难，从而使傅里叶变换近红外漫反射光谱分析（FTNIRDRSA）成为一种极有发展前途的光谱技术。FTNIR-DRSA 在理论上对发展现代光谱分析以及化学计量学、智能仪器有重要意义，在实用上是

一种理想的复杂样品分析方法。

1.1.2.5 近红外光导纤维用于实时分析

近红外谱区分析样品时，所用的溶剂和制样技术与中红外区不同。近红外谱区的仪器技术比中红外谱区简单，已在前面讨论。目前，在近红外谱区有高透过率的光导纤维已经商品化，这对于一些不能直接检测的样品（如有毒样品）提供了测试的方便，也给实时分析提供了条件。

1.2 近红外光谱分析的信息负载——吸收光谱分析

1.2.1 比耳定律

吸收光谱定量分析是根据样品对某一谱区光吸收强度与吸光粒子（低能态的分子或原子）之间的关系，并考虑到样品中吸光粒子数与样品粒子总数的关系来定量的。吸收光谱定量的基础是比耳定律。

比耳定律，也称为物质对光的吸收定律，简称吸收定律。比耳定律的表达式为：

$$A = -\lg \frac{I}{I_0} = \epsilon b C \quad (1-1)$$

其中： I_0 为波长为 λ 的平行、均匀入射光束强度； I 为透过溶液后的光束强度； A 为吸光度； ϵ 为待测组分的摩尔吸光系数； b 为光程； C 为待测组分的物质的量浓度。

比耳定律可表述为：对一定波长的单色光，物质的吸光度 A 与光程 b 及浓度 C 成正比，比例常数称为吸光系数，它与所用的浓度单位有关。若用物质的量浓度单位时，比例常数 ϵ 称为摩尔吸光系数；吸光系数与样品本性 & 波长 λ 有关。单波长点的吸光度，其信息量较小，可用于对已知组分的定量分析；由各波长点的吸光度组成的光谱，其信息量较大。

吸光度 A 是一个多元函数。确定的 ϵ （对一定波长的光与物质）可以得到 A - C 的线性关系，即定量分析的工作曲线，用于定量分析；对于一定的 b 和 C ， A - λ 间的关系即组分的一维吸收光谱，用于定性分析。

影响比耳定律偏离的主要因素有以下几个方面：

(1) 非单色光引起的偏离

从理论上来说，比耳定律只适用于单色光，但在实际工作中并非如此，绝对不可能从光学分析仪器上得到真正的单色光，而只能是波长范围很窄的光谱带。因此，进入被测试样的光仍为在一定波段内的复合光。由于物质对不同波长的光具有不同的吸光程度，故在实际工作中即使应用很高级的分光光度计、采用很窄的狭缝宽度（用波段很窄的复合光照射样品），仍会产生比耳定律偏离。假设入射光的光谱带宽 $\Delta\lambda$ 的波长范围为 $\lambda_1 \sim \lambda_2$ ，对 λ_1 设其吸光度为 A_1 ，则 $A_1 = -\lg(I_1/I_{01}) = \epsilon_1 b C$ ， $I_1 = I_{01} \times 10^{-\epsilon_1 b C}$ ；对 λ_2 设其吸光度为 A_2 ，则 $A_2 = -\lg(I_2/I_{02}) = \epsilon_2 b C$ ， $I_2 = I_{02} \times 10^{-\epsilon_2 b C}$ 。但实际测定时，入射光强度为 $I_{01} + I_{02}$ ，透射光强度为 $I_1 + I_2$ ，因此吸光度值为：

$$A = -\lg \frac{(I_1 + I_2)}{(I_{01} + I_{02})} = -\lg \frac{(I_{01} \times 10^{-\epsilon_1 b C} + I_{02} \times 10^{-\epsilon_2 b C})}{(I_{01} + I_{02})} \quad (1-2)$$

若 $\epsilon_1 = \epsilon_2 = \epsilon$ 时，上式可写成：

$$A = -\lg 10^{-\epsilon b C} = \epsilon b C$$

此时， A 与 C 成直线关系。但实际上， ϵ 是有色物质的与波长有关的特征常数，对同一有色物质而言，在其他条件一定时，不同波长有不同的 ϵ 值，故 ϵ_1 不可能完全与 ϵ_2 相等。因此，对非单一波长的入射光， A 与 C 不可能真正成直线关系，因而产生了比耳定律偏离。

(2) 化学因素引起的偏离

从理论上讲，比耳定律只能适用于均匀、相互独立的、无相互作用的吸收粒子体系，但在实际工作中并非如此。试样在测定过程中，经常会发生缔合、离解、电离、溶剂作用和产生同性异构体、组成新的络合物等化学变化，从而使吸收粒子及其相互间的平均距离发生变化，以致每个粒子都可影响其邻近粒子的电荷分布。这种相互作用，可使它们的吸光能力发生改变。以致影响比耳定律的准确性，即产生比耳定律的偏离。

(3) 杂散光引起的偏离

实践证明，杂散光是引起比耳定律偏离的主要因素之一。其原因是：①吸光物质由许多粒子组成，这些粒子会对入射光产生散射，并且随着浓度的增大，这些散射光强度会不断加强，降低透射光强度，使被测试样的吸光度增大，从而引起比耳定律偏离。②由于仪器本身的光学系统（特别是光栅）会产生杂散光，使得分析测试的吸光度减小，以致引起比耳定律的偏离。③由于不少物质在光的照射下会产生光致发光现象或产生荧光，也将严重影响比耳定律的偏离。

(4) 其他因素引起的偏离

除以上因素外，测定时的温度也可引起比耳定律的偏离。还有压力、光学传感器的非线性等都可引起比耳定律的偏离。

1.2.2 近红外光谱透射分析

近红外光谱的分析测定技术可以分为两大类：即透射光谱技术和反射光谱技术。透射光谱法是把待测样品置于作用光（光源发出的光）与检测器之间。检测器所检测到的分析光是作用光通过样品体与样品分子相互作用后的光（因而装载了样品的结构与组成信息）。若样品是透明的真溶液，则分析光在样品中经过的路程（光程）一定，透射光的强度与样品中组分浓度的关系可以由比耳定律决定。这种把透射光用于分析测定的方法称为透射分析法，也就是吸收光谱的常规分析方法。若样品是混浊的，样品中有对光产生散射的颗粒，则由于光散射的原因，光在样品中经过的路程不确定，透射光的强度与样品浓度间的关系不符合比耳定律，用于测定时称为漫透射分析测定法。

反射光谱分析时，检测器与光源置于待分析样品的同一侧，检测器所检测到的分析光是光源发出的作用光投射到物体后，以各种方式反射回来的光。物体对光的反射分为规则反射（镜面反射）与漫反射。规则反射指光在物体表面，按入射角等于反射角的反射定律发生的反射；漫反射是光投向漫反射体（常是粉末或其他颗粒组成的物体）后，在物体表面或内部发生的方向不定的反射。应用漫反射光进行的分析测定称为漫反射分析测定法。液态或固态的透明物体也可以进行漫反射光谱分析，此时在待测样品的后面放置陶瓷一类

表面漫反射体，作用光通过透明物体后在该漫反射物体上发生漫反射，漫反射光又透过样品后由检测器检测，这种分析方法适用于在漫反射装置上测定透明样品。此外还有把透射和漫反射分析结合在一起的综合漫反射分析测定，以及衰减全反射测定法。这里只介绍近红外常规分析法，即透射光谱法，在后面将介绍目前近红外光谱分析中用得最多的粉末或颗粒样品的漫反射光谱分析法。

近红外常规定量分析方法与紫外可见区定量方法类似，也是依据比耳定律，在这里只讨论近红外透射光谱法与紫外、可见透射光谱法不同之处。分子对近红外谱区的吸收不但比紫外可见区低得多，而且与中红外光谱相比，由于分子振动的倍频吸收和合频吸收发生的几率比基频吸收的几率小得多，所以近红外谱区有机分子的摩尔吸光系数一般比中红外谱区小 1~2 个数量级，大致在 $10\text{l/mol}\cdot\text{cm}$ 以下。所以近红外谱区分析样品的浓度一般应达到 $10\sim100\text{g/L}$ 。但近红外谱区可像紫外可见区那样，用 $1\sim10\text{cm}$ 光程的样品池，这比在中红外区常用的 0.1mm 样品池方便得多。由于样品池光程长，日常分析不需进行光程校正。但近红外分析所用的样品量随之也比中红外区分析所用的要多。近红外样品池的窗口材料可选用石英或氟化钙，它们的透光范围分别为 $0.18\sim3.8\mu\text{m}$ ($5555\sim2631\text{cm}^{-1}$) 及 $0.12\sim12\mu\text{m}$ ($8333\sim833\text{cm}^{-1}$)，其他窗口材料参见表 1-1。近红外区分析用的样品池常可与紫外区的样品池合用，所以制样技术大体与紫外光谱法相似，而不同于中红外区分析制样技术。

近红外区分析选用溶剂比紫外区要困难得多，在本区内透明的溶剂只有 CCl_4 和 CS_2 。图 1-1 是近红外光在一些溶剂内的透过状况。

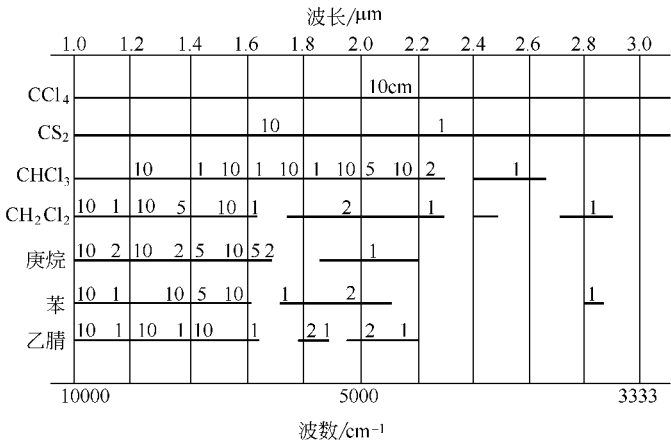


图 1-1 近红外区常用溶剂的光谱特性

近红外谱区常用于分析含 C—H、N—H、O—H 键的有机物，所以溶剂分子中不宜带有上述基团。必须使用含这些基团的溶剂时，溶剂分子中 C—H、N—H、O—H 基团内的 H 原子应用氘替代 (C—D、N—D、O—D)。C—D、N—D、O—D 的近红外吸收带比相应的 C—H、N—H、O—H 吸收带有较大的偏移，可避免对样品测定的影响。

近红外谱区吸收带的强度较弱，一方面影响了这一谱区分析的检测限，另一方面又便于对某些高聚物和含水样品的分析。近红外谱区的反射率比较高，因此有利于做反射光谱分析，因为不含 C—H、N—H、O—H 的化合物在近红外区无吸收，所以干扰该谱区分

析的杂质相对少些。

近红外谱区因为信息量小，用于结构分析不如中红外区，但对某些特殊的分析比较有效。现代近红外分析常用的技术是近红外漫反射光谱分析，将在本章后面做进一步讨论。

1.3 近红外光谱分析的信息负载——漫反射光谱分析

现代近红外光谱分析的各种技术中，漫反射分析测定占有特别重要的地位，这是因为漫反射光谱分析在近红外谱区比中红外谱区更为有利。漫反射光谱分析不需要对样品做化学处理，直接可以测定粉末状、纤维状等不规则样品，漫反射光谱分析技术也比较简单。本节讨论近红外漫反射光谱分析的理论基础。为了将漫反射光谱用于样品组分分析，首先需要了解漫反射光谱的参数、确定样品漫反射光谱参数与样品浓度间的关系，多组分样品的漫反射光谱分析还需要了解样品中各组分的含量与样品实际能测得的总的漫反射光谱参数之间的关系，以便由样品总漫反射光谱参数确定样品中各组分的含量。

1.3.1 漫反射光谱参数

1.3.1.1 粉末样品对光的漫反射

若有一束光投向由粉末组成的漫反射体，则光和物质相互作用包括反射、吸收、透射、衍射等。其中光的反射还可以分为镜面反射（符合入射角等于反射角的条件）与漫反射，见图 1-2。

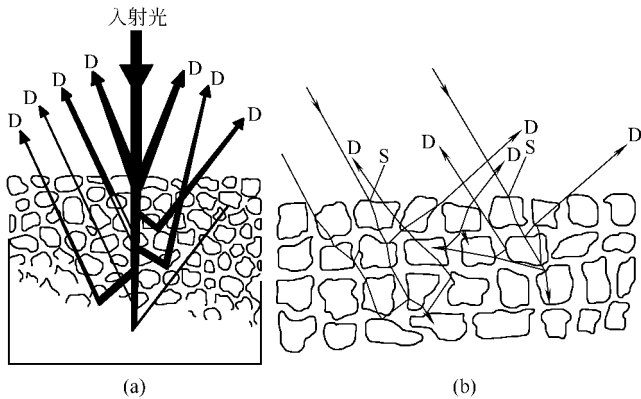


图 1-2 样品对光漫反射的示意图

图 (b) 是漫反射体中某一局部的颗粒对光的镜面反射和漫反射的示意（分别用 S、D 表示镜面反射与漫反射）。由图可见镜面反射只发生在表面颗粒的表层，因为镜面反射光束没有进入样品和颗粒的内部，未与样品内部发生作用，因此它没有负载样品的结构和组成的信息，不能用于组分分析。而漫反射光是分析光进入样品内部后，经过多次反射、折射、衍射、吸收后返回表面的光。漫反射光是分析光和样品内部分子发生了相互作用后的光，因此负载了样品的结构和组成信息，可以用于分析。固体样品的透射光虽然也负载了样品的结构和组成的信息，但透射光强受样品厚度的影响很大，应用于分析测定比较困难。

漫反射光的强度决定于样品对光的吸收，以及由样品的物理状态所决定的散射。漫反射光强度与样品组分含量不符合比耳定律。因此需研究与样品浓度成线性关系的漫反射光谱参数，才能把漫反射光谱用于定量分析。

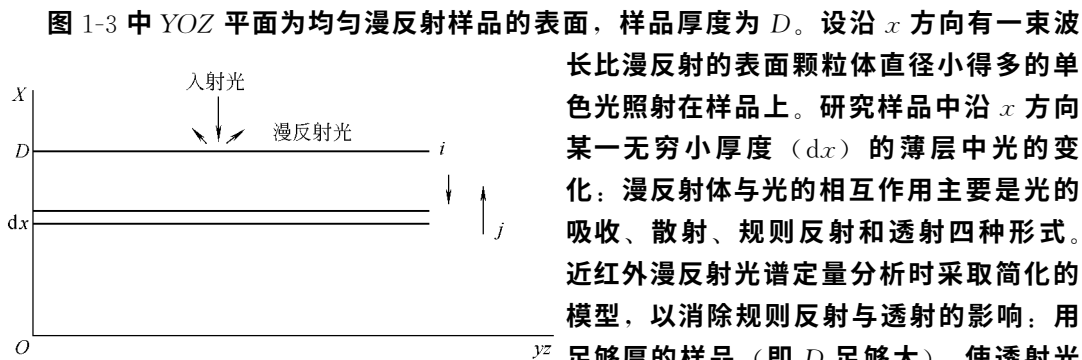


图 1-3 推导漫反射率的示意图

长比漫反射的表面颗粒体直径小得多的单色光照射在样品上。研究样品中沿 x 方向某一无穷小厚度 (dx) 的薄层中光的变化：漫反射体与光的相互作用主要是光的吸收、散射、规则反射和透射四种形式。近红外漫反射光谱定量分析时采取简化的模型，以消除规则反射与透射的影响：用足够厚的样品（即 D 足够大），使透射光可忽略；仪器检测用的特殊的设置方式，可消除规则反射光对测定的影响。因此我

们在对漫反射光谱定量分析理论讨论中可以只考虑样品对光的吸收与散射。

设样品的 YOZ 平面比入射光截面大得多，则可忽略样品边界的影响。考虑无穷小厚度的样品薄层 dx 上有强度为 i 的光沿负 x 方向入射，同时又有强度为 j 的光沿正 x 方向散射。不同物质对光的吸收与散射能力不同。若漫反射体的吸收系数为 K （由于物质对光的吸收，光在物质中经单位光程后物质对光的衰减程度为 di/dx ； K 主要决定于漫反射体的化学组成），散射系数为 S （由于物质对光的散射，光在物质中经单位光程后物质对光的衰减程度为 di/dx ； S 主要决定于漫反射体的物理特性），则在薄层 dx 中光强 i 的变化为：

$$-di = -Ki2dx - Si2dx + Sj2dx \quad (1-3)$$

其中： $2dx$ 为光通过 dx 薄层时的折合光程。

类似地，光强 j 沿正 x 方向的变化为：

$$dj = -Kj2dx - Sj2dx + Si2dx \quad (1-4)$$

由式 (1-3)、式 (1-4) 即得到描述均匀漫反射体对光吸收与反射的微分方程：

$$-di/dx = -2(K+S)i + 2Sj \quad (1-5)$$

$$dj/dx = -2(K+S)j + 2Si \quad (1-6)$$

令

$$R = j/i \text{ (} R \text{ 的物理意义为漫反射率)} \quad (1-7)$$

对 R 求导，并利用商的导数公式可得：

$$dR/Sdx = R^2(1+K/S)R + 1 \quad (1-8)$$

解此微分方程，当 $D = \infty$ ，即样品层为无穷厚（一般实际上样品厚度只需 1mm 左右即可），得样品无穷厚度绝对漫反射率：

$$R'_{\infty} = 1 + K/S - \sqrt{(K/S)^2 + 2(K/S)} \quad (1-9)$$

式 (1-9) 说明 R'_{∞} 是 (K/S) 的函数, 即依赖于 S 与 K 的比值, 而与各自的绝对值无关。

R'_{∞} 不易测定, 实际测定的是相对漫反射率。硫酸钡 (BaSO_4) 等材料在近红外谱区内的 $R'_{\infty} \approx 1$ 即 $K \approx 0$ 。在近红外漫反射光谱定量分析中, 为简化测试方法 (避免测试入射光强), 减少规则反射对测定的影响, 消除仪器的光谱特性, 取它们为参比样品, 可得到待测样品无穷厚度的相对漫反射率。可简称为漫反射率。

$$R_{\infty} = \frac{R'_{\infty}(\text{样品})}{R'_{\infty}(\text{参比})} \approx 1 + K/S - \sqrt{(K/S)^2 + 2(K/S)} \quad (1-9')$$

R_{∞} 与样品中的组分浓度不成线性关系, 常用的与组分含量成线性关系的漫反射光谱函数有两种, 即反射吸光度与 “Kubelka-Munk” 函数。

1.3.1.2 漫反射吸光度

如同透射光谱的吸光度那样, 定义漫反射吸光度:

$$A = \log[1/R_{\infty}] = -\log[1 + K/S - \sqrt{(K/S)^2 + 2(K/S)}] \quad (1-10)$$

A 称为漫反射体的反射吸光度, A 与 K/S 是一条通过零点的关系曲线, 如图 1-4 所示。

从图 1-4 可知, 当 K/S 在一定范围内, A 与 K/S 可用截距不等于零的一条直线来近似表达。若方程 (1-10) 以间隔 0.01 用下列直线拟合。

$$A = a + b(K/S) \quad (1-11)$$

当 $K/S \in (0.2, 1)$, $(0.1, 0.5)$ 与 $(0.5, 1)$ 时, 直线拟合结果如图 1-4, 图中拟合线 1 的相关系数为 0.99, 标准差为 0.01。

回归结果充分说明反射吸光度在一定范围内, 即 K/S 在一定范围内, 这种以直线代曲线方法完全可行。

对于只有一种组分的样品 (例如 BaSO_4 背景中均匀分散蛋白质的样品), 当样品浓度不高时, 吸光系数 K 与样品浓度 C 成比例。

$$K = \epsilon \cdot C \quad (1-12)$$

其中 ϵ 为摩尔吸光系数。因此若 S 为常数 (保持 S 不变)。把式 (1-11) 中的 S 及 ϵ 有关的常数都包含到 b 中, 则式 (1-11) 可改写为:

$$A = a + b \cdot C \quad (1-13)$$

即漫反射吸光度 A 与样品浓度 C 成线性关系。反射吸光度是近红外漫反射光谱分析 (NIRDRSA) 定量分析时常用的光谱线性参数, 但和透射光谱中吸光度 A 与浓度 C 的线性关系不同: 漫反射光谱中, 漫反射吸光度 A 与浓度 C 的线性关系只有在散射系数 S 保

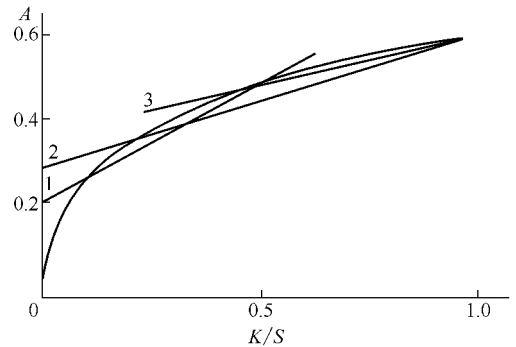


图 1-4 漫反射吸光度 A 与 K/S 的关系曲线

持不变，以及前述样品的粒度合适，透射光、规则反射光及仪器光谱特性的影响可以不计等各个条件满足时才能成立，这是 NIRD RSA 从样品制备到数据处理的全过程中必须注意的。此外，线性方程 (1-13) 是由拟合得到的，样品的浓度范围不能很大， A 与 C 的线性关系中，必然存在一个截距 a ，这是与透射光谱比耳定律的重要差别。在比耳定律中， A 与 C 不仅为线性关系，而且是成比例的（截距等于 0）；在漫反射光谱中，截距 a 、斜率 b 以及线性程度都与样品的浓度范围有关。当样品浓度较高时，可在较宽的浓度范围内保持较好的线性关系，但此时 a 较大， b 较小（见图 1-4），用于测定时灵敏度较低；当样品浓度较低时， a 较小， b 较大（见图 1-4），即用于测定样品时可以得到较高的灵敏度，这对于测定低浓度样品是有利的。

实测的结果与理论分析一致：将牛血清蛋白与赖氨酸分别分散在 BaSO_4 中，制成不同浓度的蛋白质与赖氨酸固体样品，在 NICOLET SX-170 傅里叶变换分光光度计上，蛋白质样品在 4588cm^{-1} (2180nm) 处，赖氨酸样品在 4487cm^{-1} 与 4588cm^{-1} (2229nm , 2180nm) 处用漫反射样品池测得漫反射吸光度与样品浓度 C 的关系见表 1-2 与图 1-5。

表 1-2 不同浓度蛋白质样品的 A 值与 $K-M$ 函数值					
浓度 $C\%$	A 值	$K-M$ 值	浓度 $C\%$	A 值	$K-M$ 值
0.1733	0.0312	0.00258	12.40	0.1208	0.0389
0.9477	0.0406	0.00437	15.33	0.1401	0.0525
4.529	0.0731	0.0142	相关系数	0.993	0.9986
9.495	0.1066	0.0303	标准差	0.0057	

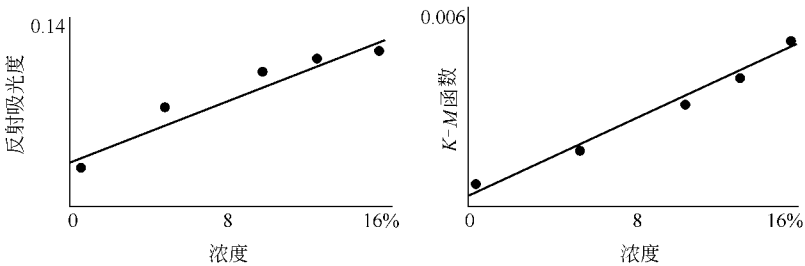


图 1-5 蛋白质漫反射光谱参数与浓度间的关系

由蛋白质浓度 C 与反射吸光度 A 之间的相关系数与标准差可知漫反射测定的准确度相当高。此外尽管反射吸光度与浓度只是近似的线性关系，仍可用微分光谱消除某些实验条件不可避免的不确定性。如含量为 $0.05\% \sim 1\%$ 的氨基酸样品 C 与 A_{4589} 的相关系数与标准差分别只有 0.876 与 0.00107，而改用微分光谱 ($A_{4589} \sim A_{4487}$) 后，浓度 C 与反射吸光度差 ($A_{4589} \sim A_{4487}$) 之间的相关系数与标准差可分别达到 0.984 与 0.00025。

1.3.1.3 漫反射光谱的 Kubelka-Munk ($K-M$) 函数

由式 (1-9')

$$R_{\infty} = 1 + K/S - \sqrt{(K/S)^2 + 2(K/S)}$$

可以解出 K/S 的表达式

$$K/S = (1 - R_{\infty})^2 / (2R_{\infty})$$

定义上式的右边为一函数

$$F(R_{\infty}) \equiv (1 - R_{\infty})^2 / (2R_{\infty}) = K/S \quad (1-14)$$

此函数即为 K - M 函数 (Kubelka-Munk function) 即 $F(R_{\infty}) = K/S$, 对某一样品组分, 当各样品间散射系数 S 一定时, 在某一波长处, $F(R_{\infty}) \propto K$, 而以式 (1-12) 代入得 K - M 函数与浓度的关系:

$$F(R_{\infty}) = K/S = \epsilon C / S = bC$$

式中: b 为常数, 与样品摩尔吸光系数及光程等有关; C 为样品组分浓度。即 K - M 函数与样品浓度之间的关系与比耳定律一样, 是正比的关系。实测的结果与理论分析一致, 见表 1-2 与图 1-5。

1.3.2 多组分样品 NIRDRS 理论

1.3.2.1 多组分样品光谱分析的一般理论

若把下标 i 记为漫反射光谱中的波长序号, 把下标 j 记为复杂样品中各组分的序号, f 记为样品中某一种成分漫反射光谱的某种线性函数, F 记为复杂物漫反射光谱的某一线性函数时, 且样品只有一种 j 成分, 其浓度为 C_j , 则样品漫反射光谱第 i 波长的线性函数值 $f_{ij} \propto C_j$, 即

$$f_{ij} = K_{0ij} + K_{1ij}C_j \quad (1-15)$$

式中 K_{0ij} 、 K_{1ij} 为系数。若光谱参数与浓度成比例, 则

$$f_{ij} = K_{1ij}C_j \quad (1-16)$$

若线性函数符合加合性, 则由 n 种成分组成的复杂物, 在第 i 波长处的线性函数 F_i 等于各组分在同一波长处线性函数值 f_{ij} 之和。由式 (1-15) 得:

$$F_i = \sum_{j=1}^n f_{ij} = \sum_{j=1}^n K_{0ij} + \sum_{j=1}^n K_{1ij}C_j \quad (1-17)$$

即常数项与变量项需要分别加合。

由式 (1-16) 得:

$$F_i = \sum_{j=1}^n f_{ij} = \sum_{j=1}^n K_{1ij}C_j \quad (1-17')$$

又若样品中各组分的光谱特性已知, 为了求出 C_j , 需要列出 j 个波长处的 F_i 表达方程, 组成 j 元联立方程, 解此联立方程可得样品中第 j 组分的浓度 C_j 。

$$C_j = K'_j + \sum_{i=1}^n K'_{ij}F_i \quad (1-18)$$

1.3.2.2 使用 K - M 函数作为光谱参数做多组分样品的 NIRDRS 分析

由于 $F(R_{\infty}) = (1 - R_{\infty})^2 / (2R_{\infty}) \equiv K/S$

得: $F(R_{\infty})_{\text{mix}} = K_{\text{mix}} / S_{\text{mix}} \quad (1-19)$

分别讨论 K_{mix} , S_{mix} :

对 n 种样品混合时, 各组分的物理状态一致, 即对光的散射系数相等时, 则:

$$S_{\text{mix}} = S_j = S \quad (1-20)$$

又对于由 n 种组分的混合样品的漫反射光谱, 由于各组分分子对光的吸收和散射是互相独立的, 因此可得吸收系数的加合性。任一波长 λ_i 处混合物样品总的吸收系数 K_{mix} 等于各组分吸收系数之和。

$$K_{\text{mix}} = K_{i1} + K_{i2} + \cdots + K_{in} = \epsilon_{i1} C_1 + \epsilon_{i2} C_2 + \cdots + \epsilon_{in} C_n = \sum_{j=1}^n K_{ij} = \sum_{j=1}^n \epsilon_{ij} C_j \quad (1-21)$$

由式 (1-20)、式 (1-21):

$$F(R_{\infty})_{\text{mix}} = K_{\text{mix}} / S_{\text{mix}} = \sum_{j=1}^n K_{ij} / S = \sum_{j=1}^n (K_{ij} / S) = \sum_{j=1}^n F(R_{\infty})_j \quad (1-22)$$

可见, 从理论上, 漫反射光谱具有光谱的 K - M 函数 $[F(R_{\infty})]$ 的加合性, 实测的结果与理论分析一致。因此可以按前面讨论的通过解联立方程得到式 (1-18) 的浓度方程。

实际的样品中各组分的光谱特性是未知的, 即方程 (1-22) 中的常数项 K_{ij} 为未知的, 因此上述方程式不能用线性代数直接解出, 只能用统计方法求得。理论上可用一批已知组分含量的样品通过自变量选择结合多元线性回归来求出式 (1-18) 中各浓度计算方程的常数。

1.3.2.3 使用反射吸光度为光谱参数做多组分样品 NIRDRS 分析

按式 (1-11) 在任一波长处的反射吸光度为:

$$A_{\text{mix}} = a_{\text{mix}} + b_{\text{mix}} (K_{\text{mix}} / S_{\text{mix}}) \quad (1-23)$$

由式 (1-20)、式 (1-21) 得到:

$$A_{\text{mix}} = a_{\text{mix}} + b_{\text{mix}} \left(\sum_{j=1}^n K_j / S \right) = a_{\text{mix}} + (b_{\text{mix}} / S) \left(\sum_{j=1}^n \epsilon_j C_j \right) \quad (1-24)$$

因为截距 a 是式 (1-10) 拟合所产生的, 拟合样品的浓度范围若不同或波长点不同, a 的值就不同, 因此 a 不具备加合性, 即

$$a_{\text{mix}} \neq \sum_{j=1}^n a_j$$

故

$$A_{\text{mix}} \neq \sum_{j=1}^n A_j$$

这就是为什么漫反射光谱的反射吸光度不具备光谱吸光度的加合性的原因, 因此漫反射光谱分析若以反射吸光度为线性参数时, 不能像经典多组分分析那样, 通过光谱吸光度的加合性建立并求解联立方程, 来实现多组分分析。但任一波长处如式 (1-24) 那样的方程确实存在 (只是不等于混合物中各组分在该波长处的实际反射吸光度之和), 由此推得混合物样品中任一组分浓度 C_j 也必然存在下列形式的方程:

$$C_j = b_0 + b_1 A_1 + b_2 A_2 + \cdots + b_n A_n \quad (1-25)$$

1.3.2.4 近红外漫反射分析的浓度计算方程

式 (1-25) 就是混合样品近红外漫反射分析中某组分的浓度计算方程, 方程中的 A 是漫反射吸光度, 实际分析时也可以用其他漫反射光谱线性函数 F 来代替 A (如 K - M 函数或反射吸光度的微分), 方程 (1-25) 改为:

$$C_j = b_0 + b_1 F_1 + b_2 F_2 + \cdots + b_n F_n = K_0 + \sum_{i=1}^n K_{ij} F_i \quad (1-26)$$

做 NIRDRSA 时为了得到这个方程, 只可能通过已知样品的光谱由多元线性回归等统计方法得到。

天然物等复杂样品常包含成百上千种成分, 对这样的多组分要解出方程 (1-26) 实际上是不可能的, 而且该浓度计算方程中每一项 F 是仪器测量值, 含有系统测量误差、非线性误差等统计误差, 因此浓度计算方程的项数越多, 计算结果的总误差也越大。实际测定包含 n 个组分的复杂样品中的每一个组分时, 并不需要测定 n 个波长位置的 F 值, 只需适当选择少数几个波长, 使样品中某一组分浓度 C 主要只与这少数几个波长处的 F 值有关。这些波长中有的波长处的值与待测成分浓度 C 关系特别密切, 即信息量特别大, 这个波长称为待测成分的灵敏波长。复杂样品在灵敏波长位置的 F 值除了与待测成分浓度有关外, 也受复杂物中其他成分的影响, 为了补偿这些影响, 还需另外选择几个波长, 这几个波长处的 F 值应覆盖所有背景对测量的影响。用这几个波长处的 F 值来校正其他成分在灵敏波长处对待测成分的影响。若一共选择了 m 个波长处的 F 值来计算复杂样品中某一组分的含量, 则这个组分的含量计算公式为:

$$C_j = K_0 + \sum_{i=1}^m K_{ij} F_i \quad (1-26')$$

式中 C_j 为样品中某待测成分的浓度, F_i 是混合物样品漫反射光谱经线性化变换后的第 i 个波长的某一光谱线性函数值。 K_0 、 K_{ij} 是浓度计算方程中的常数与系数, 没有物理意义。实际的浓度计算方程还应加上一项误差 E , 包括由于线性变换的近似性以及测量 F_i 所产生的误差 (系统误差)。

若一共选择了 $m=3$ 个波长处的 F 值来计算复杂样品中某一组分的含量, 则方程 (1-26) 可改为:

$$C_j = K_0 + \sum_{i=1}^m K_{ij} F_i + E = K_0 + K_1 F_1 + K_2 F_2 + K_3 F_3 + E(m=3) \quad (1-27)$$

1.4 近红外光谱测定的物理学基础

1.4.1 离散点测定及连续测定

1.4.1.1 近红外光谱的离散点测定

离散点的测定是测量几个特定波长点的光谱数据, 并建立样品浓度与这些光谱数据的关系。进行离散点测量的仪器通常是滤光片型的近红外仪器。

滤光片 (干涉滤光片) 是一种简单而廉价的波长选择器, 其作用是选择性地透过一定波长范围的光, 其优点是透射光的波长范围窄, 透光率高。滤光片的滤光特性用最大透光

波长（中心波长 Center Wavelength）、谱带半宽度（带宽 Bandwidth）、透过率（Percent Transmittance）来描述。中心波长是指在该波长处辐射有最大透光率，而带宽是指最大

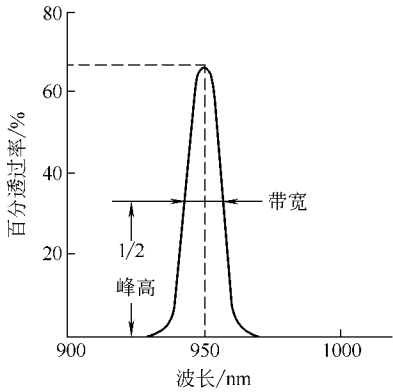


图 1-6 滤光片的中心波长、带宽、透过率示意图

透光率的一半处谱带的波长范围，以 nm 表示，带宽越小，则单色光的纯度越高，一般滤光片的带宽为 20～50nm，高的可达 10nm。透过率是指透过滤光片的光能量与入射到滤光片的光能量的百分比。例如某滤光片的中心波长为 950nm，带宽为 10nm，透过率为 65%，可记作：950nm/10nm/65%。滤光片的 3 个指标的示意图见图 1-6 所示。

图 1-7 所示为发光二极管（LED）型近红外仪器。它用 LED 作光源，由不同的二极管及干涉滤光片产生不同的波长。因为 LED 器件体积小、消耗低，仅需要几个二极管就能将光谱仪做成比其他类型的仪器更小、更价廉、更精巧的过程控制器。此类仪器适合于在过程检测中在线使用，或最适合于手提式仪器中。此类仪器的

优点是没有移动部件，是相当坚固的；缺点是带宽变化，波长数目有限，准确度和精度有限。

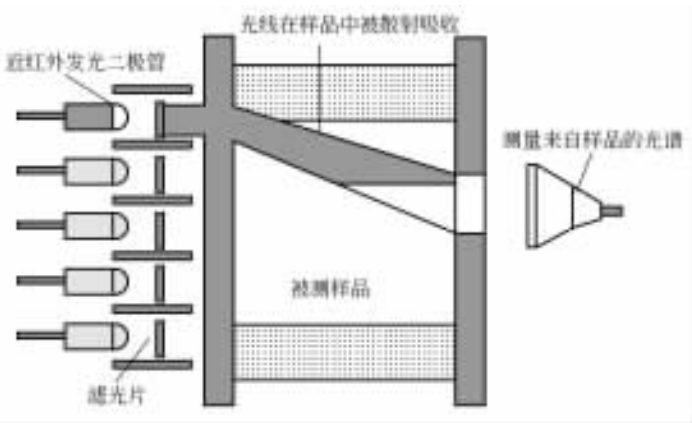


图 1-7 发光二极管（LED）型近红外仪器

1.4.1.2 近红外光谱的连续测定

近红外光谱的连续测定是测定一个波长区域范围内的连续光谱。用建立模型的各种全谱方法，来建立样品成分含量与这些光谱数据的关系。要得到连续光谱，可在频率域进行扫描，以得到光谱参数与波长的关系（光谱图），或在时间域进行扫描，得到样品的干涉图，再经傅里叶变换得到样品的光谱图。

1.4.2 频率域的直接扫描

进行频率域直接扫描的仪器通常有光栅单色器型近红外仪器、声光可调滤光片型近红外仪器等。下面简单介绍一下这些仪器的物理原理。

1.4.2.1 扫描色散型近红外光谱仪的物理原理

采用光栅单色器的仪器，通常称为色散型仪器，它是最常见的近红外仪器。这类仪器

与紫外-可见光谱仪具有通用的光学设计，只要更换光栅、滤光片和检测器，就可构成近红外光谱仪。图 1-8 是基于单色器的近红外光谱仪。

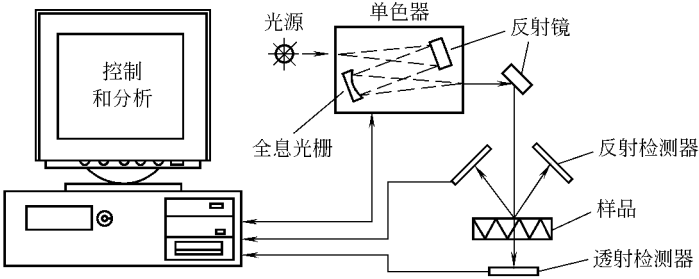


图 1-8 基于单色器的近红外光谱仪

这类仪器的单色器采用光栅，下面介绍一下光栅及光栅的色散作用。
常用的光栅元件有透射光栅和反射光栅。透射光栅是在一块玻璃或其他透明材料上刻划一系列平行的等距离的刻线（凹槽）而制成，制作这种母光栅需要精密的光学机械设备，成本高昂；利用母光栅作模子而制得复制光栅则比较便宜，其性能虽不如母光栅好，但基本上可满足一般分光仪器的要求。反射光栅是在一抛光的金属表面上刻划一系列平行而等距离刻线或在复制光栅表面喷镀一层铝薄膜而制成，反射光栅比透射光栅应用更广泛。

图 1-9 表示阶梯光栅（反射光栅）的衍射原理示意图，入射光束从凹槽表面反射，就如同一个新辐射源，不同波长的辐射以不同反射角反射而引起干涉，相邻两平行单色光束（1，2）构成干涉的必要条件是两光束的光程差等于入射光波长 λ 的整数倍，即：

$$m\lambda = \overline{CD} - \overline{AB} \tag{1-28}$$

因 $\angle CAD = \angle i$ ， $\angle BDA = \angle r$ ，故 $\overline{CD} = d \sin i$ ， $\overline{AB} = -d \sin r$ 。 d 为相邻凹槽之间的距离；又称光栅常数；整数 m 为衍射级次；负号表示入射角 i 和反射角 r 的符号相反。将 $\overline{AB}$ 和 $\overline{CD}$ 的值代入上式得到构成反射的条件：

$$m\lambda = d(\sin i + \sin r) \tag{1-29}$$

若入射角 i 保持不变，对上式微分，对任一给定的入射角，则有：

$$\frac{dr}{d\lambda} = \frac{m}{d \cos r} \tag{1-30}$$

式（1-29）说明，对给定光栅，当入射角 i 一定时，任一光栅级中，波长愈长，反射角 r 愈大。因此，光栅光谱按波长变化形成空间排列。从式（1-30）可知，光栅的角色散（ $dr/d\lambda$ ）随衍射级次的增加和光栅常数 d 的减小而增大。

色散型近红外仪器的光源为带石英外壳的 Tungsten-Halogen 灯，在 360~3000nm 的

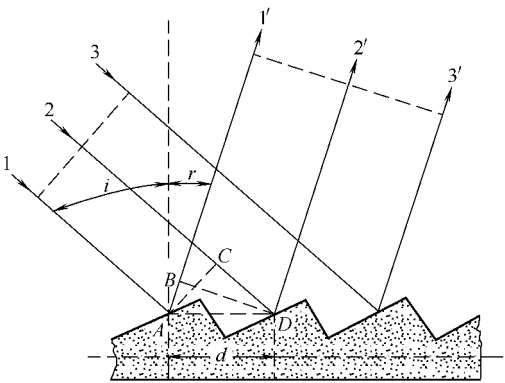


图 1-9 阶梯光栅（反射光栅）的衍射原理
1~3—入射光束 1'~3'—反射光束

区域提供高能量的输出。在灯壳里的 halide 起“沐浴”作用，保持石英窗口的洁净，因此它们的寿命很长。用于反射和透射仪器的检测器通常有：硅检测器用于 360~1000nm 的区域，PbS 检测器用于 900~2600nm 的区域。为克服硫化铅的温度漂移，常采用带制冷的 PbS 检测器。在某些情况下，也使用 InGaAs 检测器。但由于这种检测器的面积较小，不能用于收集漫反射光。

计算机成为近红外技术不可缺少的部分。用计算机来进行光谱数据的采集，用化学计量学软件来提取信息。化学计量学软件是近红外光谱仪的必要组成部分，它提供对数据进行预处理，建立校准模型，对模型进行统计检验，对未知样品进行预测，输出校正与分析报告及相关图等。

1.4.2.2 声光可调滤光片型近红外仪器物理原理

20 世纪 90 年代以来，最惊人的发展是声光可调滤光器（AOTF：Acoustooptical Tunable Filter），它是没有移动部分的单色器。

光栅是利用机械刻划或全息原理形成周期性变化的空间结构。不同波长的光通过光栅因衍射和多光束干涉而色散。声光调制器是利用交变电场作用在具有压电效应的晶体上，在晶体内部形成周期性变化的空间，从而和普通光栅一样能使光发生色散作用。改变交变电场的频率相当于改变光栅常数，从而改变声光调制器产生的波长。

声光分光是基于双折射晶体的声光衍射。声光衍射的原理类似于透射光谱。声光光栅的光栅常数等于声波的波长，因此声光光栅可看成一个色散元件。声光滤光器就是根据声光衍射原理，采用具有较高的声光品质因素和较低的声衰减的双折射单晶制成的分光器件。图 1-10 是 AOTF 的结构简图。滤光器是由双折射晶体（TeO₂ 晶体）黏结上压电转换器件组成。当高频电信号由压电换能器转换成超声信号并耦合到双折射晶体内以后，在晶体内形成一个声行波场；当一束复色光以一个特定的角度入射到声波行波长内后，经过光与声的相互作用，入射光被超声衍射成两束正交偏振的单色光。当改变入射超声频率时，衍射光波长也随之改变；连续改变超声频率，就能实现衍射光波长的快速扫描。由此可见，声光滤光器对衍射光的调谐是很方便的。

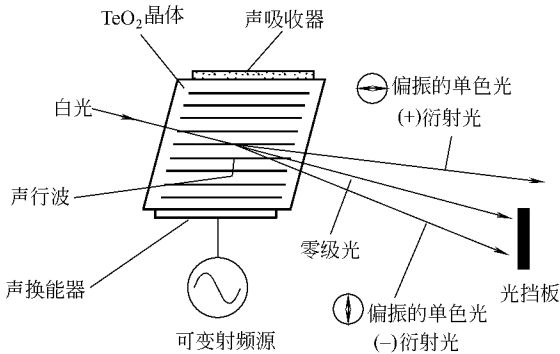


图 1-10 AOTF 结构简图

在各向异性双折射晶体中，要实现声光衍射必须满足以下动量匹配条件：

$$K_d = K_i + K_a$$

K_d 、 K_i 、 K_a 分别为衍射光波矢、入射光波矢、声波矢。各波矢的大小为： $K_d =$

$2\pi n_d/\lambda_0$, $K_i=2\pi n_i/\lambda_0$, $K_a=2\pi f_a/V_a$, 其中: n_d 为衍射光的折射率; n_i 为入射光的折射率; λ_0 为真空中的光波波长; f_a 为超声波长; V_a 为声速。

可见, 通过改变超声频率可以得到不同的衍射光波长。当超声频率升高, 衍射光波矢增大, 从而衍射光波长变小。连续改变超声频率就能实现衍射光波长的快速扫描, 超声频率与调制波长的对应关系如图 1-11 所示。

1. 4. 3 时间域的扫描

时间域的扫描是由傅里叶变换型近红外光谱仪来得到的。20 世纪 70 年代傅里叶变换技术在中红外光谱仪器上的应用使红外光谱仪器的性能得到彻底的改变, 进入 20 世纪 80 年代傅里叶变换红外光谱已成为红外光谱仪器的主导产品。借助于研制中红外光谱仪器的基础, 通过调整光源和检测器, FT-NIRS 仪器应运而生, FT-NIRS 仪器的主要光学元件是 Michelson 干涉仪, 其作用是使光源发出的光分成两束后造成一定的光程差再使之复合以产生干涉, 所得的干涉图函数包含了光源的全部频率和强度信息, 用计算机将干涉图函数进行傅里叶变换计算出原来光源的强度按频率的分布, 样品干涉图函数经傅里叶变换后与空白时光源的强度按频率分布的比值得到样品的近红外图谱。与扫描型仪器相比, FT-NIRS 仪器的扫描速度快、波长精度高、分辨率好, 由于短时间内即可进行多次扫描, 使信号做累加处理, 加之光能利用率高、输出能量大, 因而仪器的信噪比和测定灵敏度较高, 可对样品中的微量成分进行分析; 这类仪器的弱点是干涉仪中有移动性部件, 需要较稳定的工作环境。这类仪器中使用的检测器与扫描型仪器类似, 由于多在长波近红外区域使用, 一般采用 PbS 检测器。定性和定量分析采用全谱校正技术。

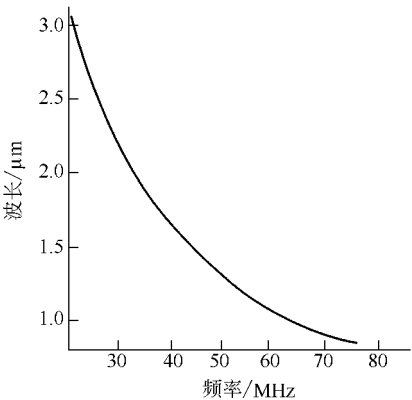


图 1-11 超声频率与调制波长的关系图

下面对光波在频率域与时间域中的表征及傅里叶变换红外光谱仪做详细的讨论。

1. 4. 3. 1 光波在频率域与时间域中的表征

周期性的运动可在两种域 (Domain) 中得到表征: 一种表征域是表现出周期性的域, 例如, 电 (磁) 场强度随时间 (空间) 的分布, 就是在时 (空) 域中表征光波的特征; 另一种表征域是运动状态按某一周期性参数 (频率、波长、波数等) 的分布, 可统称为频率域。这两种域中表征同一运动状态, 可通过傅里叶变换 (Fourier Transform, 简称 FT) 相互转变。通常所说的某种光的光谱是指该光包含的不同频率成分的强度按频率的分布, 因此光谱就是光在频率域中的表征。

初相位为零 (以下讨论均如此)、时间频率为 ν 的某种单色光, 在时间域中的表征是指在空间一定的位置, 电场强度 E_t 随时间 t 的变化: $E_t=A\cos2\pi\nu t$

空间频率为 σ 的光, 在空间域中的表征是指在某一定时间, 电场强度 E_x 随离开光源距离 x 的变化, 或称 E_x 在空间 x 的分布: $E_x=A\cos2\pi\sigma x$ 。

因为在时间域与在空间域中表征光的形式相同, 下面只是以光在空间域中的表征来讨论光在时空域和频率域表征的关系。

图 1-12 中 A、B、C 是指某种偏振光在空间域中的表征，图中 A'、B'、C' 是指对应光在频率域中的表征。

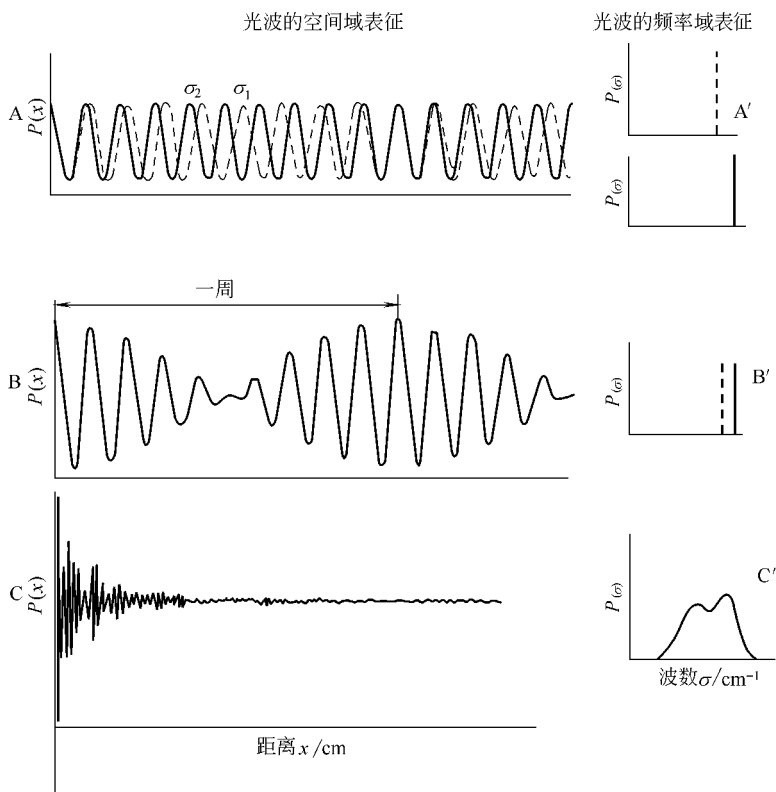


图 1-12 光在时空域和频率域的表征

图 1-12 中的 A 和 A' 分别是波数为 σ_1 和 σ_2 、电场振幅为 A 的两种一维单色平行光，在空间域与频率域中的表征。

图 1-12 中的 B' 是包含波数为 σ_1 、 σ_2 两种成分的复色光在频率域中的表征；若 1, 2 两光为相干光，空间电场强度可叠加，则空间 x 处的电场强度等于两光在该点的电场强度 E_1 、 E_2 的叠加（见图 1-12 中 B）；其空间域中的数学表达为：

$$E_x = E_1 + E_2 = A_1 \cos 2\pi \sigma_1 x + A_2 \cos 2\pi \sigma_2 x \quad (1-31)$$

同理，对于空间频率（波数）为 σ_1 、 σ_2 、 $\dots$ 、 σ_n 共 n 种成分的复色相干光，在空间域中的数学表达为：空间 x 处的电场强度 E_x 是各频率成分 σ_i ($i=1 \dots n$) 在 x 处的电场强度 E_i 的叠加：

$$E_x = \sum_{i=1}^n E_i = \sum_{i=1}^n A_i \cos 2\pi \sigma_i x \quad (1-31')$$

图 1-12 中的 C' 是波数为连续的复色光在频率域中的表征，这就是一般的光谱；图中的 C 是在空间域的表征（假定为相干光），其数学表达为：空间域 x 处的电场强度 E_x 是复色光中各频率成分（波数从 0 到 ∞ ）的光在该点 X 处电场强度的积分。

$$E_x = \int_{i=0}^{+\infty} A(\sigma) \cos 2\pi \sigma_i x d\sigma \quad (1-32)$$

因为波数为 σ 的光强度 $P(\sigma)$ 与其振幅 $A(\sigma)$ 有确定关系，因此可以把电场的振幅 $A(\sigma)$ 换成不同波数光的强度 $P(\sigma)$ ；相应 E_x 换成 $P(x)$ ，上式成为：

$$P(x) = \int_{\sigma=0}^{+\infty} P(\sigma) \cos 2\pi \sigma x d\sigma \tag{1-32'}$$

$P(x)$ 与 $P(\sigma)$ 分别是光在时间域与频率域中的表征，上述关系式就是傅里叶变换式，即通过 FT 可把光在频率域表征转换为空间域表征。又因为 FT 是可逆的，故频率域中某一频率的光强 $P(\sigma)$ ，可以从光的空间域表征形式通过 FT 逆变换得到：

$$P(\sigma) = \int_{x=0}^{+\infty} P(x) \cos 2\pi \sigma x dx \tag{1-32''}$$

式中 $P(x)$ 是复色光中各种频率成分的光在空间 x 点相干并叠加后的总强度。

以上是从光谱的物理意义上对 FT 所做的定性的讨论，上述这种 FT 转换只是理论上的分析，实际只能在 FT 光谱仪中实现。FT 的数学意义参见有关数学论著。FT 在仪器分析中有极重要的意义，不但直接用于红外光谱、核磁共振谱与质谱分析仪中，还广泛用于频谱分析、降低噪声等技术方面。

1. 4. 3. 2 傅里叶变换红外分光光度计

傅里叶变换红外分光光度计简称傅里叶变换 (FT) 光谱仪，其基本组成包括五部分：
①分析光发生系统，包括光源、样品等用以产生负载了样品信息的分析光。
②迈克尔逊干涉仪：产生空间（时间）域中表达的分析光，即干涉光（或称干涉图）。
③检测器：检测干涉光。
④采样系统：通过模数转换器 (A/D) 把干涉光数字化，并把数据采入（读入）计算机。
⑤计算机系统与显示器：把数字化了的干涉光经过 FT 数字变换，由数模转换 (D/A) 成为频率域中的样品光谱，并在显示器中显示，见图 1-13。

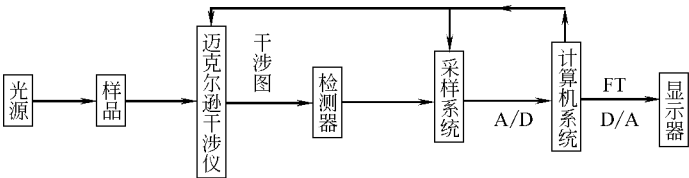


图 1-13 FT 光谱仪原理结构框图

与色散型红外分光光度计不同，实际的 FT 光谱仪中样品直接放置于检测器前，这是因为：①迈克尔逊干涉仪需要较好的平行光；②所检测的是干涉仪所产生的干涉光，样品自身产生的红外辐射对它的影响相对较小。

下面对 FT 光谱仪的检测器、迈克尔逊干涉仪与采样系统三个组成部分进行介绍。

(1) 检测器

FT 光谱仪的扫描速度快，因此色散型分光光度计常用的热电偶等检测器，因响应速度过慢而不能使用。FT 光谱仪所用检测器见表 1-3。

FT 光谱仪的检测器有两种频率特征：一是所检测光的频率；另一是被干涉仪调制成的频率（干涉光强度变化的频率），即表中的响应时间；后者决定了 FT 光谱仪的扫描速度。

表 1-3

FT-IR 中使用的检测器

名 称	类 型	工作温度/K	适用波长范围/ μm	响应时间/ μs
硫酸三甘肽(TGS)	热电型	295	2~50	1
汞镉碲(MCT)	光电导型	77	0.8~40	1
硒化铅(PbSe)	光电导型	195	1~5	2
碲化铟(InSb)	光伏型	77	1~5.5	1

(2) 迈克尔逊干涉仪

光在频率域中的表达可通过普通光谱仪得到；而光在空间域中的表达可通过迈克尔逊干涉仪得到（分析光包含的不同频率成分的干涉强度在空间的分布，并不是电场强度在空间的分布）。图 1-14 是迈克尔逊干涉仪原理图。

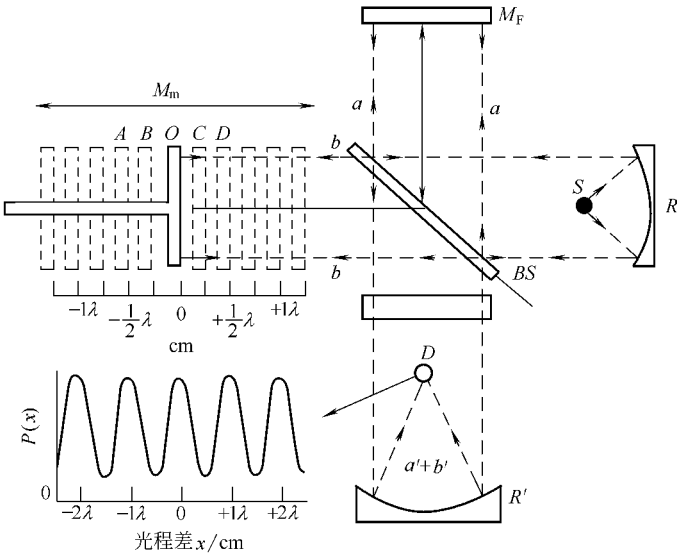


图 1-14 迈克尔逊干涉仪工作原理图

图中光源 S 产生的光经反射镜 R 成为平行光， BS 为光的分束器。入射平行光按 45° 角射到 BS 后，有 $1/2$ 强度的光被反射（图中的 a ），另外 $1/2$ 强度的光透过 BS （图中的 b ）。图中 M_F 为固定反射镜， M_m 为可移动反射镜， A 、 B 、 O 、 C 、 D 分别为移动反射镜的 5 个位置。 a 光被 M_F 反射后又透过 BS ，形成 a' ； b 光被 M_m 反射后又被 BS 反射形成 b' ， $(a' + b')$ 光被反射镜 R' 反射后由检测器 D 检测。由于 a' 光和 b' 光都是从 S 光源产生的，因此它们是相干光；相干的结果是增强还是减弱，决定于 M_m 的位置。图的左下方为 a' 、 b' 光相干后的强度随光程差的变化，称为干涉图。干涉图的横坐标是光程差，与动镜 M_m 的位置相对应。当 M_m 与 M_F 对于 BS 为对称时， a' 与 b' 两相干光所经过的光程相同，光程差为 0 ，因此在干涉图上相干后的强度为极大。

若光源产生的光是波长为 λ 的单色光，则当动镜的位置移动 $1/4\lambda$ 的距离达到 B 或 C 点（即 $OB = OC = 1/4\lambda$ ）时， a' 、 b' 两相干光的光程差为 $1/2\lambda$ ，因此相干后的强度为极小；若动镜位置继续移动到 A 或 D ， $OA = OD = 1/2\lambda$ 时， a' 与 b' 两相干光的光程差为 λ ，此时相干后的强度又达到极大。随着动镜的移动，检测器可检测到信号 $P(x)$ 随光程差 x

(等于动镜移动距离的 2 倍) 变化的正弦波:

$$P(x) = A \cos 2\pi(x/\lambda) \quad (1-33)$$

式中 A 为振幅, 这就是 λ 单色光通过干涉仪产生的干涉图, 它是 λ 单色光的空间域表达形式。

若光源产生的光包含波数为 σ_1 、 σ_2 的两种成分: 因为这两种波数成分的光, 在迈克尔逊干涉仪中分别被分成 a' 、 b' 两种相干光, 并分别因干涉而形成频率不同的余弦波 (见图 1-12 中的 A); 对于某一定的光程差 x , 检测器所检测到的光强度 $P(x)$ 是两波数成分的光, 在 x 处相干后光强度 $P(x_1)$ 、 $P(x_2)$ 的叠加 (见图 1-12 中 B); 其空间域中的数学表达同式 (1-31):

$$P(x) = P(x_1) + P(x_2) = A_1 \cos 2\pi\sigma_1 x + A_2 \cos 2\pi\sigma_2 x \quad (1-34)$$

若光源产生的光是包含波数为 σ_1 、 σ_2 、 $\dots$ 、 σ_n 共 n 种成分的光: 因为这些波数成分的光在迈克尔逊干涉仪中分别因干涉而形成频率不同的余弦波; 对于某一定的光程差 x , 检测器所检测到的光强度 $P(x)$ 是这 n 种波数成分的光, 在 x 处相干后的光强度 $P(x_i)$ 的叠加, 其空间域中的数学表达同式 (1-31'):

$$P(x) = \sum_{i=1}^n P(x_i) = \sum_{i=1}^n A_i \cos 2\pi\sigma_i x \quad (1-35)$$

若光源产生的光是波长连续的复色光, 因为每种波数成分的光在迈克尔逊干涉仪中分别因干涉而形成频率不同的余弦波; 对于某一定的光程差 x , 检测器检测到的光强度 $P(x)$ 是各种波数成分的光在 x 处相干后的光强度的积分, 其空间域中的数学表达同式 (1-32'):

$$P(x) = \int_{-\infty}^{+\infty} P(\sigma) \cos 2\pi\sigma x d\sigma \quad (1-36)$$

式中: $P(x)$ 为复色光中各种频率成分的光在空间 x 点相干叠加后的总强度; $P(\sigma)$ 为各种频率 (σ) 成分光的强度。

其中当动镜位置在 O 点时, 因为对所有波长成分都是零光程差, 相干后都得到极大, 因此这时检测器所检测到的相干光总强度特别大。动镜离开 O 点后, 相干光总强度即迅速下降。这就形成了图 1-12 中 C 这样复色光相干后的特殊形式。傅里叶变换光谱仪通过迈克尔逊干涉仪直接得到的是光的空间域表达形式, 即光的干涉图。为了把光的干涉图转变成人们习惯的频率域表达式即光谱图, 需进行傅里叶反变换, 在 FT 光谱仪中这是由计算机通过 FT 数字变换来完成的。FT 光谱仪先把干涉图数字化后存入计算机, 然后对已经数字化的干涉图进行傅里叶变换, 实际上是进行一系列的积分: 为了得到频率域表达的光谱图, 需把光谱分割成若干个波数点, 分别算得各波数 σ 处光的强度 $P(\sigma)$:

$$P(\sigma) = \int_{-\infty}^{+\infty} P(x) \cos 2\pi\sigma x dx \quad (1-37)$$

计算机积分实际上是做数值积分, 即把干涉图分解成 n 个点, 然后由计算机做下述加法处理来代替积分。

$$P(\sigma) = \sum_{i=1}^n P(x_i) \cos 2\pi\sigma x_i \quad (1-38)$$

式中： $P(x_i)$ 为当光程差等于 x_i 时检测器实际测得的干涉光总强度。把各波数 σ 处光的强度 $P(\sigma)$ 按波数大小的次序排列，就得到光谱。

(3) 采样系统

把某一光程差 x 时的相干强度 $P(x)$ 读入计算机，称为计算机对干涉仪数据的采样。FT 光谱仪为了得到准确的结果，必须严格的按光程差（或动镜移动距离）一定间距进行采样。采样的起点、采样点的密度、动镜移动的速度等参数，对 FT 光谱仪的性能都有一定影响。现代的 FT 光谱仪，采样点是用 He-Ne 激光（波数为 15798cm^{-1} 的严格的单色光）通过迈克尔逊干涉仪后产生的干涉图（为严格的正弦曲线）作为采样点的依据。

参 考 文 献

- [1] 严衍禄, 吉海彦. 傅里叶变换近红外光谱技术及应用. 见, 吴谨光主编: “近代傅里叶变换红外光谱技术及应用” (上卷). 北京: 科学技术文献出版社, 1994. 251~301
- [2] 严衍禄主编. 现代仪器分析. 北京: 北京农业大学出版社, 1995
- [3] 徐阶平. 声光器件的原理、设计与应用. 北京: 科学出版社, 1982
- [4] Hildrum K I, Isaksson T, Naes T, et al. Near Infrared Spectroscopy Bridging the gap between data analysis and NIR application. New York: Ellis Horwood, 1992
- [5] Celia M Henry. Near-IR gets the job done. Anal Chem, 1999, 71 (17): 625A~628A
- [6] W fred McClure. Near-Infrared spectroscopy: The giant is running strong. Anal Chem, 1994, 66 (1): 43A~53A
- [7] Deborch Nobl. Illuminating Near-IR. Anal Chem, 1995, 67 (23): 735A~740A
- [8] Donald A Burns, Emil W Ciurczak. Handbook of Near-Infrared Analysis. Marcel Dekker INC, 1992

2. 近红外光谱分析的化学信息基础

与其他仪器分析一样，近红外光谱分析是用自然科学的方法获得有关物质系统的信息，并且对其解释、研究和应用^[1,2]。近红外光谱分析是从近红外光谱中提取样品的信息，因此研究和掌握近红外光谱技术首先必须研究和了解近红外光谱中包含的信息以及这些信息的特点，才能正确理解和运用近红外光谱分析技术提取样品的信息、完成分析。

本章第一节介绍近红外光谱的信息源，即近红外光谱信息的来源。这是理解近红外分析的基础：近红外光谱的各种信息特征以及近红外光谱分析技术的特征都是由近红外光谱的信息源所决定的。第二节具体介绍近红外光谱中包含的化学信息，这些信息决定了近红外光谱分析应用的范围。第三节介绍近红外光谱信息的特点，正是这些特点决定了近红外光谱分析技术与方法。

2.1 近红外光谱的信息源

分子在近红外谱区的吸收主要都是由于分子内部振动状态的变化而产生的。可以从经典力学与量子学说两种观点来理解与说明分子的振动状态。

2.1.1 分子振动的基频、倍频与合频

经典力学的观点，分子的振动状态用基频、倍频与合频来描述。

按经典力学，最简单的分子即两个质量分别为 m_1 、 m_2 的原子组成的双原子分子，可模拟为由弹簧连在一起的两个小球组成的弹簧振子。由虎克定律知道，这种系统的振动属简谐振动，因此称为谐振子（Harmonic Oscillator）。谐振子机械振动的频率为：

$$\nu_m = (1/2\pi) \sqrt{K/\mu} \quad (2-1)$$

式中： K 为力常数，其值决定于组成分子的两个原子间键能量的大小（键的强度）； μ 为双原子分子的折合质量。若组成分子的两个原子的质量分别为 m_1 、 m_2 ，则

$$\mu = (m_1^{-1} + m_2^{-1})^{-1}$$

若 K 以 mdyn/nm 为单位（1 达因 = 10^{-5} 牛）， μ 取原子量单位，则双原子分子机械振动相应的波数频率为

$$\sigma_m = (1/2\pi c) \sqrt{K/\mu} = 413 \sqrt{K/\mu} \quad (2-2)$$

分子的这种振动频率为振动的基频，分子也可以按基频的整数倍 2σ 、 3σ 频率发生振动，这种频率称为分子振动的 1、2…级倍频。若分子中存在几种频率的振动，则在一定条件下两种频率的振动可以发生耦合，形成频率相当于两种频率之和的所谓合频振动。

若分子是极性分子，则分子在机械振动的同时产生电磁场的振动（即交变电场），其波数频率 σ_E 与机械振动的波数频率相同。若用 σ 表示分子电振动波数频率，则

$$\sigma = \sigma_E = \sigma_m = 413 \sqrt{K/\mu} \quad (2-3)$$

具有电磁场振动的分子能与环境中相同频率的电磁波发生相互作用（共振吸收或发射），因为有机物中不同键伸缩振动的力常数 K 约为 $40 \sim 180 \text{mdyn/nm}$ ，弯曲振动的 K 值更小。因此按式（2-3）计算的有机物基频吸收的波数范围约在中红外区，即中红外谱区的频率范围与分子振动的基频频率相一致，而近红外谱区的频率范围与分子振动的倍频、合频振动频率相一致。

由于分子振动的频率包括基频、各级倍频与各种合频，因此分子可能通过共振吸收，对环境中频率与分子的基频、倍频与合频相同的电磁波产生吸收（分子内部振动的振幅变化），这就是分子的基频吸收、各级倍频吸收与各种合频吸收。

2.1.2 分子振动的能级^[3]

按量子的观点，分子的振动状态用能级来描述，振动能级的变化是指分子在不同振动能级之间的跃迁。由量子力学可得谐振子的能级公式为

$$E_V = hc\sigma(V + 1/2) \quad (2-4)$$

式中振动量子数 $V = 0, 1, 2, 3, \dots$ 。 E_V 为 V 能级的能量值。谐振子在不同能级间跃迁时的选择定则为 $\Delta V = \pm 1$ ，即跃迁只能在相邻能级间发生。每次跃迁吸收光子的能量 E 等于 $hc\sigma$ 。

按 Poltzmenn 方程可算得在室温下分子绝大部分处于振动基态（ $V=0$ ）。由振动基态到振动第一激发态（ $V=1$ ）之间的跃迁称为基频跃迁。这种跃迁所产生对辐射的吸收即为基频吸收，其吸收电磁波的频率可按经典力学公式（2-3）计算。

实际的分子振动并不完全符合简谐振动，由分子化学键的位能曲线可知分子属非线性谐振子（Anharmonic Oscillator）。其振动量子数为 V 的能级公式为下列近似表达式：

$$E = hc\sigma[(V + 1/2) - X(V + 1/2)^2 + X(V + 1/2)^3 \dots] \quad (2-5)$$

当上式中只取 2 项，则相邻的 [振动量子数分别为 V 和 $(V-1)$] 能级间的能量差

$$\Delta E = hc\sigma(1 - 2VX) \quad (2-6)$$

式中 X 为非谐性常数（Anharmonic constants）， X 常为很小的正数，其值约 0.01。这说明 V 越大时，相邻能级的间距 ΔE 越小，因此倍频的实际频率比基频乘以倍频数的计算值略小。

非线性谐振子跃迁的选择定则是： $\Delta E = \pm 1, \pm 2 \dots$ 即除了基频跃迁外也可能发生从基态到第二或更高激发态（ $V=2, 3 \dots$ ）之间的跃迁。这种跃迁称为二级倍频或多级倍频跃迁，所产生的吸收谱带为二级倍频或多级倍频吸收，总称为倍频吸收。

分子在中红外谱区的吸收是由于振动状态的相邻振动能级之间的跃迁，即基频吸收所形成的；而近红外谱区的吸收是由于分子振动的倍频（振动状态在相隔一个或几个能级间的跃迁）或合频（分子两种振动状态的能级同时发生跃迁）吸收所造成的，这是近红外光谱的信息源。

多原子分子的振动比较复杂，但可把它们看作是许多种简单的独立振动的线性组合而成。因为各种振动不是严格简谐的，它们之间可能发生相互作用，如果电磁波光子的能量

恰好等于某两个基频跃迁所需能量的和，而这两种基频振动又具有相同的对称性，这时该能量的 1 个光子可能同时激发这两种基频跃迁。在光谱中出现的吸收峰的波数（或频率）等于这两种基频振动波数（或频率）之和，这种吸收叫合频吸收。

2.1.3 近红外光谱的信息范围

分子振动的倍频与合频的特征决定了近红外光谱分析的信息范围与特征。对一般分子，当单位取定后，式（2-3）中的 K 与 μ 分别在 $10^1 \sim 10^2$ 和 $10^0 \sim 10^1$ 的数量级，因此分子振动的基频频率 σ_E 约为 10^3 (cm^{-1}) 数量级，即在中红外区；中红外光谱可以取得分子振动基频的信息。分子振动的倍频与合频的频率比基频频率高一到几个数量级，即达到近红外谱区；近红外光谱可以取得分子振动倍频与合频的信息。理论上近红外谱区分子振动倍频与合频的信息量可能比中红外谱区基频的信息量更高，这是因为同一种分子在近红外光谱的谱峰数目比在中红外光谱的谱峰数目多。但对应分子振动倍频与合频的跃迁几率比基频跃迁几率小；大体上分子振动的倍频每提高一个级次，跃迁几率减少一个数量级，因此实际有分析意义的分子近红外光谱的谱峰受到一定的限制。总体上近红外光谱能用于分析的信息量接近中红外光谱。

因为近红外区从 4000cm^{-1} 开始到 12500cm^{-1} ，基团的基频在 2000cm^{-1} 以下谱区的一级倍频与合频的频率因为小于 4000cm^{-1} ，不属于近红外区，这些基团只有高级次的倍频与合频才能进入近红外区，但强度很弱，因此近红外谱区所包含的信息主要是某些基频在 2000cm^{-1} 以上的基团的信息。因为含氢的基团吸收强度较高，在近红外分析中具有实际意义的主要是以含 H 的基团为主（包括 O—H、N—H、C—H 等）的一些特征基团。

因为对应分子振动倍频与合频跃迁的能量间距比相邻能级的能量间距更为离散，所以分子近红外光谱的谱峰宽度比中红外光谱的谱峰宽，加上分子近红外光谱的谱峰数目比中红外光谱的谱峰多，因此即使是单纯物质的近红外光谱，其谱峰也常有重叠，复杂物质的近红外光谱，其谱峰重叠更为严重。

2.2 近红外光谱的化学信息基础^[4]

有机物在近红外谱区的吸收主要是含氢基团的各级倍频与合频的吸收。文献中常用的“ n 级倍频”，其频率相当于基频的 $(n+1)$ 倍的频率，因此也可称为“ $(n+1)$ 倍频”，例如一级倍频相当于二倍频。本书这两种名称都用。C—H 基的各级倍频在近红外区的分布见表 2-1（N—H 及 O—H 基则相应向短波方向移动）。表 2-1 中列出了各倍频区的波长范围（nm）、相对吸收强度（以基频吸收为“1”）以及作溶液分析时建议使用的光程。

表 2-1 C—H 基团各级倍频在近红外区的分布^[5]

	波长范围/nm	相对吸收强度	建议光程
基频	2900~3200	1	0.1~4mm
合频	2200~2450	0.01	0.1~2cm
一级倍频(二倍频)	1600~1800	0.01	0.1~2cm
二级倍频(三倍频)	1150~1250	0.001	0.5~5cm
三级倍频(四倍频)	850~940	0.0001	5~10cm
四级倍频(五倍频)	700~780	0.0005	10~20cm

主要基团合频与倍频吸收带的近似位置见表 2-2。

表 2-2 主要基团合频与各级倍频吸收带的近似位置

单位	cm ⁻¹				nm			
基团	C—H	N—H	O—H	H ₂ O	C—H	N—H	O—H	H ₂ O
合频	4250	4650	5000	5155	2350	2150	2000	1940
二倍频	5800	6670	7000	6940	1720	1500	1430	1440
三倍频	6500	9520	10500	10420	1180	1050	950	960
四倍频	11100	12500	13500	13300	900	800	740	750
五倍频	13300				750			

由于近红外应用专业文献中常用 nm 为单位，光谱专业的文献常用 cm⁻¹ 为单位，因此本章尽量两种单位对照使用。

主要基团合频与倍频吸收带在电磁波谱中的分布见图 2-1。图 2-1 中以 1300nm 为界，把近红外谱区分为两区：即短波近红外区（SW-近红外）与一般的近红外区（近红外），SW-近红外区主要是高倍频吸收带，后者主要是一级倍（二倍）频吸收带。因为倍频每增加一个级次，吸收强度大约降低一个数量级，因此 SW-近红外可用于长光程的透射分析，近红外区主要用于漫反射分析。

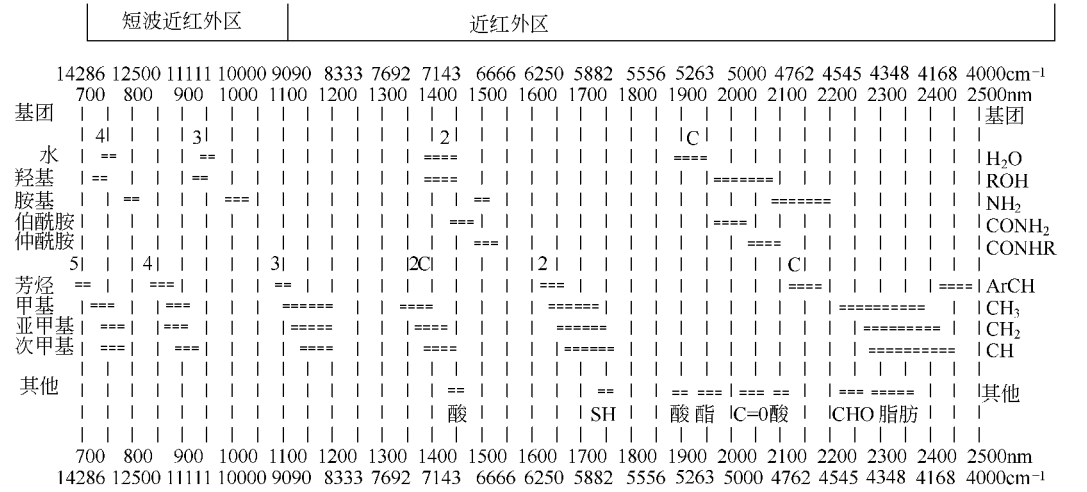


图 2-1 主要有机基团的合频与各级倍频在电磁波谱中的分布

图中：“=”：基团吸收的位置；
“C”：合频；“2C”：合频的倍频；“ArCH”：芳香环上的 CH 基团；
“2”、“3”、“4”、“5”：分别指二、三、四、五倍频，即一级、二级、三级、四级倍频。

2.2.1 C—H 键的近红外光谱

有机物中普遍存在的 C—H 键在近红外区有特征吸收带。C—H 键伸缩振动的基频吸收带在中红外区的 3000cm⁻¹ 附近，其中饱和烃（除环丙烷）的吸收带小于 3000cm⁻¹（大于 3330nm），不饱和烃中的烯烃与芳香烃伸缩振动的基频吸收带略大于 3000cm⁻¹（3330nm），炔烃伸缩振动的基频吸收带约在 3300cm⁻¹ 附近。C—H 键弯曲振动的基频吸

收带在中红外区的指纹区，其中剪式振动在 1450cm^{-1} 。考虑到非谐性常数的影响，由此推测 C—H 键伸缩振动的一级倍频吸收区约在 $1600\sim 1800\text{nm}$ ($6250\sim 5555\text{cm}^{-1}$)，二级倍频吸收带约在 $1100\sim 1200\text{nm}$ ($9090\sim 8333\text{cm}^{-1}$)。伸缩振动和弯曲振动的合频吸收带有两个区，较弱的在 $1300\sim 1400\text{nm}$ ($7691\sim 7143\text{cm}^{-1}$)，较强的在 $2000\sim 2400\text{nm}$ ($5000\sim 4166\text{cm}^{-1}$) (见表 2-2)。主要 C—H 键伸缩振动的基频吸收与倍频吸收带见图 2-1。

烃类中饱和烃（除环丙烷外）的一级倍频吸收带一般大于 1690nm （小于 5917cm^{-1} ），图 2-2 是饱和烃正己烷的近红外光谱。

图 2-2 中 $1700\sim 1800\text{nm}$ ($5882\sim 5555\text{cm}^{-1}$) 处的一组吸收带分别是甲基和亚甲基的不对称与对称伸缩振动的一级倍频吸收，它们的合频吸收带在 $2300\sim 2350\text{nm}$ ($4348\sim 4255\text{cm}^{-1}$) 附近。图中 $1150\sim 1250\text{nm}$ ($8698\sim 8000\text{cm}^{-1}$) 处是 C—H 的二倍频吸收带。

芳香烃和不饱和烃中 C—H 键的近红外吸收波长比饱和烃的稍短，其中一级倍频吸收带一般小于 1690nm 。图 2-3 是苯的近红外光谱。

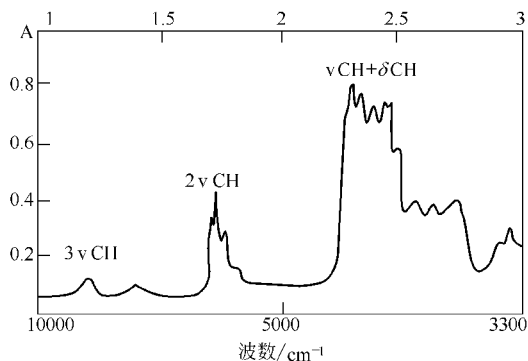


图 2-2 饱和烃正己烷的近红外光谱

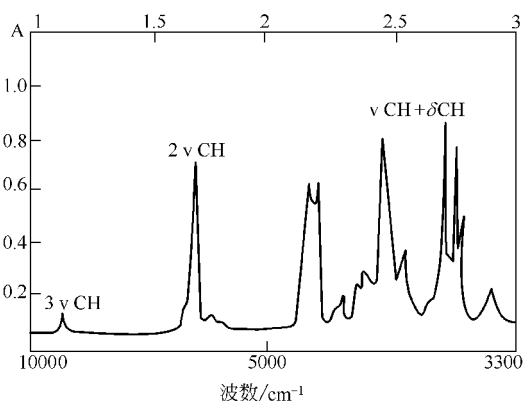


图 2-3 苯的近红外光谱

图 2-3 中 1685nm (5935cm^{-1}) 与 1145nm (8734cm^{-1}) 分别为 C—H 伸缩振动的一级倍频吸收带与二级倍频吸收带^[6]， 2150nm (4651cm^{-1})、 2460nm (4065cm^{-1}) 是合频吸收带。其中二级倍频的吸收带相当稳定，受苯环的影响很少，因此可以用于定量分析，但灵敏度较低^[7]。

近红外谱能把不同的 C—H 基团区别开，图 2-4 中的 (a)、(b)、(c) 分别是苯、异辛烷、正己烷的 SW—近红外谱^[8]。

由图 2-4 可见苯（分子中只有 ArCH）、异辛烷（分子中只有 CH_3 ）与正己烷（1 个分子中有 2 个 CH_3 与 4 个 CH_2 ）的短波近红外谱有明显的差别：由 ArCH、 CH_3 、 CH_2 各基团的吸收波长递减，见表 2-3。

因此利用近红外光谱可以区别有机分子直链与支链的结构，这是用近红外光谱测定汽油的辛烷值、测定直链淀粉与支链淀粉比值的基础，此外近红外光谱还可以测定物质的密度、冰点、闪点、馏分等物理参数，也可以测定芳烃、烯烃等化学参数。

以烃类作为溶剂时，为了避免 C—H 键对分析的干扰，可以对烃类氘化。由于 C—D 键的折合质量大约等于 C—H 键的两倍，而这两种键的力常数 K 值大致相同，由此可算得 C—H 对近红外光吸收的波数大约是 C—D 吸收的 1.4 倍（波长变小）。饱和烃 C—D 一

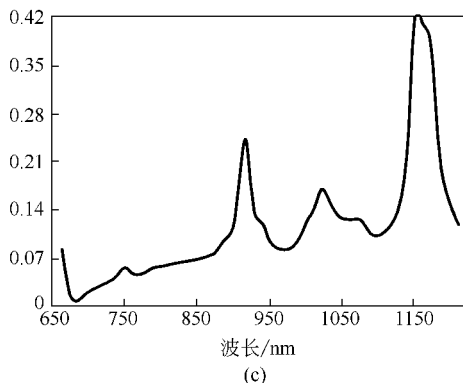
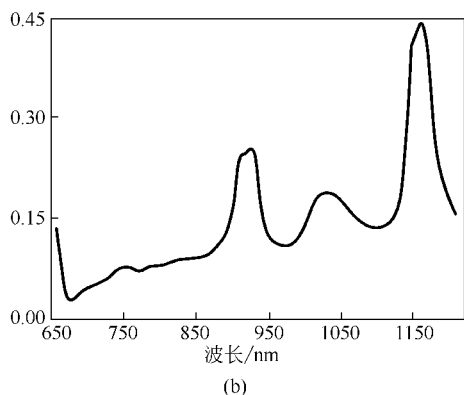
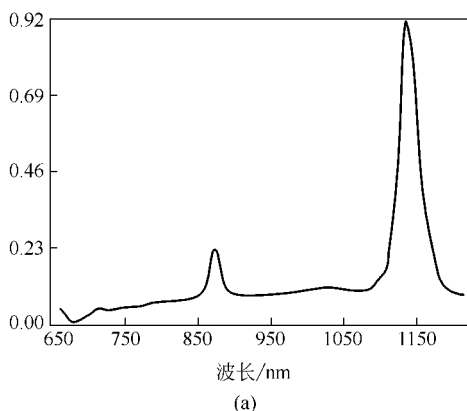


图 2-4 苯 (a)、异辛烷 (b)、正己烷 (c) 的 SW—近红外谱

表 2-3 ArCH、CH₃、CH₂ 在短波近红外谱区的吸收波长 单位：(nm)

基 团	ArCH	CH ₃	CH ₂	基团	ArCH	CH ₃	CH ₂
合频		1015	1053	五倍频	714	746	762
三倍频	1145	1190	1210	六倍频			715
四倍频	875	913	934				

级倍频的吸收带由 C—H 键 1700~1800nm (5882~5555cm⁻¹) 的吸收带移到了 2450nm (4081cm⁻¹) 附近；C—D 二级倍频吸收带则由 C—H 的 1100~1250nm (9090~8000cm⁻¹) 移到 1550nm (6452cm⁻¹) 附近；三级倍频的吸收带相应移到 1150nm (8700cm⁻¹)，其吸收强度很低^[9]。

含氧有机物中的 C—H 与 C=O 伸缩振动的合频吸收带对结构分析很有用。例如醛类在 2210nm (4525cm⁻¹) 与 2250nm (4444cm⁻¹) 的两个吸收带分别是 C—H 的伸缩振动与 C=O 伸缩振动的合频吸收带^[10]。生物体中广泛存在的各种脂肪类分子是高级脂肪酸的甘油酯，其含氢基团只有 C—H 基团。脂肪分子中 C—H 基的伸缩振动和弯曲振动的基频吸收分别在 2900~3000cm⁻¹ 与 1300~1400cm⁻¹，因此近红外谱区的合频吸收带在 4200~4400cm⁻¹。

图 2-5 是几种有机物的近红外光谱图^[2]，其中光谱“2”是以 BaSO₄ 为背景花生油的

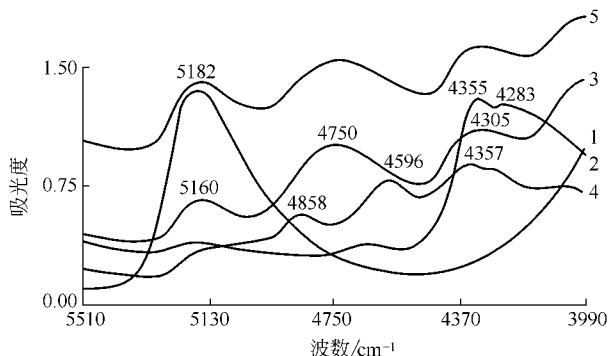


图 2-5 谷物中主要组分的近红外漫反射光谱

近红外 DR (近红外漫反射) 谱, 图 2-5 中 $4355 \sim 4283\text{cm}^{-1}$ ($2296 \sim 2335\text{nm}$) 是脂肪分子中 C—H 键的合频吸收带。

2.2.2 N—H 键的近红外光谱

N—H 键是有机分子中的一种重要结构。许多重要有机物, 以及与生命有关的有机物如胺类、酰胺类、氨基酸及蛋白质分子中都含有 N—H 键。N—H 键的振动形式与次甲基 C—H 及羟基的振动形式类似; NH_2 和 NH_3 的振动形式与亚甲基的振动形式类似。N—H 键的力常数 K 较 C—H 键的大, 它们的质量不同, 在中红外区 N—H 伸缩的波数比 C—H 高, 约在该区的 $3250 \sim 3350\text{cm}^{-1}$, 弯曲振动中的剪式振动约在 1600cm^{-1} 附近, 由此可推得 N—H 伸缩振动的倍频吸收与合频吸收的波长比 C—H 的小, 分别在 2150nm (4650cm^{-1}) 与 1500nm (6666cm^{-1}) 附近。

蛋白质分子是由氨基酸以肽键结合在一起的生物大分子, 分子中的含氢基团有 C—H、N—H 与 O—H。N—H 基团伸缩振动和弯曲振动的合频吸收带在 4600cm^{-1} (2174nm) 左右。图 2-5 中的光谱“4”是 BaSO_4 为背景的血清蛋白的近红外漫反射谱, 图中的 4596cm^{-1} (2176nm) 的吸收峰是蛋白质分子中 N—H 键的合频吸收带。蛋白质及氨基酸在近红外区的吸收, 在农产品分析中很有用处。

伯酰胺 $\text{R}-\text{CO}-\text{NH}_2$ 、仲酰胺 ($\text{R}-\text{CO}-\text{NH}-\text{R}$) 的近红外光谱与对应的胺类物质相似, 其一级倍频吸收带在 1500nm (6666cm^{-1}), 合频吸收带在 2000nm (5000cm^{-1}) 附近, 参见文献^[11]。

2.2.3 O—H 键的近红外光谱

O—H 键是有机分子中另一种重要结构。有机物中的醇、酚、酸及生命活动中起重要作用的各种单糖、多糖都含有 O—H 键基。O—H 键的振动形式与次甲基 C—H 及仲胺基 N—H 相似, 但 C、H、O 三种元素中氧原子的电负性最高, O—H 键的力常数 K 比 N—H 和 C—H 的高, 因此 O—H 键的基频频率也较高。在酚和醇中游离羟基伸缩振动的基频吸收约在中红外的 3650cm^{-1} , 面内弯曲振动约在 1300cm^{-1} , 由此可推得在近红外区伸缩振动一级倍频吸收约在 1430nm (6993cm^{-1}) 附近, 二级倍频吸收约在 950nm (10530cm^{-1}) 附近, 合频吸收约在 2000nm (5000cm^{-1}) 附近。图 2-6 是乙醇的近红外光谱图, 图中可看到上述各种吸收带。

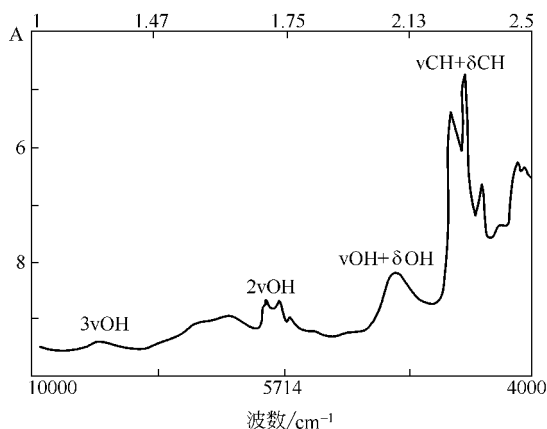


图 2-6 乙醇的近红外光谱图

伯醇、仲醇和叔醇中的游离 OH 基伸缩振动的基频吸收约在中红外区的 3640cm^{-1} 、 3630cm^{-1} 、 3610cm^{-1} （即 2247nm 、 2755nm 、 2770nm ）。与此相对应的它们的倍频吸收分别为在近红外区的 $1408\sim 1409\text{nm}$ 、 $1412\sim 1416\text{nm}$ 、 $1418\sim 1419\text{nm}$ ^[12]。

OH 基团有很强的极性，因此含 OH 基的化合物可通过氢键发生各种形式的缔合，其中多个分子之间的缔合可使吸收峰向长波长（短波数）方向偏移 $5\%\sim 10\%$ ，而分子内部 OH 基与其他基团如 C=O 等基团缔合而形成的吸收峰

有更大的偏移。分子与溶剂分子形成的氢键也影响吸收峰的位置，由此对近红外光谱的影响见本章参考文献中^[13]。因此醇类有机物游离 OH 基的近红外光谱应采用稀溶液测定。

淀粉、纤维素等生物大分子属于多糖，含有大量羟基，分子中含氢基团有 O—H、C—H 键，其特征吸收为 O—H 的伸缩振动和弯曲振动的合频吸收带，图 2-5 中的光谱“3”是以 BaSO_4 为背景，淀粉的近红外漫反射谱，图中的 5169cm^{-1} 吸收峰是淀粉分子中 O—H 基的合频吸收带。

含 OH 基的物质中水是最普遍的一种，水分子的极性强，其振动在中红外区有很强的吸收，对中红外光谱分析的影响极大。在近红外区水分子的吸收比中红外区弱得多，该谱区可以测定水溶液样品。水在近红外区内有一些特征性很强的合频吸收带（其他各种分子的倍频与合频吸收相对的较弱），这些都使近红外光谱研究水分子的结构和测定物质中水分的含量较为方便。

液态水由于分子的缔合，吸收带的细节消失，形成较宽的谱带。纯水分子 OH 及伸缩振动的一级倍频约在近红外区的 1440nm (6944cm^{-1})，二级倍频在 960nm (10420cm^{-1})，其合频吸收带主要有两个，较强的在 1940nm (5155cm^{-1})，较弱的 1220nm (8197cm^{-1})。图 2-5 中的光谱“1”是水的近红外 DRS 谱，光谱中 5182cm^{-1} (1930nm) 是水分子中 O—H 基的合频吸收带。

除醇类和水外，含 OH 基的有机物还有羧酸和过氧化物等，它们的近红外光谱在分析中的作用不如水和醇。羧酸中非缔合的 OH 基伸缩振动的倍频吸收在近红外区的 1450nm (6897cm^{-1})，伸缩振动和弯曲振动的合频吸收带在 2100nm (4762cm^{-1}) 左右。羧酸中羰基的一级倍频吸收带在中红外区，二级倍频吸收在 1900nm (5263cm^{-1})。若羧酸分子发生缔合会提高 OH 振动吸收波长。因为羧酸分子内和分子之间键合对其近红外吸收的强度和位置有很大的影响，所以羧酸的 OH 吸收带在分析上的作用不大。

过氧化物的稀溶液中的 OH 基的倍频与合频的吸收带分别在 1450nm (6897cm^{-1}) 和 2070nm (4831cm^{-1})。由于过氧化物在中红外的振动光谱与醇类和酚类已经有明显的差别，这就使强度弱得多的近红外光谱研究过氧化物的价值大为降低。

2.2.4 其他物质在近红外区的吸收

其他物质在近红外区的吸收可参看 Sadtler Research Laboratories 在 1981 年发表的 1000 种有机物的近红外图谱集^[14]。近红外区与中红外区高聚物分析的比较研究参看本章参考文献^[15]。

近红外谱区常用于分析含 C—H、N—H、O—H 键的有机物，所以分析时所用的溶剂分子中不宜带上述基团。因此，近红外区分析选用溶剂比紫外区困难得多。在本区内透明的溶剂，只有 CCl₄ 和 CS₂。图 2-7 是近红外光在一些溶剂内的透过状况参看本章参考文献^[16]。

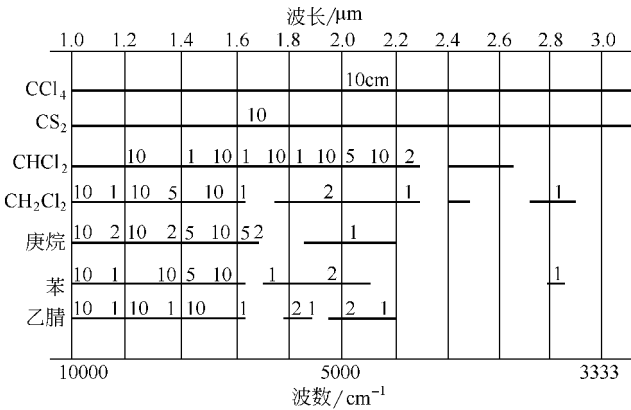


图 2-7 近红外谱区常用溶剂的光谱特性

图 2-7 中的数值是指在所示谱区用 CCl₄ 和 CS₂ 作为溶剂时，样品池最大可用的光程 (cm) 值。如果必须使用含 C—H、N—H、O—H 基团的溶剂时，则溶剂分子中这些基团内的 H 原子应用氘代替成 (C—D、N—D、O—D)。C—D、N—D、O—D 的近红外吸收带比相应的 C—H、N—H、O—H 吸收带有较大的偏移，可避免对样品测定的影响。

因为不含 C—H、N—H、O—H 的化合物在近红外区无吸收，所以干扰该谱区分析的杂质相对少些，有利于复杂物的分析。近红外谱区吸收带的强度较弱，一方面影响了这一谱区分析的检测限，另一方面又便于对某些高聚物和含水样品的分析。

2.2.5 近红外光谱区的信息分布^[17]

分子中原子间振动的合频与各级倍频大多在近红外谱区，因此本谱区包含的信息极为丰富，大体与中红外谱区相近。前已论及：在近红外光谱分析具有实际意义的主要以含 H 的基团为主 (包括 O—H、N—H、C—H 等)，也有其他一些基团 (如 C=O 等) 的信息，但强度较弱 (见图 2-1)。这些基团是有机物中的一些重要基团。近红外技术可以分析测定与这些基团有关的成分以及物理、化学性质。近红外谱区不但可以分析测定与这些基团有关的成分如蛋白质、淀粉、脂肪、氨基酸等，还可以分析物质的密度、黏度、颗粒的大小以及有关样品的电学、热学、力学性质等等。近红外谱区丰富的结构与组成信息，加上近红外谱区信号的特征，这是近红外分析应用广泛的基础。

图 2-8 与表 2-4 中列出了可在近红外区进行分析的一些常见基团在该谱区的谱峰位置，并列出了产生这些谱峰的结构。

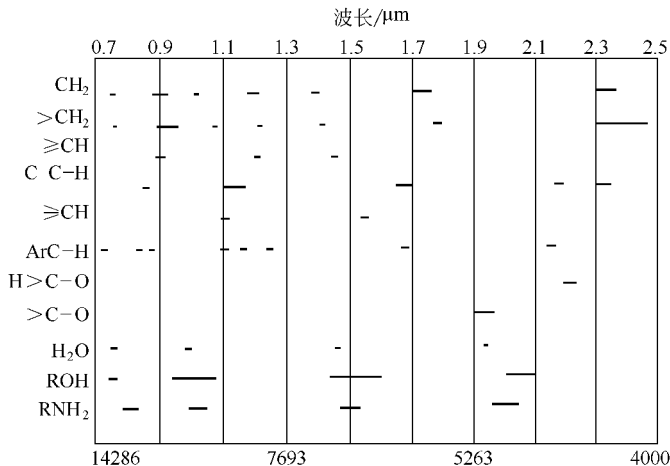


图 2-8 重要有机基团在近红外区的吸收带

表 2-4 长波近红外谱区内主要吸收谱带的来源^[16,18]

波长/nm (cm^{-1})	振动方式	结构	波长/nm (cm^{-1})	振动方式	结构
713(14030)	5×C—H str.	苯	1080(9259)	2×C—H str. + 2×C—C str.	苯
738(13550)	4×O—H str.	CH ₃	1097(9116)	2×C—H str. + 2×C—C def.	环丙烷
740(13510)	5×C—H str.	CH ₂	1143(8749)	3×C—H str.	芳烃
746(13410)	5×C—H str.	CH ₂	1152(8681)	3×C—H str.	CH ₃
747(13390)	4×O—H str.	ArOH	1170(8547)	3×C—H str.	HC=CH
760(13160)	4×O—H str.	H ₂ O	1195(8368)	3×C—H str.	CH ₃
762(13120)	5×C—H str.	CH ₂	1215(8231)	3×C—H str.	CH ₂
779(12840)	4×N—H str.	RNH ₂	1225(8163)	3×C—H str.	CH
790(12660)	4×N—H str.	ArNH ₂	1360(7353)	2×C—H str. + C—H def.	CH ₃
806(12407)	4×N—H str.	RNH ₂	1395(7168)	2×C—H str. + C—H def.	CH ₂
808(12376)	2×N—H str. + 2×N—H def...	RNHR'	1410(7092)	2×O—H str.	ROH
815(12270)	4×N—H str.	CONH ₂	1415(7067)	2×C—H str. + C—H def.	CH ₂
832(12020)	2×N—H str. + 2×N—H def...	RNHR'	1417(7057)	2×C—H str. + C—H def.	苯环
840(11905)	3×C—H str. + 2×C—C str.	苯	1420(7042)	2×O—H str.	ArOH
874(11442)	4×C—H str.	苯	1430(6993)	2×O—H str.	CONH ₂
880(11364)	4×C—H str.	CHCl ₃	1440(6944)	2×O—H str.	糖类
900(11111)	4×C—H str.	CH ₃	1440(6944)	2×C—H str. + C—H def.	CH
910(10990)	4×C—H str.	蛋白质	1446(6916)	2×C—H str. + C—H def.	苯环
913(10953)	4×C—H str.	CH ₃	1450(6897)	2×O—H str.	水, 糖类
928(10776)	4×C—H str.	油	1460(6849)	2×N—H str.	CONH ₂
938(10662)	4×C—H str.	CH ₂	1471(6798)	2×N—H str.	CONHR
970(10310)	3×O—H str.	ROH, H ₂ O	1480(6757)	2×O—H str.	葡萄糖
990(10101)	3×O—H str.	淀粉	1483(6743)	2×N—H str.	CONH ₂
1000(10000)	3×O—H str.	ArOH	1490(6711)	2×N—H str.	CONHR
1015(9852)	2×C—H str. + 3×C—H def.	CH ₃	1490(6711)	2×N—H str. (内氢键)	CONH ₂
1020(9804)	2×N—H str. + 2×amide	蛋白质	1490(6711)	2×O—H str.	纤维素
1020(9804)	3×N—H str.	ArNH ₂	1492(6702)	2×N—H str.	ArNH ₂
1030(9709)	3×N—H str.	RNH ₂	1500(6667)	2×N—H str.	NH
1037(9643)	2×C—H str. + 2×C—H def.	油	1510(6623)	2×N—H str.	蛋白质
1053(9497)	2×C—H str. + 2×C—H def.	CH ₂	1520(6579)	2×O—H str.	ROH
1060(9434)	3×N—H str.	RNH ₂	1520(6579)	2×N—H str.	CONH ₂

续表					
波长/nm (cm^{-1})	振动方式	结构	波长/nm (cm^{-1})	振动方式	结构
1528(6545)	$2\times\text{O—H str.}$ (内氢键)	淀粉	2030(4926)	$3\times\text{C=O str.}$	CONH_2
1530(6536)	$2\times\text{N—H str.}$	RNH_2	2050(4878)	$\text{N—H sym. str.} + \text{amide II}$	蛋白质
1533(6523)	$2\times\text{C—H str.}$	$=\text{CH}$	2050(4878)	$\text{N—H asym. str.} + \text{amide II}$	CONH_2
1540(6494)	$2\times\text{O—H str.}$ (内氢键)	淀粉	2080(4808)	$\text{O—H str.} + \text{O—H def.}$	ROH
1570(6370)	$2\times\text{N—H str.}$	CONH	2100(4762)	$2\times\text{O—H def.} + 2\times\text{C—O str.}$	淀粉
1580(6329)	$2\times\text{O—H str.}$ (内氢键)	淀粉	2110(4739)	$\text{N—H sym. str.} + \text{amide III}$	$\text{CONH}_2, \text{CONHR}$
1620(6173)	$2\times\text{C—H str.}$	$=\text{CH}_2$	2132(4690)	$\text{N—H str.} + \text{C=O str.}$	氨基酸
1645(6079)	$2\times\text{C—H str.}$	R—CH—CH	2140(4673)	$=\text{C—H str.} + \text{C=C str.}$	HC=CH
1660(6024)	$2\times\text{C—H str.}$	cis—RCH=CHR	2150(4651)	$2\times\text{amide I} + \text{amide III}$	CONH_2
1685(5935)	$2\times\text{C—H str.}$	Ar.	2160(4630)	$2\times\text{amide I} + \text{amide III}$	CONHR
1695(5900)	$2\times\text{C—H str.}$	CH_3	2180(4587)	$2\times\text{amide I} + \text{amide III}$	蛋白质
1705(5865)	$2\times\text{C—H str.}$	CH_3	2190(4566)	$\text{CH}_2 \text{ asym. str.} + \text{C= str.}$	HC=CH
1725(5797)	$2\times\text{C—H str.}$	CH_2	2200(4545)	$\text{C—H str.} + \text{C=O str.}$	—CHO
1740(5747)	$2\times\text{S—H str.}$	—SH	2242(4460)	$\text{N—H str.} + \text{NH}_3 \text{ def.}$	氨基酸
1765(5666)	$2\times\text{C—H str.}$	CH_2	2252(4440)	$\text{O—H str.} + \text{O—H def.}$	淀粉
1780(5618)	$2\times\text{C—H str.}$	纤维素	2276(4394)	$\text{O—H str.} + \text{C—C str.}$	淀粉
1820(5495)	$\text{O—H str.} + 2\times\text{C—O str.}$	纤维素	2280(4386)	$\text{C—H str.} + \text{C—H def.}$	CH_3
1900(5263)	$\text{O—H str.} + 2\times\text{C—O str.}$	淀粉	2294(4359)	$\text{N—H str.} + \text{C=O str}$	氨基酸
1900(5263)	$3\times\text{C=O str.}$	—COOH	2310(4329)	$\text{C—H str.} + \text{C—H def.}$	CH_2
1908(5241)	$2\times\text{O—H str.}$	ROH	2323(4305)	$\text{C—H str.} + \text{C—H def.}$	CH_2
1920(5208)	$2\times\text{C=O str.}$	CONH	2336(4281)	$\text{C—H str.} + \text{C—H def.}$	纤维素
1940(5155)	$\text{O—H str.} + \text{O—H def.}$	H_2O	2347(4261)	$\text{CH}_2 \text{ sym. str.} + =\text{CH}_2 \text{ def.}$	HC=CHCH_2
1950(5128)	$3\times\text{C=O str.}$	—COOR	2352(4252)	$3\times\text{C—H def.}$	纤维素
1960(5102)	$\text{N—H asym. str.} + \text{amide II}$	CONH	2380(4202)	$3\times\text{O—H def.}$	ROH
1980(5051)	$\text{N—H asym. str.} + \text{amide II}$	蛋白质	2461(4063)	$\text{C—H str.} + \text{C—C str.}$	淀粉
2000(5000)	$2\times\text{O—H def.} + \text{C—O def.}$	淀粉	2488(4019)	$\text{C—H str.} + \text{C—C str.}$	淀粉
2000(5000)	$\text{N—H sym. str.} + \text{amide II}$	$\text{CONH}_2, \text{CONHR}$	2500(4000)	$\text{C—H str.} + \text{C—C str.}$	淀粉

表中符号说明：

“ $2\times$ ”，“ $3\times$ ”：分别为一级倍频和二级倍频（振动频率为基频频率的 2 或 3 倍）；

“str.”，“def.”：分别为伸缩振动和变形振动；

“sym”，“asym”：分别为对称振动和不对称振动；

“Ar”：为苯环；“cis”：为顺式结构；“...”为省略的结构；

“amide I（Ⅱ，Ⅲ）”：分别为酰胺分子中羰基与胺基不同的耦合方式。

2.3 近红外光谱分析信息提取的难点^[19]

近红外光谱分析信息源的特点带来了该技术的应用广泛、操作简便，另一方面也决定了该技术具有一系列难点。

2.3.1 近红外光谱的复杂性和重叠性

由于近红外光谱测定的样品经常是不经过提纯等预处理的复杂样品，在这些样品中待测成分一起，同时存在高强度的复杂背景，图 2-5 中光谱 5 是谷子粉样的近红外光谱，图中 1、2、3、4 分别是谷子的组成成分——水、油、淀粉与蛋白质的近红外光谱。测定谷子中的蛋白质受到背景包括水、油、淀粉等组分的影响，近红外分析难以用常规方法解

析图谱必须解决图谱背景复杂影响分析的困难，这是复杂物近红外分析的第一个难点。

为了消除这些背景的影响，需要取得这些背景的信息，因此解决图谱背景复杂的困难需要收集各种类型的样品，选取代表性样品来取得背景的信息；近红外光谱分析建立数学模型所用校正样品集应当有代表性，即应包含待测样品的各种背景，由这样的标准样品所建立的数学模型能够校正背景对测定的影响。因此收集样品资源是近红外光谱分析的一个基础。

近红外光谱的谱峰宽度比中红外谱区宽得多，物质分子在近红外谱区吸收峰的分布范围也较宽，同一种基团可以在本谱区的二、三、四……不同倍频区出现吸收峰（见图 2-1）；由于信息的分布范围广，而谱峰的宽度较宽，因此复杂物同一近红外谱区内常有不同分子、多种基团的谱峰重叠在一起。如图 2-5 谷子粉样的近红外光谱中各种有机成分的谱峰重叠严重，以致不可能用常规的方法对复杂样品谱峰重叠的光谱（如图中的谷子粉样光谱）进行直接谱峰解析。多组分样品的近红外光谱在每一个波长位通常有多个谱峰的重叠，这是复杂物近红外分析的第二个难点。

对应某一个波长点虽然可能有多种组分光谱的重叠，难以区分不同组分；但任意两种组分在许多波长点或者在近红外的全谱区内一般却不可能完全相同，可以用多波长点的特征来区分不同组分。解决近红外光谱分析谱峰重叠的困难是采用多元分析的方法，在全谱区内或某些特定波长点的组合来提取需要的信息。

近红外光谱分析是复杂、重叠光谱的分析，在进行多波长校正时经常需要应用光谱中“肩部”的信息。因此，近红外分析对仪器的波长和吸光度的准确度有特殊的要求。Stark 指出近红外光谱仪的波长重复性误差应小于 0.02nm ，吸光度测量精度应达 10^{-5} 吸光度单位^[16]；现代近红外仪器通常具有波长自动校正和用标准样品校正吸光度的功能。

2.3.2 近红外光谱的变动性

近红外分析用于不经前处理、非破坏性整体测定样品的技术也带来分析上的一系列其他困难：由于样品未经预处理，因此样品的状态、测定的方式以及测定的条件都会影响测量的结果。特别是在近红外漫反射光谱分析时，因为吸光度 A 受样品散射系数 S 的影响，因此在漫反射测量过程中，所有影响散射系数 S 的因素都影响近红外谱测定的稳定性以及定量分析的结果，这就是近红外漫反射光谱测定的不稳定性。漫反射光谱测定的不稳定性造成近红外光谱的失真，主要分成两类失真：系统误差（光谱的平移、线性误差等）与随机误差（噪声、散射误差等）。近红外光谱经常出现这样的情况：同一样品不同时间下测定的图谱有明显的差别，甚至表现的差别比不同样品在同一条件下测定的差别还要大，见图 2-9。

图 2-9 中光谱 1 与 2 分别是两种蛋白质含量为 10.7% 与 15.8% 的谷子粉样在同一日期测定的光谱；光谱 1 和 3 是同一样品（蛋白质含量为 10.7%）在不同日期测定的光谱。由图可见在不同日期测定的同一样品，由于不可避免的某些测定条件的变化，造成的图谱改变（平移等），比蛋白质含量不同的样品在同一日期测定的光谱差异更加明显。近红外漫反射光谱测定的不稳定性主要是测定条件统计性随机波动造成的，可以称为统计波动性（见 7.4.2，简称近红外光谱的不稳定性或波动性），近红外光谱的波动性是近红外分析的第三个难点。

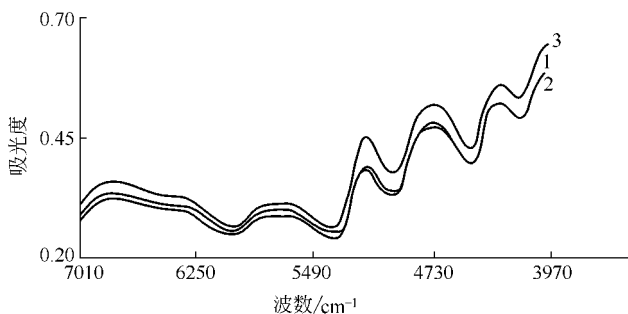


图 2-9 近红外漫反射光谱测定的不稳定性

为了降低近红外图谱测定不稳定对分析的影响，在硬件方面可以运用旋转的样品池、大面积移动式样品池以及积分球等特殊的样品池进行测量，以减少样品的统计波动性对测量的影响；在软件技术方面还可以通过一些校正系统误差的信息处理技术如矢量归一化、多元散射校正以及导数光谱等，以校正样品光谱的失真；并建立规范的测试过程以减少光谱的不稳定。

2.3.3 近红外光谱包含的信息强度低

有机物近红外谱区各级倍频与合频的吸收强度大体上依次比基频或上一级倍频低一个数量级（见图 2-10）。

图 2-10 中 $\times$ 表示倍频， $\times$ 前面的数字表示倍频的倍次；C 表示合频。图中 A、B、C 这三个谱区是用三种不同的光程——10cm、2cm、1mm 测得的。由图可见 C—H 基团的

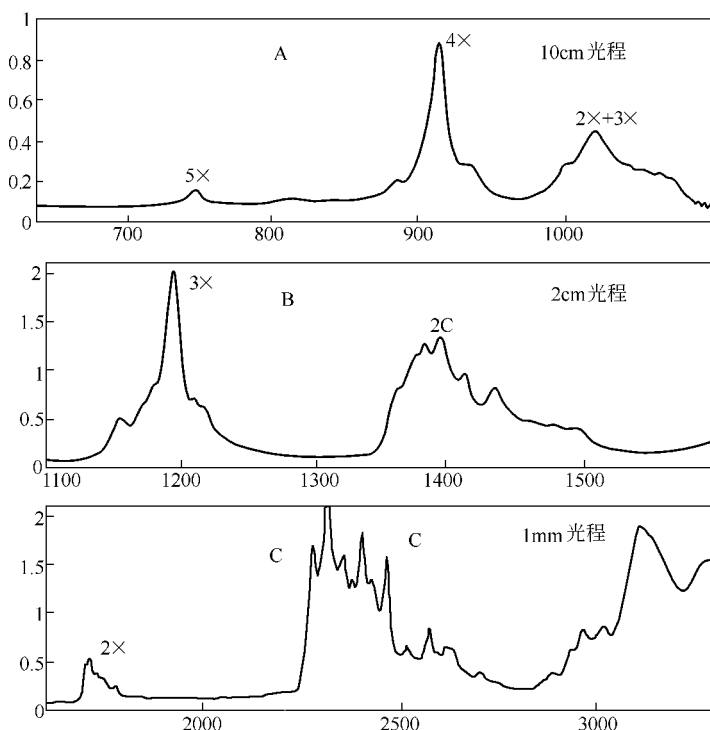


图 2-10 烃类在近红外区的吸收光谱

吸收由合频到各级倍频的吸收强度越来越低, 倍频大约每增高一级, 相对峰高约降低一个数量级。因此, 短波近红外区 ($0.8\sim1.1\mu\text{m}$) 适合做长光程的透射光谱分析, 长波近红外区 ($1.1\sim2.5\mu\text{m}$) 适合做短光程的透射光谱或漫反射光谱分析。

由于有机物在本谱区吸收弱, 近红外光可以深入样品内部, 取得深层样品的物质信息, 因此近红外分析可以在分析中用大光程的样品池或不经预处理或稀释的样品; 还可以运用光导纤维、积分球以及其他测量简便的各种漫反射测定方法。运用这些方法使近红外分析的样品可以不需任何物理、化学制备与其他预处理、后处理。一旦近红外分析的数学模型建立后, 对进行分析的操作人员知识背景与经验背景的要求可以大幅度降低。近红外技术可以直接分析含水样品、水溶液、黏稠样品、固体样品, 甚至玻璃瓶内的样品与生物活体组织、完整的种子等复杂样品, 加上后面要提到的近红外信号的一些特点, 这是近红外谱区分析“操作简单”的原因。

另一方面, 由于近红外光谱中信息强度低, 因此组分在近红外光谱中谱峰强度较弱, 近红外谱区吸收强度低的特点影响近红外测量的检测限 (约比中红外区的检测限差一至几个数量级), 这是近红外分析的第四个难点。

为了将近红外谱区用于分析, 必须从硬件与软件两方面解决近红外分析吸收强度低这个难点。在硬件方面需要提高仪器的信噪比, Stark 指出近红外仪器的信噪比应达 10^5 ^[16]。现代近红外仪器运用各种校正随机误差、降低噪声的方法与技术, 可以将信噪比提高一个数量级左右, 因此对仪器硬件的信噪比要求实际可以有所降低。

概括说, 近红外光谱分析的难点是从复杂、重叠、变动的背景中提取弱信息。为了解决这个分析的难点, 在硬件上近红外仪器应有较高的波长准确度和吸光度准确度。在软件方面近红外分析必须依靠现代化学计量学的算法与计算机技术, 利用标准样品近红外光谱的全谱区信息, 即运用每个样品光谱全部数据点组成的光谱数据点群体, 建立样品光谱和待测量间的数学关系, 即所谓的数学模型, 运用此数学模型对样品的近红外光谱进行运算, 算出样品的待测量。

近红外光谱分析技术解决这些困难的方法与策略详见第 7 章。

参 考 文 献

- [1] Douglas A Skoog. Principles of Instrumental Analysis. CBS College Publishing, 1986
- [2] 严衍禄主编. 现代仪器分析. 北京: 北京农业大学出版社, 1995
- [3] 吴瑾光主编. 近代傅里叶变换红外光谱技术及应用. 北京: 科学技术文献出版社, 1994
- [4] Weler J G. Appl Spectroscopy Riv, 1985, 21 (1, 2): 1~43
- [5] Jerome J Workman, JR. Appl Spectroscopy Riv, 1996, 31 (3): 251~320
- [6] Surhman R, Klein P. Z Phys Chem. Liepzig, 1941, B50: 23~73
- [7] Lemer H J, Schmidt J. Chem Technol, 1973, 25 (2): 99~100
- [8] Jeffrey J Keiiy. Anal Chem, 1989, 61 (1): 313~320
- [9] Powell D L, Darland E J, T. R. Williams, et al. 1971, 4 (7): 479~483
- [10] King J F, Vig B. Can J Chem, 1962 (40): 1023~1025
- [11] Krikorian S E, Mahpour M. J Pharm Sci, 1973, 29A: 1233, 1246
- [12] OCriser R, Burrill A M. Anal Chem. 1959, 31 (12): 2055~2057
- [13] Trott G F, Woodside E E, Taylor K G. Carbohydr Res, 1973 (27): 415~435

- [14] The Atlas of Near Infrared Spectra. Research Laboratories, Philadelphia, 1981
- [15] Kathryn A, Bunding Lee. Appl Spectroscopy Riv, 1993, 28 (3): 231~284
- [16] Stark E, Luchter K. Appl Spectroscopy Riv, 1986, 22 (4): 335~399
- [17] **林水, 吴平等主编. 实用傅里叶变换红外光谱学. 北京: 中国环境科学出版社, 1991. 318~319**
- [18] Osborne B G, Fearn T. Near-Infrared Spectroscopy in Food Analysis. Longman Scientific and Technical. New York, 1986
- [19] **严衍禄, 赵龙莲等. 现代近红外光谱分析的信息处理技术. 光谱学与光谱分析, 2000, 20 (6): 777~780**
- [20] K A Martin. Appl Spectroscopy Riv, 1992, 27 (4): 325~383
- [21] **张录达, 严衍禄等. 北京农业大学学报, 1990, 16 (增): 18~25**
- [22] Shenk J S, M O Westerhaus. Crop Sci, 1991 (31): 469~474
- [23] Jerome J Workman, JR. Appl Spectroscopy Riv, 1996, 31 (1, 2): 73~124
- [24] Paul R Mobley, Bruce R Kowalski, Appl Spectroscopy Riv, 1996, 31 (4): 334~368
- [25] **严衍禄, 高文淑等. 北京农业大学学报, 1990, 16 (增): 26~36**

3. 近红外光谱分析技术的数学基础

3.1 近红外光谱分析的统计学基础

近红外光谱（NIR）分析属于分析化学的范畴。化学计量学是数学、统计学在分析化学中的应用，本节介绍 NIR 分析的化学计量基础方法。

3.1.1 分布函数

分析测量的过程中，由于种种原因，测量结果总有一定变异性，即测量结果是一个随机变量 X ，它可用如下分布函数 $F(x)$ 来描述。

$$F(x) = P(X \leq x) \tag{3-1}$$

分布函数表达随机变量 X 取值不大于实数 x 的概率，图 3-1 是一个随机变量 X 的分布函数的一般模型。

分布函数具有性质

$$0 \leq F(x) \leq 1 \tag{3-2}$$

如果 $F(x)$ 为一个可微分函数，则

$$f(x) = \frac{dF(x)}{dx} = F'(x) \tag{3-3}$$

为随机变量 X 的概率密度。当随机变量 X 具有概率密度 $f(x)$ 时

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \tag{3-4}$$

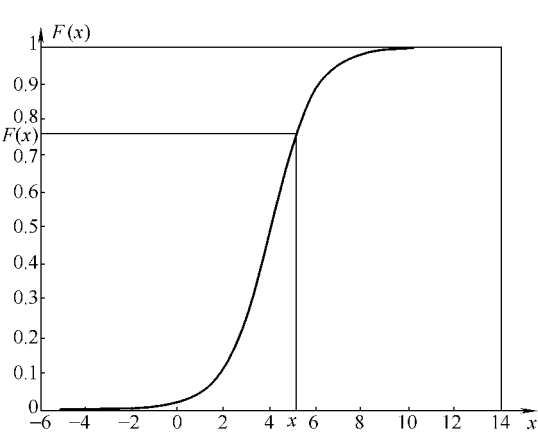


图 3-1 随机变量的分布函数模型

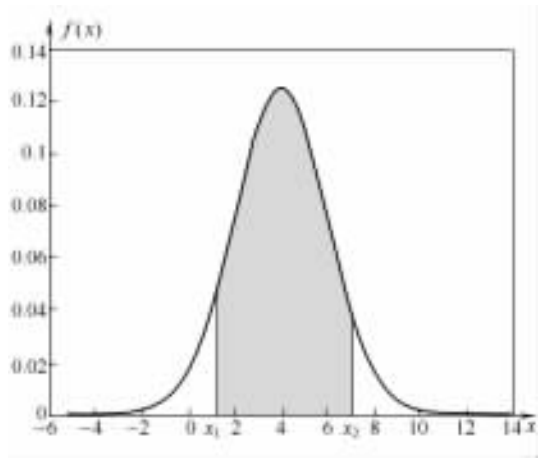


图 3-2 随机变量的概率密度模型

且随机变量 X 落在 (x_1, x_2) 之间的概率为

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x) dx \tag{3-5}$$

如图 3-2 阴影部分所示。

随机变量 X 的两个重要的数字特征数学期望（均值） μ 和方差 σ^2 如下定义：

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx = E(X) \tag{3-6}$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = D(X) \tag{3-7}$$

3.1.2 正态随机变量及其分布

正态分布随机变量是分析测量中基本的随机变量，正态随机变量 X 的概率密度为

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < +\infty \tag{3-8}$$

如图 3-3 所示。

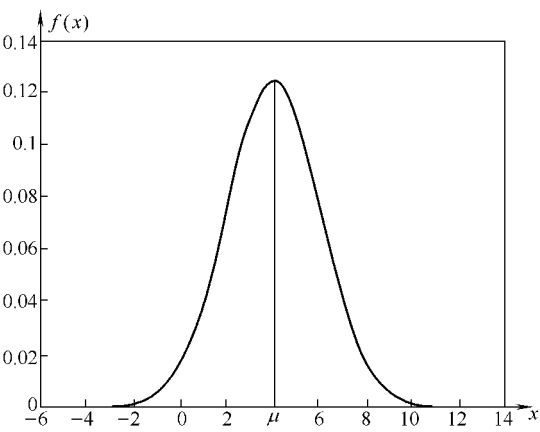


图 3-3 正态分布随机变量的概率密度

如果随机变量 X 为正态随机变量，记为 $X \sim N(\mu, \sigma^2)$ 。

正态随机变量的密度函数 $f(x, \mu, \sigma^2)$ 具有性质：

- ① 在 $x = \mu$ 处有最大值， μ 的变化决定了均值的位置；
- ② σ^2 的变化决定了峰的宽窄。

如果随机变量 $X \sim N(X, \sigma^2)$ ，定义 X 的标准化随机变量为

$$Z = \frac{X - \mu}{\sigma} \tag{3-9}$$

该随机变量有， $E(Z) = 0$ ， $D(Z) = 1$ ，即 $Z \sim N(0, 1)$ 。 Z 有密度函数为

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < +\infty \tag{3-10}$$

相应的分布函数为

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt \quad -\infty < z < +\infty \quad (3-11)$$

对给定的 z ，概率

$$\Phi(z) = P(Z \leq z) \quad (3-12)$$

如图 3-4 所示。

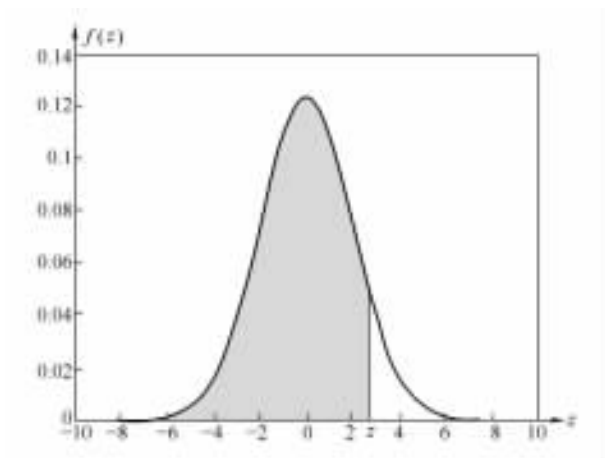


图 3-4 标准正态随机变量的概率密度

在概率论中已经证明（中心极限定理）：若随机变量是大量独立随机变量之和，而每一个随机变量对此和所起的作用都很小，则总和随机变量服从正态分布或近似服从正态分布。

在分析测量中，如果无过失误差出现，则测量结果将服从正态分布或近似服从正态分布。

3.1.3 其他分布

下面概述在近红外光谱分析中经常使用的其他分布函数。

(1) χ^2 分布

如果随机变量 X 的概率密度为

$$f(x, n) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad 0 \leq x < +\infty \quad (3-13)$$

则称 X 服从自由度为 n 的 χ^2 分布，记为 $X \sim \chi^2(n)$ ， X 的分布函数为

$$F(x, n) = \int_{-\infty}^x f(y, n) dy \quad x \geq 0 \quad (3-14)$$

(2) t 分布

如果随机变量 X 的概率密度为

$$f(t, n) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad -\infty < t < +\infty \quad (3-15)$$

则称 X 服从自由度为 n 的 t 分布，记为 $X \sim t(n)$ ， X 的分布函数为

$$F(x, n) = \int_{-\infty}^x f(t, n) dt \quad (3-16)$$

可以证明, 当 $n \rightarrow \infty$ 时, t 分布趋向于正态分布 $N(0, 1)$ 。

(3) F 分布

如果随机变量 X 的概率密度为

$$f(x, n_1, n_2) = \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} \quad 0 < x < +\infty \quad (3-17)$$

则称 X 服从自由度为 n_1, n_2 的 F 分布, 记为 $X \sim F(n_1, n_2)$, F 分布随机变量 X 的分布函数为

$$F(x, n_1, n_2) = \int_0^x f(y, n_1, n_2) dy \quad 0 < x < +\infty \quad (3-18)$$

3.1.4 抽样分布及应用

设 $X_1, X_2, \dots, X_n$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的一个容量为 n 的样本, 如果记统计量

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3-19)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3-20)$$

分别称为样本均值与样本方差。则有如下结论:

- ① $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;
- ② $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;
- ③ $\bar{x}$ 与 S^2 相互独立;
- ④ $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t(n-1)$ 。

如果, $X_1, X_2, \dots, X_{n_1}$ 与 $Y_1, Y_2, \dots, Y_{n_2}$ 分别为取自于总体 $N(\mu_1, \sigma^2)$ 与总体 $N(\mu^2, \sigma^2)$ 的两个独立样本, 记 $\bar{x}, \bar{y}, S_1^2, S_2^2$ 分别为两样本的样本均值与样本方差, 则随机变量

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \quad (3-21)$$

其中

$$S_w = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$$

关于以上各抽样分布在参数估计与假设检验中的各种应用如表 3-1, 表 3-2 所示。

表 3-1		正态总体参数 μ, σ^2 的区间估计		
参数	条件	选用随机变量	置信度	置 信 区 间
μ	σ^2 已知	$u = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$	$1 - \alpha$	$(\bar{x} - u_{\alpha/2} \sigma/\sqrt{n}, \bar{x} + u_{\alpha/2} \sigma/\sqrt{n})$
	σ^2 未知	$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t(n-1)$	$1 - \alpha$	$(\bar{x} - t_{\alpha/2}(n-1) S/\sqrt{n}, \bar{x} + t_{\alpha/2}(n-1) S/\sqrt{n})$
σ		$x^2 = \frac{(n-1)S^2}{\sigma^2} \sim x^2(n-1)$	$1 - \alpha$	$\left(\frac{(n-1)S^2}{x_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{x_{1-\alpha/2}^2(n-1)} \right)$

表 3-2 正态总体参数的假设检验法（显著水平 α ）

一 个 总 体	原假设	检验统计量	H_0 为真时统计量的分布	拒绝域
	$\mu = \mu_0$ (σ^2 已知)	$u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$N(0, 1)$	$ u > u_{\alpha/2}$
	$\mu = \mu_0$ (σ^2 未知)	$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$	$t(n-1)$	$ t > t_{\alpha/2}(n-1)$
	$\sigma^2 = \sigma_0^2$	$x^2 = \frac{(n-1)S^2}{\sigma^2}$	$x^2(n-1)$	$x^2 > x_{\alpha/2}^2(n-1)$ 或 $x^2 < x_{1-\alpha/2}^2(n-1)$
两 个 总 体	$\mu_1 = \mu_2$ (σ_1^2, σ_2^2 已知)	$u = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0, 1)$	$ u > u_{\alpha/2}$
	$\mu_1 = \mu_2$ (σ_1^2, σ_2^2 未知)	$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}}$	$t(n_1 + n_2 - 2)$	$ t > t_{\alpha/2}(n_1 + n_2 - 2)$
	$\sigma_1^2 = \sigma_2^2$ (μ_1, μ_2 未知)	$F = \frac{S_1^2}{S_2^2}$	$F(n_1 - 1, n_2 - 1)$	$F > F_{\alpha/2}(n_1 - 1, n_2 - 1)$ 或 $F < F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$

其中 $\bar{x}, \bar{y}, S_1^2, S_2^2, S_w^2$ 为样本均值与方差， $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ 。

3. 2 近红外光谱定量分析基础模型

通过近红外光谱对样品某组分含量进行定量分析是近红外光谱分析的一个重要应用。它属于多元回归方法的应用范畴，本节介绍近红外光谱定量分析的基础模型，即多元线性回归模型。

3. 2. 1 多元线性回归模型

设 $x_{i1}, x_{i2}, \cdots, x_{ip}$ 为第 i 个样品在 p 个波长点处的吸光度值（光谱参数），而第 i 个样品的待研究组分的含量为 y_i ，则定量分析模型为

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i (i = 1, 2, \cdots, n)$$
 (3-22)

其中 ϵ_i 为测量随机误差。通常假定 $E(\epsilon_i) = 0, D(\epsilon_i) = \sigma^2$ ，且 $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$

相互独立。

将式 (3-22) 写为矩阵形式即为：

$$Y = X\beta + \varepsilon$$

其中

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

确定模型 (3-22) 中的回归系数 β_i ，通常采用最小二乘法，即求 $\beta_0, \beta_1, \dots, \beta_p$ 使

$$Q = \sum_{i=1}^n [y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}]^2 \quad (3-23)$$

达到最小。由多元函数极值存在的必要条件知使上式达最小的 $\beta_0, \beta_1, \dots, \beta_p$ 满足

$$\frac{\partial Q}{\partial \beta_j} = 0, j = 0, 1, 2, \dots, p$$

整理即得正规方程组

$$\begin{cases} n\beta_0 + \left(\sum_{i=1}^n x_{i1}\right)\beta_1 + \left(\sum_{i=1}^n x_{i2}\right)\beta_2 + \cdots + \left(\sum_{i=1}^n x_{ip}\right)\beta_p = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_{i1}\right)\beta_0 + \left(\sum_{i=1}^n x_{i1}^2\right)\beta_1 + \left(\sum_{i=1}^n x_{i1}x_{i2}\right)\beta_2 + \cdots + \left(\sum_{i=1}^n x_{i1}x_{ip}\right)\beta_p = \sum_{i=1}^n x_{i1}y_i \\ \cdots \\ \left(\sum_{i=1}^n x_{ip}\right)\beta_0 + \left(\sum_{i=1}^n x_{ip}x_{i1}\right)\beta_1 + \sum_{i=1}^n \left(x_{ip}x_{i2}\right)\beta_2 + \cdots + \left(\sum_{i=1}^n x_{ip}^2\right)\beta_p = \sum_{i=1}^n x_{ip}y_i \end{cases} \quad (3-24)$$

则式 (3-24) 的矩阵表示为：

$$X'X\beta = X'Y \quad (3-25)$$

如果 $X'X$ 满秩，则 $(X'X)^{-1}$ 存在，这样由式 (3-25) 可得到 β 的最小二乘估计 $\beta = b$

$$b = (X'X)^{-1} X'Y \quad (3-26)$$

式 (3-26) 给出 β 的最小二乘解，但为了计算的方便往往不是求 $X'X$ 的逆矩阵，而是通过解线性方程组 (3-24) 求得 b 。

式 (3-24) 是一个含 $p+1$ 个未知数的线性方程组，它的第 1 个方程为

$$\beta_0 = \bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2 - \cdots - \beta_p \bar{x}_p \quad (3-27)$$

其中

$$\begin{cases} \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} & j = 1, 2, \dots, p \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases} \quad (3-28)$$

将式 (3-27) 代入式 (3-24) 的其余方程，整理可得

$$\begin{cases} S_{11}\beta_1 + S_{12}\beta_2 + \cdots + S_{1p}\beta_p = S_{1y} \\ S_{21}\beta_1 + S_{22}\beta_2 + \cdots + S_{2p}\beta_p = S_{2y} \\ \cdots \\ S_{p1}\beta_1 + S_{p2}\beta_2 + \cdots + S_{pp}\beta_p = S_{py} \end{cases} \quad (3-29)$$

其中

$$\begin{cases} S_{ij} = \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \\ = \sum_{k=1}^n x_{ki} x_{kj} - \frac{1}{n} \left(\sum_{k=1}^n x_{ki} \right) \left(\sum_{k=1}^n x_{kj} \right) \\ S_{iy} = \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_k - \bar{y}) \\ = \sum_{k=1}^n x_{ki} y_k - \frac{1}{n} \left(\sum_{k=1}^n x_{ki} \right) \left(\sum_{k=1}^n y_k \right) \end{cases} \quad (3-30)$$

这样不仅将 $p+1$ 个未知数的线性方程组 (3-24) 化成了有 p 个未知数的线性方程组 (3-29)，简化了计算，而且在讨论估计值 $\beta=b$ 的性质时也将带来很大方便。若记方程组 (3-29) 的系数矩阵为

$$S = (S_{ij})_{p \times p}, B = \begin{bmatrix} S_{1y} \\ S_{2y} \\ \vdots \\ S_{py} \end{bmatrix}$$

则方程组 (3-29) 可简化为

$$S\beta = B \quad (3-31)$$

由此，若 $|S| \neq 0$ 则

$$\hat{\beta} = b = S^{-1}B \quad (3-32)$$

由式 (3-32) 与式 (3-27) 确定模型式 (3-22) 的 $p+1$ 个参数后则可得回归模型

$$y = b_0 + b_1 x_1 + \cdots + b_p x_p \quad (3-33)$$

在 NIR 分析中式 (3-33) 常称为定量预测方程。

若将回归模型的残差平方和式 (3-23) 由矩阵表示，则：

$$\begin{aligned} Q_{\text{残}} &= Q(b) = (Y - Xb)'(Y - Xb) \\ &= (Y - X(X'X)^{-1}X'Y)'(Y - X(X'X)^{-1}X'Y) \\ &= Y'Y - Y'X(X'X)^{-1}X'Y - Y'X(X'X)^{-1}X'Y \\ &\quad + Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &= Y'Y - Y'X(X'X)^{-1}X'Y \\ &= Y'(1 - X(X'X)^{-1}X')Y \end{aligned}$$

令 $H = X(X'X)^{-1}X'$ ，易见 $H' = H$ ， $H^2 = H$ 。

3.2.2 多元线性回归模型最小二乘估计的性质

关于多元线性回归模型最小二乘估计有如下性质。

(1) σ^2 的估计

在式 (3-22) 中 ε_i 的方差 σ^2 一般是未知的，由对随机误差的假定则

$$\begin{aligned} E(Y) &= X\beta \\ D(Y) &= \sigma^2 I_n \end{aligned}$$

由此可验证

$$E(Q(b)) = (n - p - 1)\sigma^2$$

而

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} Q(b) = \frac{1}{n - p - 1} Q_{\text{残}}$$

为 σ^2 的无偏估计。

关于 $Q_{\text{残}}$ 可如下计算：先计算回归平方和 $Q_{\text{回}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{j=1}^n b_j S_{jy}$ ，然后计算总

平方和， $Q_{\text{总}} = \sum_{i=1}^n (y_i - \bar{y})^2$ 则残差平方和

$$Q_{\text{残}} = Q_{\text{总}} - Q_{\text{回}}$$

(2) β 与 σ^2 估计的性质

可以证明， β 与 σ^2 有如下性质：

- ① $E(b) = \beta$;
- ② $\text{Cov}(b) = \sigma^2 (X'X)^{-1}$;
- ③ 设 $y \sim N(X\beta, \sigma^2 I_n)$ ，则 $b \sim N(\beta, \sigma^2 (X'X)^{-1})$;
- ④ 同样在③的假设之下， b 与 $\hat{\sigma}^2$ 相互独立；
- ⑤ $\frac{1}{\sigma^2} (b - \beta)' X'X (b - \beta) \sim \chi^2(p + 1)$;
- ⑥ $\frac{Q_{\text{残}}}{\sigma^2} \sim \chi^2(n - p - 1)$ 。

3.3 近红外光谱定量分析模型的评价

上一节我们已通过最小二乘法建立了近红外光谱定量分析确定组分含量的预测方程。

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_p x_p \tag{3-34}$$

但是我们事先并不能断定组分含量 y 与所选 p 个波长点处的光谱数据 $x_1, x_2, \cdots, x_p$ 之间是否严格存在线性关系。另外用模型 (3-22) 对未知样品进行定量预测事先也并不知预测效果的好坏。基于这方面的原因，我们认为对模型 (3-22) 完全有进一步评价的必要性。下面分别讨论关于模型 (3-22) 本身的评价，及通过对模型 (3-22) 的应用效果的评价进一步评价模型 (3-22)。

3.3.1 回归方程和回归系数的显著性检验

关于模型 (3-22)，确定变量 y 与 $x_1, x_2, \dots, x_p$ 之间是否存线性关系及判断每一个变量 x_j 对 y 的影响是否一样重要可进行如下两种检验。

(1) 回归方程的显著性检验

对回归方程进行的显著性检验是对假设

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0 \tag{3-35}$$

进行检验。若 H_0 被接受，则表明 y 与 $x_1, x_2, \dots, x_p$ 关系由式 (3-34) 表达不合适。为了检验 H_0 ，如下构造统计量，可以证明，当 $y \sim N\left(\sum_{i=0}^p \beta_j x_j, \sigma^2\right)$ ，而 H_0 为真时，统计量为

$$F = \frac{Q_{\text{回}}/p}{Q_{\text{残}}/(n-p-1)} \sim F(p, n-p-1) \tag{3-36}$$

用统计量 F 可对 H_0 进行假设检验。对给定的水平 α 由 F 分布表可确定临界值 $F_\alpha(p, n-p-1)$ ，如果

$$F > F_\alpha(p, n-p-1)$$

则认为在显著水平 α 下， y 对 $x_1, x_2, \dots, x_p$ 有显著的线性关系，即回归方程显著；反之则认为回归方程不显著。这个检验过程可总结在方差分析表 3-3 中进行。

表 3-3 方差分析

方差来源	平方和	自由度	均方差	F
回归	$Q_{\text{回}} = \sum_{j=1}^p b_j s_{jy}$	p	$Q_{\text{回}}/p$	$\frac{Q_{\text{回}}/p}{Q_{\text{残}}/(n-p-1)}$
剩余	$Q_{\text{回}} = Q_{\text{总}} - Q_{\text{回}}$	$n-p-1$	$Q_{\text{残}}/(n-p-1)$	
总和	$Q_{\text{总}}$	$n-1$		

(2) 回归系数的显著性检验

回归方程显著并不意味着每个自变量对 y 的影响都显著。因此需要对每个自变量进行考查。显然，如果某个自变量 x_j 对 y 的作用不显著，那么在回归方程中，它的系数 β_j 将取值接近于零。因此，检验 x_j 是否显著就等价于检验假设

$$H_0 : \beta_j = 0 \tag{3-37}$$

由上节多元回归模型最小二乘估计的性质知 b 与 $\hat{\sigma}^2 = Q_{\text{残}}/(n-p-1)$ 相互独立，并且 $b \sim N(\beta, \sigma^2 (X'X)^{-1})$ ， $E(b_j) = \beta_j$ ， $D(b_j) = C_{jj}\sigma^2$ ，其中 C_{jj} 为 $S^{-1} = (X'X)^{-1} = C$ 对角线上第 j 个元素，于是有

$$\frac{b_j - \beta_j}{\sqrt{C_{jj}\sigma^2}} \sim N(0, 1)$$

$$\text{又 } \frac{Q_{\text{残}}}{\sigma^2} \sim \chi^2(n-p-1), \frac{(b_j - \beta_j)^2}{C_{jj}\sigma^2} \sim \chi^2(1),$$

由此

$$F = \frac{(b_j - \beta_j)^2 / C_{jj}}{Q_{\text{残}} / (n - p - 1)} \sim F(1, n - p - 1)$$

$$t = \frac{(b_j - \beta_j) / \sqrt{C_{jj}}}{\sqrt{Q_{\text{残}} / (n - p - 1)}} \sim t(n - p - 1)$$

因此，在假设式 (3-37) 下，可用统计量

$$F = \frac{b_j^2 / C_{jj}}{Q_{\text{残}} / (n - p - 1)}$$

或

$$t = \frac{b_j}{\sqrt{C_{jj} Q_{\text{残}} / (n - p - 1)}}$$

来检验 β_j 是否为零，即检验 x_j 对 y 的影响是否显著。

当接受假设式 (3-37) 时，即说明变量 x_j 可有可无时，则应在回归方程中将变量 x_j 去掉，重新用最小二乘法估计回归系数，建立新的回归方程。为了减少计算量，人们找到了新老回归系数间的关系。

$$b_i^* = b_i - \frac{C_{ij}}{C_{jj}} b_j \quad (i \neq j) \quad (3-38)$$

式中 b_i^* 为去掉变量 x_j 之后其余 $p - 1$ 个变量中 x_i 的新的回归系数， b_j 是被剔除变量 x_j 的原回归系数。

在剔除变量时，每次只剔除一个，如果有几个变量经检验都不显著，则先剔除其中 F 值或 $|t|$ 值中最小的一个变量，然后再对求得的新的回归方程进行检验，有不显著的变量再剔除，直到被保留的变量都显著为止。一般需多次剔除变量时，采用下节介绍的逐步回归法建立回归方程。

3.3.2 回归模型预测效果的检验

上面的两种检验法是对回归模型的直接检验法，下面介绍评价模型 (3-22) 预测效果的评价法。

(1) 相关系数

设有 m 个样品，其定量分析的标准测量值为 $y_1, y_2, \dots, y_m$ 。而由近红外光谱法通过模型 (3-22) 定量结果为 $z_1, z_2, \dots, z_m$ ，定义 m 个样品两种定量结果的相关系数如下：

$$R = \frac{\sum_{i=1}^m (z_i - \bar{z})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} \quad (3-39)$$

其中 $\bar{z} = \frac{1}{m} \sum_{i=1}^m z_i, \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$ 。

相关系数是描述 2 个定量结果相关程度的一个统计量，但是当一种定量方法结果存在系统误差时，则相关系数 R 不能完全用于评价模型 (3-22) 的预测结果的好坏，因此介绍

下面的评价方法。

(2) 回归参数检验法

将 $(z_i, y_i), i=1, 2, \dots, m$ 描点于坐标纸上, 一个直观的意义是: 当同一个样品的两种定量结果 z_i 与 y_i 相差不大时, 则各点应基本位于一三象限的角平分线上, 由此, 如果以 $(z_i, y_i) (i=1, 2, \dots, m)$ 建立回归方程

$$\hat{y} = a + bz \quad (3-40)$$

则 a 与 0, b 与 1 相差不应太大。这种认识形成了评价模型 (3-22) 预测效果的回归参数检验法。

设以 $(z_i, y_i) (i=1, 2, \dots, m)$ 经最小二乘法确定式 (3-39) 中参数 a, b 的最小二乘估计

$$\hat{b} = \frac{\sum_{i=1}^m z_i y_i - n \bar{z} \bar{y}}{\sum_{i=1}^m z_i^2 - n \bar{z}^2}, \hat{a} = \bar{y} - \hat{b} \quad (3-41)$$

其中 $\bar{z} = \frac{1}{m} \sum_{i=1}^m z_i, \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i$ 。

则检验假设

$$H_{a0}: \hat{a} = 0 \quad (3-42)$$

$$H_{b0}: \hat{b} = 1 \quad (3-43)$$

的工作可通过分别构造如下统计量进行, 记

$$T_a = \left[\frac{m(m-2) \sum_{i=1}^m (z_i - \bar{z})^2}{\sum_{i=1}^m Z_i^2} \right] \frac{\hat{a} - 0}{\sqrt{\sum_{i=1}^m (y_i - \hat{a} - \hat{b} z_i)^2}}, T_b = \left[\frac{(m-2) \sum_{i=1}^m (z_i - \bar{z})^2}{\sum_{i=1}^m (y_i - \hat{a} - \hat{b} z_i)^2} \right]^{\frac{1}{2}} (\hat{b} - 1)$$

可以证明, 在当式 (3-40), 式 (3-42) 为真时, 统计量

$$T_a \sim t(m-2)$$

$$T_b \sim t(m-2)$$

如果在给定的概率水平上两个假设全部被接受, 则认为样品组分的两种测定值无显著差异, 由此进一步认为待测组分的预测模型 (3-22) 有较好的预测效果; 否则, 则认为已建立的预测模型预测效果不好。

3.3.3 模型异常点与高杠杆点的判定

一个试验点 $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$, 它的残差 e_i 如果比其他数据点上的残差大得多, 我们称这样的点为异常点。在建立回归模型的实际应用中, 探查异常点, 判断并剔除异常点, 则更有利于研究回归模型的预测效果。本部分内容我们给出探查异常点的一种检验方法。

设给定线性回归模型为

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \quad i=1, 2, \cdots, n \quad (3-44)$$

如果 $(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i)$ 是一个异常点, 那么它的残差之所以很大是因为 $E(y_i) \neq \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$, 而是 $E(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \eta_i$, 而 η_i 是一个非随机量。因此, 若 $y_{i_1}, y_{i_2}, \cdots, y_{i_k}$ 对应于 k 个异常点, 则回归模型为

$$y_i = \begin{cases} \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i & \bar{i} \in I \\ \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \eta_i & i \in I \end{cases} \quad (3-45)$$

记 $d_i = (0, \cdots, 0, 1, 0, \cdots, 0)'$, 其中第 i 个元素为 1, 其余 $n-1$ 个元素全为 0, 又记 $D = (d_{i_1}, d_{i_2}, \cdots, d_{i_k})$, $\eta = (\eta_{i_1}, \eta_{i_2}, \cdots, \eta_{i_k})'$ 则以上模型为

$$Y = X\beta + D\eta + \epsilon \quad (3-46)$$

要判定 $(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i)$, $i \in I$ 是不是异常点, 等价于检验假设

$$H_0: \eta = 0 \quad (3-47)$$

可以证明式 (3-47) 为真时, 统计量

$$F = \frac{(n-p-k-1)Q(I)}{k[(N-P-1)\hat{\sigma}^2 - Q(I)]} \sim F(k, n-p-k-1) \quad (3-48)$$

其中 $Q(I) = e'_I (I - H_I)^{-1} e_I$, e_I 为残差向量 e 的第 $i_1, i_2, \cdots, i_k$ 分量组成的子向量, $H_I = X_I (X'_I X_I)^{-1} X'_I$, X_I 的定义类似于 e_I 。

对给定的水平 α 如果经检验, 假设 $\eta = 0$ 被拒绝, 则推断在 k 组数据 $(x_{i1}, \cdots, x_{ip}, y_i)$ $i = i_1, \cdots, i_k$ 中, 至少有一组是异常点。特别对 $K=1$ 可采用如下统计量进行假设检验。

$$F_i = \frac{(n-p-2)r_i^2}{(n-p-1)\hat{\sigma}^2 - r_i^2} \sim F(1, n-p-2)$$

或

$$t_i = r_i \left(\frac{n-p-2}{(n-p-1)\hat{\sigma}^2 - r_i^2} \right)^{\frac{1}{2}} \sim t(n-p-2)$$

其中, $r_i = \frac{e_i}{\hat{\sigma} \sqrt{1-h_{ii}}}$ 称为学生化残差;

$$h_{ii} = (1, x_{i1}, \cdots, x_{ip})(X'X)^{-1} \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$$

对给定的点 $(x_{i1}, \cdots, x_{ip}, y_i)$ 如果 $|t_i| > t_{\alpha/2}(n-p-2)$, 则以水平 α 拒绝假设 $H_0: \eta_i = 0$, 即认为 $(x_{i1}, \cdots, x_{ip}, y_i)$ 是一个异常点。

在回归诊断中对任意的样本点 $(x_{i1}, \cdots, x_{ip}, y_i)$, $0 \leq h_{ii} \leq 1$ 。当 $h_{ii} \approx 1$ 时, 我们称该点为高杠杆点, 此时残差 e_i 的方差 $\text{Var}(e_i) = \sigma^2(1-h_{ii}) \approx 0$, 即 $e_i \approx 0$, 即不论实验值 y_i 等于多少, 拟合值 $\hat{y}_i \approx y_i$, 即点 $(x_{i1}, \cdots, x_{ip}, y_i)$ 将回归线(面)拉向自己, 这种点对回归系数的最小二乘估计影响很大。如何诊断高杠杆点呢? 可以证明

$$F = \frac{n-p-1}{p} \frac{h_{ii} - \frac{1}{n}}{1-h_{ii}} \sim F(n, n-p-1)$$

给定水平 $\alpha=0.05$ (0.01)，如果 $F > F_{\alpha}(n, n-p-1)$ ，认为 h_{ii} 很大，则在水平 α 下判定 $(x_{i1}, \dots, x_{ip}, y_i)$ 为高杠杆点。

3.4 优化定量分析模型的逐步回归法

在实际建立近红外光谱定量分析模型时，选择合适的波长点处的光谱参数（自变量）建立回归方程不是一件容易的事，如果选择变量时选用了对组分 y 影响小的变量，而遗漏了重要的变量，可能导致回归系数 β_j 的估计量产生偏倚性和不一致性。因此希望得到“最优”回归方程，所谓“最优”回归方程，就是包含所有对 y 影响比较显著的变量，而不包含对 y 影响不显著的变量。本节介绍在建立“最优”回归方程时最常用的方法——逐步回归法。

3.4.1 逐步回归过程

逐步回归法建立回归方程的过程如下：

第一步在全部 (p 个) 自变量中选一个对 y 作用显著的变量 X_{k_1} 建立一元线性回归方程

$$y = b_0 + b_{k_1}^{(1)} x_{k_1}$$

第二步，在余下的 $p-1$ 个自变量中选一个对 y 作用显著的变量 x_{k_2} ，由 x_{k_1} 和 x_{k_2} 建立二元回归方程

$$y = b_0 + b_{k_1}^{(2)} x_{k_1} + b_{k_2}^{(2)} x_{k_2}$$

检验 x_{k_1} 的作用是否仍然显著，如不显著马上剔除，如仍然显著考虑下一步引入变量。继续下去。

第 k 步，假如前 $k-1$ 步引入了 l ($l \leq p-1$) 个变量建立了一元回归方程

$$y = b_0 + b_{k_1}^{(k-1)} x_{k_1} + b_{k_2}^{(k-1)} x_{k_2} + \dots + b_{k_l}^{(k-1)} x_{k_l}$$

并检验 l 个因子都重要，则在第 k 步考虑从剩下的变量中引入新变量 $x_{k_{l+1}}$ 。

如果引入 $x_{k_{l+1}}$ 时所有变量都显著，则考虑引入下一个变量；要有不显著的变量，即将其剔除，直到每个变量都显著才考虑引入下一个变量。

如果到某一步，在剩下的变量中没有作用显著的可引入，则停止，此时得到回归方程即为所求“最优”回归方程。

3.4.2 Sweep 变换

逐步回归引入、剔除变量的过程是对方阵进行 Sweep 变换的过程，下面介绍 Sweep 变换的概念与其性质。

(1) Sweep 变换的定义

设同阶方阵 $A = (a_{ij})$ 和 $B = (b_{ij})$ 有如下关系成立

$$b_{ij} = \begin{cases} a_{kj}/a_{kk} & (i=k, j \neq k) \\ a_{ij} - a_{ik}a_{kj}/a_{kk} & (i \neq k, j \neq k) \\ 1/a_{kk} & (i=k, j=k) \\ -a_{ik}/a_{kk} & (i \neq k, j=k) \end{cases}$$

则称为矩阵 B 由矩阵 A 经 Sweep 变换 T_k 而得到，简记为 $B = T_k A$ ，或 $A \xrightarrow{T_k} B$ ，称 T_k 是以 A 的元素 a_{kk} 为枢轴的第 k 步 Sweep 变换。

(2) Sweep 变换的性质

Sweep 变换有如下性质

① 可交换性：对任意的自然数 $k \neq p$ 有

$$T_k T_p A = T_p T_k A;$$

② 自反性：设 k 为任意自然数，则

$$T_k T_k A = A。$$

3.4.3 变量的选入和剔除

假定对每个变量已进行了显著性检验，且不妨设对 y 的作用大小如下顺序排列：

$$x_1, x_2, \dots, x_p$$

以下讨论如何引入和剔除变量。

(1) 变量的引入

我们通过用 Sweep 变换求解 3.2 中方程组 (3-29)，实现变量的引入。记增广矩阵

$$A = \begin{pmatrix} s_{11} & \cdots & s_{1p} & s_{1y} \\ s_{21} & \cdots & s_{2p} & s_{2y} \\ \cdots & \cdots & \cdots & \cdots \\ s_{p1} & \cdots & s_{pp} & s_{py} \end{pmatrix}$$

对 A 施行 Sweep 变换。

第一步对 A 施行 s_{11} 为枢轴的 Sweep 变换 T_1 有：

$$A_1 = T_1 A = \begin{pmatrix} 1/s_{11} & s_{11}/s_{11} & \cdots & s_{1p}/s_{11} & s_{1y}/s_{11} \\ -s_{21}/s_{11} & s_{22} - s_{12}s_{21}/s_{11} & \cdots & s_{2p} - s_{1p}s_{2p}/s_{11} & s_{2y} - s_{1y}s_{21}/s_{11} \\ -s_{21}/s_{11} & s_{22} - s_{12}s_{21}/s_{11} & \cdots & s_{2p} - s_{1p}s_{2p}/s_{11} & s_{2y} - s_{1y}s_{21}/s_{11} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -s_{p1}/s_{11} & s_{p2} - s_{1p}s_{p1}/s_{11} & \cdots & s_{pp} - s_{1p}s_{pp}/s_{11} & s_{py} - s_{1p}s_{p1}/s_{11} \end{pmatrix}$$

A 中最后一列的第一个元素

$$s_{1y}^{(1)} = \frac{s_{1y}}{s_{11}}$$

即为一元回归方程的回归系数的估计值。

若再以 A_1 中元素 $s_{22}^{(1)} = s_{22} - s_{12}s_{21}/s_{11}$ 为枢轴做 Sweep 变换得 A_2 。根据 Sweep 变换的性质知 A_2 的最后一列的前 2 个元素即为具有 2 个变量 x_1, x_2 的回归方程

$$y = b_0 + b_1^{(2)} x_1 + b_2^{(2)} x_2$$

的回归系数 $b_1^{(2)}, b_2^{(2)}$ 的估计值。

一般地, 在第 k 步拟把变量 $x_{k_{l+1}}$ 引入方程, 只要对 A_{k-1} 施行一次以 $S_{k_{l+1}k_{l+1}}^{(k-1)}$ 为枢轴的 Sweep 变换即可, 而 A_k 的最后一列中第 k_{l+1} 个分量 $S_{k_{l+1}y}^{(k)}$ 即为该步对应的回归方程 $x_{k_{l+1}}$ 的回归系数 $b_{k_{l+1}}^{(k)}$ 。

(2) 变量的剔除

由 Sweep 变换的自反性, 可以用 Sweep 变换实现变量的剔除。比如执行 $T_{k_1}, T_{k_2}, \dots, T_{k_l}$ 引入变量 $x_{k_1}, x_{k_2}, \dots, x_{k_l}$, 此时

$$A_l = T_{k_l} \cdots T_{k_2} T_{k_1} A$$

如果检验变量 x_{k_h} 不显著需要剔除, 则只要以 A_l 中元素 $S_{hh}^{(l)}$ 为枢轴施行一次 Sweep 变换即可。

3.4.4 变量的显著性检验

前面说过, 每一步选入 (剔除) 变量都是在对自变量进行显著性检验后进行的, 下面介绍显著性检验在计算机上的具体实现过程。

(1) 显著性的度量——方差贡献

设已由 l 个变量 $x_{k_1}, x_{k_2}, \dots, x_{k_l}$ 建立了回归方程

$$\hat{y}_k = b_0 + b_1 x_{k_1} + \cdots + b_l x_{k_l} \quad k = 1, 2, \dots, n \quad (3-49)$$

则

$$Q_{\text{回}} = \sum_{k=1}^n (\hat{y}_k - \bar{y})^2 = \sum_{i=1}^l b_i S_{iy} \quad (3-50)$$

$$Q_{\text{总}} = \sum_{k=1}^n (y_k - \bar{y})^2 = S_{yy}$$

$$Q_{\text{残}} = Q_{\text{总}} - Q_{\text{回}} = S_{yy} - Q_{\text{回}}$$

经一次 Sweep 变换从式 (3-49) 中剔除因子 x_{jk} , 则方程变为

$$\hat{y}'_k = b'_0 + b'_1 x_{k_1} + \cdots + b'_{j-1} x_{k_{j-1}} + b'_{j+1} x_{k_{j+1}} + \cdots + b'_l x_{k_l} \quad (3-51)$$

$$k = 1, 2, \dots, n$$

记式 (3-51) 的回归平方和与残差平方和分别为 $Q_{\text{回}}$ 和 $Q_{\text{残}}$, 则

$$\bar{Q}_{\text{回}} = \sum_{\substack{i=1 \\ i \neq j}}^l b'_i S_{iy} \quad (3-52)$$

$$\bar{Q}_{\text{残}} = S_{yy} - \sum_{\substack{i=1 \\ i \neq j}}^l b'_i S_{iy} \quad (3-53)$$

记

$$\begin{aligned} Q_j &= Q_{\text{回}} - \bar{Q}_{\text{回}} \\ &= \sum_{i=1}^n b_i S_{iy} - \sum_{\substack{i=1 \\ i \neq j}}^n b'_i S_{iy} \\ &= \sum_{\substack{i=1 \\ i \neq j}}^n (b_i - b'_i) S_{iy} + b_j S_{jy} \end{aligned}$$

称 Q_j 为偏回归平方和。

由 Sweep 变换过程可证

$$b_i - b'_i = C_{ij} b_j / C_{jj} \quad (3-54)$$

$i=1, 2, \dots, l$, 且 $i \neq j$, 由此

$$\begin{aligned} Q_j &= \sum_{\substack{i=1 \\ i \neq j}}^l \frac{C_{ij} b_j}{C_{jj}} S_{iy} + b_j S_{jy} \\ &= \frac{b_j}{C_{jj}} \sum_{i=1}^l C_{ij} S_{iy} = \frac{b_j}{C_{jj}} b_j = \frac{b_j^2}{C_{jj}} \geq 0 \end{aligned}$$

又

$$\begin{aligned} \bar{Q}_{\text{残}} - Q_{\text{残}} &= (S_{yy} - \bar{Q}_{\text{回}}) - (S_{yy} - Q_{\text{回}}) \\ &= Q_{\text{回}} - \bar{Q}_{\text{回}} = Q_j \geq 0 \end{aligned}$$

由此 Q_j 可用于度量变量 x_j 对变量 y 的作用或贡献的大小。

(2) 变量贡献的计算

我们仍讨论 p 个变量的回归问题, 并假设前 l 个变量选入了回归方程, 如 x_j 已引入方程, 且其贡献为

$$Q_j^{(l)} = \frac{(b_j^{(l)})^2}{C_{jj}^{(l)}} = (S_{jy}^{(l)})^2 / S_{jj}^{(l)} \quad (3-55)$$

其中 $S_{jy}^{(l)}$, $S_{jj}^{(l)}$ 为 Sweep 变换过程中 S_{jy} 与 S_{jj} 的当前值, 要从方程中剔除变量, 应剔除贡献最小的 l 个。如 x_j 未引入回归方程, 且要作为第 $l+1$ 变量引入, 同式 (3-55) 得引入后的贡献为

$$Q_j^{(l+1)} = (S_{jy}^{(l+1)})^2 / S_{jj}^{(l+1)} \quad (3-56)$$

由 Sweep 变换过程知

代入式 (3-56) 得

$$\begin{aligned} S_{jy}^{(l+1)} &= \frac{S_{jy}^{(l)}}{S_{jj}^{(l)}}, S_{jj}^{(l+1)} = \frac{1}{S_{jj}^{(l)}} \\ Q_j^{(l+1)} &= \frac{(S_{jy}^{(l)})^2}{S_{jj}^{(l)}} = Q_j^{(l)} \end{aligned} \quad (3-57)$$

此式表明，对于未引入的变量若把它引入，则其贡献的计算方式与已引入方程的变量完全一样。在 S_{ij} 当前值的基础上，可容易地把 $p-1$ 个未引入变量的贡献分别算出。如果要选 1 个引入回归方程，应选贡献最大的一个。

(3) 检验统计量的构造

偏回归平方和 Q_j 一方面与变量 x_j 有关，但又与当前存在于回归方程中的其他因子有关，因此选取如下统计量

$$F = \frac{Q_j/1}{Q_{\text{残}}/(n-p-1)}$$

进行显著性检验。可以证明，在 $H_0: \beta_j = 0$ 为真下，

$$F \sim F(1, n-p-1) \quad (3-58)$$

3.4.5 逐步回归的计算步骤

逐步回归的计算分以下 3 步进行。

(1) 建立正规方程组

正规方程组如 3.2 中式 (3-29) 所示。为了使计算有更好的数字效果，可把正规方程组改为

$$\begin{cases} r_{11}\tilde{b}_1 + r_{12}\tilde{b}_2 + \cdots + r_{1p}\tilde{b}_p = r_{1y} \\ r_{21}\tilde{b}_1 + r_{22}\tilde{b}_2 + \cdots + r_{2p}\tilde{b}_p = r_{2y} \\ r_{p1}\tilde{b}_1 + r_{p2}\tilde{b}_2 + \cdots + r_{pp}\tilde{b}_p = r_{py} \end{cases} \quad (3-59)$$

其中， r_{ij} 为相关系数。

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}} \sqrt{S_{jj}}} \quad i, j = 1, 2, \cdots, p, y \quad (3-60)$$

显然 $r_{ii} = 1$ ($i = 1, 2, \cdots, p, y$)。式 (3-59) 与原正规方程组的解有如下关系：

$$b_i = \tilde{b}_i \sqrt{S_{yy}} / \sqrt{S_{ii}} \quad i = 1, 2, \cdots, p \quad (3-61)$$

相关系数阵 (r_{ij}) 的逆矩阵 ($\tilde{C}_{ij}$) 与离差阵 (S_{ij}) 的逆矩阵 (C_{ij}) 元素间有如下关系

$$C_{ij} = \frac{\tilde{C}_{ij}}{\sqrt{C_{ii}} \sqrt{C_{jj}}} \quad (3-62)$$

而

$$Q_{\text{回}} = \tilde{Q}_{\text{回}} S_{yy}, Q_{\text{总}} = \tilde{Q}_{\text{总}} S_{yy}, Q_{\text{残}} = \tilde{Q}_{\text{残}} S_{yy}。$$

(2) 逐步计算

设计算进行了 1 步 (包括 $l=0$) 引入了 l 个变量建立了回归方程 (3-49)，第 $l+1$ 步的计算内容如下：

(a) 由式 (3-55) 和式 (3-57) 算出全部变量的贡献

$$\tilde{Q}_j^{(l)} = (r_{jy}^{(l)})^2 / r_{jj}^{(l)} = \tilde{Q}^{(l+1)} \quad (3-63)$$

其中前一等号可理解为方程中剔除变量 x_j 所损失的贡献，后一等号理解为方程中未引入变量 x_j 被引入时所增加的贡献。

(b) 在已引入的变量中剔除不显著的变量，确定 k 使

$$\tilde{Q}_k^{(l)} = \min_i (\tilde{Q}_i^{(l)})$$

计算

$$F = \frac{\tilde{Q}_k^{(l)} / 1}{\tilde{Q}_{\text{残}}^{(l)} / (n - l - 1)}$$

若 $F \leq F_\alpha$ ，则剔除 x_k [见 (c)]；

若 $F > F_\alpha$ ，则考虑引入变量。确定 k 使

$$\tilde{Q}_k^{(l+1)} = \max_i (\tilde{Q}_i^{(l+1)})$$

计算

$$\begin{aligned} F &= \frac{\tilde{Q}_k^{(l+1)} / 1}{\tilde{Q}_{\text{残}}^{(l+1)} / [n - (l + 1) - 1]} \\ &= \frac{[n - (l + 1) - 1] \tilde{Q}_k^{(l+1)}}{\tilde{Q}_{\text{残}}^{(l)} - \tilde{Q}_k^{(l+1)}} \end{aligned}$$

若 $F > F_\alpha$ ，引入 x_k [见 (c)]；否则逐步计算结束，进入 (3) 结尾阶段。

(c) 剔除或引入变量 x_k ，作 l 次 Sweep 变换

$$r_{ij}^{(l+1)} = \begin{cases} r_{kj}^{(l)} / r_{kk}^{(l)} & (i = k, j \neq k) \\ r_{ij}^{(l)} - r_{ik}^{(l)} r_{kj}^{(l)} / r_{kk}^{(l)} & (i \neq k, j \neq k) \\ 1 / r_{kk}^{(l)} & (i = k, j = k) \\ -r_{ik}^{(l)} / r_{kk}^{(l)} & (i \neq k, j = k) \end{cases}$$

这时，对已引入回归方程的变量 x_j ，其回归系数

$$\tilde{b}_j^{(l+1)} = \tilde{r}_{jy}^{(l+1)}$$

而由式 (3-61) 得

$$b_j^{(l+1)} = r_{jy}^{(l+1)} \sqrt{S_{yy}} \sqrt{S_{ij}} \quad (3-64)$$

至此，第 $l+1$ 步计算结束，重复 (a) ~ (c) 进行下一步计算，在既不能剔除，也无法引入变量的情况下，逐步计算结束。

(3) 结尾

计算 b_0 ，残差 e_k ，以及复相关系数 R ，偏相关系数 R_j ，回归系数 t ，检验统计量 t_j ，其中

$$b_0 = \bar{y} - \sum_j b_j \bar{x}_j$$

$$e_k = y_k - \hat{y}_k = y_k - (b_0 + \sum_j b_j x_j)$$

其中 j 为已选入变量的序号。

$$R = \sqrt{1 - \frac{Q_{\text{残}}}{S_{yy}}}$$

$$R_j = \frac{b_j}{\sqrt{C_{jj}Q_{\text{残}} + b_j^2}}$$

$$t_j = \frac{b_j}{\sqrt{C_{jj}S_{yy}}}$$

说明：逐步回归选变量的方式是逐个增加的，选取第 l 个变量时，仅要求它与前面已选的 $l-1$ 个变量配合起来有最小的残差平方和，因此最终选出的 l 个变量有时候可能不是使残差平方和最小的。

3.5 主成分回归

在 3.2 建立回归模型 (3-33) 时，由于 NIR 光谱各波长点处的光谱信息重叠严重，再者当波长点数大于样品个数时则光谱阵 S 不满足 $|S| \neq 0$ ，因此由 3.2 中式 (3-22) 不能确定模型 (3-33) 中回归参数 β 的最小二乘估计 $\beta = b$ 。此外当 $\epsilon_i \sim N(0, \sigma^2)$ ，且相互独立时， b 的均方误差为

$$L(b) = \sigma^2 \text{tr}(S^{-1})$$

$$= \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

其中 λ_j 为 S 的第 j 个特征根。当其中 1 个特征根很小时，则估计系数 b 的均方误差 $L(b)$ 将相当大，这种情况下，最小二乘估计 b 不会给出 β 的良好估计，此时，模型 (3-33) 用于预测时将有非常不好的预测结果。正是由于以上原因，一些研究者提出了其他一些定量分析模型的建立方法。本节介绍主成分回归法。

3.5.1 总体主成分的概念

主成分就是 p 个变量 $X_1, X_2, \dots, X_p$ 的一些特殊组合，在几何上，这些组合正是把 $X_1, X_2, \dots, X_p$ 构成的坐标系旋转产生的新坐标系，新坐标轴代表了具有最大变差的方向，同时提供了协方差阵的简洁结构表示。

设随机向量 $X' = (X_1, X_2, \dots, X_p)$ 的协方差阵为 Σ ，而 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ 为 Σ 的特征值，考虑线性组合

$$Z_1 = l'_1 X = l_{11} X_1 + l_{21} X_2 + \dots + l_{p1} X_p$$

$$Z_2 = l'_2 X = l_{12} X_1 + l_{22} X_2 + \dots + l_{p2} X_p$$

$$Z_p = l'_p X = l_{p1} X_1 + l_{p2} X_2 + \dots + l_{pp} X_p$$

则

$$D(Z_i) = l'_i \Sigma l_i \quad i=1, 2, \dots, p$$

$$\text{COV}(Z_i, Z_k) = l'_i \Sigma l_k \quad i=1, 2, \dots, p$$

所谓找主成分就是确定线性组合 $Z_1, Z_2, \dots, Z_p$, 它们互不相关, 且使方差 $D(Z_i) = l_i' \Sigma l_i$ ($i=1, 2, \dots, p$) 达到最大。

由于 $D(Z_i)$ 大小与 l_i 的取法有关, 为了消除这种不确定性, 全部限制 Z_i 的系数向量 l_i 为单位向量, 并定义

第一主成分=线性组合 $l_1' X$, 满足 $\|l_1\| = 1$ 。

且使 $D(l_1' X)$ 达到最大。

第二主成分=线性组合 $l_2' X$, 满足 $\|l_2\| = 1$ 和 $\text{Cov}(l_1' X, l_2' X) = 0$, 且使 $D(l_2' X)$ 达到最大。

一般地, 第 i 个主成分=线性组合 $l_i' X$, 满足 $\|l_i\| = 1$ 及对任意 $k < i$, $\text{Cov}(l_i' X, l_k' X) = 0$, 且使 $D(l_i' X)$ 达到最大。

3.5.2 NIR 光谱的主成分提取——样本主成分

设

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} X_1' \\ X_2' \\ \cdots \\ X_{nl}' \end{pmatrix}$$

为 n 个样品 p 个波长上的光谱信息阵, 记 X 的样本均值、样本协差阵和相关阵分别为 $\bar{X}$, S , R , 我们从样品信息阵 X 出发构造原各波长点处光谱信息变量的不相关的线性组合, 且具有最大的样本方差, 即构造样本主成分。

与总体主成分类似, 我们限制所讨论的 $X_1, X_2, \dots, X_p$ 的线性组合, $l' X$ 满足 $\|l\| = 1$, 用 j 表示第 j 次观测, 样本主成分定义如下:

第一样本主成分=线性组合 $l_1' X_j$, 满足 $\|l_1\| = 1$, 且 $l_1' X_j$ 的样本方差最大;

第二样本主成分=线性组合, $l_2' X$, 满足 $\|l_2\| = 1$ 及 $l_2' X_j$ 与 $l_1' X_j$ 样本协方差为 0, 且使 $l_2' X_j$ 的样本方差最大;

一般地, 第 i 个样本主成分=线性组合 $l_i' X_j$, 满足 $\|l_i\| = 1$ 得 $l_i' X_j$ 与 $l_k' X_j$ ($k < i$) 的样本协方差为 0 且 $l_i' X$ 的样本方差最大。

比较样本主成分与总体主成分易见用 S 代替 Σ 有类似的结论。

设 $S = (S_{ik})$ 为 $p \times p$ 阶样本协方差阵,

$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ 为其特征值, $\hat{l}_1, \hat{l}_2, \dots, \hat{l}_p$ 为相应的标准正交特征向量, 则第 i 个样本主成分为

$$\hat{Z}_i = l_i' X = \hat{l}_{i1} x_1 + \hat{l}_{i2} x_2 + \cdots + \hat{l}_{ip} x_p \quad i=1, 2, \dots, p \text{ 且 } \hat{Z}_i \text{ 的样本方差} = \hat{\lambda}_i;$$

$$i=1, 2, \dots, p$$

$$\hat{Z}_i \text{ 与 } \hat{Z}_k \text{ 的样本协方差} = 0, i \neq k$$

$$\text{总样本方差} = \sum_{i=1}^p S_{ii} = \sum_{i=1}^p \hat{\lambda}_i$$

3.5.3 主成分提取的 Jacobi 方法

Jacobi 法是求实对称矩阵全部特征根与特征向量的方法，是一种解特征根问题的变换法，其基本思想是把对称矩阵 S 经一系列正交相似变换化为一个对角阵，对角阵的对角线上的元素即为 S 的特征根，而由所作正变换的矩阵的乘积可求得特征向量。在介绍 Jacobi 法之前先定义如下矩阵

$$R(p, q, \theta) = \begin{matrix} & \begin{matrix} (g) & (h) \end{matrix} \\ \begin{pmatrix} p \\ q \end{pmatrix} & \begin{bmatrix} 1 & \cdots & \sin\theta \\ \vdots & & \vdots \\ -\sin\theta & \cdots & \cos\theta \\ & & 1 \end{bmatrix} \end{matrix} \begin{matrix} (g) \\ (h) \end{matrix}$$

称为旋转矩阵，显然旋转矩阵是一个正交阵，即满足

$$R'R = I, R^{-1} = R'$$

Jacobi 方法就是通过一系列旋转变换，把 S 化为对角阵，从而求得特征根、特征向量的方法。

(1) Jacobi 算法原理

设 S 为实对称矩阵，记 $S_0 = S$ ，对 S 作一系列旋转相似变换

$$S_1 = R_1 S_0 R_1', S_2 = R_2 S_1 R_2', \dots, S_k = R_k S_{k-1} R_k', \dots \text{旋转阵 } R_k \quad (k=1, 2, \dots)$$

的选取原则是：选取 g, h 使 $|S_{gh}^{(k-1)}| = \max_{1 \leq i, j \leq p} |S_{ij}^{(k-1)}|$ ，选取 θ 使 $|S_{gh}^{(k)}| = 0$ 。

经旋转相似变换得 S_k ，仍保持对称性，并且由 S_{k-1} 变到 S_k 时，只有 S_{k-1} 的 g 行 h 列， h 行 g 列发生变化而其他元不变，计算公式为：

$$\begin{cases} S_{gj}^{(k)} = S_{gj}^{(k-1)} \cos\theta + S_{hj}^{(k-1)} \sin\theta = S_{jg}^{(k)} \\ S_{hj}^{(k)} = -S_{gj}^{(k-1)} \sin\theta + S_{hj}^{(k-1)} \cos\theta = S_{jh}^{(k)} & j \neq g, h \\ S_{gg}^{(k)} = S_{gg}^{(k-1)} \cos^2\theta + 2S_{gh}^{(k-1)} \sin\theta \cos\theta + S_{hh}^{(k-1)} \sin^2\theta \\ S_{hh}^{(k)} = S_{hh}^{(k-1)} \sin^2\theta - 2S_{gh}^{(k-1)} \sin\theta \cos\theta + S_{gg}^{(k-1)} \cos^2\theta \\ S_{gh}^{(k)} = (S_{gg}^{(k-1)} - S_{hh}^{(k-1)}) \sin\theta \cos\theta + S_{gh}^{(k-1)} (\cos^2\theta - \sin^2\theta) = S_{hg}^{(k)} \\ S_{ij}^{(k)} = S_{ij}^{(k-1)} & i, j \neq g, h \end{cases} \quad (3-65)$$

按 R_k 的选取要求 $S_{gh}^{(k)} = 0$ ， θ 应满足，

$$\tan 2\theta = \frac{2S_{gh}^{(k-1)}}{S_{gg}^{(k-1)} - S_{hh}^{(k-1)}}$$

其中限制

$$|\theta| \leq \frac{\pi}{4}, \text{ 若 } S_{gg}^{(k-1)} - S_{hh}^{(k-1)} = 0, \text{ 当 } S_{gh}^{(k-1)} > 0 \text{ 时, 取 } \theta = \frac{\pi}{4}; \text{ 当 } S_{gh}^{(k-1)} < 0 \text{ 时, 取 } \theta = -\frac{\pi}{4}$$

在式 (3-65) 中只要计算 $\sin\theta, \cos\theta$ 即可。若令

$$y = |S_{gg}^{(k-1)} - S_{hh}^{(k-1)}|, x = \sin(S_{gg}^{(k-1)} - S_{hh}^{(k-1)}) \cdot 2S_{gh}^{(k-1)}$$

根据

$$\cos\theta = \frac{1}{\sqrt{1 + \tan^2 2\theta}}, \sin 2\theta = \tan 2\theta \cdot \cos 2\theta, \text{ 可得}$$

$$\begin{cases} \cos 2\theta = \frac{y}{\sqrt{x^2 + y^2}} \\ \cos \theta = \sqrt{\frac{1 + \cos 2\theta}{2}} \\ \sin \theta = \frac{x}{\sqrt{x^2 + y^2}} \\ \sin 2\theta = \frac{\sin 2\theta}{2 \cos \theta} \end{cases} \quad (3-66)$$

可以证明：当 $k \rightarrow \infty$ 时， $S_k \rightarrow \Lambda$ ， Λ 为对角阵， Λ 对角线上的元素即为 S 的特征根。

但是，一般地，把 S 化为对角阵实际上不能在有限步完成，只能通过迭代步骤求近似值，只要 k 充分大时，则可取

$$S_{ii}^{(k)} \approx \lambda_i \quad (i=1, 2, \dots, p)$$

控制迭代结束的信号，可由

$$\sum_{\substack{i,j=1 \\ i \neq j}}^p |S_{ij}^{(k)}| < \epsilon$$

给出。

关于 S 的特征向量可以与特征根同时求得。设经 m 次旋转变换迭代结束，则

$$S_m = R_m R_{m-1} \cdots R_1 S R_1' \cdots R_{m-1}' R_m'$$

记

$T_m = R_1', \dots, R_{m-1}' R_m'$ ，则 T_m 仍为正交阵，此时

$$S_m = T_m' S T_m \approx \Lambda$$

而 T_m 的第 j 列就是 S 的对应于特征根 λ_j 的标准正交特征向量的近似值。实际计算时不必保留 R_k ，可逐步形成 T_m 。令 $T_0 = I$ ， $T_k = T_{k-1} R_k'$ ($k=1, 2, \dots, m$)，每步计算公式为：

$$\begin{cases} t_{ig}^{(k)} = t_{ig}^{(k-1)} \cos \theta + t_{ih}^{(k-1)} \sin \theta \\ t_{ih}^{(k)} = -t_{ig}^{(k-1)} \sin \theta + t_{ih}^{(k-1)} \cos \theta \\ t_{ij}^{(k)} = t_{ij}^{(k-1)} \quad j \neq g, h \\ i=1, 2, \dots, p \end{cases} \quad (3-67)$$

(2) Jacobi 方法的计算步骤

Jacobi 方法的计算步骤如下：

- ① 确定 (g, h) 后由式 (3-66) 计算 $\sin \theta, \cos \theta$ ；
- ② 由式 (3-65) 计算 S_{gi}, S_{hj} ($j=1, 2, \dots, p$)；
- ③ 由式 (3-67) 计算特征向量；
- ④ 若 $\sum_{\substack{i,j=1 \\ i \neq j}}^p (S_{ij}^{(k)})^2 < \epsilon$ ，终止迭代，并取 $S_{ii} \approx \lambda_i$ ($i=1, 2, \dots, p$)，否则转①。

(3) 样本主成分的计算

当样本协差阵 S 的特征根与特征向量确定之后，即可以确定样本主成分。设 S 的特

征根由大到小的排列为

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 对应的特征向量为 $T_1, T_2, \dots, T_p$, 如果记 $Z_1, Z_2, \dots, Z_p$ 为 p 个主成分, 则

$$Z_i = XT_i \quad i=1, 2, \dots, p。$$

通常不要求全部主成分, 只需求其中前几个, 一般地如果前 k 个主成分的贡献率为

$$\sum_{i=1}^k \lambda_i \geq 85\%$$

则取前 k 个主成分就足够了, 因为这时它们的方差之和已占总方差的 85% 以上。

3.5.4 主成分的迭代算法

当 p 较大时, 用 Jacobi 方法求特征根, 特征向量在计算机上往往很难实现, 下面给出求主成分的迭代算法:

- ① 取 X 的任一列作为主成分 Z 的初始值 (向量);
- ② 令 $q' = Z'X/Z'Z$;
- ③ 归一化: $q' = q' / \|q'\|$;
- ④ 由 q 计算 Z : $Z = Zq/q'q$;
- ⑤ $\|Z_{\text{new}} - Z_{\text{old}}\| \geq \epsilon$, 转②, 否则转⑥;
- ⑥ $X = X - Zq'$, 转①计算下一主成分。

3.5.5 回归系数的主成分估计

对于模型

$$Y = \beta_0 + X\beta + \epsilon$$

若记 $\lambda_1 \geq \dots \geq \lambda_p$ 为 X 的协差阵 S 的特征根, $T_1, \dots, T_p$ 为相应的标准正交化特征向量, 记

$T = (T_1 \dots T_p)$ 则

$$T'ST = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_p \end{pmatrix} = \Lambda$$

这样通过正交阵 T , 回归模型可如下式所示

$$Y = \gamma_0 + Z\gamma + \epsilon \quad (3-68)$$

其中 $Z = XT$, $\gamma = T'\beta$, 即原来线性回归模型就变为以原回归变量 $x_1, x_2, \dots, x_p$ 的主成分 $Z_1, Z_2, \dots, Z_p$ 为新自变量的回归模型。

如果当 S 呈病态时, 则 S 的特征根 $\lambda_1, \lambda_2, \dots, \lambda_p$ 中有一部分将很小, 不妨设后 $p-k$ 个很小, 即 $\lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_p \approx 0$, 这时后 $p-k$ 个新自变量 (即主成分) $Z_{k+1}, Z_{k+2}, \dots, Z_p$ 在 n 次试验中取值变化将相对很小, 因此这些新自变量的作用就可并入常数项; 也就是说, 新自变量 $Z_{k+1}, Z_{k+2}, \dots, Z_p$ 可以从模型中剔除。基于上述思想, 若 $\lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_p \approx 0$, 将 Λ, γ, Z 和 T 作相应的分块表示:

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}, \text{其中 } \Lambda_1 : k \times k$$

$$\gamma = \begin{pmatrix} \gamma_{(1)} \\ \gamma_{(2)} \end{pmatrix}, \text{其中 } \gamma_{(1)} : k \times 1$$

$$Z = (Z_{(1)} Z_{(2)}), \text{其中 } Z_{(1)} : n \times k$$

$$T = (T_{(1)} T_{(2)}), \text{其中 } T_{(1)} : p \times k$$

于是式 (3-68) 变形为

$$y = \gamma_0 + Z_{(1)} \gamma_{(1)} + Z_{(2)} \gamma_{(2)} + e$$

剔除 $Z_{(2)} \gamma_{(2)}$, 即用 $\hat{\gamma}_{(2)} = 0$ 估计 $\gamma_{(2)}$, 然后求得 $\gamma_{(1)}$ 的最小二乘估计

$$\hat{\gamma}_{(1)} = \Lambda_1^{-1} Z'_{(1)} Y \quad (3-69)$$

最后利用关系 $\beta = T\gamma$, 得到 β 的估计

$$\tilde{b} = \tilde{\beta} = T \begin{pmatrix} \hat{b}_{(1)} \\ 0 \end{pmatrix} = T_{(1)} \hat{\gamma}_{(1)} = T_{(1)} \Lambda_1^{-1} Z'_{(1)} Y \quad (3-70)$$

$\tilde{b}$ 称为 β 的主成分估计。

对于主成分估计, 保留主成分个数问题, 应用上常通过所给数据信息确定, 一般是以方差贡献率的大小来确定所选主成分个数。

3.6 近红外光谱定量分析中的几种常用算法

应用于光谱定量分析中的化学计量学方法主要是多元校正方法。前面已经介绍了多元线性回归 (MLR) 方法、逐步回归法和主成分回归 (PCR) 方法, 本节继续给出常用的几种定量分析方法。

3.6.1 偏最小二乘法 (PLS)

3.6.1.1 基本原理

PLS 法首先将 n 个样品 m 个组分的浓度矩阵 $Y = (y_{ij})_{n \times m}$, 和仪器测定 n 个样品 p 个波长点处吸光度矩阵 $X = (x_{ij})_{n \times p}$ 分解成特征向量形式:

$$Y = UQ + F \quad (3-71)$$

$$X = TP + E \quad (3-72)$$

其中 U 和 T 分别为 n 行 d 列 (d 为抽象组分数) 的浓度特征因子矩阵和吸光度特征因子矩阵, Q 为 $d \times m$ 阶浓度载荷阵, P 为 $d \times p$ 阶吸光度载荷阵, F 和 E 分别为 $n \times m$, $n \times p$ 阶浓度残差阵和吸光度残差阵。

PLS 法是根据特征向量的相关性分解 Y 和 X , 建立回归模型

$$U = TB + E_d \quad (3-73)$$

其中 E_d 为随机误差阵, B 为 d 维对角回归系数阵。

对待测样品, 如果吸光度向量为 x , 则浓度为

$$y = x(UX)^'BQ \quad (3-74)$$

3.6.1.2 交叉证实

在 PLS 算法中抽象组分数 d 的确定一般采用交叉证实法, 即将 X 中部分样品数据, (通常取 $1/4$) 暂时排斥在一维 PLS 建模计算之外, 由保留下来的数据计算模型参数向量, 然后, 以光谱值和计算模型预测“被剔除”样品的浓度 $\hat{y}$ 值, 得到预测 $\hat{y}$ 值和其实测值 y 的“Press”, 再对 X 中的第二个 $1/4$ 部分进行删除, 重复计算, 直到许多轮后, 预测平方和中包含了每个样品一次。而后抽象组分数增 1。循环继续, 循环次数可以人为确定, 也可根据预测平方和 Press 的变化 (先大到小, 后从小到大的一般变化) 确定最佳维数 d 。

3.6.1.3 PLS 算法

(1) 校准部分

- ① 将吸光度矩阵 X 和浓度矩阵 Y 标准化, 即中心化和方差归一化;
- ② 令维数 $k=0$, 迭代交叉回归计算开始;
- ③ 令 $k=k+1$, 取 Y 中的某列作为 U 的初始向量;
- ④ 求 X 的权向量 $W_k: W'_k = U'X$, 将 W_k 标准化 $\|W_k\| = 1$;
- ⑤ 计算 $t: t = XW'_k$;
- ⑥ 计算 Y 的载荷向量 $q: q = U'Y$;
- ⑦ 确定 Y 的新特征向量 $U = Y_q q$;
- ⑧ 若 $\|t - t_{\text{old}}\| > 10^{-6} \|U\|$, 转④; 否则执行以下步骤;
- ⑨ 计算关联 X 和 Y 的特征向量 t 和 u 的系数 B_k (第 k 个元素): $B_k = t'U / \|t\|$;
- ⑩ 计算 X 的载荷向量 $P: P = t'X$;
- 11 计算 X 和 Y 中形成残差, 将其作为下一维中新的 X 和 Y :

$$X = X - tP, Y = Y - B_k t q$$

- 11 由交叉证实法确定 k 是否为最佳维数 d , 若达到最佳维数, 停止迭代, 否则转③。

(2) 待测样品中组分的确定

根据校准模型和待测样品的吸光度向量 x 确定样品各组分浓度 y 。

- ① 将待测样品吸光度向量 x 按校准过程中进行标准化处理;
- ② 令 $k=0, y=0$;
- ③ $k=k+1, t = x \cdot W'_k; y = y + B_k t q$;
 $x = x - tP$ 。
- ④ 若 $k < d$ 转③, 否则迭代停止。

由于测得浓度值是以标准化的形式给出的, 故需进行逆运算求得待测样品的浓度值。

3.6.2 非线性偏最小二乘法 (NLPLS)

偏最小二乘法是一种很有效的多元统计方法, 它适合于光谱分析中的线性模型, 近年来受到化学计量学工作者的高度重视。但当因变量和自变量不完全呈线性关系时, 采用线性处理可能会给分析结果带来一定的偏差, 以致偏最小二乘法将不能充分发挥其作用。

S. Wold 等人提出了一种非线性偏最小二乘模型，以便改善分析结果的准确度。下面介绍这种 NLPLS 方法的基本原理和算法。

3.6.2.1 基本原理

NLPLS 法与 PLS 法相同，首先分解浓度阵与吸光阵为

$$Y = UQ + F$$

$$X = TP + E$$

因 U 与 T 不再满足线性模型 $U = TB + E_d$ ，非线性偏最小二乘法将 U 与 T 之间线性模型扩展到非线性模型

$$u = f(t) + h \quad (3-75)$$

其中 u, t 为 U 与 T 的对应列分量， h 为残差向量，通常 $f(t)$ 是多项式形式，一般取

$$u = c_0 + c_1 t + c_2 t^2 + h \quad (3-76)$$

NLPLS 法与 PLS 法的主要区别为，NLPLS 所建立的浓度主成分特征向量与吸光度主成分特征向量间的关系模型以非线性（二次）函数所示，以下基于此原理介绍 NLPLS 算法。

3.6.2.2 NLPLS 算法

(1) 校准部分

将吸光度矩阵 X 和浓度矩阵 Y 的数据标准化。

步骤 0. 开始：以 Y 的任一列作为 U 的初始向量，用 PLS 算法估计向量 U 和 T 。

$$\textcircled{1} W' = U'X$$

$$\textcircled{2} W = W / \|W\| \quad (\text{归一化})$$

$$\textcircled{3} t = XW / W'W$$

$$\textcircled{4} q' = U'Y / U'U$$

$$\textcircled{5} q = q' / \|q\| \quad (\text{归一化})$$

$$\textcircled{6} U = Yq / q'q$$

$\textcircled{7}$ 若 $\|t - t_{\text{old}}\| / \|t_{\text{old}}\| \geq 10^{-6}$ 转 $\textcircled{1}$ 否则执行以下步骤。

步骤 1. 采用多项式拟合法估计系数 c_0, c_1, c_2 。

$$U = c_0 + c_1 t + c_2 t^2 + h$$

$$\text{记估计值 } r = \hat{u} = c_0 + c_1 t + c_2 t^2$$

$$\text{计算 } q' = r'Y / r'r$$

$$q = q' / \|q\| \quad (\text{归一化})$$

$$\text{步骤 2. } U = Yq / q'q$$

用新的 U 重新计算系数 c_0, c_1, c_2 。

步骤 3. 根据下式采用非线性拟合估计向量 W

$$U = c_0 + c_1 (X_W) + c_2 (Z_{2W})^2 + h$$

$\textcircled{1}$ 构造矩阵 Z

$$Z_{ij} = (c_1 + 2c_2 t_i) x_{ij}$$

$$i=1, 2, \dots, n; j=1, 2, \dots, p$$

$$Z_{ip+1}=1, i=1, 2, \dots, n$$

$$Z_{ij}=t_i^{j-p-1}, i=1, 2, \dots, n; j=p+2, p+3$$

$$\textcircled{2} V'=U'Z/U'U$$

$$\textcircled{3} V=V/\|V\| \quad (\text{归一化})$$

$$\textcircled{4} S=ZV/V'V$$

$$\textcircled{5} b=S'U/S'S$$

$$\textcircled{6} \Delta_w=bV \quad (V \text{ 取前 } p \text{ 个元素})$$

$$\textcircled{7} W_{\text{新}}=W_{\text{老}}+\Delta_w$$

$$\textcircled{8} W_{\text{新}}=W_{\text{老}}/\|W_{\text{老}}\| \quad (\text{归一化})$$

$$\text{步骤 } 4. t=XW/W'W$$

步骤 5. 如果 $\|t-t_{\text{old}}\|/\|t_{\text{old}}\| \geq 10^{-10}$ 或者迭代次数小于约定次数，则转步骤 1，否则转步骤 6。

步骤 6. 用新的 t 向量计算最终的向量 r, q 及 U ，并重新计算系数 c_0, c_1, c_2 。

$$\text{步骤 } 7. P'=t'X/t't$$

$$\text{步骤 } 8. X_e=X-tq', Y_e=Y-rq'$$

步骤 9. 用 X_e, Y_e 分别代替 X 和 Y 进行下一主成分计算，直到整个循环次数等于主成分数。主成分数可通过交叉证实法确定。

(2) 待测样品中组分的确定

将待测样品的吸光度向量 x 标准化，得标准化数据

$$\text{步骤 } 0. i=0, y=0$$

$$\text{步骤 } 1. i=i+1, t_i=Xp_i/p'_ip_i$$

$$r_i=c_{0i}+c_{1i}t_i+c_{2i}t_i^2$$

$$y=y+r_iq'$$

$$X_{\text{新}}=X_{\text{旧}}-t_iP'_i$$

直至 i 等于主成分数，否则转步骤 1，由此得到标准化形式的浓度矩阵 Y

步骤 2. 将标准化形式的 y 进行标准化运算的逆运算，求得待测样品的浓度值。

3.6.3 局部权回归算法

局部权回归是充分利用距样品最近的那些样品的信息建立回归方程的一种计算方法，其算法如下：

$$\textcircled{1} \text{ 读入校正样品集的光谱信息 } X_{n \times p} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \text{ 浓度向量 } Y_{n \times 1}, \text{ 读入预测样品集的光}$$

$$\text{谱信息阵 } X_{N \times p}^* = \begin{pmatrix} X_1^* \\ X_2^* \\ \vdots \\ X_N^* \end{pmatrix}$$

- ② 令 $k=k+1$
- ③ 对第 k 个样品 X_k^* ，计算 X_k^* 与 X_i ($i=1, 2, \dots, n$) 的欧氏距离（或马氏距离） ρ_{ki} ($i=1, 2, \dots, n$)
- ④ 将 ρ_{ki} 从小到大排列， $\rho_{ki_1} \leq \rho_{ki_2} \leq \dots \leq \rho_{ki_q}$ ，取 $d_k = \rho_{ki_q}$ （前 q 个中最大者）
- ⑤ 计算 $u_{ij} = \frac{\rho_{ki_j}}{d_k} \quad j=1, 2, \dots, q$
- ⑥ 计算权系数

$$w_{ij} = W(u_{ij}) = \begin{cases} (1 - u_{ij}^3)^3 & j=1, 2, \dots, q \\ 0 & j=q+1, \dots, n \end{cases}$$

构成对角形权矩阵 $W = \text{diag} (w_1, w_2, \dots, w_n)$

- ⑦ 计算广义逆矩阵 $(WX)^+$
- ⑧ 确定回归系数 $\hat{b}_k = (WX)^+ WY$
- ⑨ 预测样品 X_k^* 的浓度值， $y_k^* = X_k^* \hat{b}_k$ ，输出 y_k^*
- ⑩ 若 $K=N$ 计算停止，否则转②

3.6.4 人工神经网络的 BP 算法

前馈型 BP 网络即误差拟传播神经网络是实现映射变换的前馈型网络中最常使用的一类网络，也是人们研究最多的、认知最清楚的一类网络。主要介绍 BP 网络实现映射变换的功能结构和典型的学习算法。

3.6.4.1 BP 网络的数学模型

一个典型的三层前馈型 BP 网络的拓扑结构如图 3-5 所示。从结构上讲，三层 BP 网络是一个典型的前馈型层次网络，它被分为输入层 LA、隐含层 LB、输出层 LC。同层节点间无关联，异层网络间向前连接。其中，LA 层含 m 个节点，对应于 BP 网络可感知的 m 个输入；LC 层含有 n 个节点，与 BP 网络的 n 种输出相对应；LB 层节点的数目 u 可根据需要设置。

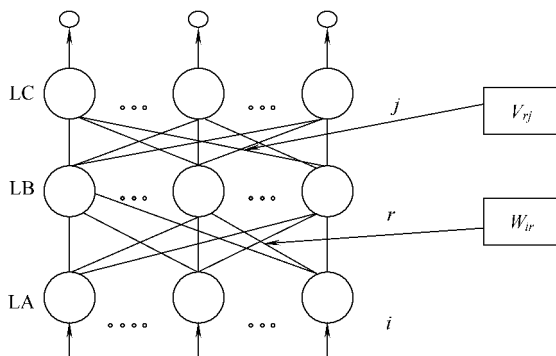


图 3-5 三层 BP 网络结构

令 LA 层节点 a_i 到 LB 层节点 b_r 间的连接权为 W_{ir} ，LB 层节点 b_r 到 LC 层节点 c_j 的连接权为 V_{rj} ， T_r 为 LB 层节点的阈值， q_j 为 LC 层节点的阈值，则 LB 层中节点的输出函数为

$$b_r = f\left(\sum_{i=1}^m W_{ir} \cdot a_i + T_r\right) \quad (r=1,2,\dots,u)$$

$$c_j = f\left(\sum_{r=1}^n V_{rj} \cdot b_r + \theta_j\right) \quad (j=1,2,\dots,n)$$

LC 层中节点的输出函数为

其中, $f(\cdot)$ 为 S 型函数 $f(x) = (1 + e^{-x})^{-1}$ 。

对于有限数目的二值型或连续性模式

$$(A^{(k)}, c^{(k)}) \quad (k=1,2,\dots,p)$$

一般来说, 三层前馈型 BP 网络可按给定的 BP 算法进行存储, 以实现预定的映射

$$A^{(k)} \rightarrow c^{(k)},$$

只要 $A^{(k)}$ 可表达为

$$A^{(k)} = (a_1^{(k)}, a_2^{(k)}, a_m^{(k)})$$

$c^{(k)}$ 可表达为

$$c^{(k)} = (c_1^{(k)}, c_2^{(k)}, c_n^{(k)})$$

且 $a_i^{(k)} \in R, c_j^{(k)} \in R, (R \text{ 为实数集})$ 。

3.6.4.2 BP 网络的学习算法

三层前馈型 BP 网络存储知识 (即调整网络连接权值及节点阈值) 时采用的 BP 方法, 即误差逆传播学习方法, 是一种典型的误差修正方法。其基本思路是: 把网络学习时输出层出现的与“事实”不符的误差, 归结为连接层中各节点间连接权及阈值 (有时将阈值作为特殊的连接权并入连接权) 的“过错”, 通过把输出层节点的误差逐层向输入层逆向传播以“分摊”给各个连接点, 从而可算出各个连接点的参考误差, 并据此对各连接权进行相应的调整, 使网络适应要求的映射。BP 算法的学习过程如下所述。

- ① 给 W_{ir}, T_r, V_{rj}, q_j 随机赋予一个较小的值;
- ② 对每一模式对 $(A^{(k)}, c^{(k)})$, $(k=1, 2, \dots, p)$, 进行下列操作;

$$b_r = f\left(\sum_{i=1}^m W_{ir} \cdot a_i + T_r\right) \quad (r=1,2,\dots,u)$$

$$c_j = f\left(\sum_{r=1}^n V_{rj} \cdot b_r + \theta_j\right) \quad (j=1,2,\dots,n)$$

- a. 将 $A^{(k)}$ 的值 $(a_i^{(k)})$ 输入 LA 层节点, 据 LA 层节点激活值 $a_i^{(k)}$, 依次正向计算
- b. 计算 LC 层节点输出 c_j 与期望输出值 $c_j^{(k)}$ 的误差, 令

$$d_j = c_j(1 - c_j)(c_j^{(k)} - c_j)$$

- c. 向 LB 层节点反向分配误差, 令

$$e_r = b_r(1 - b_r)\left(\sum_{j=1}^n V_{rj} \cdot d_j\right)$$

- d. 调整 LB 层与 LC 层节点间连接权 V_{rj} 及 LC 层节点阈值 q_j ;

$$\begin{aligned} V_{rj} &= V_{rj} + a \cdot b_r \cdot d_j \\ q_j &= q_j + a \cdot d_j \end{aligned} \quad (0 < a < 1)$$

- e. 调整 LA 层与 LB 层节点间连接权 W_{ir} 及 LB 层节点阈值 T_r ;

$$\begin{aligned}W_{ir} &= W_{ir} + b \cdot a_i \cdot e_r \\T_r &= T_r + b \cdot e_r\end{aligned}\quad (0 < b < 1)$$

③ 重复步骤②，直至对于 $i=1, 2, \dots, n$, $j=1, 2, \dots, p$, 误差 d_j 变得足够小或变为零。

三层前馈型 BP 网络是一个典型的映射网络。对三层前馈 BP 网络，人们已经证明，假定网络中隐含单元 b_r 可以根据需要自由设定，那么，一个三层前馈 BP 网络可以实现以任意精度逼近任一连续映射函数 φ ：

$$\varphi(X) \rightarrow Y$$

3.7 聚类分析

聚类分析 (Cluster Analysis) 是应用多元统计分析原理研究分类问题的一种数学方法，近年来已被广泛用于地质勘探、天气预报、作物分类等许多方面。在聚类方法中，系统聚类是一种广为流传的方法，这种方法开始把每个样品都看成一类， n 个样品视为 n 类，然后按一定原则缩小类数，直到所有样品并为一类为止。

3.7.1 聚类分析基础

聚类分析都是依据一定的条件进行的，把这些条件叫做指标或变量，而把要进行分类的对象叫样品。为了根据变量对样品进行分类，就要研究样品间的关系，而描述这种关系的方法通常有两种，一种是距离法；另一种为相似系数法。样品间的距离与相似系数又有多种不同定义方法，具体方法将在后面介绍。依据变量对样品进行分类时，在计算距离或相似系数时，一般与变量的取值关系很大，因此经常将数据进行一些适当的处理。

3.7.1.1 数据变换

在实际中，不同的变量一般取值的量纲（单位）不同，为了使不同量纲的变量进行比较，消除聚类时量纲对聚类结果的影响，经常对原始数据做变换。常用的变换方法有：

(1) 中心化变换

中心化是一种标准化处理方法，它是先求出每个变量的样本平均值，再从原始数据中减去该变量的均值就得到中心化数据。

设原始测量数据为

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} n \times p \quad (3-77)$$

n 为样本容量（试验观测次数）， p 为变量个数。

设中心化后的数据为 x_{ij}^* 则

$$x_{ij}^* = x_{ij} - \bar{x}_j \quad (i=1, 2, \dots, n, j=1, 2, \dots, p) \quad (3-78)$$

其中

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (j=1, 2, \dots, p)$$

(2) 标准化变换

标准化变换的变换公式为：

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, (i=1, 2, \dots, n; j=1, 2, \dots, p) \quad (3-79)$$

其中

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (3-80)$$

s_j 为第 j 个变量的标准差, $\bar{x}_j$ 为第 j 个变量的样本平均值。

(3) 极差标准化变换

极差标准化变换的变换公式为

$$x_{ij}^* = \frac{x_{ij} - \min_{1 \leq i \leq n} \{x_{ij}\}}{\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}} \quad (i=1, 2, \dots, n; j=1, 2, \dots, p) \quad (3-81)$$

经以上变换后, 将每列的最大数据变为 1, 最小数据变为 0 而

$$0 \leq x_{ij}^* \leq 1$$

(4) 对数变换

对数变换公式为：

$$x_{ij}^* = \lg(x_{ij}) \quad (x_{ij} > 0) \quad (i=1, 2, \dots, n; j=1, 2, \dots, p) \quad (3-82)$$

对数变换可将具有指数特征的数据结构化为具有线性结构的数据。

3.7.1.2 相似系数与距离的定义

目前已设计了多种相似系数和距离, 下面介绍在聚类分析中常用的几种。

(1) 距离

设

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip}), i=1, 2, \dots, n.$$

为从总体中取得的一样本容量为 n 的样本, 每个样本点为 p 维空间中一个向量, 用 d_{ij} 表示 x_i 与 x_j 的距离, 则常用的距离有以下几种:

① 明科夫斯基 (Minkowski) 距离

$$d_{ij}(q) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^q \right]^{\frac{1}{q}} \quad (i, j=1, 2, \dots, n) \quad (3-83)$$

② 欧氏距离

$$d_{ij}(2) = \left[\sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right]^{\frac{1}{2}} \quad (i, j=1, 2, \dots, n) \quad (3-84)$$

欧氏距离是在聚类分析中用得最广泛的距离。

③ 绝对值距离

$$d_{ij}(1) = \sum_{k=1}^P |x_{ik} - x_{jk}| \quad (i, j=1, 2, \dots, n) \quad (3-85)$$

④ 切比雪夫距离

$$d_{ij}(\infty) = \max_k |x_{ik} - x_{jk}| \quad (i, j=1, 2, \dots, n) \quad (3-86)$$

以上各种距离是假定变量之间相互独立，即在正交空间中讨论的距离，当变量之间存在相关性，可以采用如下马氏距离。

⑤ 马氏距离 (P. C. Mahalanobis)

样品 X_i, X_j 的马氏距离定义为

$$d_{ij} = (X_i - X_j)' S^{-1} (X_i - X_j) \quad (3-87)$$

其中 S^{-1} 为样本协差阵的逆矩阵，设 $S = (S_{ij})_{p \times p}$ 则

$$S_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad (i, j=1, 2, \dots, p)$$

(2) 相似系数

聚类分析除了研究样品的分类外，有时也需要对变量分类。在对变量进行分类时，通常采用相似系数表示变量之间的亲疏程度。常用的相似系数有以下几种：

① 夹角余弦

设 X_i, X_j 为 n 维空间中两个向量（表示两个样品或两个变量的 n 次观测值），其夹角为 α_{ij} 。则夹角余弦为：

$$\cos \alpha_{ij} = \frac{X'_i X_j}{\sqrt{X'_i X_j} \cdot \sqrt{X'_j X_j}} = \frac{\sum_{k=1}^n x_{ki} x_{kj}}{\sqrt{\left(\sum_{k=1}^n x_{ki}^2 \right) \cdot \left(\sum_{k=1}^n x_{kj}^2 \right)}} \quad (3-88)$$

如果 X_i 与 X_j （两样品）比较相似，则 X_i 与 X_j 的夹角接近于 0，从而 $\cos \alpha_{ij}$ 接近于 1。相似系数的数值范围为 $[0, 1]$ 区间。

② 相关系数

设 r_{ij} 为 n 维向量 X_i 与 X_j 之间的相关系数，则

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \cdot \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (3-89)$$

距离与相似系数的选择是一个比较复杂，带主观性的问题，一般需做具体分析，在多次聚类分析过程中，可多试探几种距离进行聚类，从中总结经验，以选择合适的距离。

3.7.2 系统聚类

系统聚类法是将每个样品视为一类，然后定义样品间的距离（或相似系数）和类与类间的距离，聚类过程是首先选择距离最小的两类合并为一类，再按类间距离的定义，计

算新类与其他类间的距离，再将距离最近的两类合并，如此继续，直至所有样品归为一类。

类与类间的距离又有不同的定义方法，因此产生了系统聚类的不同方法，而常用的有 8 种方法：最短距离法、最长距离法、中间距离法、重心法、类平均法、可变类平均法、可变法和离差平方和法。下面分别介绍这些方法。

3.7.2.1 聚类方法

(1) 最短距离法

设 $G_1, G_2, \dots, G_n$ 表示 n 类， d_{kl} 表示样品 k 和 l 间的距离， D_{ij} 表示类 G_i 和 G_j 间距离，则

$$D_{ij} = \min_{\substack{k \in G_i \\ l \in G_j}} \{d_{kl}\} \quad (3-90)$$

最短距离法的聚类步骤为：

- ① 计算所有样品间的距离，得距离矩阵 $D_{(0)}$ ，各样品自成一类，此时 $D_{ij} = d_{ij}$ 。
- ② 在 $D_{(0)}$ 非对角线元素中选取最小元素，设为 D_{ij} ，将 G_i 与 G_j 合并为一类，记作 G_r 。则 $\{G_i, G_j\}$ 即 G_r 中样品为 G_i, G_j 中全部样品。
- ③ 计算新类 G_r 与其他类 G_s 间的距离 D_{rs} 。

$$\begin{aligned} D_{rs} &= \min_{\substack{k \in G_r \\ l \in G_s}} \{d_{kl}\} \\ &= \min \left\{ \min_{\substack{k \in G_i \\ l \in G_s}} d_{kl}, \min_{\substack{k \in G_j \\ l \in G_s}} d_{kl} \right\} \\ &= \min \{D_{is}, D_{js}\} \end{aligned} \quad (3-91)$$

由此得到距离矩阵 $D_{(1)}$

- ④ 对 $D_{(1)}$ 重复 $D_{(0)}$ 的步骤得 $D_{(2)}$ ，如此继续下去直到所有样品都归为一类为止。

(2) 最长距离法

最长距离法规定两类间的距离为两类中最远样品间的距离，即：

$$D_{ij} = \max_{\substack{k \in G_i \\ l \in G_j}} \{d_{kl}\} \quad (3-92)$$

同最短距离法步骤一样，只是在两类 G_i, G_j 合并为新类 G_r 时， G_r 与其他类 G_s 间距离为：

$$D_{rs} = \max \{D_{is}, D_{js}\} \quad (3-93)$$

(3) 中间距离法

中间距离法定义类间距离采用介于最短距离与最长距离法之间的距离。

设某一步将 G_i 与 G_j 合并为 G_r ，则 G_r 与其他类 G_s 间的距离定义为：

$$D_{rs} = \sqrt{\frac{1}{2} D_{si}^2 + \frac{1}{2} D_{sj}^2 - \frac{1}{4} D_{ij}^2} \quad (3-94)$$

几何意义如图 3-6 所示。

中间距离法可推广到更一般的形式

$$D_{rs} = \sqrt{\frac{1}{2}D_{si}^2 + \frac{1}{2}D_{sj}^2 + \beta D_{ij}^2} \quad (3-95)$$

其中 $-\frac{1}{4} \leq \beta \leq 0$

(4) 重心法

重心法在定义两类之间的距离时，考虑了每一类中所包含的样品数，即以两类重心之间的距离，做为两类之间的距离，设 G_i 与 G_j 类中分别含 n_i, n_j 个样品，其重心分别为 $\bar{X}_i, \bar{X}_j$ ，将 G_i 与 G_j 合并为 G_r 类，则 G_r 中含 $n_r = n_i + n_j$ 个样品，其重心为：

$$\bar{X}_r = \frac{1}{n_r}(n_i\bar{X}_i + n_j\bar{X}_j) \quad (3-96)$$

如果类 G_s 的重心为 X_s ，则类 G_s 与 G_r 的距离平方为

$$D_{rs}^2 = \frac{n_i}{n_r}D_{si}^2 + \frac{n_j}{n_r}D_{sj}^2 - \frac{n_i}{n_r} \cdot \frac{n_j}{n_r}D_{ij}^2 \quad (3-97)$$

(5) 类平均法

类平均法以两类样品两两之间的距离平方和的平均值确定两类之间的距离平方，即：

$$D_{ij}^2 = \frac{1}{n_i n_j} \sum_{\substack{k \in G_i \\ l \in G_j}} d_{kl}^2 \quad (3-98)$$

如果 G_i 与 G_j 类合并为 G_r ，则新类 G_r 与其他类间的距离平方如下计算

$$\begin{aligned} G_{rs}^2 &= \frac{1}{n_r n_s} \sum_{\substack{k \in G_s \\ l \in G_r}} d_{kl}^2 = \frac{1}{n_r n_s} \left[\sum_{\substack{k \in G_s \\ l \in G_i}} d_{kl}^2 + \sum_{\substack{k \in G_s \\ l \in G_j}} d_{kl}^2 \right] \\ &= \frac{n_i}{n_r} D_{si}^2 + \frac{n_j}{n_r} D_{sj}^2 \end{aligned} \quad (3-99)$$

(6) 可变类平均法

在类平均法递推公式中 G_i, G_j 之间的距离没有反映进去，为修改公式 (3-99) 得到递推公式

$$G_{rs}^2 = \frac{n_i}{n_r}(1-\beta)D_{si}^2 + \frac{n_j}{n_r}(1-\beta)D_{sj}^2 + \beta D_{ij}^2 \quad (3-100)$$

式 (3-100) 中 $\beta > 0$ 时，有空间压缩作用，当 $\beta = 0$ 时，聚合空间守恒，当 $\beta < 0$ 时，有空间扩张作用，一般情况下， β 常取负值且 $\beta = -\frac{1}{4}$ 。

(7) 可变量

可变法的递推公式为：

$$D_{rs}^2 = \frac{1-\beta}{2} [D_{si}^2 + D_{sj}^2] + \beta D_{ij}^2 \quad (3-101)$$

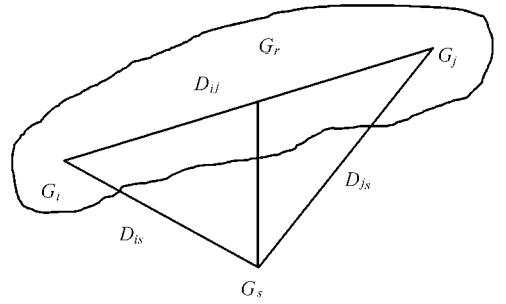


图 3-6 中间距离法的几何意义

其中 $\beta < 1$ ，且常取 $\beta = -\frac{1}{4}$ 。

(8) 离差平方和法

假定已将 n 个样品分为 k 类： $G_1, G_2, \dots, G_k$ 。 X_{ij} 表示 G_j 类中第 i 个样品，记 n_j 为 G_j 中样品数， $\overline{X}_j$ 表示 G_j 的重心，则 G_j 中样品的离差平方和为：

$$S_j = \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)' (X_{ij} - \overline{X}_j)$$

(3-102)

全部 k 类的平方和

$$S = \sum_{j=1}^k S_j$$

(3-103)

离差平方和法是先将 n 个样品自成一类，然后每次缩小一类，每缩小一类离差平方和要增大。该方法缩小类的原则是选择使 S 增加最小的两类进行合并，直至所有样品为一类为止。

3. 7. 2. 2 系统聚类的统一公式

前述 8 种聚类方法中，它们并类的原则和步骤是相同的，不同之处在于计算类与类间的距离时选用了不同的定义，从而有不同的递推公式。维希特（Wishart）1969 年提出了统一的递推公式，这就是

$$D_{rs}^2 = \alpha_i D_{si}^2 + \alpha_j D_{sj}^2 + \beta D_{ij}^2 + \gamma | D_{si}^2 - D_{sj}^2 |$$

(3-104)

其中系数 $\alpha_i, \alpha_j, \beta, \gamma$ 对不同的聚类方法取不同的数值，表 3-4 列出了前述 8 种聚类方法中 $\alpha_i, \alpha_j, \beta, \gamma$ 在公式（3-104）中的取值。

表 3-4 系统聚类法统一公式参数表

聚类方法	α_i	α_j	β	γ
最短距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
最长距离法	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
中间距离法	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
重心法	$\frac{n_i}{n_r}$	$\frac{n_j}{n_r}$	$\frac{-n_i n_j}{n_r^2}$	0
类平均法	$\frac{n_i}{n_r}$	$\frac{n_j}{n_r}$	0	0
可变类平均法	$(1-\beta)n_i/n_r$	$(1-\beta)n_k/n_r$	<1	0
可变法	$(1-\beta)/2$	$(1-\beta)/2$	<1	0
离差平方和法	$(n_s + n_i)/(n_s + n_r)$	$(n_s + n_j)/(n_s + n_r)$	$\frac{-n_s}{n_s + n_r}$	0

3. 8 贝叶斯判别分析

判别分析是多元统计分析中用于判别样品所属类型的一种统计分析方法。它所要解决

的问题是在一些已知样品用某种方法已分成若干类的情况下，确定新的观测样品属于已知类别中的哪一类。判别分析按判别的组数来分，可分为两组判别分析和多组判别分析；按区分不同总体用的数学模型来分有线性判别和非线性判别；按判别对所处理的变量方法不同有逐步判别和序贯判别。判别分析可以从不同角度提出问题，因此有不同的判别准则。本节介绍贝叶斯准则下的多组判别分析。其基本思想是把 p 维空间划分为互不相交的多个区域，使对待判样品，判别归类时错判的平均损失最小。多个区域互不相交是指区域间彼此互不重叠，每个样品只能归属于多个区域中的某一个区域，而不能同时落在两个或更多个区域之中。而对一个待判样品 $(x_1, x_2, \dots, x_p)$ ，判别它属于已知的组中任何一组的方法是，通过计算该样品属于第 l 组的条件概率 $P(l/X)$ ($l=1, 2, \dots, m$)，比较这 m 个概率的大小，将其归入概率最大的一组。

在做贝叶斯判别分析时，首先假设各组内的观测值的分布服从多维正态分布，且各组的协方差阵没有显著差异。

3.8.1 贝叶斯判别方法的原理

设有 m 组样品，第 l 组 ($l=1, 2, \dots, m$) 样品数为 n_l ，每个样品有 p 个指标 (判别变量) 第 l 组的原始数据为：

$$\begin{cases} (x_{11}^{(l)}, x_{12}^{(l)}, \dots, x_{1p}^{(l)}) \\ (x_{21}^{(l)}, x_{22}^{(l)}, \dots, x_{2p}^{(l)}) \\ \dots \\ (x_{n_l 1}^{(l)}, x_{n_l 2}^{(l)}, \dots, x_{n_l p}^{(l)}) \end{cases} \quad l=1, 2, \dots, m \quad (3-105)$$

其中 $x_{kj}^{(l)}$ 表示第 l 组的第 k 个样品第 j 个变量的观测数据，($l=1, 2, \dots, m$; $k=1, 2, \dots, n_l$, $j=1, 2, \dots, p$)

假设各组样品的观测数据相互独立且

$$(x_{k1}^{(l)}, x_{k2}^{(l)}, \dots, x_{kp}^{(l)}) \sim N(\mu_l, \Sigma_l)$$

其中 μ_l 为第 l 组 p 个变量的数学期望， Σ_l 为协差矩阵。在贝叶斯判别分析中，假设：

$$\Sigma_1 = \Sigma_2 = \dots = \Sigma_m$$

令 $n_1 + n_2 + \dots + n_m = n$ ，以样本均值估计各组总体均值

$$\hat{\mu}_l = \bar{X}_l = \begin{pmatrix} \bar{x}_1^{(l)} \\ \bar{x}_2^{(l)} \\ \vdots \\ \bar{x}_p^{(l)} \end{pmatrix} \quad l=1, 2, \dots, m \quad (3-106)$$

其中 $\bar{x}_j^{(l)}$, $j=1, 2, \dots, p$ 为 l 组中 j 个变量的均值。

记各组样本离差阵为 $S_l = (s_{ij}^{(l)})_{p \times p}$ ，其中

$$S_l^{(l)} = \sum_{k=1}^{n_l} (x_{ki}^{(l)} - \bar{x}_i^{(l)}) (x_{kj}^{(l)} - \bar{x}_j^{(l)})$$

以

$$D = \hat{\Sigma} = \frac{1}{n-m} \sum_{l=1}^m S_l \quad (3-107)$$

估计各组总体的总的协差矩阵。则第 l 组总体密度函数为

$$f_l(x) = \frac{|D|^{\frac{1}{2}}}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}(x-\bar{x}^{(l)})'D^{-1}(x-\bar{x}^{(l)})} \quad l=1,2,\dots,m \quad (3-108)$$

若对于样品 $X = (x_1, x_2, \dots, x_p)$ ，如果属于第 k 组时而判别确实属于第 k 组时，记损失为 0，而被误判为其他组时，记损失为 1，则对一个样品 X ，其属于第 g 组的贝叶斯判别公式为：

$$q_g f_g(x) = \max_l q_l f_l(x) \quad (3-109)$$

其中 q_l 以各组的样品出现的频率作为其估计值，即

$$q_l = \frac{n_l}{n} \quad (l=1,2,\dots,m) \quad (3-110)$$

在公式 (3-109) 中，若对 $q_l f_l(x)$ 取对数得

$$\begin{aligned} \ln(q_l f_l(x)) &= \ln(q_l) + \ln f_l(x) \\ &= \ln(q_l) + \ln \frac{|D|^{\frac{1}{2}}}{(\sqrt{2\pi})^p} - \frac{1}{2}(x-\bar{x}^{(l)})'D^{-1}(x-\bar{x}^{(l)}) \\ &= \ln(q_l) + \ln \frac{|D|^{\frac{1}{2}}}{(\sqrt{2\pi})^p} - \frac{1}{2}x'D^{-1}x + x'D^{-1}\bar{x}^{(l)} - \frac{1}{2}(\bar{x}^{(l)})'D^{-1}\bar{x}^{(l)} \end{aligned} \quad (3-111)$$

对给定的样本，式 (3-111) 中 2、3 项与组号无关，因此比较 1、4、5 项的和

$$Q_l(x) = \ln(q_l) + x'D^{-1}\bar{x}^{(l)} - \frac{1}{2}(\bar{x}^{(l)})'D^{-1}\bar{x}^{(l)} \quad (3-112)$$

的大小即可确定所给样品 X 的归属类别。即多组判别的贝叶斯公式为：

$$Q_g^{(x)} = \max_l Q_l(Z) \quad (3-113)$$

3.8.2 贝叶斯判别的计算步骤

设 n 个样品已按某种方法分为了 m 组，如式 (3-105) 所示，而样品 $X = (x_1, x_2, \dots, x_p)$ 为一个待判样品，贝叶斯判别的计算步骤如下：

① 计算式 (3-105) 中各个判别变量的组内均值和总均值

$$\bar{x}_j^{(l)} = \frac{1}{n_l} \sum_{k=1}^{n_l} x_{kj}^{(l)} \quad j=1,2,\dots,p, l=1,2,\dots,m$$

$$\bar{x}_j = \frac{1}{n} \sum_{l=1}^m \sum_{k=1}^{n_l} x_{kj}^{(l)} \quad j=1,2,\dots,p$$

$$n = n_1 + n_2 + \dots + n_m$$

② 计算各组的离差阵 S_l 和总的协方差阵 D

离差阵

$$S_l = (s_{ij}^{(l)}) \quad l=1, 2, \dots, m; i, j=1, 2, \dots, p$$

$$s_{ij}^{(l)} = \sum_{k=1}^{n_l} (x_{ki}^{(l)} - \bar{x}_i^{(l)}) (x_{kj}^{(l)} - \bar{x}_j^{(l)})$$

总的协方差阵

$$D = \frac{1}{n-m} \sum_{l=1}^m S_l$$

③ 计算 D 的逆矩阵 D^{-1}

④ 计算判别函数

由式 (3-110) 计算各组的先验概率 q_l ，由式 (3-112) 计算各组的判别函数 $Q_l(x)$ 值。

⑤ 辅助性检验

为了了解 p 个变量对 m 个组的区分能力，计算马哈拉诺比斯 x^2 统计量

$$x^2 = \sum_{l=1}^m n_l [(\bar{x}^{(l)} - \bar{x})' D^{-1} (\bar{x}^{(l)} - \bar{x})]$$

其中

$$\bar{x}^{(l)} = (\bar{x}_1^{(l)}, \bar{x}_2^{(l)}, \dots, \bar{x}_p^{(l)})', \bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$$

则 $x^2 \sim x^2(p, (m-1))$ ，对给定显著水平 α ，若 $x^2 > x_{\alpha}^2(p, (m-1))$ ，则判定 p 个变量能够鉴别 m 个组。否则认为 p 个变量不能鉴别 m 个组。

⑥ 对待判样品的判别

将待判样品 X 代入各组判别函数式 $Q_l(x)$ 计算对应于各组的判别函数值，然后将其判属于函数值最大的一组。

参 考 文 献

- [1] 中科院概率统计组编. 概率统计计算. 北京: 科学出版社, 1979
- [2] 陈希儒等著. 近代回归分析. 合肥: 安徽教育出版社, 1987
- [3] 裴鑫德编著. 多元统计分析及其应用. 北京: 北京农业大学出版社, 1991
- [4] 肖云茹编著. 概率统计计算方法. 天津: 南开大学出版社, 1994
- [5] 高惠璇编著. 统计计算. 北京: 北京大学出版社, 1998
- [6] 曾善玉, 张录达, 杨丽明编. 应用概率统计. 北京: 科学出版社, 2000
- [7] 俞汝勤著. 化学计量学导论. 长沙: 湖南教育出版社, 1991
- [8] 陆晓华主编. 化学计量学. 武汉: 华中理工大学出版社, 1997
- [9] 刘树深等编著. 基础化学计量学. 北京: 科学出版社, 1999
- [10] 梁逸曾, 俞汝勤主编. 分析化学手册 (第十分册) —— 化学计量学. 北京: 化学工业出版社, 2000
- [11] 陆婉珍, 袁洪福等编著. 现代近红外光谱分析技术. 北京: 中国石油化工出版社, 2000
- [12] 朱尔一等著. 化学计量学技术及应用. 北京: 科学出版社, 2001
- [13] 焦李成著. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1990
- [14] 刘增良, 刘有才著. 模糊逻辑与神经网络——理论与研究与探索. 北京: 北京航空航天大学出版社, 1996

4. 近红外光谱分析仪器

4.1 概述

最早近红外光谱仪器是一台摄谱仪，大约在 20 世纪初，人们采用摄谱的方法首先获得了有机化合物的近红外光谱，并对有关基团的光谱特征进行了解析，预示近红外光谱可能作为一种新的分析技术得到应用^[1]。

直到 20 世纪 50 年代中期，Kaye 首先研制出透射式近红外光谱仪器^[2,3]，一些厂家也相继开始开发商品化的近红外光谱仪器。但早期近红外光谱仪器都是采用紫外-可见光谱仪，再配上适当的近红外检测器扩展而成，如 Cary-2300 型（现 Varian 公司）、Lambda-9 型（Perkin-Elmer 公司）等，这些仪器因噪声高，数据处理系统不完善，很难满足近红外分析要求。

20 世纪 60 年代，由于当时商用近红外仪器噪声很高，无法满足农副产品分析要求，Karl Norris 决定自己设计一台近红外光谱仪器^[4,5]。Norris 开始研制一种用近红外透射法测定谷物中水分的仪器。工作中，他发现谷物中的其他成分如脂肪、蛋白质和淀粉等对测定结果干扰严重，利用计算机相关技术，他选定了一套波长，通过样品对这些波长的吸收和反射值的测定，不但能避免干扰，而且还能测定这些干扰成分的含量，从而真正开始了近红外光谱分析的应用，他选定的这套波长直到今天仍被大家所采用。

由于 Norris 的出色工作，1970 年，美国伊利诺伊州农业部在全美招标制造用于测定大豆中水分、蛋白和脂肪含量的近红外光谱仪器，这些仪器的设计原理都由 Karl Norris 提供。当时 Dickey-John 公司和 Neotec 公司由于在光电技术方面的优势而中标。1971 年，Dickey-John 公司生产了第一台商用近红外光谱仪器（该仪器于 1973 年 10 月获得美国专利，项目名称为 Grain Analysis Computer，专利号为 3776642），这台仪器光源用碘钨灯，分光用 6 个窄带干涉滤光片，检测器用硫化铅。不久，另一个中标厂家 Neotec 公司也生产出 3 个滤光片的近红外谷物分析仪。这些早期的近红外仪器使用模拟电路计算成分的含量，每台仪器需现场校正，校准数据需通过选择不同的电阻器来输入，同今天的仪器相比，这些仪器在硬件上、可靠性上和可操作性上都很原始，且样品必须先干燥，并粉碎到一定的粒度。

1975 年，Dickey-John 公司和 Technicon 公司合作生产了一台近红外光谱分析仪 InfraAnalyzer 2.5 型，Neotec 公司也开发了 Neotec 31EL 型。这时仪器具有温度补偿功能，同时密封的光学部件增加仪器的稳定性，但仍需费时而繁琐的校准，使得操作人员更像一个电子工程师而不是粮食分析专家。尽管如此，一个未经训练的操作人员在几分钟（包括制样）内就可用这种仪器得到一个可与在实验室内经费时费力的漫长化学分析而得到的结果相媲美的结果，从而使他们从国家谷物检测中心得到了大量仪器订单，并将近红外光谱分析方法确认为一种标准方法^[6]。

20 世纪 70 年代末 80 年代初, 由于应用了微处理器, 使仪器在性能上有了很大的提高, 仪器的稳定性和测量的精确度大为改善, 功能也加强了, 仪器有自诊断系统和偏差自动校正系统, 并用微处理器实现数据处理、存储、打印, 使用非常方便。当时的代表产品有 Dickey-John 公司的 GACⅢ 型, Technicon 公司的 InfraAnalyzer400 型和 Neotec 公司的 Neotec101 型。

20 世纪 80 年代中期, 由于高强度光源、低散射光栅、固态高灵敏检测器的研制成功, 尤其是计算机硬件和软件技术的发展及其广泛应用, 为近红外光谱仪器的开发研制提供了良好的基础。这些技术提高了仪器的各项性能指标和运行的可靠性, 简化了仪器的结构, 增强仪器的功能, 达到数据的实时采集与处理, 大大提高了仪器的自动化、数字化与智能化的程度, 同时降低了仪器的成本^[7]。这一时期, 各仪器厂家纷纷研制各种类型的近红外光谱仪器, 出现了高分辨率的傅里叶变换近红外光谱仪器。同时, 使光栅型近红外光谱仪器的性能也有了很大的提高。竞争使各种新技术不断涌现, 也使仪器的性能不断完善, 这时近红外光谱仪器已经完全成熟, 近红外光谱分析技术迅速得到推广应用。

进入 20 世纪 90 年代, 声光可调滤光型近红外光谱仪器的出现、多通道检测器性能的提高和价格的降低, 使得多通道型近红外光谱仪器的大量研制开发, 为近红外光谱仪器的家族增添了不少色彩。同时, 随着光纤技术的发展, 光导探头在近红外测样技术得到广泛应用, 它使近红外光谱采集更加便利, 光纤的远距离传输使近红外光谱仪器广泛地用于在线过程分析中。现在, 近红外光谱仪器已具有较高的信噪比、波长精度和分辨率, 仪器的信噪比已超过 10^5 , 噪音水平降至 10^{-5} A, 波长精度达 0.02nm, 重现性达 0.01nm, 分辨率优于 0.2nm, 各种仪器附件可以很方便地对不同环境下不同物态的样品进行分析。除此之外, 随着化学计量学技术的发展, 功能强大的光谱数据处理软件采用全谱校正已使用户可以方便地进行定性和定量分析。

我国在 20 世纪 80 年代初就进行了近红外光谱技术的应用研究, 但对近红外光谱仪器的研制起步较晚, 大约在 20 世纪 90 年代中后期, 通过一些厂家和科研单位的积极努力, 在近红外光谱仪器的研制、软件开发方面取得了一定的成绩。如北京第二光学仪器厂研制了傅里叶变换近红外辛烷值分析仪^[8], 石油化工科学研究院研制了采用电荷耦合检测器 (CCD) 的多通道近红外光谱仪器^[9], 中国农业大学研制了滤光片型漫透射近红外谷物品质分析仪^[10], 天大光谱公司研制了声光可调滤光 (AOTF) 型近红外乳品成分快速分析仪-MILK2001 型。

近年来, 近红外光谱作为迅速崛起的光谱分析技术在分析测试领域中所起的作用越来越引起人们关注。由于样品在分析时基本不需处理, 且不破坏和消耗样品, 自身又无环境污染, 近红外光谱分析技术堪称是绿色分析仪器的典型代表, 近红外光谱分析技术作为一种无损、快速检测技术正越来越多被大家认同和应用。将来, 近红外光谱仪器研制除了继续改善仪器的性能, 提高仪器的稳定性和信噪比外, 一方面, 应该在低价位的专用仪器开发上多做努力, 普及近红外光谱技术的应用; 另一方面, 近红外光谱图像处理方面的仪器开发也将会是一个热点, 由于二维面阵列检测器的运用, 以及微机性能价格比的大幅度提高, 做近红外光谱的空间各层次结构分析的需求也将越来越多。

4.2 近红外光谱仪器的基本组成与结构

近红外光谱仪器的基本结构与一般光谱仪器一样，都是由光源系统、分光系统、样品室、检测器、控制和数据处理系统及记录显示系统组成，如图 4-1 所示。光源 1 发出的光经分光系统 2 成为单色光；单色光与样品室 3 中的样品作用后，一部分被吸收，一部分透

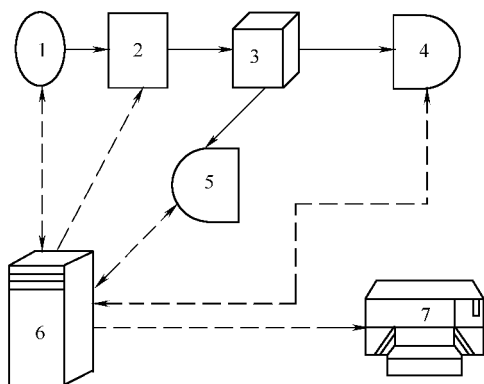


图 4-1 近红外光谱仪器的基本结构

1—光源系统 2—分光系统 3—样品室 4—透射检测器
5—漫反射检测器 6—控制和数据处理系统 7—记录显示系统（实线为光路，虚线为电路）

射，还有一部分被反射；通过透射检测器 4 测定的透射光，就可得到近红外吸收光谱，通过漫反射检测器 5 测定的反射光，就可得到近红外漫反射光谱；光谱经数据处理系统 6 后，结果可用显示器或打印机等记录显示系统 7 记录显示；整个系统都是在控制系统 6 的作用下协同完成的；一般控制系统和数据处理系统都是由微处理器或计算机完成，所以把它们合为一个系统。

4.2.1 光源系统

近红外光谱仪器的光源系统主要由光源和光源稳压电路组成。对光源系统基本要求是在测量谱区有足够的强度和非常高的稳定性，因为光的强度和稳定性直接影响仪器的信噪比，是仪器设计中非常重要的一环。

光源的发光范围决定仪器的工作波长范围，在近红外光谱仪器中常用的光源为钨灯和溴钨灯，它们的光谱覆盖整个近红外谱区，强度高，性能稳定，寿命也较长。在一些专用仪器，尤其是便携式仪器中普通发光二极管用得最多，因为普通发光二极管功耗低，性能稳定，寿命长达几万小时，价格低廉，且很容易调制和控制。近年来，激光发射二极管作为一种新型光源也很快在近红外光谱仪器中得到应用，激光发射二极管因为发射的光带宽窄，不需分光系统，在一些专用仪器的设计上得到了很好的应用。但激光发射二极管的稳定性较差，往往需要功率补偿、光强监控补偿等措施来弥补。

为提高光源的稳定性，光源供电必须用高性能的稳压电路，另外，可通过调制光源、监测光源强度反馈补偿或增加参比光路来提高光强测量的准确性，从而提高仪器的信噪比。

4.2.2 分光系统

分光系统的作用是将光源发射的连续光变成单色光，对分光系统的要求是获得的单色光波长准确、单色性好，它关系到近红外光谱仪器的分辨率、波长准确性和波长重复性，是近红外光谱仪器的核心部分。近红外光谱仪器的分光方式主要有滤光片、光栅、傅里叶变换和声光可调滤光等 4 种类型，早期与紫外-可见连用的近红外光谱仪器也有用棱镜作为分光系统的，但随着光栅技术的发展，光栅性能的完善，现在这类仪器已很少用棱镜作

为分光系统。

滤光片型分光系统一般是用干涉滤光片作为分光器件，干涉滤光片价格低廉，仪器设计简单，所以特别适合用于专用、便携式仪器；但滤光片的光学性能如带宽、峰值波长和透过率受温度、湿度等影响很大，仪器在设计上要有有效的校正系统。

光栅作为分光器件的近红外光谱仪器所占比例也很大，由于使用全息光栅，使光栅的质量大大提高，没有鬼线，杂散光很低，面积可以制造很大，使光栅分光系统的光学性能有很大的提高；同时配上多通道检测器，或使用大口径振动凹面光栅，改变传统的扫描方式使扫描速度也大大提高，且价格非常低廉，对近红外光谱技术的普及与推广将起很大作用。

傅里叶变换分光系统是由迈克逊干涉仪和数据处理系统组合而成，由于这类仪器信噪比高，波长精度达 0.02nm，波长重现性达 0.01nm，仪器分辨率达 0.2nm，扫描速度快，备受大家青睐，但由于价格较高，主要应用于实验室研究。

声光可调滤光分光系统是根据声光衍射原理制成的，详见 4.7 所述，它是由晶体和键合在其上的电声转换器构成，电声转换器将高频的驱动电信号转换为在晶体内的超声波振动，超声波产生了空间周期性的调制，其作用像衍射光栅，当入射光照射到此光栅后将产生布拉格（Bragg）衍射，其衍射光的波长与高频驱动电信号的频率有着一一对应关系，只要改变驱动电信号的频率，即可改变衍射的波长，从而达到分光的目的；由于采用了声光可调滤波器分光方式，扫描速度快，波长精度和波长重现性较好，仪器内无可移动的机械部件，保证了仪器的稳定性，所以这类近红外光谱仪器自 20 世纪 90 年代推出以来，在全世界得到了越来越广泛的应用。

4.2.3 检测器

检测器的作用是把光信号转变为电信号，检测器一般由光敏元件构成，光敏元件的材料不同，其工作范围也不相同，从而决定了仪器的检测波长范围。常用的光敏材料及波长范围如表 4-1 所示。

表 4-1 常用的检测器光敏材料及波长范围

光敏材料	波长范围/nm	光敏材料	波长范围/nm
Si	700~1100	InSb	1000~5000
Ge	700~2500	InAs	800~2500
PbS	750~2500	InGaAs	800~2500

另外，检测器按工作方式可分为单通道和多通道 2 种类型。单通道检测器只有一个检测单元，一次只能接受一个光信号，得到全谱需经过光谱扫描；多通道检测器有许多个检测单元排列在检测面上，可同时接收检测面上不同波长的光信号，不需扫描，速度很快。多通道检测器的类型主要有二极管阵列（PDA）和电荷耦合器件（CCD），其材料决定工作波长范围，检测面上阵列的数目决定仪器的波长分辨率，常见的阵列有 256×1、512×1、1024×1、2048×1，阵列的数目越大分辨率越高，但价格也越高，因此选用多通道检测器阵列的数目要与光栅的分光性能相匹配。检测器的性能直接影响仪器的信噪比，设计时，除尽可能提高检测器的信号强度，还应尽量降低检测器的噪声并消除各种干扰。

4.2.4 样品室

样品室是用来放置样品的。由于近红外光谱样品的差异很大，要求各异，因此各种近红外光谱仪器的样品室设计也不一样。总体而言，材料一般用玻璃或有机玻璃即可，形状根据具体样品而定。此外，根据需要可加恒温、低温、旋转或移动等装置。与一般光谱仪器不同，近红外光谱仪器中常用光纤和积分球直接作为测样器件，在此对光纤和积分球的结构做简单介绍。

4.2.4.1 光纤探头

使用光纤作为测样探头，充分发挥了近红外光谱分析技术的优势。通过光纤探头可实现长距离的在线检测，仪器可远离采样现场达 200m 进行实时测量；通过光纤探头，可以方便地进行定位分析，甚至可以实现体内、样品内部定位分析；此外，通过光纤探头，可以在恶劣、危险的环境中采样分析，如有毒、易爆、易燃、复杂的分析现场，用光纤可直接测定。因此，现在大多数商品近红外仪器都配有光纤探头。

近红外光谱仪器的光纤探头主要有 2 种类型：透射式探头和漫反射式探头，结构示意图如图 4-2、图 4-3 所示。

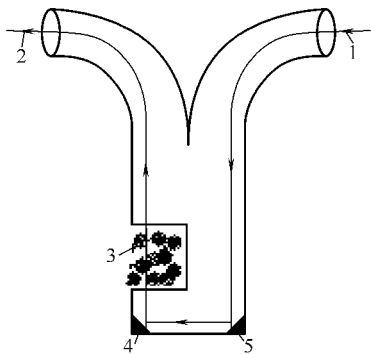


图 4-2 透射式光纤探头示意图

1—入射光纤 2—出射光纤 3—样品 4、5—反射镜

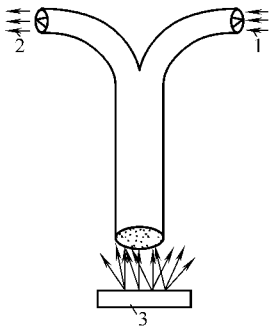


图 4-3 漫反射式光纤探头示意图

1—入射光纤 2—出射光纤 3—样品

由图可以看出，透射式探头和漫反射式探头都由入射光纤和出射光纤构成。漫反射式探头光纤芯由多根光纤集束而成，传输距离较短，一般不超过 3m；透射式探头光纤芯既可采用多根光纤集束，也可用单根光纤，单根光纤的传输距离较长。光纤芯的直径从 50 μm 到 1500 μm 均有。光纤材料主要有 2 类，一类为用于短波近红外区的低 OH^- 的石英光纤，另一类为用于长波近红外区的氟化锆光纤。

光纤探头在使用时应避免强光照射，同时，注意其允许的最小弯曲半径，防止光纤被损坏。

4.2.4.2 积分球

最早的商用近红外光谱仪器使用积分球作为测样器件。漫反射光谱所用的积分球的结构如图 4-4 所示，圆球为积分球球体，内涂白色的硫酸钡，其反射率高达 96% 以上，当样品光束照射到样品上时，被样品漫反射的光经球体内部多次反射，绝大部分进入检测器被接收。由于收集了绝大部分的分析光，首先，大大增加了信号强度，提高了信噪比；其

次，消除了由于光的反射、散射、偏振的干扰，降低了由于入射光的形状和角度等变化产生的影响，使仪器更稳定、更可靠；第三，降低了由于样品的不均匀和空间位置变化等对测量产生的影响，提高了测量的重复性。因此，用积分球作为测样器件使近红外光谱分析技术在分析水平上有了很大的提高。

另外，使用积分球很容易在单光束仪器上实现类似双光束的测量。如图 4-4 所示，只要稍微改变入射光的角度，让它照射在球体的空白部分，如图中的参比光束，即可测量背景信号，因此，只要交替切换光的入射方向，就可很方便地测量背景和样品，实现双光束的测量。

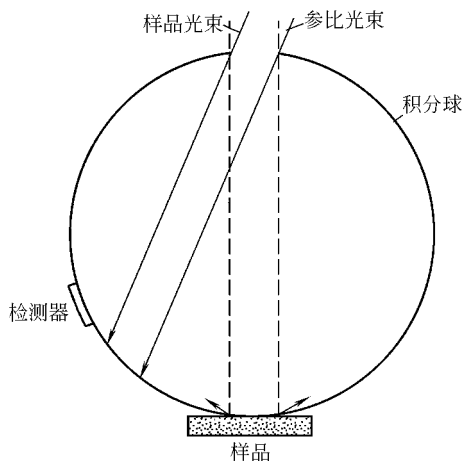


图 4-4 积分球示意图

4.2.5 控制和数据处理系统

一台近红外光谱仪器的各个系统能很好地协同工作，完全是控制系统的功劳。它控制仪器各个部分的工作状态，如控制光源系统的发光状态、调制或补偿，控制分光系统的扫描波长、扫描速度，控制检测器的数据采集、A/D 转换，有时还要控制样品室的旋转、移动或温度等。控制系统一般是由微处理器或计算机配以相应的软件和硬件组成的。

数据处理系统是一套计算机软件，主要对所采集的光谱进行分析处理，实现定性或定量分析。近红外光谱仪器的数据处理软件通常由光谱数据预处理、校正模型建立和未知样品分析三大部分组成，其核心是校正模型建立部分软件。由于近红外光谱特征峰差别不明显，需要通过光谱数据预处理减少各干扰因素对光谱信息提取的影响，常用的预处理方式有光谱平滑、光谱求导、光谱归一化、多元散射校正和小波变换滤波等。校正模型建立部分软件是光谱信息提取的手段，直接影响到分析结果的准确性，一些好的软件，都有其独到的建立校正模型的算法，以便尽可能准确地提取样品信息，常用的校正模型的算法有适合定性分析的主成分分析、聚类分析，适合定量分析的多元线性回归、主成分回归、偏最小二乘法、人工神经网络等。

4.2.6 记录显示系统

显示或打印样品光谱或测量结果。

4.3 近红外光谱仪器的主要类型

近红外光谱仪器种类繁多，根据不同角度有多种分类方式。从应用角度分类，近红外光谱仪器被分为在线的过程监测仪器、专用仪器、通用仪器和图像仪器等。从仪器获得的光谱信息来看，有只测定几个波长的专用仪器，也有的可以测定整个近红外谱区信息的研究型仪器，有的专用于测定短波近红外信息，有的适用于测定长波近红外信息。根据仪器分光方式不同，可分为滤光片型、色散型（光栅、棱镜）、傅里叶变换型和声光滤光型等

4 种类型。本章将以仪器分光方式不同，介绍各类仪器。

进入 20 世纪 80 年代中后期，近红外光谱作为一门现代测试技术，研究日趋活跃，应用领域迅速扩大，为适应不同的需要，各仪器厂家争先研制各种近红外光谱仪器，据不完全统计，目前近红外光谱仪器生产厂达 60 多家，网站 <http://users.pandora.be/roy.de.maesschalck/NIRproviders01.html> 提供了大量国外近红外通用仪器、专用仪器、图像仪器和近红外仪器附件的生产厂家。表 4-2 列出了二十几个生产厂家的近红外光谱仪器的主要技术指标，并在参考文献中提供了这些厂家的网址，但由于近年来国际上各公司合并严重，很多公司的名称也已变化，如 Bomen 已在 ABB 旗下，Nicolet 也已改为 Thermo Nicolet，Foss 和 NIRSystem 也已合并，有关这些公司及其产品的详细信息可从提供的网址查阅。

在本章的 4.4、4.5、4.6、4.7 节将详细介绍各类仪器的原理、结构、技术特点。

表 4-2 近红外光谱仪器的生产厂家及技术指标

生产厂家(仪器类型)	波长范围/nm	分辨率/nm	检测器	采集速度
AnalytiChem(光栅)	1000~2500	2	InAs	2s
Bomem(ABB)(FT)	1100~2500	0.2	InGaAs	0.8 次/s
	400~1100	0.2	Si	0.8 次/s
Bran+Luebbe(滤光片)	1000~2500	10	恒温 PbS	10s
(AOTF)	900~1700	3~8	InGaAs	2000 波长/s
Brimrose(AOTF)	650~1050	1~3	Si	4000 波长/s
	900~2200	2~10	InGaAs	4000 波长/s
Bruker(FT)	650~2500	0.3	Ge、PbS	2 次/s
Buhler(FT)	1000~2500	0.3	PbS、InGaAs	0.3 次/s
Dikey-John(滤光片)	1400~2400	10	PbS	10s
Foss(光栅)	800~1100	6	Si	45s
LTIndustries(光栅)	1200~2400	1	PbS、InGaAs	5 次/s
LUMEXLtd.(FT)	800~1100	0.1	Si	2min
Nicolet(FT)	400~2500	0.1	PbS	5 次/s
OceanOptics(光栅)	1000~3000	6	CCD256 阵列	10 次/s
PerkinElmer(光栅)	185~3300	5	PbS	100 波长/s
(FT)	700~3000	0.2~6.4	PbS、InGaAs	5 次/s
PerstorpAnalytical(光栅)	400~2500	2~10	Si、InGaAs	1.8 次/s
Perten(滤光片)	1400~2500	10	PbS	20s/样品
Zeltex(滤光片)	800~1100	10	Si	1 次/s
深圳英贤(光栅)	700~1100	0.5	CCD2048	50 次/s
北京第二光学仪器厂(FT)	770~4000	0.5	Ge、InGaAs	0.2~1.5cm/s
天津先石公司(AOTF)	1100~2200	4	InGaAs	4000 波长/s
四平金国公司(滤光片)	800~1100	10~15	Si	1 次/s

4.4 滤光片型近红外光谱仪器

滤光片型近红外光谱仪器就是以滤光片作为仪器的分光系统，仪器工作过程是：由光

源发出的光经滤光片得到一定带宽的单色光，通过与样品作用后由检测器检测。该类仪器的特点是设计简单、成本低、光通量大、信号记录快、坚固耐用；且可根据需要在固定几个波长下进行测量，灵活方便；但这类仪器单色光的带宽较宽，波长分辨率差，如遇样品基体或温湿变化较大，往往会引起较大的测量误差，需要完善的校正系统，且所选滤光片的波长也需通过扫描型仪器对样品的全谱扫描分析才能确定。随着滤光片性能的提高和校正技术发展，这类仪器已广泛用于专用或便携式仪器上，也是近红外光谱技术普及应用的重要发展方向^[11]。滤光片型近红外光谱仪器可分为波长固定滤光片和波长可扫描滤光片两种形式。而用得较多的是波长固定滤光片型，它又可分为滤光片固定不动设计方式和通过旋转滤光片架切换波长设计方式。下面就对这三种类型的仪器做一介绍。

通过旋转滤光片架切换波长的波长固定滤光片型近红外光谱仪是近红外光谱仪器的最早设计形式，第一台商用近红外光谱仪器就是在 1971 年由帝强公司（Dickey-John）生产，使用了 6 个固定波长滤光片。图 4-5 为帝强公司的 Instalab600 型的原理图，光源 1 使用溴钨灯，光经过斩光器 2 在时间上被分为 2 束，斩光器的结构如图 4-6 所示，其结构为每隔 1/4 圆分别为能吸收全部光的吸收部分和空心的透射部分交替，当斩光器以一定的频率旋转时，光交替透过或被吸收，吸收部分可作为零点自动校正之用，透过部分经滤光片后成为单色光，单色光射到样品 4 上，用与光线成一定角度的检测器 5 检测光信号，转动滤光片架 3，使光从不同波长的滤光片透过，即可得到各个不同波长的信号，从而得到离散的光谱，可以进行分析。这是滤光片型近红外光谱仪中用得最多的一种。

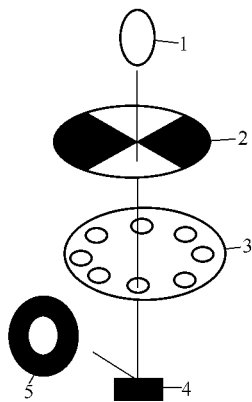


图 4-5 Instalab 600 型的原理图

1—红外光源 2—斩光器 3—滤光片架
4—样品 5—检测器

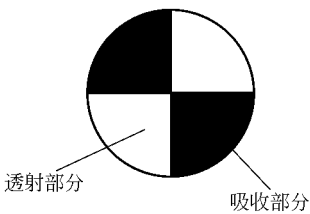


图 4-6 斩光器结构图

另外一种波长固定滤光片型近红外光谱仪使用多个光源，即每个滤光片配一个光源，从而使滤光片架不必旋转，整台仪器没有可移动件，仪器相对比较稳定结实，体积也可设计得很小，尤其适用设计便携式仪器。如 Zeltex 公司的 Zx202C 手持式辛烷值分析仪、国产四平金国公司的 Kx2000 型谷物分析仪都是采用这种设计。这类仪器的原理如图 4-7 所示，光源采用普通发光二极管，它功耗低、寿命长、价格便宜，同时很容易控制，适合多光源，光源发光采用轮流发光的方式，每次只有一个灯亮，测定的是该光源对应滤光片波长的信号，轮流测定一遍即得一个离散的光谱。

20 世纪 70 年代还出现了可做波长扫描的滤光片型近红外分析仪器，图 4-8 为这类仪器的光路示意图。由图看见，光源产生的光经透镜准直成为一束平行光，投射到滤光片上。滤光片是装在有 3 片滤光片组成的转轮上，转轮转动时，滤光片和平行光之间的夹角发生变化，因此透过滤光片的近红外光波长随转轮的转动发生连续的变化。通过适当的组合，使转轮转动时在近红外谱区内波长连续变化的单色光投射到样品上，这样就实现了对样品的波长扫描。

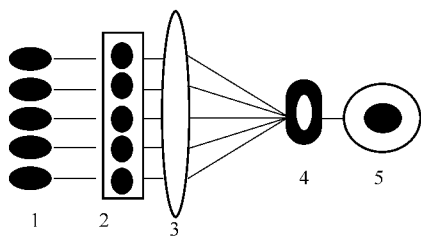


图 4-7 固定滤光片型近红外光谱仪光路原理图

1—光源 2—滤光片架 3—聚光镜
4—样品室 5—检测器

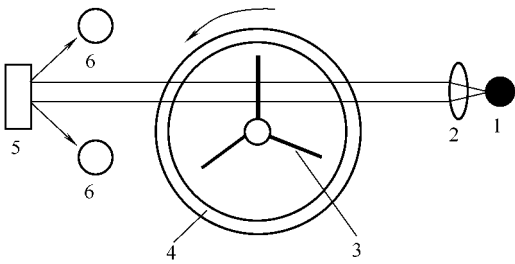


图 4-8 波长可扫描的滤光片型近红外光谱仪器的光路示意图

1—光源 2—准直透镜 3—滤光片 4—可旋转的滤光片轮
5—样品 6—检测器

4.5 色散型近红外光谱仪器

由于近红外光谱的早期应用主要集中在粮食和食品领域，用户有限，加之近红外光谱的光路系统与紫外-可见光谱仪器的设计基本相同，出于商业利益的考虑，早期近红外测试都是在紫外-可见仪器上的延伸，配上适当的近红外检测器，即形成紫外-可见-近红外分光光度计。例如 Cary-2300（Varian 公司）型、Lambda-9（Perkin-Elmer）型和上海分析仪器厂的 710 型紫外-可见-近红外分光光度计。现在这种设计方式仍被广泛采用，如北京普析通用公司的 TU-1800 系列，波长从 200~1100nm；日立公司的 U-4100 系列，波长从 240~2600nm。在这些仪器中，近红外谱区的光源与可见区共用钨灯，单色器采用谱区扩展的光栅和棱镜系统。可见区和近红外谱区分别采用光电倍增管与 PbS 检测器，由继电器或步进电机与谱区同步切换检测器。图 4-9 是 TU-1800 分光光度计的光学系统示意图。

图中检测器有 2 个（光电倍增管和 PbS 检测器）。从可见光区到近红外谱区变换时通过继电器改变反射镜 M_{13} 的角度，更换使用检测器。

这类仪器的价格较贵，购用这类仪器大部分用于紫外可见区，近红外谱区只是作为辅助设备使用；同时，这类仪器由于光谱范围较宽，从光源到分光元件的配置，缺乏对近红外区域的整体优化，且对专门从事近红外光谱分析用户，这种设计方式使仪器的造价大大提高。

20 世纪 60 年代末 70 年代初，随着近红外光谱技术在农产品分析中应用的增多，近红外光谱的研究和应用日趋活跃，各仪器厂家也开始争先研制专用的近红外光谱仪器。为获得较高的分辨率，现代扫描型仪器中多使用全息光栅作为分光元件，通过光栅的转动，

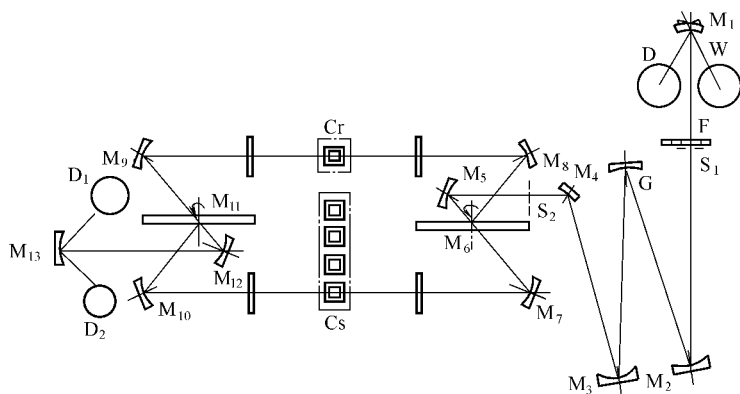


图 4-9 TU-1800 分光光度计的光学系统示意图

D—氙灯 W—钨灯 G—全息光栅 F—滤光片 $M_1 \sim M_{13}$ 中除 M_6 、 M_{11} 是旋转镜，其余为反射镜
Cr、Cs—参比与测量池 S_1 、 S_2 —进、出光狭缝 D_1 —光电信放管 D_2 —PbS 检测器

使单色光按波长高低依次通过测样器件，进入检测器检测。检测器根据扫描的光谱范围，一般在长波近红外区域采用 PbS 光敏元件，短波近红外采用硅基光敏元件。这类仪器的结构与普通扫描型紫外可见分光光度计相差不多，故在此不做过多介绍。

这类仪器的特点是可进行全谱扫描，分辨率较高，仪器价位适中，便于维护。主要缺点是扫描速度慢，但随着光谱技术的不断发展，如使用大口径振动凹面光栅或多通道检测器，使这类仪器的扫描速度已与傅里叶变换型相差不多。

图 4-10 为 Foss 公司的 NIRSystem 系列 Model 5000 型光学系统示意图。图中，从光源发出的光经反射准直，投射在大口径全息凹面光栅上，该光栅是以中心面为轴、以固定的频率摆动，光栅摆动一次即完成一次扫描，所以扫描速度很快，光栅上的衍射光透过狭缝直接射在样品上，被检测器检测。该系统的关键技术是波长校准系统，由于光栅是在快

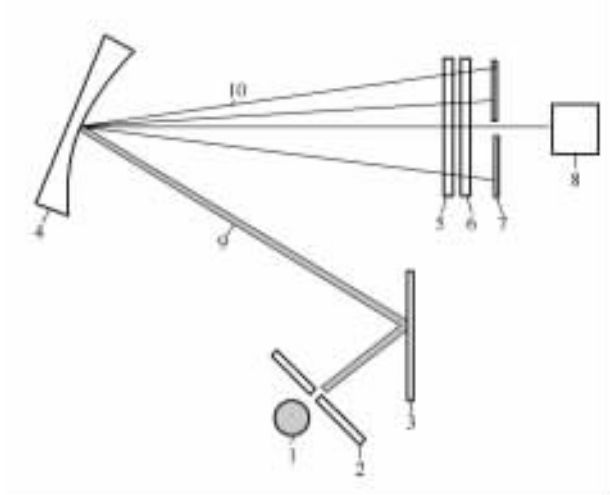


图 4-10 Model 5000 型光学系统示意图

1—光源 2—入射狭缝 3—反射镜 4—大口径振动凹面光栅
5—波长参比 6—滤光片 7—出射狭缝 8—样品
9—非相干光（白光） 10—相干光

速摆动，如何确定每一时刻进入检测器的光的波长是该仪器设计的关键，这也是 Foss 公司的专利技术。采用这种设计，一是扫描速度很快，达 1.8 次/s；二是光通量大，不仅提高了信噪比，而且使被测的样品面积增大，对测定相对不很均匀的农产品，大大提高了测量的准确性。

固定光路多通道检测近红外光谱仪器是 20 世纪 90 年代发展起来的一类仪器，仪器构造示意图见图 4-11。

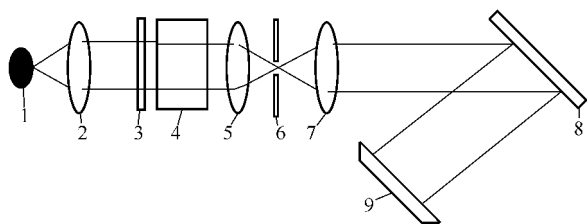


图 4-11 多通道检测器型仪器结构示意图

1—光源 2—准直镜 3—滤光片 4—样品室 5—透镜
6—狭缝 7—准直镜 8—光栅 9—多通道检测器

其原理是光源发出的光先经过样品池，再由光栅分光，光栅不需转动，经光栅色散的光聚焦在多通道检测器的焦面上，同时被检测。在短波近红外区域使用的多通道检测器有 2 种，一种是二极管阵列（PDA）检测器，另一种是采用电荷耦合器件（CCD）作检测器。这类仪器采用全息光栅分光，加之检测器的通道数达 1024 或 2048 个，可获得很好的分辨率。由于检测器对所有波长的单色光同时检测，在 1s 内可完成几十次或上百次的扫描累加，从而得到较高的信噪比和灵敏度。采用全谱信息，可以方便地进行定性和定量分析。由于仪器光路固定，整个仪器内无移动性部件，仪器波长精度和重现性得到保证，使用的耐久性和可靠性得到提高。因此，这类仪器也很适合作为现场分析仪器和在线分析仪器使用。

4.6 傅里叶变换型近红外光谱仪器

20 世纪 70 年代傅里叶变换技术在中红外光谱仪器上的应用使红外光谱仪器的性能得到革命性的变化，进入 20 世纪 80 年代傅里叶变换红外光谱已成为红外光谱仪器的主导产品。借助于研制中红外光谱仪器的技术，通过调整光源、分束器和检测器，并配合适当的软件，傅里叶变换型近红外光谱仪器应运而生。

傅里叶变换型近红外光谱仪的主要光学部件是迈克尔逊干涉仪，有关迈克尔逊干涉仪的结构和工作原理见第 1 章所述。

傅里叶光谱仪要在近红外区工作时，需选用合适的光源、分束器与检测器。在近红外谱区工作时常用钨灯作为光源；常用分束器有石英分束器、CaF₂-Si 分束器、KBr-Ge 分束器等；常用检测器有 PbS、InSb、InAs、Si、Ge 检测器等。

现在，生产傅里叶近红外光谱仪的厂家很多，结构和设计也各不相同，图 4-12 为一般傅里叶近红外光谱仪的工作原理图。图中，由光源 1 发出的光经准直系统变为平行光进

入干涉仪，经干涉仪调制得到干涉光；干涉光通过样品达到检测器变成电信号，该电信号是一时间函数，绘出的就是干涉图，其横坐标是动镜移动的时间或动镜移动的距离；对这种包含有光谱信息的时域干涉图，很难直接进行光谱分析，因而通过模数转换送入计算机，由计算机进行傅里叶变换的快速计算，获得以波数（或波长）为横坐标的频域光谱，即一般的光谱图。

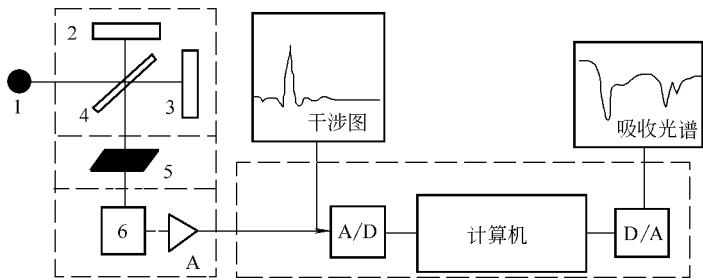


图 4-12 傅里叶近红外光谱仪的工作原理图

1—光源 2—定镜 3—动镜 4—分束器 5—样品 6—检测器
A—放大器 A/D—模数转换器 D/A—数模转换器

傅里叶变换型近红外光谱仪器与其他类型仪器相比，由于具有信噪比高、分辨率高、波长准确且重复性好、稳定性好等优点，而往往作为研究型仪器的首选。

4.7 声光可调滤光型近红外光谱仪器

声光可调滤光型近红外光谱仪器被认为是 20 世纪 90 年代近红外光谱技术最突出的进展。由于采用声光器件分光，该仪器的最大特点是无机械移动部件，测量速度快、精度高、准确性好，提高工作的可靠性和维修费用，可以稳定地长时间工作。它的分辨率也很高，目前可以达到 0.01nm，波长调节速度快，一般 4000nm/s。

声光可调滤波器分光系统是根据各向异性双折射晶体声光衍射原理制成的，如图 4-13 所示^[12]。

它是由双折射晶体、射频辐射源、电声转换器和声波吸收装置组成。双折射晶体一般采用具有较高的声光品质因素和较低的声衰减的双折射晶体作为分光器件，常用的双折射晶体有 TeO_2 、石英和锗等，而 TeO_2 由于具有较高的声光品质因素被广泛采用；射频辐射源提供频率可调的高频辐射输出；将高频的驱动电信号转换为在晶体内的超声波振动，它是键合在晶体上的；声波吸收装置吸收通过晶体后的声波。

电声转换器将高频的驱动电信号转换为在晶体内的超声波振动，超声波振动使晶体的格子发生变化，使晶体产生了空间周期性的调制，其作用像衍射光栅，当入射光照射到此光栅后将产生布拉格（Bragg）衍射，入射光被分成三束，一束为振动方向与晶体表面平行的衍射光（图中的 +1 级衍射光），另一束为振动方向与晶体表面垂直的衍射光（图中的 -1 级衍射光），还有一束为未被衍射的光，其中两束衍射光的波长与高频驱动电信号的频率有着——对应关系，只要改变驱动电信号的频率，晶体的格子就会发生相应的变化，从而改变衍射的波长，达到分光的目的。一般只用一束衍射光用来分析，所以需用挡

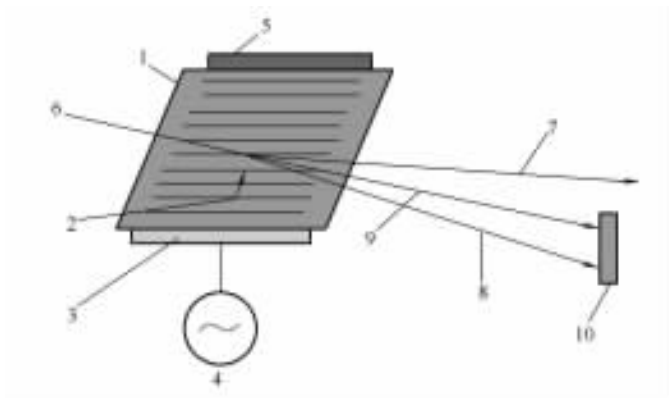


图 4-13 声光可调滤波器分光系统的工作原理

- 1—AOTF 晶体 2—传输的声波 3—电声转换器 4—可调射频源
5—声波吸收器 6—入射光 7、8—+1、-1 级衍射光
9—未衍射的 0 级光 10—挡光板

光板（图中 10）吸收去掉另两束光。

设可变射频输出频率为 f ，晶体内的超声波的速度为 V ，则被衍射光的波长 λ 为

$$\lambda = \Delta n \cdot \alpha \cdot V / f \quad (4-1)$$

式中 Δn 为晶体双折射系数， α 为由晶体的物理设计所决定的参数。为了得到不同波长的衍射光，只要改变射频的输出频率即可，一般射频的输出频率改变后 $20\mu\text{s}$ 时间内晶体的格子就会发生变化，因此声光近红外光谱仪器有很快的扫描速度。被分光的光束带宽一般是由晶体的尺寸、射频输出功率、射频频率的带宽等来决定的。

虽然声光效应在 20 世纪 30 年代就已经开始研究，但直到 20 世纪 90 年代才被应用到近红外光谱技术中，近几年取得了很大的进展。如德国 BRAN+LUEBBE 公司推出的声光可调滤光型近红外光谱仪器成功地用于在线分析，其波长范围 $900\sim 1700\text{nm}$ 。Brimrose 公司推出的声光可调滤光型近红外光谱仪器波长范围为 $650\sim 2200\text{nm}$ ，扫描速度达 4000 波长点/s。天津市先石光学技术有限公司推出的乳品成分快速分析仪和近红外成分监测仪声光可调滤光型近红外光谱仪器光谱范围 $1100\sim 2200\text{nm}$ ，光谱分辨率 4nm ，波长精确度 0.5nm ，波长重复性 0.01nm ，波长转换时间 $<250\mu\text{s}$ ，扫描速度 4000 波长点/s^[13~16]。另据报道^[17]，美国 Brimrose 公司和 Jet Propulsion 实验室联合设计一种微型电晶体近红外光谱仪。这种基于 AOTF 的反射型近红外微型光谱仪主要用于航天领域，使用发光二极管（LED）阵列作为光源，光纤作为光波传输介质，该光谱仪重量 $<250\text{g}$ ，尺寸 $9.2\text{cm} \times 5.4\text{cm} \times 3.2\text{cm}$ ，扫描速度达 4000 波长点/s。

4.8 近红外光谱仪器的主要性能指标

分析仪器的目的是快速、准确地得到测试样品的分析结果，分析仪器的性能就体现在其快速、准确地得到测试样品的分析结果的能力。分析测试结果的准确性，首先取决于近红外光谱仪器获得测定样品光谱的质量，即保真度，这与近红外光谱仪器的单色光带宽、

波长准确度、波长精确度（波长重复性）、吸光度准确度、信噪比、杂散光强度等有关，其次取决于软件数据处理能力，即从光谱信息中准确提取样品分析结果的能力。分析速度主要取决于近红外光谱仪器的光谱扫描速度。此外，分析仪器都有一定的分析范围，近红外光谱仪器的分析范围取决于工作波长范围。在此，对常用的 10 种近红外光谱仪器主要性能指标做一介绍。

4.8.1 波长范围

近红外的整个谱区的波长范围根据 ASTM 定义为 780~2526nm，而在一般应用中大家往往把 700~2500nm 或 700~2600nm 作为近红外谱区，并通常把它分为 2 段，700~1100nm 的短波近红外谱区和 1100~2600nm 的长波近红外谱区。短波近红外谱区更适合做透射分析，故又叫近红外透射区，长波近红外谱区更适合做反射或漫反射分析，也称之为近红外反射区。

仪器的波长范围指该近红外光谱仪器所能记录的光谱范围，它影响能实现分析测试的项目，主要取决于仪器的光源种类、分光系统、检测器类型和透光材料。专用的近红外仪器往往只覆盖某一波段，如 Zeltex 的 ZX101 型手持式辛烷值分析仪用 700~1100nm 的短波近红外谱区，AGMED 公司的土壤快速分析仪用的 1650~2650nm 的长波近红外谱区；而通用型的近红外仪器往往覆盖整个近红外谱区。

4.8.2 仪器的分辨率

近红外光谱仪器的分辨率是指仪器对于紧密相邻的峰可分辨的最小波长间隔，表示仪器实际分开相邻两谱线的能力，往往用仪器的单色光带宽来表示，它是仪器最重要的性能指标之一，也是仪器质量的综合反映。

仪器的分辨率主要取决于仪器的分光系统的性能。对色散型仪器而言，还与光源的强度、检测器的灵敏度有关，光源的强度大、检测器的灵敏度高可减小狭缝宽度，降低单色光带宽，提高仪器的分辨率。而对用多通道检测器的仪器，仪器的分辨率与检测器的像素有关，单位长度像素越多分辨率越高。对于滤光片型近红外光谱仪器，滤光片的带宽就是仪器的分辨率。仪器的分辨率主要影响光谱仪器获得测定样品光谱的质量，从而影响分析的准确性，对于一台仪器的分辨率是否满足要求，这与待测样品的光谱特征有关，有些物质光谱重叠、特征复杂，要得到满意的分析结果，就要求较高的仪器分辨率。

4.8.3 波长的准确度

波长准确度是指仪器所显示的波长值和分光系统实际输出单色光的波长值之间相符的程度。波长准确度可用波长误差，即上述两值之差来表示。保证波长准确度是近红外光谱仪器能够准确测定样品光谱的前提，是保证分析结果的准确度前提。近红外分析结果一般是通过用已知化学值的标准样品建立的模型来分析待测样品，如果波长准确度不能保证，整组数据就会因波长平移而使每个数据出现偏差，造成分析结果的误差。波长准确度主要决定于光学系统的结构，此外还受温度的影响。傅里叶变换近红外光谱仪器因内部有波长校正系统，所以波长准确度很高；而色散型近红外光谱仪器和滤光片型近红外光谱仪器的波长准确度相对低些，需用已知波长值且性质比较稳定的标准物质经常进行校正。

4.8.4 波长的精确度

波长精确度又称波长重复性，是指对同一样品进行多次扫描，光谱谱峰位置间的差异程度或重复性，通常用多次测量某一谱峰所得波长的标准差来表示。波长精确度是体现仪器稳定性的一个重要指标，取决于光学系统的结构，与波长准确度一样，也会影响分析结果的准确性。如果仪器的光学系统全部设计成固定不动，则仪器的波长的精确度就会很高。

4.8.5 光度准确度

光度准确度是指仪器对某物质进行透射或漫反射测量时，测得的光度值与该物质真实值之差。主要是由检测器、放大器、信号处理电路的非线性引起。它会直接影响近红外定量分析结果的准确度。

4.8.6 信噪比

信噪比就是样品吸光度与仪器吸光度噪声的比值。仪器吸光度噪声是指在一定的测量条件下，在确定的波长范围内对样品进行多次测量，得到光谱吸光度的标准差。仪器的噪声主要取决于光源的稳定性、放大器等电子系统的噪声、检测器产生的噪声及环境噪声，如电子系统设计不良、元件质量低劣、仪器接地不良、工作环境潮湿、外界电磁干扰多会使仪器噪声增大。信噪比是近红外光谱仪器非常重要的一项指标，直接影响分析结果的准确度与精确度；因为近红外光谱分析是一门弱信号提取技术，在一个很强的背景信号下提取出相对很弱的有用信息，得到分析结果，所以信噪比对近红外光谱仪器尤为重要。对于高档仪器，一般要求信噪比达到 10^5 。

4.8.7 杂散光强度

杂散光是指分析光以外被检测器接收的光，主要是由于光学器件表面的缺陷、光学系统设计不良或机械零部件表面处理不良与位置不当等引起的，尤其是光栅型近红外光谱仪器的设计中，杂散光的控制非常关键，往往是导致仪器测量出现非线性的主要原因。杂散光对分析测量的影响在分析高吸光度样品时更为明显。

4.8.8 分析速度

近红外光谱仪器往往被用于实时、在线的品质检测和监测，分析样品的数量往往比较多，所以分析速度也是值得注意的一项重要指标。仪器的分析速度主要由仪器的扫描速度决定。仪器的扫描速度是指在仪器的波长范围内，完成一次扫描得到一个光谱所需要的时间。不同仪器类型扫描速度相差很大，如多通道仪器因同时接收全部的光信息，速度取决于电子电路对信息的处理时间上，所以速度很快，一般为几十毫秒；傅里叶变换近红外光谱仪器的扫描速度一般为 1s 左右；而传统的光栅型近红外光谱仪器的扫描速度相对较慢，一般需几分钟，而利用大口径振动凹面光栅，如 4.5 所述，NIR System 公司设计的光栅型近红外光谱仪器，扫描速度达 1.8次/s 。AOTF 型近红外光谱仪器由于采用声光调制产生单色光，所以扫描速度也非常快，一般达 5000个波长点/s 。

4.8.9 软件功能及数据处理能力

“软件就是仪器，仪器技术就是软件技术”，软件是近红外光谱仪器的主要组成部分。近红外光谱仪器的软件一般由两部分组成，一部分是仪器控制平台软件，它控制仪器的硬件，进行光谱数据采集，这部分各个厂家差别不大，并且有可能发展形成一个通用仪器操作平台软件；另一部分是数据处理软件，近红外光谱仪器的数据处理软件通常由光谱数据预处理、校正模型建立和未知样品分析三大部分组成，其核心是校正模型建立部分软件，它是光谱信息提取的手段，直接影响到分析结果的准确性，一些好的软件，都有其独到的建立校正模型的算法，以便尽可能准确地提供样品信息。有关信息可参见技术篇的相关章节。

参 考 文 献

- [1] W Abney, E R Festing. Near-Infrared Spectral of Organic Liquids. Philos Trans R Soc, 1881 (172): 887
- [2] W Kaye. Near Infrared Spectroscopy①: Spectra Identification and Analytical Applications. Spectrochim Acta, 1954 (6): 257
- [3] W Kaye. Near Infrared Spectroscopy②: Spectra in di-and tri-atomic molecules. Spectrochim Acta, 1955 (7): 181
- [4] J R Hart, K H Norris, C Golumbic. Determination of the Moisture Content of Seeds by Near-infrared Spectrophotometry of Their Methanol Extracts. Cereal Chem, 1962 (39): 94~99
- [5] D R Massie, K H Norris. The Spectral Reflectance and Properties of Grain in the Visible and Near Infrared. Trans Am Soc Agric Eng, 1965 (8): 598
- [6] L G Weger. Near-Infrared Spectroscopy of Organic Substance. Applied Spectroscopy Review, 1985 (21): 1
- [7] 范世福. 现代分析仪器发展的前沿技术和新思想. 现代科学仪器, 2000, 3
- [8] 武惠忠, 王百华, 高学军. 国产傅里叶变换红外光谱仪研制动向. 现代仪器, 2001, 5: 14
- [9] 袁洪福, 龙义成, 徐广通等. 近红外光谱仪的研制. 分析化学, 1999, 5
- [10] 张晔晖, 劳彩莲, 吉海彦等. 近红外玉米品质分析仪的研制与应用. 现代仪器, 2001, 5: 11
- [11] 蒋士强. 农业现代化中科学仪器和分析测试技术的应用与展望. 第九届北京分析测试学术报告会及展览会交流论文, 2001
- [12] <http://www.mecasys.com.cn/china/research/research.htm>
- [13] 何嘉耀等. AOTF 在光谱分析中的应用研究 1: AOTF 可见分光光度计的硬件和软件设计. 光谱学与光谱分析, 2002, 22 (1)
- [14] 彭荣飞等. AOTF 在光谱分析中的应用研究 2: 自装 AOTF 可见分光光度计的应用. 光谱学与光谱分析, 2002, 22 (1)
- [15] 陈泽勇等. AOTF 在光谱分析中的应用研究 3: AOTF 在原子发射光谱分析中的应用. 光谱学与光谱分析, 2002, 22 (3)
- [16] 孙振东, 万峰, 赵玉春等. 采用 AOTF 技术的新型红外单色系统的研制. 光学技术, 2002, 28 (1)
- [17] http://www.jicheng.net.cn/special_topic/auto_news/lunwenjingcui/

5. 近红外光谱分析的信号处理

信号是分析过程得到的原始数据，所有分析信号都是由散布在一定时间、波长范围或距离（空间）内的一系列数据组成的。对分析化学家来说，并非每一个数据点或每一个波动都是同等重要的，某些数据仅是噪声或基线。在基线和峰信号上面，分布着噪声，在总信号很难辨认之处，噪声在总信号中或多或少占优势。这种原始数据的质量常用信噪比（SNR）来描述，信号处理的重要目的之一是通过去除噪声，从而区分开信号的有用部分和噪声，提取出有用的信号并因此可增强分析信息，获得高信噪比、低背景干扰的分析信号。

通常广义噪声包括仪器的随机噪声（机械和电子）、基线漂移、测定环境背景（如散射）及干扰组分背景光谱等，分析信号处理包含有分析检测与放大校正、滤波与背景扣除、信号分离等手段，可采用硬件和软件两种手段进行处理和校正。硬件处理涉及电子方面知识，而且实现原理在许多方面与软件方法相似，如微分、积分、平滑和累加等，因此，本章不再介绍。

最重要的信号处理方法有变换、平滑、相关、卷积、微分和积分。对于所有这些方法，目前一般是将连续模拟信息（记录在长图记录仪上的信息）数字化以便可将这些信息存储在计算机内，这些信息的提取则随后通过数字法来实现。

采用软件方法来进行分析信号的处理，尤其是后处理越来越受到重视。常用的有：①滤波方法有积分法、累加/平均法、平滑和厢车平均法（Boxcar）；②背景扣除及信号分离的方法有峰高比法、差谱法、导数法等；③基线扣除和校正方法有线性漂移校正、附加散射背景校正（MSC）等；④数据统计过程还采用中心化（Mean Center）、自动标尺放大（Autoscale）和归一化处理（如 SNV）等。除外还可以采用傅里叶变换、小波变换、主成分分析等方法把信号与噪声背景分离，达到提高分析信号质量的目的。本章将介绍在近红外光谱分析中常用的几种信号处理方法。

5.1 近红外光谱中噪声和背景的扣除方法

常见的高频随机噪声的去除可采用硬件的方法来达到。随着现代分析仪器扫描速度和自动化程度的提高，多次累加和平滑技术已成为常见的一种硬件降噪方式，但它是以增加硬件成本、机时和人力为代价的；而软件降噪（或数字滤波）方法仍不失为一种重要的低成本和快速的处理方法，当然它是以牺牲一定的分辨率和时间作为代价的。

高频随机噪声对定性、定量分析结果的影响因建模算法的不同而不同。一般认为，多元线性回归（MLR）抗噪能力较差，PCR 和 PLS 的抗噪能力较强，这与回归变量选择的不同有直接关系，MLR 是直接光谱变量回归，而 PLS 和 PCR 是潜变量回归。

5.1.1 光谱数据的平滑

光谱平滑的基本思路是在平滑点的前后各取若干点来进行“平均”或“拟合”，以求

得平滑点的最佳估计值，消除随机噪声，这一方法的基本前提是随机噪声在处理“窗口”内的均值为 0。对光谱数据进行平滑处理方法也叫数字滤波器。平滑算法一般可用下式表示：

$$y_k^* = \frac{\sum_{i=k-r}^{i=k+r} a_i y_i}{\sum_{i=k-r}^{i=k+r} a_i}$$

其中， y_k ， y_k^* 分别为第 k 点（中心点）的平滑前、后的值，而 a_i 为平滑系数（或权重），“窗口”宽度为 $2r+1$ 个点，在窗口内进行加权平均，平均区段是逐点后移的。这种平滑处理方法可有效平滑高频噪声，提高信噪比。但由于它是 $2r+1$ 点的平均平滑，对有用分析信号也进行平滑，造成信号失真，降低了光谱分辨率，而且数据最前 r 点和最后 r 点不能做平滑，因此存在一定的局限。

数学上数据的平滑拟合方法有许多，本节主要介绍几种在近红外光谱分析中常用的方法。

5.1.1.1 厢车平均法

厢车（Boxcar）平均是一种常见的平滑方法，此方法把光谱数据分成几个等分，然后用每等分段的中心平均值来代表原等分段的值，如图 5-1 所示，显然，分辨率会有所损失，为保证光谱分辨率，必须要求采集频率大于光谱信号的变化率，而在每个厢车内随机噪声的平均值理论上为 0；每个厢车内数据点的个数越多，则随机噪声平均值越接近于 0，此时光谱分辨率下降光谱失真也会越严重。因此每个车厢的大小（数据点数）要依实际情况而定。这一点从图 5-1 可以看出。

5.1.1.2 移动窗口平均法

这种方法是厢车平均法的改进，也可以称为最小二乘拟合的一种特例，如图 5-2 所示，在移动窗口平均法中，不同厢车的跨度相互重叠，就像窗口依次后移来平滑。假设厢车宽度为 5 个点： X_1, X_2, X_3, X_4, X_5 ，用 X_3^* 来代替 X_3 值， X_3^* 的求法用上 5.1.1

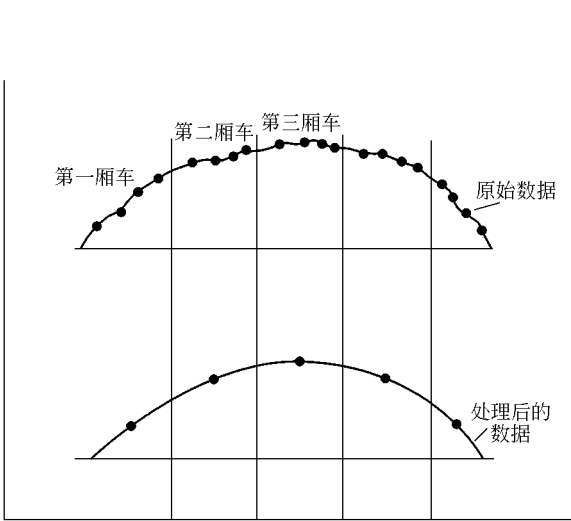


图 5-1 累加平均原理示意图

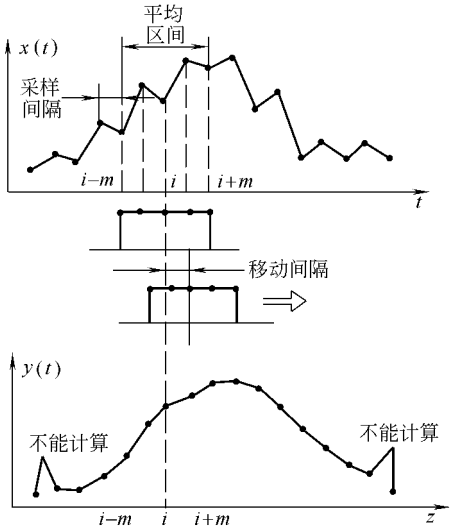


图 5-2 单纯移动平均的原理

中公式求解；而下一个厢车是 X_2, X_3, X_4, X_5, X_6 ，再下一个是 $X_3, X_4, X_5, X_6, X_7, \dots$ ，直至所有数据，最后 N 个原光谱数据被 $N-2r$ 个平滑数据所代替，移动窗口宽度是 $2r+1$ 。这种方法的分辨率损失远小于厢车平均法。

移动窗口平均法计算过程在数学中称为卷积，平滑处理后的光谱 X^* （假设为一向量）可以看成是用滤波函数 W 来对原光谱（向量） X 的卷积结果，即：

$$X^* = W * X$$

对于 5 点移动平滑而言， $W=[1, 1, 1, 1, 1]$ 。

从以上公式中可以看出，移动窗口平均法把窗口内的点平等看待，将导致信号的失真。

5.1.1.3 最小二乘拟合法

1964 年，Savitzky 和 Golay 提出用最小二乘拟合系数作为数字滤波响应函数来对原光谱进行卷积平滑处理的方法，这种方法目前广泛应用于信号的滤波处理。

一般的平滑可用下式表示：

$$X_i^* = \sum_{j=-r}^r X_{i+j} W_j / \sum_{j=-r}^r W_j$$

这里 X_i^* ， X_i 分别是平滑后和平滑前光谱数据向量中的一个元素， W_j 是移动窗口平滑中的权重因子（窗口长度为 $2r+1$ ），对于简单平滑过程， $W_j=1$ ，就是移动窗口平滑；而多项式平滑的原理与之相似，不同的是 W_j 的取值为多项式的拟合系数。

下面以 5 点平滑为例来说明平滑过程，把光谱一段区间的等波长间隔的 5 个点记为 X 集合，多项式平滑就是利用在波长点 $\lambda_{j-2}, \lambda_{j-1}, \lambda_j, \lambda_{j+1}, \lambda_{j+2}$ 的数据的多项式拟合值来取代 X_j ，对于二次曲线拟合，模式函数为：

$$X^* = a_0 + a_1 \lambda + a_2 \lambda^2$$

这里 a_0, a_1, a_2 是待定系数或权重因子。

应用最小二乘法，计算最小误差 ϵ ：

$$\epsilon = \sum (X_j^* - X_j)^2 = \left[a_0 + a_1 \sum_{j=-n}^n \lambda_j + a_2 \sum_{j=-n}^n \lambda_j^2 - \sum_{j=-n}^n X_j \right]^2$$

并令 $\frac{\partial \epsilon}{\partial a_i} = 0$ ，则有方程组

$$\begin{cases} a_0 \sum_{j=-n}^n (1) + a_1 \sum_{j=-n}^n \lambda_j + a_2 \sum_{j=-n}^n \lambda_j^2 = \sum_{j=-n}^n X_j \\ a_0 \sum_{j=-n}^n \lambda_j + a_1 \sum_{j=-n}^n \lambda_j^2 + a_2 \sum_{j=-n}^n \lambda_j^3 = \sum_{j=-n}^n X_j \lambda_j \\ a_0 \sum_{j=-n}^n \lambda_j^2 + a_1 \sum_{j=-n}^n \lambda_j^3 + a_2 \sum_{j=-n}^n \lambda_j^4 = \sum_{j=-n}^n X_j \lambda_j^2 \end{cases}$$

因为 λ_j 值为等间隔， $\Delta\lambda = \lambda_j - \lambda_{j-1}$ 是常数

$$\lambda_j = j \cdot \Delta\lambda,$$

即有：

$$\sum \lambda_j^1 = \Delta\lambda \sum j^1 = 0$$

$$\sum \lambda_j^2 = \Delta\lambda \sum j^2 = 10 \cdot \Delta\lambda$$

$$\sum \lambda_j^3 = \Delta\lambda \sum j^3 = 0$$

$$\sum \lambda_j^4 = \Delta\lambda \sum j^4 = 34 \cdot \Delta\lambda$$

代入方程组有：

$$5a_0 + 10\Delta\lambda^2, a_2 = \sum X_j$$

$$10\Delta\lambda \cdot a_1 = \sum X_j\lambda_j = -2X_{j-2} - X_{j-1} + X_{j+1} - 2X_{j+2}$$

$$10a_0 + 34\Delta\lambda^2 a_2 = \sum X_j\lambda_j^2 = 4X_{j-2} - X_{j-1} + X_{j+1} + 4X_{j+2}$$

整理得：

$$\begin{cases} a_0 = (-3X_{j-2} + 12X_{j-1} + 17X_j + 12X_{j+1} - 3X_{j+2})/35 \\ a_1 = (-2X_{j-2} + X_{j-1} + X_j + 12X_{j+1} - 2X_{j+2})/(10\Delta\lambda) \\ a_2 = (2X_{j-2} - X_{j-1} - 2X_j - X_{j+1} + 2X_{j+2})/(14\Delta\lambda) \end{cases}$$

在移动窗口中心，有 $\lambda_j=0$ 和 $X_j^*=a_0$ ，从上方程组解得：

$$W=[-3,12,17,12,-3]$$

其他不同窗口宽度的权重值见表 5-1。

表 5-1 Savitgky-Golay 系数（权重因子）表

点数	25	23	21	19	17	15	13	11	9	7	5
−12	−253										
−11	−138	−42									
−10	−33	−21	−171								
−09	62	−2	−76	−136							
−08	147	15	9	−51	−21						
−07	222	30	34	24	−6	−78					
−06	287	43	149	89	7	−13	−11				
−05	342	54	204	144	18	42	0	−36			
−04	387	63	241	189	27	87	9	9	−21		
−03	422	70	284	224	34	122	16	44	14	−2	
−02	447	75	309	249	39	147	21	69	39	3	−3
−01	462	78	324	264	42	162	24	84	54	6	12
00	467	79	329	269	43	167	25	89	59	7	17

应用 S-G 方法进行光谱平滑的结果见图 5-3，而且窗口宽度不同，平滑效果不同。在实际应用中，宽度的选择有一定的经验性，它与分析的要求、仪器的条件（如仪器分辨率等）、测试参数及样品的光谱特性等因素有关。

光谱平滑的本质是对光谱曲线进行低通滤波，去掉高频成分，保留有用低频信息。滤波的效果取决于低通滤波器的频谱特性。当权因子不同时低通滤波器的频谱特性不同，滤波的效果也不同。这一点可以从图 5-4 中看出。

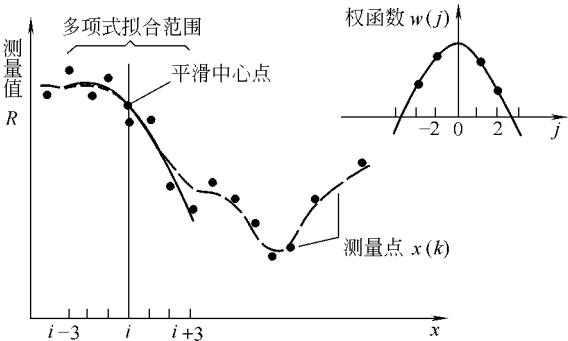


图 5-3 二次、三次多项式拟合平滑原理

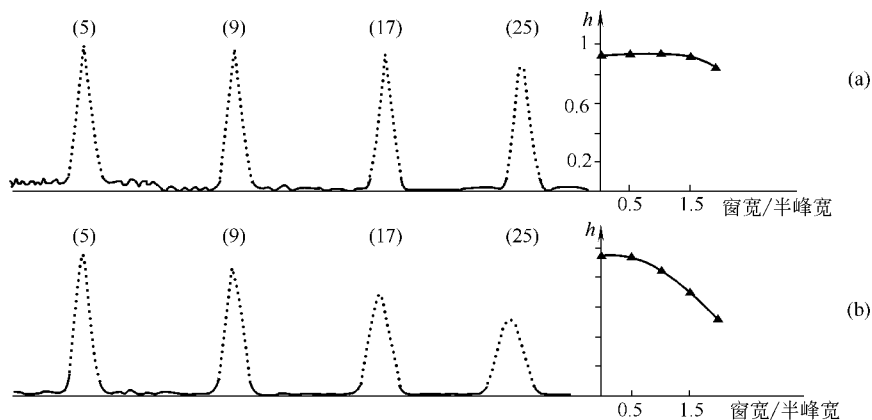


图 5-4 两种卷积平滑方法的比较

(a) S-G 卷积平滑 (b) 移动平均平滑

5.1.2 导数

导数光谱法作为一种常用的光谱处理方法广泛应用于光谱分析中，导数光谱可以利用按波长扫描的光谱数据的微分（逐点相减）而得到，也可用微分电路来实现。而现代光谱分析一般都是用数学方法计算来得到一阶、二阶……导数光谱。如图 5-5 所示。

导数光谱既可以消除基线漂移或平缓背景干扰的影响，也可以提供比原光谱更高的分辨率和更清晰的光谱轮廓变化。如图 5-6 所示，原光谱中 2 个高斯峰重叠严重，但在一阶导数光谱中二峰更清楚。另外在确定肩峰的大小中，二阶导数光谱也比原光谱更清楚（见图 5-7），求导后肩峰放大了。

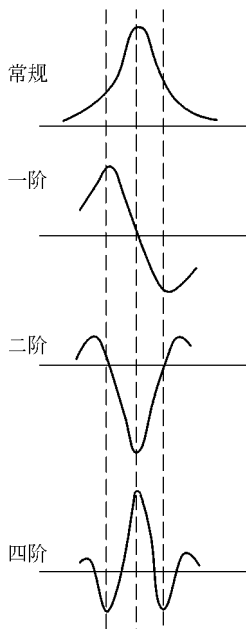


图 5-5 高斯曲线导数光谱

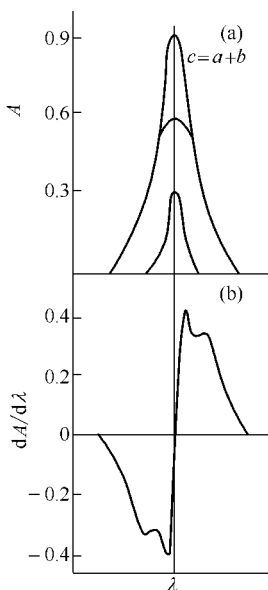


图 5-6 重叠吸收曲线及其透过率的一阶导数曲线

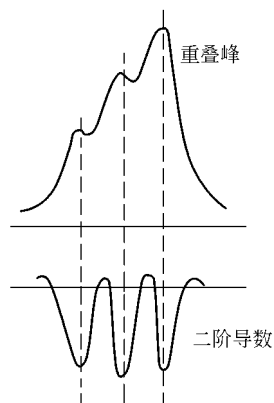


图 5-7 二阶微分使重叠峰分离为单峰

当然，在上述讨论中忽略了噪声的存在。如果考虑噪声的影响，情况将更加复杂，因为随机噪音一般都是高频信号，导数将进一步放大噪声信号，因此，必须要求原光谱有很高的信噪比。

光谱数据的求导方法很多，假设原光谱数据的 X 轴（波长）是等间隔的，则一阶导数为：

$$\frac{dy}{d\lambda} = \frac{y_{i+1} - y_i}{\Delta\lambda} \quad \text{或} \quad \frac{dy}{d\lambda} = \frac{y_{i+1} - y_{i-1}}{2\Delta\lambda}$$

二阶导数为：

$$\frac{d^2y}{d\lambda^2} = \frac{y_{i+1} - 2y_i + y_{i-1}}{\Delta\lambda^2}$$

其中 y 为吸光度。采用 Savitzky Golay 多项式拟合方法，利用权重系数确定平滑函数来对原光谱进行卷积处理；若以 5 点二次平滑方法，则有如下公式：

$$\frac{dy}{d\lambda} = \frac{1}{10\Delta\lambda} (-2y_{i-2} - y_{i-1} + y_{i+1} + y_{i+2})$$

$$\frac{d^2y}{d\lambda^2} = \frac{1}{7\Delta\lambda^2} (-2y_{i-2} - y_{i-1} + 2y_i + y_{i+1} + 2y_{i+2})$$

上述公式利用多个波长点处的信息，因此拟合效果较好。采用平滑拟合一阶导数各系数见表 5-2。

表 5-2 最小二乘移动平滑法计算平滑谱的一阶导数公式中的 N_k 值和 $C_{k,j}$ 值

j $2K+1$ $C_{k,j}$	17	15	13	11	9	7	5
—8	748						
—7	—98	12922					
—6	—643	—4121	1133				
—5	—930	—14150	—660	300			
—4	—1002	—18334	—1578	—294	86		
—3	—902	—17842	—1796	—532	—142	22	
—2	—673	—13843	—1489	—503	—193	—67	1
—1	—358	—7506	—832	—296	—126	—58	—8
0	0	0	0	0	0	0	0
1	358	7506	832	296	126	58	8
2	673	13843	1489	503	193	67	—1
3	902	17842	1796	532	142	—22	
4	1002	18334	1578	294	—86		
5	930	14150	660	—300			
6	643	4121	—1133				
7	98	—12922					
8	748						
N_k	—23256	334152	24024	5148	1188	252	12

5.2 傅里叶变换滤波

5.2.1 连续傅里叶变换

傅里叶变换是时间（或空间）域函数 $f(t)$ 与频率域函数 $F(v)$ 间的数学关系，

即有：

$$f(t) = \int_{-\infty}^{\infty} F(v) e^{-2\pi jtv} dv$$

或：

$$F(v) = \int_{-\infty}^{\infty} f(t) e^{-2\pi jtv} dt \quad \text{其中 } (j = \sqrt{-1})$$

$f(t)$ 称为时（空）域函数， $F(v)$ 称为频率域函数，即时空域与频率域函数可通过一个变换来统一。一般地， $F(v)$ 即是常见的谱图，而 $f(t)$ 为波形图（干涉图），在傅里叶型仪器中，直接采集到的是 $f(t)$ 信号，然后通过快速傅里叶变换（FFT）而得到谱图 $F(v)$ 。傅里叶变换在数据处理中应用的主要目的是加快提取信息的过程，通过压缩数据使信息提取更方便有效，同时可以去掉干扰信号和噪声。图 5-8 是利用傅里叶变换进行信号滤波的实例。

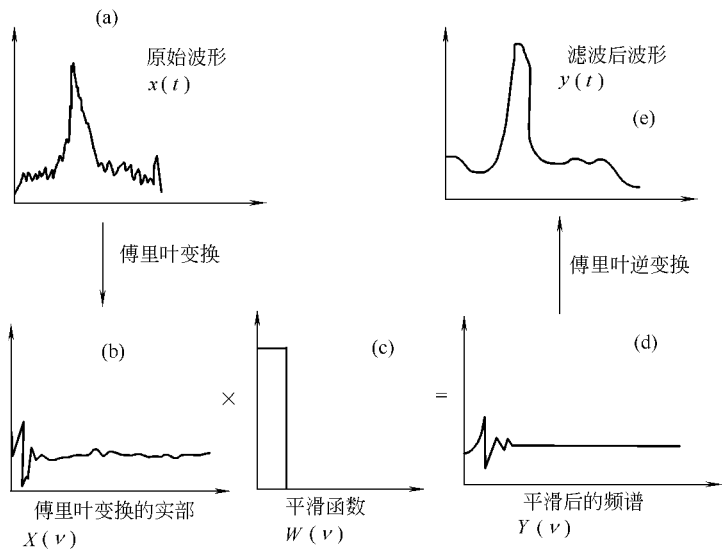


图 5-8 傅里叶变换滤波示意图

5.2.2 离散傅里叶变换

通常检测器采集到的 $f(t)$ 是由有限长度、等间隔的数据点 $f(t_n)$ 构成的序列，即是数字信号，有：

$$f(t) \in \{f(t_n)\}, n=0, \dots, N-1$$

离散傅里叶变换可写成：

$$F(W_m) = \sum_{n=0}^{N-1} f(t_n) e^{-2\pi jnm/N} \tag{5-1}$$

其逆变换为：

$$f(t_n) = \frac{1}{N} \sum_{m=0}^{N-1} F(\omega_m) e^{2\pi jnm/N} \quad (n = 0, \dots, N-1) \tag{5-2}$$

直接计算上述一个方程需 N^2 次乘法运算。1965 年，Cooley 和 Turkey 发表快速傅里叶算法，下面以 $N=4$ 说明这个算法过程，由式 (5-1) 有：

$$\begin{bmatrix} F(0) \\ F(1) \\ F(2) \\ F(3) \end{bmatrix} = \begin{bmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{bmatrix}$$

其中 $W_4^1 = W_4^{1+4N}$ ，又 $e^{jQ} = \cos Q - j \sin Q$ ，因此有：

$$\begin{bmatrix} F(0) \\ F(2) \\ F(1) \\ F(3) \end{bmatrix} = \begin{bmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^2 \\ 0 & 0 & 1 & W^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ f(2) \\ f(3) \end{bmatrix}$$

注意等号左边的变化，计算上述结果需 4 次乘法和 8 次加法，比原来 16 次乘法，12 次加法少多了，尤其是当 N 很大时，这种算法非常有效。对于等号左边，上半部为偶序数，下半部为奇序数，光谱还原时必须注意这一特点。

5.2.3 傅里叶变换在光谱处理中的应用

傅里叶变换在光谱信号处理中的应用主要有两方面：压缩数据/特征提取和滤波处理。通常 $f(t_n)$ 信号的数据量很大，如近红外光谱，一般都有上千个数据点，这给数据处理增加很大的计算量，但是，如果把 $f(t_n)$ 变换为 $F(W_n)$ 来处理，数据点可以少得多，因为原光谱主要信息集中在 $F(W_n)$ 的前若干个值上，如图 5-9 所示，前 6 个点的数据即可

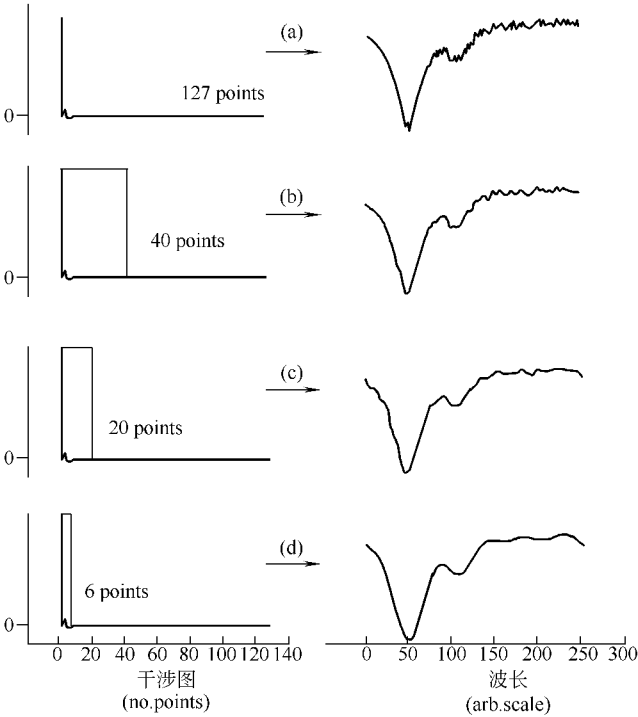


图 5-9 离散傅里叶变换特征提取示意图

代表原光谱的主要轮廓，即代表原光谱的主要信息特征，因此，用傅里叶变换系数来进行光谱分析计算将显著降低计算量，McClure 把此方法应用于近红外光谱的特征提取，应用傅里叶变换系数与化学成分进行回归，确定烟草的化学成分值，取得满意结果。

在傅里叶变换滤波方面的应用很多，从图 5-9 可以看出，通过选不同宽度的脉冲函数来卷积，得到的光谱分辨率是不同的，而高频随机噪声可通过选用较窄的脉冲函数来卷积而被滤掉，脉冲函数宽度的选择要依靠实际经验而定。

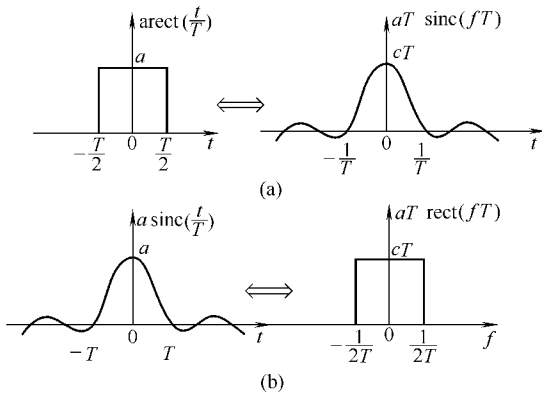


图 5-10 矩形脉冲 (a) 和 sinc 函数 (b) 以及它们的傅里叶变换

(带通) 影响着滤波的效果。通过高频滤波可以去掉高频噪声，而低频滤波器则可以扣除漂移。在实际算法实现中，一般采用离散快速傅里叶变换算法。

实际上，时间域的脉冲函数与频率域的 sinc 函数是对应的（对偶），在时间域用方波函数卷积也就相当于在频率域乘以 sinc 函数，而且时间域的脉冲函数越宽，频率域的 sinc 函数越窄，反之亦然，如图 5-10。尺度变换意味着当 $f(t)$ 在时域中越宽，则其频谱越窄，反之亦然，如图 5-11，我们自然会想到对于一门脉冲来说，一个信号的带宽与其时域的持续时间成反比，因为门脉冲的宽度为 τ ，其带宽是 $1/\tau$ （赫兹），当然，这一结论对于其他信号也是实用的。因此 sinc 函数是一个窗口平滑函数，它的宽窄

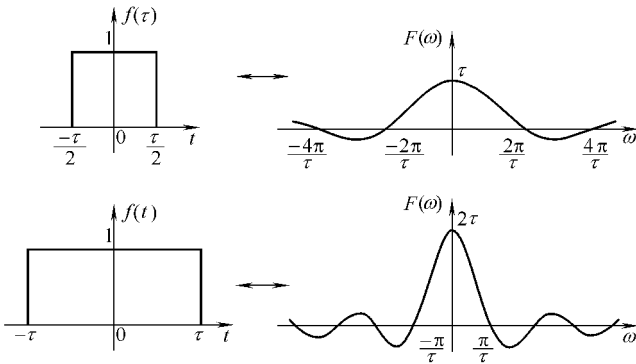


图 5-11 傅里叶变换尺度变换（波形缩放特性）形象图形

5.3 小波变换滤波

小波（又称子波）分析是近年来国际上掀起热潮的一个前沿领域，被认为是现代傅里叶分析半个世纪以来的工作结晶，在其他领域，特别是在信号分析、图像处理、语音合成、模式识别、量子物理和非线性科学等方面，它被看作近年来在工具和方法上的重大突破。

小波变换在信号分析中有广泛应用，可用于数据压缩和滤波处理，小波变换的思想与“加窗”傅里叶变换类似，只是小波变换中，“窗口”大小是随频率变化的。在上节中，傅里叶变换可把时间（或空间）域函数变成频率域来描述，我们可以得到原函数的频率分布情况，但具体每个频率分量对应原光谱的哪些位置（波长或波数）则不能得到。而小波变换可把原光谱按不同的分辨率分解成一系列不同频率的块信号，原光谱的每一个波长点的各频率分量一目了然，给信号处理带来极大方便，依经验可对特殊位置或特殊频率范围进行处理。下面介绍用来滤波的样条小波变换的算法思想。

5.3.1 小波变换的理论基础

傅里叶变换把时域特征和频域特征联系起来研究，既可把时域函数 $f(t)$ 变换为频域函数 $F(\omega)$ （正变换）来研究函数的特征，又可把频率域函数 $F(\omega)$ 经过反变换（重构）回原来函数，在反变换前，我们也可以对 $F(\omega)$ 进行特殊“修饰”分析和处理，傅里叶变换成为信号分析和处理的基本数学手段。从数学讲，傅里叶变换相当于把函数 $F(\omega)$ 表示为函数 $f(t)$ 在一函数簇 $\{1/(2\pi)e^{-i\omega t}\}$ 为规范正交基所构成的复空间 $L_2(0, 2\pi)$ 上的投影。

然而，傅里叶变换只能提供信号（或函数）的整体特征，如 $f(t)$ 的傅里叶变换 $F(\omega)$ 只能提供 $f(t)$ 中各个频率分量的强度，我们不能从 $F(\omega)$ 知道某时刻某频率分量的强度，即 $f(t) \sim \omega - t$ 之间缺乏一一对应关系。为此，Gabor D. 在 1946 年引进窗口（或短时）傅里叶变换，即：

$$F(\omega, \tau) = \int_{-\infty}^{+\infty} f(t)g(t - \tau)e^{-i\omega t} dt \tag{5-3}$$

其中函数 $g(t)$ 是权函数，称之为窗口函数（即 $g(t)$ 在一定范围内有值，范围外的取值为 0）。随 τ 的变化，窗口 $g(t)$ 在 t 轴上平移，使 $f(t)$ “依次”进入被分析状态，如图 5-12 所示， $F(\omega, \tau)$ 反映了 $f(t)$ 在 τ 时刻，频率为 ω 的信号相对强度，把时频对应起来，称时频分析法（或时频局部化方法），其逆变换公式（重构）为：

$$f(t) = \int_{-\infty}^{+\infty} F(\omega, \tau)g(t - \tau)e^{-i\omega t} d\tau d\omega \tag{5-4}$$

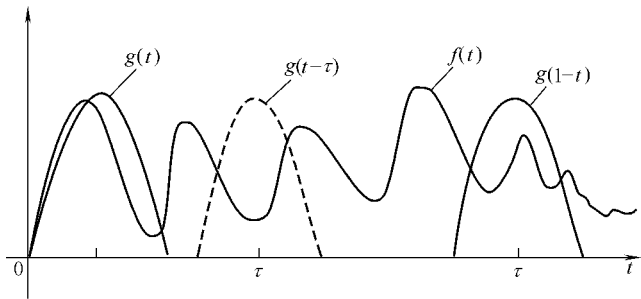


图 5-12 加窗傅里叶变换的模式

加窗傅里叶变换由于所选的窗口大小固定，当原函数 $f(t)$ 是慢变化时，因窗口太小而只能“看到”局部变化，而当 $f(t)$ 变化激烈时，又因窗口太宽而分辨不清，实际上窗口函数可被理解为有一定带宽的“滤波器”，其能通过的频率是一定的。

近年来,小波分析的出现,为我们提供了灵活可变的窗口,对任意连续信号 $f(t)$,连续小波变换可定义为:

$$Wf(s, \tau) = \frac{1}{\sqrt{S}} \int_{-\infty}^{+\infty} f(t) \cdot \Psi_{s, \tau} \frac{ds dt}{S^2} \quad (5-5)$$

其中 S 称伸缩系数 (也叫尺度), τ 称平移因子。 $\Psi\left(\frac{t-\tau}{S}\right)$ 是母小波 $\Psi(t)$ 经 τ 平移, S 伸缩后的小波函数, 记 $\Psi_{s, \tau}$ 。通常使用的小波函数有 Danbueshies 小波、Haar 小波、二进小波等。

对连续小波变换进行二进制离散化处理, 即令: $S=2^m (m \in Z)$, $\tau=n\tau_0$, 则离散二进小波函数为: $\Psi_{m, n}(t) = 2^{-m/2} \Psi(2^{-m}t - n\tau_0)$, 理论证明, 这样构造的小波函数是两两正交的。不同的 m 取值即形成伸缩的窗口 (频率成分相当于 2^{-m}), 不同的 n 取值形成窗口的不断平移, 而使信号 $f(t)$ 进入“分析”状态, 因此, 小波变换即是将交织在一起的不同频率组成的混合信号用分辨率不同的窗口分解成对应的不相同频率的块信号, 并对大小不同的频率成分采用相应粗细的时域 (或空域) 取样步长, 从而能够不断“聚焦”对象的任意微小细节。

与 FT 离散变换类似, 对 $f(t)$ 的数字信号 $\{C^0\} \in 1$, 其中二进小波变换的 Mallat 分解算法为:

$$C_n^j = S_2^j f(n) = \sum_{k \in z} C_k^{j-1} h_{k-2n} \quad (5-6)$$

$$d_n^j = W_2^j f(n) = \sum_{k \in z} C_k^{j-1} g_{k-2n} \quad (5-7)$$

且有 $S_2^0 \cdot f(n) = C_n^0$ 即原始数字信号 S_2^j 称光滑算子, 即为 $\{C_n^0\}$ 的 j 尺度分辨下的模糊信号; W_2^j 称小波变换, $\{d_n^j\}$ 为 $\{C_n^0\}$ 的 j 尺度下的小波变换信号。

从式 (5-5)、式 (5-6) 可以看出, 小波变换分解并不涉及小波函数的具体形式, 只要有 $\{h_k\}$ 和 $\{g_k\}$ 的值就可以了。在对小波变换的分解信号进行处理 (如去噪) 后再重构, 得到原信号经处理后的重构信号 $\{\tilde{C}_n^0\}$, 公式为:

$$\tilde{C}_n^{j-1} = \sum_{k \in z} C_k^j h_{k-2n} + \sum_{k \in z} d_k^j g_{k-2n} \quad (5-8)$$

通过小波变换可把各种频率组成的混合信号按不同的分辨尺度分解成一系列不同频率的块信号。根据这一特点, 可对特殊频率范围的噪声或慢背景进行滤波处理, 如高频随机噪声的滤波, 可设定一个阈值, 令高频部分小于阈值的小波信号 $W_2^j f(n)$ 为 0, 然后再重构, 即可滤掉随机噪声, 同样, 也可以对白噪声或慢漂移进行扣除。如果需要, 加以特殊的“修饰”, 甚至可以对倍频等其他干扰进行平滑处理, 小波变换灵活滤波的特点是其他方法所不能以拟的。

5.3.2 小波变换在分析信号处理中的应用

分析化学中所采集的分析信号通常涉及峰形 (频谱) 特征的描述、信息提取光谱数据的压缩处理及改善分析信号的质量——即提高信号的信噪比 (S/N) 和复杂背景的校正, 以提高分析的准确度和精密度等, 由于小波分析是时频分析, 信号的局域行为与小波系数

的行为之间存在一定的联系，如果信号 $f(t)$ 是光滑的（慢变化），则相应的小波系数很小；如果 $f(t)$ 含有一个奇异点（激烈变化），则该奇异点附近的小波系数急剧增大。由于小波变换可以看作是信号 $f(t)$ 按不同的频率段分解的，这样，我们就可以只对某频率段或特定时刻（位置）的局部信号进行特殊的频谱处理，而不影响总体信号，重构变形很小。另外，还可以依据先验知识，很方便地对高频噪声、慢漂移背景及工频、倍频等噪声进行滤波处理，提高信噪比。由于小波变换滤波几乎不影响分辨率，且变形小，相对其他滤波方法具有明显的优点。目前，小波变换在以下几个方面得到应用：

① 改善分析信号的质量：目前主要用于滤波处理。使用不同的正交小波基对烟草的近红外漫反射光谱信号进行滤波处理，极大地提高 S/N 。王洪等人还将小波变换用于对 HPLC 分析中的基线漂移的校正；另外，我们还将小波变化应用于对近红外光谱进行不同降噪要求的分段滤波并重构成功。

② 光谱数据压缩：以最小的信息损失获得最大的压缩比，考虑计算机的容量及运行速度，光谱数据压缩在光谱数据存储、检索及光谱分析计算中有着重要（甚至是决定性）的作用，小波分析在电生理信号特征的描述及其在医学诊断上成功的应用，预示着运用小波变换对光谱特征信号的信息提取及应用必将有重大进展，随着小波分析在分析化学中应用的深入，它在二维乃至多维光谱图像的压缩及处理等方面具有巨大的潜力是显而易见的。

③ 小波变换用于光谱的特征提取：小波系数是原分析信号的主要特征参数，因此可以代表原分析信号的主要特征，因此小波变换的信号分解也是信息特征提取的过程。目前，小波变换算法已经成功的应用于近红外光谱波长范围的优化，小波系数直接用于近红外分析定量分析模型的建立和定性分析的模式识别判别模型研究中。

5.4 光谱基线的校正与数据规范化处理

5.4.1 附加散射校正（Multiplicative Scatter Correction, MSC）

在漫反射光谱测量中，由于样品的不均匀性（粒度分布）常导致所测的样品光谱具有很大的差异性，不同批号的同一种样品的所测的光谱从表面看也可能有很大的差异。在许多情况下，散射引起的光谱变化可能要大于样品成分引起光谱的变化。散射的程度与光的波长、颗粒度和样品的折射指数等有关，因此在整个光谱范围内，散射的强度是不同的，通常表现为基线的平移、旋转、二次和高次曲线，在长波近红外情况下这种影响表现的尤为明显。

附加散射校正（Multiplication Scatter Correction）是由 Geladi 等人提出的，目的是校正每个光谱的散射并获得较“理想”的光谱。附加散射校正法假定与波长有关的散射对光谱的贡献和成分的贡献是不同的，理论上说，通过光谱上许多点的数据分析，可以把这两部分分开。MSC 方法认为每一条光谱都应该与“理想”光谱成线性关系，而真正的“理想”光谱无法得到，可以用校正集的平均光谱来近似。因此每个样品的任意波长点下反射吸光度值与其平均光谱的相应吸光度的光谱是近似线性关系，直线的截距和斜率可由光谱集线性回归获得，并用以校正每条光谱，截距（即附加效应）大小反映样品独特反射

作用而斜率大小则反映样品的均匀性。具体计算公式如下：

平均光谱：

$$\overline{A}_j = \sum_{i=1}^n A_i / n$$

线性回归：

$$A_i = m_i \overline{A}_i + b_i$$

MSC 校正：

$$A_{i(MSC)} = \frac{(A_i - b_i)}{m_i}$$

上述公式中， A 是校正集的光谱矩阵， A_i 为第 i 个样品的光谱， m_i 和 b_i 是第 i 个光谱 A_i 与平均光谱 A 的线性回归的斜率与截距，都是列向量。通过调整 m_i 和 b_i 的不同，使得在减少光谱差异的同时尽量保留原有的与化学成分有关的信息。这并不是说经 MSC 校正的光谱就是样品的真实光谱，只能说通过这样的校正，随机变异得到最大可能的扣除。在光谱与浓度线性关系较好和化学性质相似的情况下，MSC 校正的效果较好，因此，光谱纵坐标应为 $\text{Log}(1/R)$ 或 Kubelka-Munk。如果校正集样品的光谱间由于化学组成的变化而引起的光谱出现显著性的变化，采用 MSC 校正的效果一般不会理想。

5.4.2 变量标准化 (Standard Normalized Variate, SNV)

与 MSC 类似，SNV 校正也可用来校正样品间因散射而引起的光谱的误差。但是，两者的算法又有所不同。SNV 校正认为每一个光谱中，各波长点的吸光度值应满足一定的分布（如正态分布），通过这一假设对每一条光谱进行校正，无需“理想”光谱。SNV 是原光谱减去该光谱的平均值 μ 后，再除以该光谱数据的标准偏差 σ （标度化，scaling），实质是使原光谱数据标准正态化，即：

$$Z_i = \frac{x_i - \mu}{\sigma}$$

而多项式拟合可以扣除漫反射光谱中粒度和装样的误差，Detrending 算法采用最小二乘法对原光谱进行线性拟合，从而扣除原光谱中的线性漂移。一般单独应用 SNV 的载荷向量阵的解释较容易，Detrending 一般与 SNV 一起使用，很少单独使用。见图 5-14 (b)。

与 MSC 校正类似，SNV 校正也只适用于扣除样品光谱中的线性平移的影响（如用来校正样品的颗粒度和附加散射的影响），因此光谱也必须是以 $\text{Log}(1/R)$ 或 K-M 函数表示。由于 SNV 是对每条光谱单独进行校正，因此一般认为它的校正能力比 MSC 要强，尤其是样品组分变化较大的时候。

图 5-13 (a) 是一部分烟草样品的原光谱；图 5-13 (b)、图 5-14 (a)，(b) 和图 5-15 (a)，(b) 是分别用 MSC、SNV、SNV+Detrending 和导数方法处理后的效果。

5.4.3 数据规范化

(1) 中心化 (Mean Center)

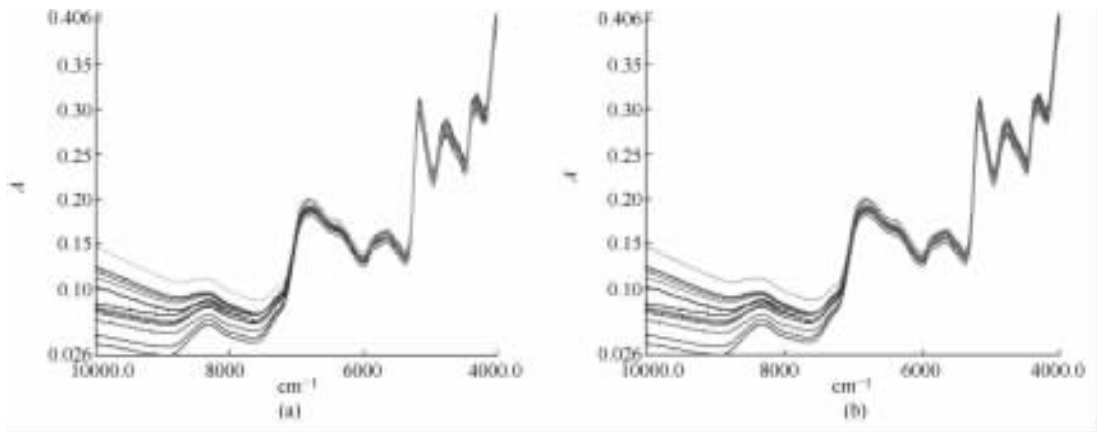


图 5-13 烟草样品的原光谱与校正结果

(a) 原光谱 (b) MSC 校正结果

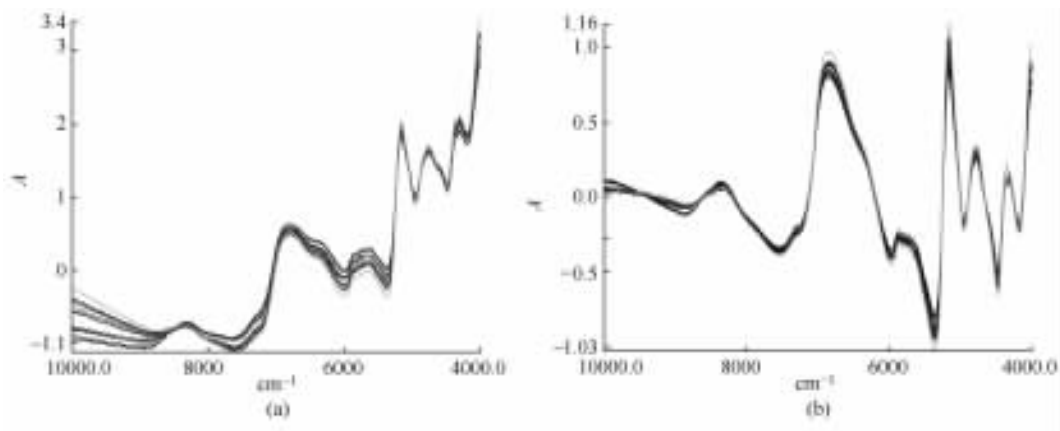


图 5-14 SNV、SNV+Detrending 校正结果

(a) SNV 校正结果图 (b) SNV+Detrending 校正结果

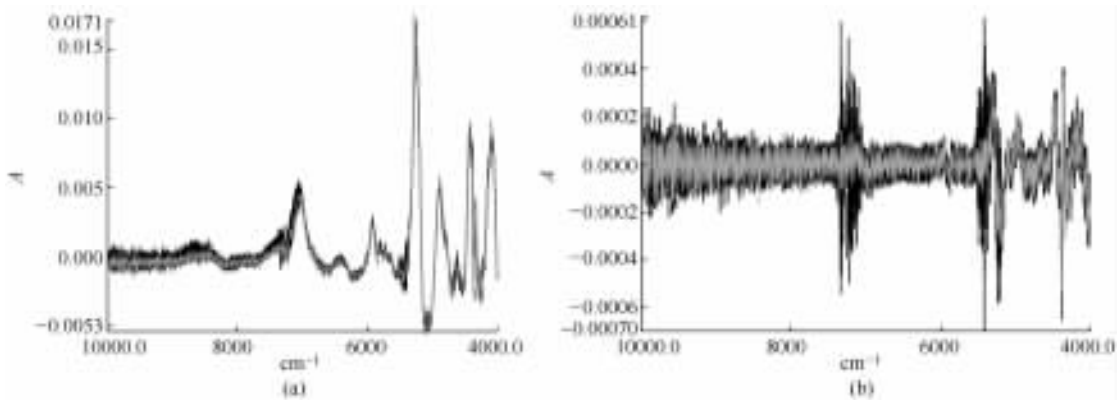


图 5-15 导数方法处理效果

(a) 一阶导数 (b) 二阶导数

数据规范化方法的目的是改变数据集空间的坐标和原点，通常原点取数据集的平均值。因此第 i 个样品的第 k_0 个测量值 X_{ik} 的中心化数据 X_{ik}^* 为：

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{S_k}$$

且

$$S_k = \left[\frac{1}{NP-1} \sum_{i=1}^{NP} (X_{ik} - \bar{X}_k)^2 \right]^{1/2} \quad (NP \text{ 为总样品数})$$

通过以上变换，把数据阵 X 变为 X^* ，使得数据的变化以平均值为原点，变化的特征性更一目了然，这种变换称中心化处理（Mean Center）。

(2) 归一化（Normalization）

标准化处理设定数据集中所有的数据向量的长度相同，每个向量中各元素的平方和相似，令：

$$\sum_{k=1}^{NV} X_{ik}^2 = C_i$$

为把数据向量归一化为 N ，将向量中的每个元素乘以 $N/\sqrt{C_i}$ 。令向量为单位长度或单位面积是常用的方法，即归一化方法；归一化方法使向量的方差为 1，可有效消除一组测量中的随机误差。向量归一化常用于对近红外光谱的标准化处理。

(3) 标尺缩放（Scaling）

在画图时，通常都希望充分利用纸面空间，使图尽量充满整图；但在数据处理中，标尺缩放的方法要严格得多。如果事先并不知道数据集的情况，我们常依据数据集的方差来作为缩放的标准（Autoscaling），另外也可依据数据的最大值（max）与最小值（min）间距来作为缩放标准（Range Scaling），但这种方法受异常值的影响很大。

自动缩放（Autoscaling）由于采用随机标尺而降低偶然失误造成的影响，相对于范围缩放（Range Scaling）而言对异常值不那么敏感，通常有 2 种自动缩放方法，都包括中心化处理，缩放因子不一样，但效果是相同的。

最常用的方式，自动缩放即：

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{\left[\sum_{i=1}^{NP} (X_{ik} - \bar{X}_k)^2 \right]^{1/2}}$$

以向量最大变化为缩放因子的缩放称极差缩放或范围缩放，即进行以下坐标变换计算：

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{\max\{X_{ik}\} - \min\{X_{ik}\}}$$

近红外光谱校正还可以使用正交信号校正（Orthogonal Signal Correction, OSC），该方法使用偏最小二乘法对原光谱矩阵中与浓度（或特性参数）不相关（即正交）的部分信号求出并予以扣除，然后以校正后的信号和浓度（或特性参数）重新进行方差最大相关回归，可以获得更稳定的近红外分析定量模型，详细内容参见有关文献。

技 术 篇

近红外光谱分析是一种需要软件支撑的实验技术，特别是建立数学模型的过程，通常需要由专业人员来掌握各种近红外光谱分析数据处理的度（如建模样品数目的多少、光谱区的宽窄、异常样品的取舍、模型适配范围的大小等等），以最大限度地挖掘所需信息，建成数学模型；这个过程需要一系列的反复试建、测试、检验、比较、试用、判断和选定，其中一些步骤需要人为的干预，才能最后锁定分析模型的各项参数。这些事实说明：近红外光谱分析是一门还在继续发展的技术，进行近红外光谱分析不但要有好的仪器，还要有能指导分析的优秀软件，此外还必须具备近红外光谱分析的样品资源，尤其是近红外分析的专业人员需要了解该技术并有一定经验，这些是实际应用近红外光谱分析的困难。

本篇共 8 章，介绍了样品的分析测试技术到近红外光谱分析的资源共享等内容，希望对该技术目前的状况做一些概括。其中也介绍了作者们的一些经验，以及某些有待和读者继续探讨的看法。本篇可以为从事近红外光谱分析的人员进一步应用与发展近红外技术提供参考。

6. 近红外光谱分析测试技术

近红外（NIR）光谱分析技术作为近年来迅速发展起来的一种高新分析技术，在分析测试技术上与其他技术相比具有如下特点^[1]：

（1）样品一般不需预处理

近红外谱区的信息是分子振动频率（中红外谱区的信息）的合频和倍频，摩尔吸光系数较小，一般较中红外基频吸收低 1~3 个数量级，因此样品不需要像中红外或其他分析技术那样需要溶解、消化、萃取等一系列预处理过程，近红外光可以直接穿透样品，样品一般不需预处理，这是近红外光谱分析技术的一大优点，正是这一优点使得近红外技术适于快速、无损分析。

但也正因为近红外技术的上述特点，使得近红外分析的检测限要差于其他分析技术，通常难以检测含量低于 0.1% 的微量物质。

（2）可以利用漫反射技术

近红外光谱分析技术的样品一般不做化学预处理，样品颗粒大，近红外光的波长远小于颗粒直径，光在样品中传播时散射效应大，而且能够穿透到样品内部，携带内部信息，这使得近红外光谱技术可以用漫反射技术对样品直接测定，而漫反射分析的样品可以是形状各异的，如水果、蔬菜、谷物、固体药丸等都可以直接测定，大部分固体样品的在线检测也都选用漫反射技术。这一特点使得近红外分析便于检查样品的各个部位或单籽粒的质

量，也非常适合在线分析。

特点（1）和（2）使得近红外技术极大地方便了用户，样品的分析不再像传统意义上的化学分析，变得简便、快速。

（3）属绿色分析技术

近红外分析样品不需要化学预处理，因此不消耗化学试剂，不会造成环境污染。此外，近红外光光子能量低，约为 $1.65 \sim 0.5 \text{ eV}$ ，不会对实验者造成伤害，因此属绿色分析技术。

（4）一般不用于无机分析

近红外谱区的信息是 C—H 、 N—H 和 O—H 等含氢基团的合频与倍频吸收，因此近红外光谱技术几乎可以用于所有与含氢基团有关的样品化学性质与物理性质分析，一般不用于无机分析。但也有用近红外方法测定无机成分的文献报道，如测定烟草中的总氯、总钾等，近红外光谱中并没有这些元素的直接信息，计算利用的可能是间接信息。

近红外分析技术的成功应用依靠优秀的数学模型，合理的实验设计则是建立优秀数学模型的基础。近红外光谱分析技术在测试技术上的上述特点，使近红外光谱分析在分析测试过程中与其他分析技术相比有自己的特色，现就近红外光谱分析中应该注意的分析测试技术方面的问题，按照分析流程，从仪器的选择、工作方式的选择，校正样品集的选择，测量条件对测试结果的影响，样品状态对测试结果的影响等几个方面讨论如下。

6.1 仪器的选择

20 世纪 60 年代后，由于计算机技术的发展，以及许多多元校正方法的运用，NIR 技术的应用得到了迅猛发展，近红外仪器也随着 NIR 技术的完善、NIR 应用领域的拓宽以及实际工作的需要逐渐发展，并由此诞生了一系列 NIR 分析仪器。

生产近红外光谱仪的厂家有国外的，也有国内的；近红外光谱仪器的类型有傅里叶变换型，也有滤光片式的；有通用近红外光谱仪，也有针对特殊应用的专用近红外仪；有实验室用近红外光谱仪，也有在线检测用近红外光谱仪；每一类型各有所长，也各有所短。到底选择什么样的近红外光谱仪器好呢？是选价位较低、分离波长、便携式的专用近红外光谱仪，还是选择价位较高、连续波长、信噪比、分辨率高的通用近红外光谱仪？是实验室用仪器还是现场分析用仪器？都要根据需解决的具体问题和具体情况（如财力、精度要求等）来定，并不一定价位最高的就是好仪器，只有适合自己要求的仪器才是“最好”的仪器。

下面就 5 种类型的近红外光谱仪从仪器的选择角度做简单介绍，关于仪器及工作原理的介绍详见第 4 章内容。

① 滤光片型近红外光谱仪：又分固定滤光片和可调滤光片，这类仪器单色光的带宽较宽、波长分辨率较差，主要作专用分析仪器，如粮食水分测定仪、油品分析仪、水果品质分析仪等。这种类型的仪器只能获得几个特定波长的光谱数据，结构简单，一般仪器价位较低，有些仪器还配备数学模型，使用简单方便，只能用于特定的应用领域。日常分析蛋白质、水分、油分、淀粉这些组分一般选用该类型仪器就足够了。但也有通用的滤光片型近红外光谱仪，如 BRAN+LUEBBE 公司和 PERTEN 公司都开发了此类仪器。这类仪器其滤光片的选择基本定位是农产品分析，因此农业领域应用可选用此类仪器，需要用户

自己建模，价位中上。

② 光栅扫描式近红外光谱仪：采用全息光栅分光、PbS 或其他光敏元件作检测器，可在一定波长范围内进行全谱扫描，波长准确度较高、信噪比较好，一般扫描时间较长。由于仪器有移动部件（如光栅轴），存在磨损问题，从而影响光谱采集的可靠性，一般不适合在线分析，但也有这类仪器用于生产线检测的报道。

③ 傅里叶变换近红外光谱仪：是目前国内近红外光谱仪器的主导产品，波长精确度、准确度高，信噪比好，分辨率高，扫描速度快。此类仪器一般属通用型仪器，需要用户自己建模，使用范围广，价位较高。

④ 声光可调滤光器（AOTF）近红外光谱仪：是 20 世纪 90 年代初出现的新型分光器件，采用双折射晶体，改变射频频率调节扫描的波长，无移动部件，扫描速度快，具有较好的仪器稳定性，特别适用于在线分析，但分辨率相对较低，价格也较高。

⑤ 多通道检测器近红外光谱仪：随着多通道检测器件生产技术的日趋成熟，采用固定光路、光栅分光、多通道检测器构成的 NIR 仪器，以其性能稳定、扫描速度快、分辨率高、性能价格比好等特点正越来越引起人们的重视，在与固定光路相匹配的多通道检测器中，常用的有二极管阵列（Photodiode-array 简称 PDA）和电荷耦合器件（Charge Coupled Devices 简称 CCD）两种类型。

若测试要求精度高，或进行科学研究，可以选用上述第②、③、④、⑤类仪器，这样可以在整个近红外波长范围内进行 PCR/PLS 回归及其他处理；日常常规分析 5 种类型的仪器都可以选用。

6.2 工作方式的选择

近红外光根据波长的长短分为长波近红外光（1100~2600nm）和短波近红外光（700~1100nm）。长波近红外的信息源是分子内部原子间振动的合频（Combination Tones）和一、二级倍频（Over Tones）（又称二倍频、三倍频），样品在此谱区较短波近红外摩尔吸光系数大，吸收较强，穿透力较弱；短波近红外的信息源是分子内部原子间振动的三级、四级倍频（又称四倍频、五倍频），摩尔吸光系数较小，吸收弱，短波近红外光在有些样品内可穿透几厘米。长、短波近红外的信息特点决定了应用时应采用不同的工作方式。

近红外光谱的常规分析技术有反射光谱法和透射光谱法两大类。一般情况下，比较均匀透明的液体选用透射光谱法。固体样品（粉末或颗粒）在长波近红外区一般选用漫反射工作方式，在短波近红外区也可以选用透射工作方式。下面就几种工作方式分别做简单介绍。

6.2.1 漫反射光谱法

6.2.1.1 漫反射光谱法原理

一般情况下，固体样品（粉末或颗粒）在长波近红外区（1100~2600nm）适合做漫反射测定。

样品在长波近红外区的摩尔吸光系数较大，吸收较强，光的穿透力较弱，如果仪器的工作谱区是长波近红外区，一般可以选用漫反射的工作方式。漫反射工作方式的检测器与

光源同侧，检测的是样品的漫反射光，其检测的原理是：波长比样品颗粒直径小得多的近红外光照射到样品上，样品作为漫反射体存在，漫反射体与光的相互作用主要有光的全反射、漫反射、散射、吸收和透射等几种形式，如图 6-1 所示^[2]。全反射（又称镜面反射）光因不携带样品的内部信息，在检测时应尽量避免，常用的方法是让检测器与入射光之间成一定夹角（如 45°），使其检测不到镜面反射光，即镜面反射不会影响测定；样品无限厚时透射光可忽略；漫反射光的强度决定于样品对光的吸收以及样品物理状态决定的散射；这样，近红外漫反射光谱定量分析可以只考虑样品对光的吸收、散射和漫反射。在漫反射过程中，

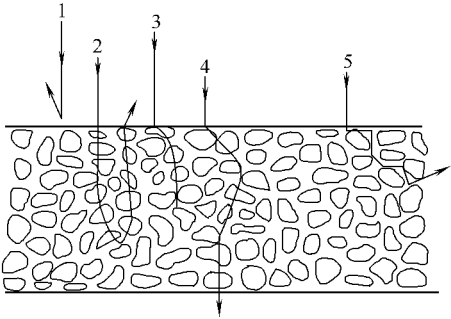


图 6-1 漫反射体与近红外光的相互作用

1—全反射 2—漫反射 3—吸收 4—透射 5—散射

分析光与样品表面或内部作用，光传播方向不断变化，最终携带样品信息又反射出样品表面，然后由检测器进行检测，这是固体样品中最常见的一种测试方式。

近红外漫反射这种工作方式进行定量分析的依据是 Kubelka-Munk 定律（详见第 1 章），

即

$$A = -\log\left(1 + \frac{K}{S} - \sqrt{\left(\frac{K}{S}\right)^2 + \frac{2K}{S}}\right)$$

其中：A 为反射吸光度；K 为样品的吸收系数；S 为样品的反射系数。当 K/S 在一定范围内时可近似用 $A = a + b(K/S)$ 来表示，当样品的浓度不高时，可改写为：

$$A = a + b \times c \quad (c \text{ 为样品某一成分的含量})$$

从而建立了反射吸光度 A 与浓度 c 之间的关系。

6.2.1.2 漫反射光谱的几种测定方法

漫反射有多种测定方式，普通的漫反射方式收集漫反射光的效率低，但结构简单，价格便宜。为了提高光谱的信噪比，更多地收集各个方向的漫反射光，最常用的测样器件是积分球。积分球也有多种方式，图 6-2 是其中一种，当近红外光束照射到样品 S 上时，被样品漫反射的光经球体内部多次反射，绝大部分进入检测器 D 被接收。这种积分球漫反射法测定简单，但要注意保持每次装样的一致性（如样品的松紧度），而且要保证样品厚度对近红外光来说是无穷厚。粉末或较小颗粒状样品在进行漫反射测定时测样器件适合用积分球，图 6-3 为在 Bruker 公司的 MATRIX™-I 上对玉米颗粒样品进行积分球漫反射测定^[3]。因为漫反射光谱分析的样品是粉末状态或颗粒状态，为了减少样品的不均匀性给分析带来的影响，实验时通常采用旋转样品或光源的方式。

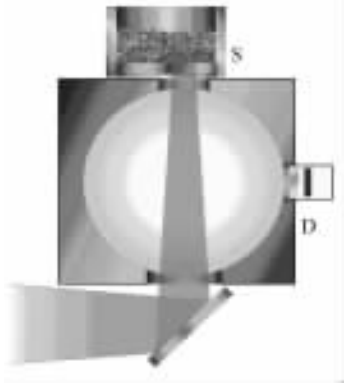


图 6-2 积分球测样示意图

由积分球测得的样品光谱通常信噪比高，重复性好。图 6-4 为小麦粉末样品在傅里叶变换近红外光谱仪上由积分球测得的漫反射光谱图，谱区范围是 4000~12000cm⁻¹ (2500~833nm)，光谱反映的信息主要是长波近红外区的信息。



图 6-3 玉米样品的积分球漫反射测定

漫反射光谱分析常用的另外一种方法是利用光纤探头，如图 6-5，直接将光纤探头对准固体样品，样品可以是任何形状，如水果、谷物或固体药丸，便于检查各部位或每一颗粒的质量。近红外光通过入射光纤照射到样品上，漫反射光又通过出射光纤返回到检测器。利用光纤探头进行的漫反射分析可以实现样品的原位分析，极大地方便了近红外用户。图 6-6 为日本研制的光纤式漫反射法测柑橘糖度的近红外仪器^[4]，只要扣动光纤探头上的按钮或计算机上的快捷键即可开始扫描，并显示结果，操作非常简单。

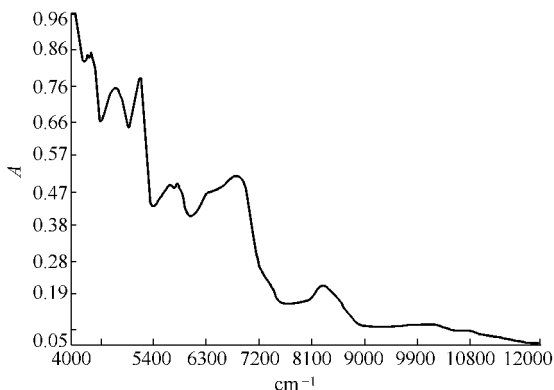


图 6-4 小麦粉末样品的漫反射光谱图

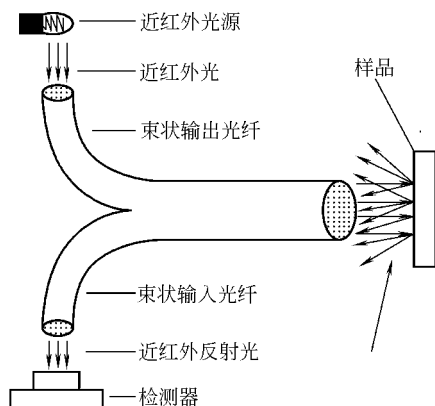


图 6-5 漫反射光纤探头示意图



图 6-6 近红外光纤式漫反射法测柑橘糖度

6.2.1.3 透反射

漫反射工作方式的一种特殊情况是透反射，如图 6-7。一般液体样品（如牛奶）测定时可以选用这种工作方式。工作时，在待测样的后面放置陶瓷类表面漫反射体，光源通过样品后在该漫反射体上发生漫反射，漫反射光又透过样品后返回检测器检测，这种方式的

检测器与光源位于同侧，相当于增加了样品池光程。透反射方式工作时也可以利用光纤探头直接插入液体样品，而且探头与漫反射体之间的光程是可以调节的，如图 6-8。

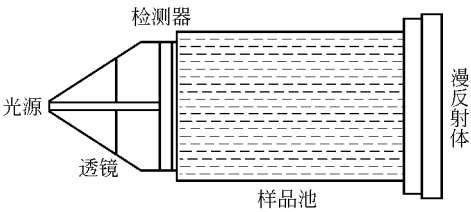


图 6-7 透反射工作方式示意图

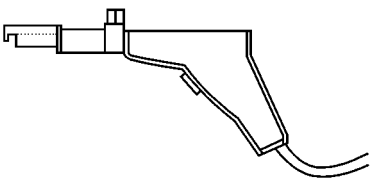


图 6-8 光纤式透反射工作方式

多种漫反射工作方式各有特点，具体测试时选择哪种方式可依据上述内容，根据样品本身的状态以及仪器的结构和附件情况来定。

6.2.2 透射光谱法

虽然短波近红外区的信息量和信息强度远比长波近红外区小，但该谱区内的一些关键器件如光源、检测器等比较廉价易得，适合于开发价格低廉的专用便携式仪器，因此该段谱区的研究和应用也成为热点。

样品在短波近红外区（700~1100nm）常采用透射分析方法。透射法测定的基本方式如图 6-9。近红外光源经透镜后以平行光照射到样品上，作用光经过样品的吸收、散射等过程到达检测器检测。

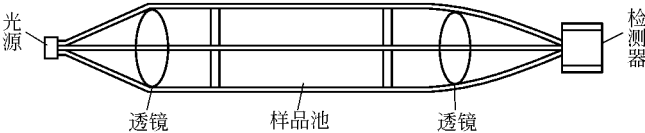


图 6-9 透射测定法基本方式

透射光谱法进行定量分析的依据是比尔定律。对于均匀透明的液体样品来说：当溶液浓度在一定范围内时，吸光度与溶质的摩尔吸光系数 ϵ 、溶质浓度 c 以及光程 l 三者成正比，即： $A = -\log(T) = \epsilon \cdot c \cdot l$ 。但当溶液比较浑浊、颗粒比较大或者固体样品采用透射方法测定时，比尔定律只是近似成立。

6.2.2.1 透射分析方法

（1）液体样品：一般情况下，透明液体样品常采用透射分析的方法。

透射分析除了可用一般的透射式样品池外，还可以采用光纤探头，图 6-10 为透射式光导纤维探头示意图，探头前方开窗作为液体样品池，如果仪器工作在长波近红外区，窗体应开得较窄，如果仪器工作在短波近红外区，窗体要开得较宽。工作时，入射光通过光纤导入照射到液体样品上，透射光经过出射光纤进入检测器检测。图 6-11 是利用光导纤维在近红外区采集液体近红外光谱的示意图^[5]。

如上所述，在短波近红外区采集液体样品的透射光谱，通常需要较长光程，图 6-12 为汽油样品在短波近红外区的透射光谱，光程为 40mm，如果在长波近红外区（1100~2500nm）扫描该样品的光谱，为了使该谱区汽油的吸光度在比尔定律应用范围内，光程应大大减小。

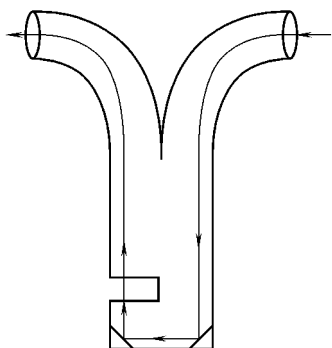


图 6-10 透射光纤探头示意图



图 6-11 透射光纤采集液体近红外光谱

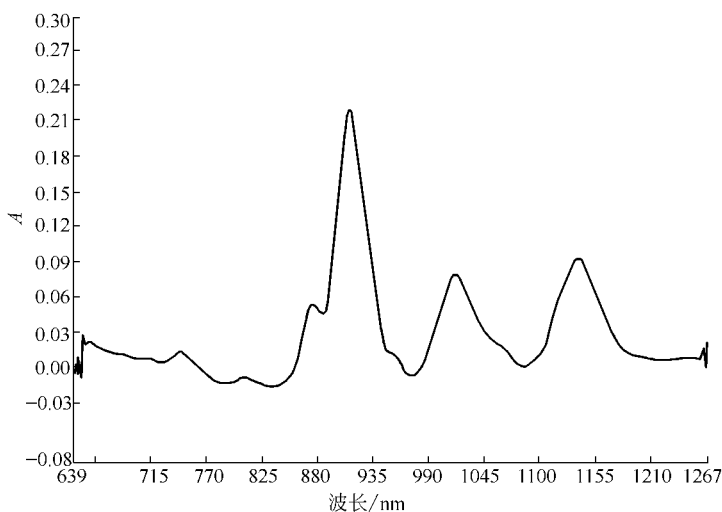


图 6-12 汽油样品在短波近红外区的透射光谱

需要注意的是，在建立局部模型 (Local Model) 时，即建立与特定仪器状态、特定操作条件、特定样品信息相对应的近红外数学模型时，样品池一定要清洗干净；此外，为提高分析速度，常采用多个样品池交替使用，一定要注意样品池间的差别，因为样品池尺寸上的微小差别对分析光在样品池中的折射效应将产生明显的影响，最终将影响测量的光谱。当然，要建立全局模型 (Global Model) 时则不同。

(2) 固体样品：固体样品在短波近红外区也可以用透射分析方法。

样品对短波近红外光 (700~1100nm) 吸收较弱，近红外光可以直接穿透某些固体样品，取得样品深层的信息，所以某些固体样品在短波近红外区也适合用透射分析方法，尤其是对体型较大的样品，如苹果、梨、柑橘等。此外，该谱区的一些光学元件如检测器等比较廉价易得，这为研发低成本、低价位、便携式仪器提供了方便，此类仪器的开发也为该谱区的应用提供了条件。

粉末状样品或颗粒较小的固体样品 (如油菜籽、小麦籽粒、玉米籽粒等) 可以在短波区近红外仪器上做透射分析。图 6-13、图 6-14 分别为小麦颗粒和粉末样品 (过 60 目筛) 在短波近红外区的光谱图。由图可见，虽然都是在短波近红外区，但粉末样品与颗粒样品

所需的光程不同，即近红外光透过样品的厚度与样品的状态有关系。此外，图 6-14 与图 6-4 相比，图 6-14 是小麦粉末样品在短波近红外区的透射光谱，光谱的信息主要是含氢基团的三、四级倍频吸收，信息量较少；而图 6-4 是小麦粉末样品在长波区的漫反射光谱，光谱的信息主要来自含氢基团的合频与一、二级倍频吸收，信息量丰富。

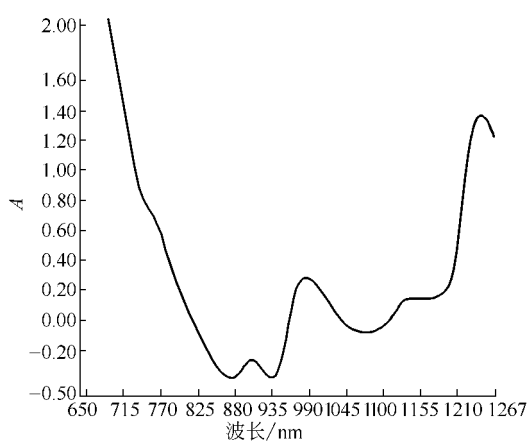


图 6-13 小麦颗粒样品在短波近红外区的透射光谱，光程为 16mm

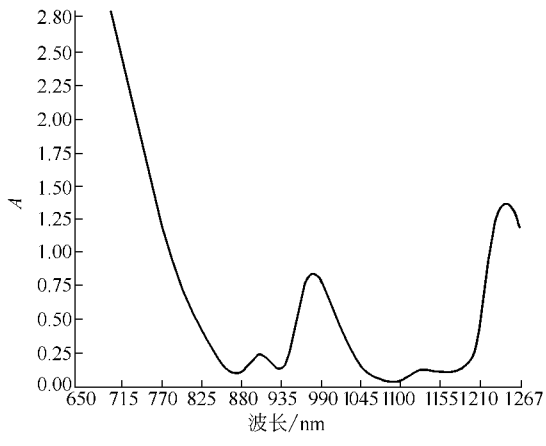


图 6-14 小麦粉末样品在短波近红外区的透射光谱，光程为 8mm

常规透射测定时，透射光的强度易受样品的厚度及透射过程光路的不规则性影响较大，因此装样过程中要注意装样的样品量以及松紧程度等。

对体型较大的样品，如苹果、梨、番茄、柑橘等水果，或其他形状较规则的样品进行内部品质快速测定时，近红外漫反射分析方法在有些国家（如日本）已实用化，但漫反射法存在以下几个问题^[4]：

- ① 只能测定光照射的部分；
- ② 内部信息只是靠近果皮的部分，深层的信息难以取得；
- ③ 内部漫射光受表面反射光的影响。

因此，漫反射技术对水果的测定只限于测定果皮比较薄的如桃子、梨和苹果等，对果皮较厚的柑橘等则测定误差较大。1996 年由 FANTEC 开发了一种透射式近红外在线测定水果内部品质的技术，测定精度比以往的反射式高，能同时测定传送带上水果的成熟度、含糖度、含酸度以及苹果中的甜果和有病斑的水果等多种成分。中国农业大学与北京世农绿方高新技术中心研发了利用短波近红外区透射技术测定果品内部品质的便携式检测仪^[6]，用以实现苹果水心病、糖度、硬度以及番茄等其他果蔬成分的分析检测，图 6-15 和图 6-16 为原理图与其外观图，透射法测量时，果实大小的变化对测定结果有较大影响，在仪器设计或建立模型时需要考虑这些因素。

6.2.2.2 样品池光程的选择

透射光谱法与漫反射法不同，漫反射测定时样品为无穷厚度，因此对样品池的光程基本无具体要求。但透射测定时，一般样品的光程应保持一致，而且要根据样品的具体情况和使用的谱区来选择不同光程的样品池，这是因为近红外区各级倍频吸收的摩尔吸光系数依次下降一个数量级，而根据文献 [2]，当测定吸光度为 0.434 时浓度的误差最小，因此

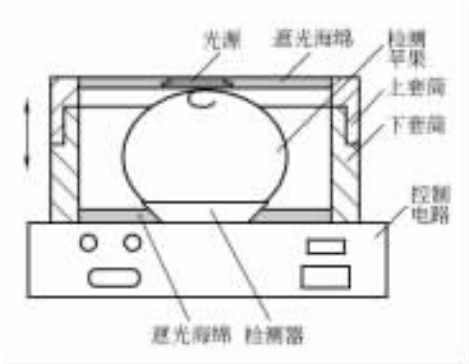


图 6-15 苹果内部品质检测仪原理图



图 6-16 果品内部品质近红外便携式检测仪

根据所使用的光谱范围，可以确定最佳的样品池长度。设 C—H 基团的基频强度为 1，表 6-1 给出了 C—H 基团基频（中红外区）与各级倍频（近红外区）吸收下的波长范围、相对吸光强度，以及所需样品池尺寸的大小，供参考选择。

表 6-1 C—H 基团吸收谱带在中红外及近红外区的差别

谱带	波长范围/nm	相对强度	常用样品池光程/mm
基频	3380~3510	1	0.1~4
一级倍频(2ν)	1690~1755	0.01	1~20
二级倍频(3ν)	1127~1170	0.001	5~50
三级倍频(4ν)	845~875	0.0001	50~100

6.2.3 近红外在线检测技术

由于经济的高速发展，现在许多行业特别是农产品加工（如食品、果品、酒类、烟草、油脂、饲料等）、石油、化工、中西药行业等对生产过程的监控有巨大需求，而传统分析方法实现这些行业生产线的在线检测、监控是相当困难的，近红外技术是最佳选择之一。

近红外在线检测技术与日常的近红外分析技术相比，虽然也是分为透射和反射两大类，上面提到的同样适用，但在测试技术上仍有许多不同之处，本书第 24 章会对此做详细介绍，在此不再赘述。

6.3 样品的收集与制备

6.3.1 样品的收集与代表性样品的选择

6.3.1.1 样品的收集

建立近红外数学模型的过程中，收集样品是最困难、也是最花费时间的。

收集样品时必须要考虑样品种类、化学组成、物理结构、颗粒大小、颜色、样品的表面特征、加工条件及其他因素。对农作物来说，因为生长地区和生长季节强烈地影响着农作物的化学组成和结构，这两个因素也是收集样品时需要考虑的重要变量，而生长地区这一因素又包含土壤情况、肥料的使用、灌溉、耕种等田间管理情况，还有病虫害和其他因

素等。可以说，影响模型的因素非常多，因而在建立模型时，需要根据自己的建模目标：是建立适应全国的还是局部范围的模型，是建立某一特定条件下的还是适应所有因素的模型等，来选择所要收集的样品范围。总之，建模样品集中要包含所有可预期到的变换因素下的样品。大量样品的收集需要很长的时间，建模也需要逐年改进。若没有收集全代表各种变化因素的样品就会导致以后的分析中出现超差数据。

收集到的样品应冷藏保存，有关样品资料的记录应准确、完整，以免引起混淆。

6.3.1.2 代表性样品的挑选

将各种因素考虑在内收集到的样品量一定很大，样品信息也可能会有重叠。建立模型时如果全部样品都测定其化学值是不现实的，需要的时间长、花费大。此外近红外数学模型包含的样品数也不是越多越好，样品数越多分析引入误差的机会就越多。因此必须从大量的样品中挑选部分样品，这部分样品既是化学分析所能承受的范围，又能充分代表原来样品的全部背景信息。

近红外分析很重要的实验技术之一就是挑选有代表性的建模样品。挑选代表性样品的方法有多种，可以根据样品的产地、品种、年份、颜色等，在挑选建模样品时应尽可能的增大这些因子的变异范围，以得到代表性尽可能好的建模样品集；也可以按照相应组分的化学值进行挑选；还可以用数学方法根据大量样品的近红外光谱特征来进行挑选，如聚类法、主成分因子法、网格法、Kennard-Stone 法等。挑选代表性样品的方法，以及挑选多少样品来建模才算最好详见第 8 章相应内容。

6.3.2 确定或建立参比分析方法

建立数学模型的化学值来自参比分析方法，近红外分析是以参比分析方法为基础的。参比方法通常是在采用 NIR 技术之前已经使用的方法，一般选国标法或常用的方法较好，特殊情况下，还需要建立专属于该应用项目的参比分析方法。表 6-2 列举了一些农产品品质分析中适用的参比方法^[7]。其他领域也都各有自己的国标或行标方法，用时可查阅有关书籍。

表 6-2 农产品分析的参比方法	
组 分	方 法
水分	鼓风烘箱法、两级烘箱法、真空烘箱法、卡尔费休法(KF)
油/脂肪	戈德菲什抽提法、索氏抽提法
蛋白质	凯氏定氮法、燃烧法
粗纤维	抽提法
氨基酸	离子交换法
矿物质(总量)	灼烧法
丹宁	比色法
淀粉	酶法

近红外分析是以参比分析方法为基础的，其分析的精确度（也叫重现性）与准确度依赖于参比分析方法的精确度与准确度，讨论近红外测定结果的精确度与准确度需要预先知道参比分析方法的精确度与准确度，而这些数据常受仪器状态和操作人员经验的影响，因

此进行近红外分析前需要验证参比分析方法的准确度和精确度，其中准确度可以通过实验室与实验室之间的相互比较来确定，精确度可以通过平行实验来确定，这些基本数据是讨论近红外分析结果的基础。

6.3.3 确定或建立样品制备技术

NIR 分析中由于制样引起的分析误差可占总误差的 60%~70%。样品的制备就是把样品转化成 NIR 分析所能用的形式，如把大的颗粒状样品粉碎成粉末样品，在这个过程中不应出现样品的组成或功能的变化。

制样时如果需要粉碎样品，则样品磨^[7]的选择很重要，如蚕豆很硬，且颗粒大，无法进入旋风磨之类的磨碎室，可使用锤磨预先粉碎一次，虽然这样得到的样品颗粒还较大，不能用于准确和精密分析，但是，已可以加入到旋风磨中进行细磨以给出实验室用的最终样品。油菜籽在粉碎时易结团，可用高速叶片磨来粉碎，而饲料类样品可用切削式磨粉碎。

此外，样品中不同成分的含量也会影响粉碎效果。如用旋风磨粉碎燕麦或玉米时，应每磨完一个样品清磨一次，但在磨小麦和大麦时则不用清磨，因为在磨这些样品时，磨子可以自动清理干净。即使像含油量为 5% 这样低的样品在粉碎时也有结团的趋势，从而影响样品的粉碎，在磨高含油量的样品如葵花籽或 Canola（卡诺拉油菜籽）时，油分引起样品粉碎后结团，使样品颗粒特性发生变化，以致在几天之内对一个样品进行复测，测定结果也不一致。纤维素和纤维素类物质可改变颗粒的形状和松紧密度。纤维素很轻，所以含纤维越多的物质、粉碎后的样品就越轻，样品密度也越小。另外，纤维素含量高的颗粒比纤维素含量低的颗粒具有较大的长宽比。富含纤维素的样品经常产生静电，粉碎后的样品会从刮刀上“飞”起附在样品盒壁上，并且在样品盒内定向分布。由于样品中的蛋白质对静电有不同的反应，所以蛋白质也会影响样品的颗粒特性。如燕麦及一些豆类就是这样。在某些情况下，其他一些组分也会影响样品粉碎后的颗粒特性，如矿物质含量很大程度地影响着颗粒的松紧密度及粉碎时的行为。

粉碎后样品的混匀也是制样过程中的重要一步。不论什么样品在粉碎过程中都会出现分层现象，所以分析之前必须充分混匀，尤其是对粒度很小的样品，例如小麦粉；另外，液体和糊状物分层现象很严重，所以在分析之前也要充分混匀。样品混匀后就可以进行近红外光谱的测量，测量过程中仪器和样品状态、扫描参数、装样条件等都会对测定结果产生影响，下面做具体讨论。

6.4 仪器状态、扫描参数的设定

近红外光谱仪的工作状态和扫描时各参数的设定对测量过程和分析结果均有一定影响，如波长的准确度、仪器的分辨率、噪声、扫描次数的设定等等。有些参数若实验过程中设置得当，会有助于得到准确可靠的近红外分析结果，有些参数虽然仪器出厂时已经确定下来，用户没有办法修改，但用户充分了解其对测定结果的影响后，也可在测试过程中充分注意其作用，分析误差来源，有助于准确的近红外测定。

下面具体分析仪器状态与扫描参数的设定对近红外分析的影响。

6. 4. 1 波长准确度对分析结果的影响

波长准确度是衡量仪器的重要指标之一，保证波长准确度是近红外光谱仪器能够准确测定样品光谱、从而保证分析结果的前提。波长准确度主要决定于光学系统的结构，一般傅里叶变换近红外光谱仪器波长准确度很高，而色散型近红外光谱仪器波长准确度相对较低些，滤光片型近红外光谱仪器会因为温度等原因使波长发生漂移，需用已知波长值且性质比较稳定的标准物质经常进行校正。如果波长准确度不能保证，光谱数据就会因波长漂移而出现偏差，带入原来模型后造成分析结果的误差。下面举例说明波长的漂移对分析结果的影响。

在傅里叶变换近红外光谱仪上扫描 66 个小麦样品的漫反射光谱，其中 44 个样品用来建立蛋白含量的数学模型，并用该模型预测另外 22 个样品（对应光谱称为“原光谱”，且化学值已知），预测结果见表 6-3 中的“原光谱”部分；为了模拟波长的漂移，人为将 22 条“原光谱”的所有光谱点左移一个数据点（约 4cm^{-1} ）而反射吸光度保持不变，得到的光谱称为“模拟光谱”，再用原建立的同一模型预测波长漂移后的“模拟光谱”，结果见表 6-3 中的“模拟光谱”部分^[8]。

由表 6-3 可见，波长漂移 4cm^{-1} 后，决定系数 R^2 由 97.53% 降为 14.77%；相对标准差 $RSD\%$ 由 1.85 变为 10.87，可见波长的漂移会给近红外分析结果带来很大的影响。

不同厂家的近红外光谱仪器一般都有自己的波长校正方法，以保证波长的准确度，如有些傅里叶变换近红外光谱仪采用 He-Ne 激光器的激光频率作为波长校正系统，如果激光频率发生变化，将会影响仪器的波长准确度，进而影响到光谱数据。为了确定激光频率的改变对样品光谱波长准确度的影响，赵丽丽等^[9]以水蒸气在 7306.74cm^{-1} 处的特征峰为基准考察由于仪器激光频率的漂移引起的波长准确度的变化。实验采用三种激光频率：15798.546 cm^{-1} 、15803 cm^{-1} 和仪器当前的激光频率 15799.608 cm^{-1} ，根据以上三种激光频率确定的水蒸气峰位置如表 6-4 所列。

表 6-3 波长的漂移对分析结果的影响			表 6-4 激光频率的改变对水蒸气峰位置的影响	
	原光谱	模拟光谱	激光频率/ cm^{-1}	水蒸气峰位置/ cm^{-1}
$R^2/\%$	97.53	14.77	15798.546	7306.3
$RMSEP$	0.29	1.68	15799.608	7306.7
$RSD/\%$	1.85	10.87	15803	7308.3

由表 6-4 可见，若将激光频率设定为 15803 cm^{-1} ，而实际仪器扫描时仍按 15799.608 cm^{-1} 确定的波长进行扫描时，由此引起的水蒸气峰波长偏差 1.6 cm^{-1} 。参考文献 [9] 还同时采用如下实验方案研究小麦蛋白模型在这三个激光频率设定值下给分析结果带来的影响。

在仪器激光频率为 15799.608 cm^{-1} 条件下扫描 30 个小麦样品的近红外光谱，并建立蛋白含量数学模型。为了模拟激光频率的改变也即波长准确度对分析的影响，分别在三种激光频率 15798.546 cm^{-1} （激光频率设定值 1，记为 HeNe¹）、15799.608 cm^{-1} （激光频率设定值 2，记为 HeNe²）和 15803 cm^{-1} （激光频率设定值 3，记为 HeNe³）下扫描三个高、

中、低蛋白质含量的样品，每个样品在三种激光频率设计值下分别重复扫描 10 次，然后用在 15799.608cm^{-1} 下获得的模型预测这批样品光谱，结果如表 6-5 所列。

表 6-5 三个高、中、低蛋白含量样品在三种激光频率下的预测值

样品	低蛋白含量样品 (12.14%)			中蛋白含量样品 (14.38%)			高蛋白含量样品 (18.33%)		
激光频率	HeNe ¹	HeNe ²	HeNe ³	HeNe ¹	HeNe ²	HeNe ³	HeNe ¹	HeNe ²	HeNe ³
近红外 预测值	12.293	12.405	12.573	14.596	14.642	14.798	18.463	18.331	18.644
	12.350	12.431	12.675	14.597	14.534	14.755	18.472	18.359	18.521
	12.392	12.419	12.661	14.482	14.526	14.863	18.359	18.468	18.542
	12.447	12.443	12.623	14.673	14.660	14.732	18.473	18.435	18.589
	12.404	12.497	12.697	14.622	14.642	14.726	18.413	18.437	18.613
	12.469	12.336	12.589	14.589	14.574	14.808	18.466	18.49	18.465
	12.336	12.440	12.817	14.539	14.511	14.754	18.328	18.486	18.590
	12.364	12.383	12.759	14.499	14.437	14.656	18.349	18.530	18.719
	12.283	12.312	12.572	14.54	14.425	14.674	18.453	18.503	18.503
	12.357	12.361	12.598	14.574	14.458	14.746	18.332	18.433	18.693

对这三种激光频率下的分析结果进行方差分析, 结果如表 6-6 所列。

表 6-6 三种激光频率下分析结果的方差分析

样 品	HeNe ² 与 HeNe ¹ 的 F 值	HeNe ² 与 HeNe ³ 的 F 值
低含量样品	1.64	64.89
中含量样品	0.84	39.03
高含量样品	1.43	18.48

在 $\alpha=0.05$ 水平上, $F_{(1,18)}=4.41$, 上述方差分析的结果表明: 当 He-Ne 激光频率由 15799.608cm^{-1} (HeNe^2) 变化到 15798.546cm^{-1} (HeNe^1) 时, 对应水蒸气峰波长偏差约 0.4cm^{-1} , 此时三个高、中、低含量样品的预测值在两种激光频率下的差异均不显著 ($F < F_{(1,18)}$); 而当激光频率由 15799.608cm^{-1} (HeNe^2) 变化到 15803cm^{-1} (HeNe^3), 对应水蒸气峰波长偏差约 1.6cm^{-1} 时, 三个样品预测值的差异均达显著水平 ($F > F_{(1,18)}$)。说明激光频率变化较大, 也即波长准确度变化较大时会给近红外分析结果带来显著影响。参考文献 [9] 中实验所用近红外仪器 4 年内激光频率由 15798.546cm^{-1} 变化到 15799.608cm^{-1} , 根据以上实验结果这种变化对小麦蛋白模型的影响不显著, 说明波长的漂移程度很小。即使波长有些微变化, 也可以根据水蒸气峰位置调整 He-Ne 激光频率, 从而起到波长校正作用。

6.4.2 仪器分辨率的设定

仪器的分辨率影响样品光谱的质量, 最终影响分析的准确性。根据经验, 一般要求仪器分辨率应相当于样品光谱半高宽的 $1/5 \sim 1/10$ 。而样品在近红外谱区漫反射光谱的半高宽约为 $150 \sim 400 \text{ cm}^{-1}$, 因此理论上用 15 cm^{-1} 的分辨率对近红外分析来说已经足够, 实际

考虑样品中个别成分的反射吸收峰更窄，或实验对分辨率有较高要求，在仪器能够满足的基础上可以将分辨率设的更高些。但分辨率越高，扫描速度越慢，扫描时间延长，而且图谱所占的存储空间增大。因此，在不影响图谱质量的前提下，分辨率可以不取太高。下面以实验数据说明仪器分辨率的设定对近红外数学模型性能的影响。

烟草粉末样品共 51 个，选取 40 个作为建模样品，另外 11 个为检验样品，分别取 4cm^{-1} 、 8cm^{-1} 、 16cm^{-1} 分辨率扫描样品图谱，并建立烟草 4 种成分的数学模型，用决定系数 (R^2) 与预测标准差 (SEP) 来评价模型，结果如表 6-7。

表 6-7 不同分辨率扫描烟草样品图谱所建数学模型预测烟草四种成分的结果				
扫描分辨率/ cm^{-1}	烟碱(R^2/SEP)	总糖(R^2/SEP)	还原糖(R^2/SEP)	总氮(R^2/SEP)
4	99.24/0.113			79.37/0.158
8	98.59/0.153	89.21/1.15	75.37/1.58	86.53/0.127
16	98.12/0.177	91.18/1.04	78.32/1.48	79.60/0.157

将实验所用三种分辨率所建数学模型预测烟草四种成分的结果在 $\alpha=0.05$ 水平上进行 F 检验，结果表明三种分辨率下模型的性能相当，差异在统计上不明显，未见分辨率提高，对数学模型的预测精度有明显的改善，可见分辨率由 16cm^{-1} 提高到 4cm^{-1} 并未明显改善数学模型的性能。

实验同时还考察了在这三种分辨率下扫描一个样品所用的时间及所产生的文件大小，结果如表 6-8。

表 6-8 三种分辨率下扫描一个样品所用的时间及所产生的文件大小		
分辨率	扫描所用时间/(s/条光谱)	文件大小/K
4cm^{-1}	70	18
8cm^{-1}	50	10
16cm^{-1}	35	5

由结果可知分辨率为 4cm^{-1} 时所产生的光谱文件较大，所用时间较长。

综合考虑以上两个因素，又考虑到样品光谱中个别谱峰较窄，实际测量烟草样品时建议分辨率选 8cm^{-1} 。实验测试过程中，分辨率到底该如何设定，需视具体情况而定，如果是日常快速分析，分辨率可以较低；如果是为了研究，如建立标准光谱库或需要细节的判别分析，则可适当提高扫描光谱的分辨率。

6.4.3 仪器信号能量的影响

虽然大部分近红外仪器能量较为稳定，但随着测量仪器的预热程度不同，环境温度不同以及测量时间的不同，仪器能量也会有较大的差别。所以不同时间扫描所得到的光谱图不易完全重合，参考文献 [10] 设计了两种不同信号能量状态下，扫描同一样品池中的同一样品的实验，得到如图 6-17 光谱曲线。图中曲线 1 为仪器信号能量在 3.9V 下的光谱图，2 为仪器信号能量在 4.9V 下的光谱图。

由图可见，不同信号能量对光谱图没有太大的影响，只是由于能量的变化带来了谱图基线的平移。可见，不同信号能量对光谱图没有太大的影响。但实际测试过程中，仪器信

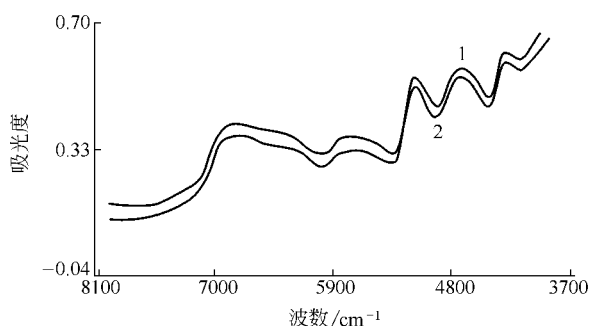


图 6-17 不同信号能量下的漫反射光谱图

1—能量 3.9V 下的光谱图 2—能量 4.9V 下的光谱图

号能量太低会影响光谱的信噪比，而且如果仪器能量的降低是因为仪器光学器件位置（如反射镜）的变化引起，则一定会影响模型的稳定性能。

6.4.4 分析光束截面积大小对分析的影响

由于时间的推移、仪器的振动或搬移等原因会导致仪器的光学器件发生位置移动或变形，从而使得照射到样品上的分析光束截面积发生变化。如果分析光束截面积大小变化，光束截面内样品的粒子数发生变化，从而在扫描过程中会引起随机误差。为了说明上述问题，严衍禄等^[10]在 Nicolet 傅里叶变换近红外光谱仪上设计了以下实验：在 6 个样品池里用同样方法装上相同样品，改变分析光束截面积 3 次，面积分别为 15.1mm²、7mm² 和 3.1mm²，在光谱仪上扫描其漫反射光谱图，而且每种光束截面积条件下都对每个样品池做重复 6 次的扫描，均得 36 张图谱，并取吸收波峰 4495cm⁻¹ 和波谷 4746cm⁻¹ 两点处的吸光度差值作参数来分析（因为这 2 个波数点是特征区的 2 个具有代表意义的波数点）。

把分析光束截面积变化作为对误差影响的 A 因素，分析过程中除光束截面积变化以外的其他随机因素作为对误差影响的 B 因素，用方差分析的方法，分析 A、B 两因素的影响的显著程度，并同时分析 A、B 两因素的交互作用对误差的影响，结果如表 6-9。

表 6-9 分析光束面积大小对误差影响的分析

方差来源	自由度	平方和	均方差	F 值	$F_{0.05}$
A 因素	2	8.876E-5	6.619E-3	4.373	3.15
B 因素	5	1.400E-4	5.292E-3	2.795	2.37
A 与 B 交互作用	10	4.619E-5	2.149E-3	0.4611	1.86
误差 E	90	9.017E-4	3.165E-3		

由以上结果可见，A 因素（分析光束截面积大小）和 B 因素的（其他随机因素）对误差的影响都是显著的（ $F \text{ 值} > F_{0.05}$ ），而且 A 因素和 B 因素的交互作用对误差的影响不显著，即它们对误差的影响是独立的。

为了更明确知道分析光束截面积对误差影响的定量概念，实验分析了三种分析光束截面积下所有光谱（两个波数点吸光度差值）的标准差，并从 6 个样品池中任取一个（3 号池），也对其样品光谱求标准差，结果见表 6-10。

表 6-10

不同分析光束截面积下图谱的标准差

分析光束截面积/mm ²	所有光谱标准差	同一池样品光谱的标准差
3.14×4.8 ²	1.359E-3	5.622E-4
3.14×1.5 ²	1.361E-3	9.621E-4
3.14×1 ²	5.236E-3	3.057E-3

由表 6-10 结果可见，所有光谱的标准差，以及任意选定样品池的光谱标准差，都随着分析光束截面积的变小越来越大。这与理论分析结果一致。为了分析比较光束截面积变化给光谱带来的变化，将上述 3 种分析光束截面的光谱又做了差谱处理，见图 6-18。图中曲线 1、2 分别代表分析光束截面积为 3.1mm²、7mm² 与 15.1mm² 的样品光谱的差谱。

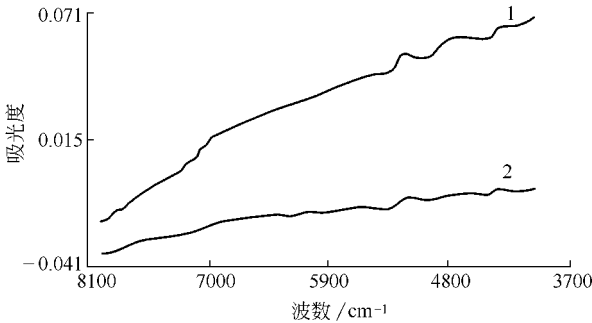


图 6-18 分析光束截面积不同时样品光谱的差谱图

1—小光束（3.1mm²）~大光束（15.1mm²）的差谱 2—中光束（7mm²）~大光束（15.1mm²）的差谱

由图可见，由于分析光束截面积大小的变化，光谱中出现了基线平移、基线旋转以及差谱带有非均匀变化等变化，而且在水吸收峰 5200cm⁻¹ 附近变化尤为突出，分析原因如下。

分析光束截面积大小发生变化实际是因为光线聚焦程度不同引起的。分析光束截面积不同，检测器中检测到的作用光的多少就不同，引起了光谱基线的平移。检测到的光线发生了变化，也就是取得的样品信息不同，加之样品散射系数 S 随波数的减小稍有下降的原因，从而使图谱的基线发生旋转，并使差谱带有非均匀变化。另外由于聚焦程度的不同，红外光的光热效应现象较明显，加之，水分在近红外区是最敏感的，有最强的吸收峰，这最容易引起水分吸收峰的变化，所以差谱中水的吸收峰的变化也较为明显。

鉴于以上原因，在近红外分析测试过程中应注意分析光束截面积的不变性，即要保持仪器硬件性能的稳定性，这样可降低近红外光谱测量的误差，从而保证分析结果的准确性。

6.4.5 扫描次数的设定

平均技术是提高光谱信噪比的一种传统有效方法，假定测量过程中近红外信号与背景噪声是相互独立的，且是加性关系，即 $x_i(k)=s(k)+n_i(k)$ $i=1, 2, \cdots, N$ ，其中 $s(k)$ 是样品在近红外区的真正信息， $n_i(k)$ 表示第 i 次扫描时的背景噪声， $x_i(k)$ 表示第 i 次测量的表观数据，则经过 N 次叠加以后（即扫描次数为 N 次）：

$$\overline{x}(k) = s(k) + \frac{1}{N} \sum_{i=1}^N n_i(k)$$

与信号无关的噪声逐渐消减，而信号以代数和的方式逐渐增大，除以叠加次数 N 以后，则信号保持原貌，而噪声减小 $\sqrt{N}$ 倍，即信噪比原来提高了 $\sqrt{N}$ 倍。增加样品的扫描次数就是增加光谱的平均次数，所以，样品的扫描次数增加可提高信噪比，从而改善近红外分析结果。例：相同的 32 个番茄样品，分别用透射法扫描 100 次（约 1s）与扫描 20 次（约 0.2s）建立的模型结果如表 6-11。

表 6-11 扫描 100 次与扫描 20 次的模型比较		
扫描次数	相关系数	标准差
20 次	0.67	0.662
100 次	0.82	0.433

由上表结果可见，扫描次数由 20 次增加到 100 次，相应模型的性能有所改善，这是因为扫描次数增加提高了光谱的信噪比。

但不是扫描次数越多越好。因为，一方面，随着扫描次数的增加，扫描一个样品所需的时间会延长，尤其是对光栅型近红外光谱仪，扫描一次所需时间较长，对扫描次数的控制应满足用户对速度的要求，尤其是实时在线分析对速度的要求；另一方面，通过扫描次数的增加来消减噪声，只是在次数较低时作用明显，次数越多噪声衰减的程度越不明显。下面以傅里叶变换近红外光谱仪扫描烟草样品为例说明此问题。

烟草粉末样品共 51 个，选取 40 个为定标样品，另外 11 个为检验样品，在傅里叶变换近红外光谱仪上扫描这些样品的图谱，扫描次数分别为 32 次、64 次、96 次，建立烟草 4 种成分的数学模型，并用决定系数 (R^2) 与预测标准差 (SEP) 来评价模型，结果见表 6-12。

表 6-12 不同扫描次数扫描烟草样品所建数学模型预测烟草 4 种成分的准确度				
扫描次数	烟碱(R^2 /SEP)	总糖(R^2 /SEP)	还原糖(R^2 /SEP)	总氮(R^2 /SEP)
32	98.15/0.175	88.57/1.18	76.62/1.54	85.41/0.133
64	98.59/0.153	89.21/1.15	75.37/1.58	86.53/0.127
96	97.72/0.195	87.88/1.21	73.57/1.64	87.88/0.121

实验所用 3 种扫描次数所建数学模型预测 4 种成分的结果在 $\alpha=0.05$ 水平上进行 F 检验差异不显著，扫描次数增加（由 32 次增加到 96 次）并没有明显提高数学模型的预测精度。图谱扫描时发现，用 64 次扫描光谱信噪比已可达 10^4 ，图谱的质量已足够近红外分析烟草样品的要求。因此烟草样品的扫描次数一般为 64 次扫描取平均，快速分析时可用 32 次。

所以，测试过程中扫描次数的设定应以提高信噪比和满足扫描速度为准则。

6.5 样品状态、装样条件对测量结果的影响

近红外谱区自身信息的特点决定了其吸收强度弱、信噪比低；一般情况下，近红外分析不需对样品做化学预处理，因此测定时背景复杂；以上两个原因使得近红外光谱易受样品状态、装样条件等的影响，测定条件不易完全一致，造成光谱的不确定性。与溶液的透射光谱分析相比，漫反射光谱分析样品是粉末或颗粒状态，其光谱受样品状态与装样条件

的影响更大，由此产生测定误差。下面以漫反射为例，简单分析样品的差别与装样条件对图谱造成的影响^[11]，见图 6-19。

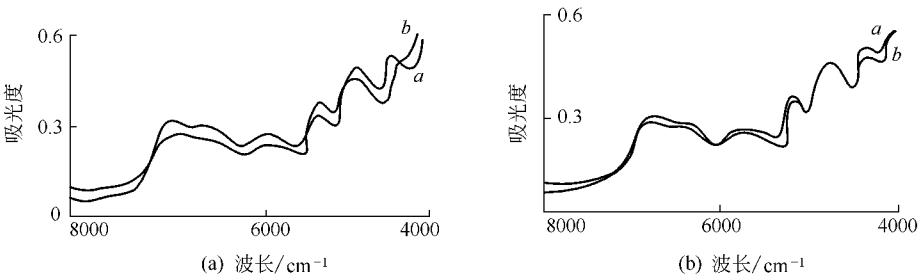


图 6-19 样品的差别与装样对图谱造成的影响

图 6-19 (a) 中 *a* 与 *b* 是同一个样品两次装样所得到的漫反射光谱，由于装样条件不易控制到完全相同，加上粉末样品粒度分布等随机的差异使 2 次装样的光谱图发生了明显的差异。图 6-19 (b) 中 *a* 与 *b* 是两个蛋白质含量分别为 11.3% 和 15.6% 的小米样品，该两样品的蛋白质含量虽然有较大差异，但其漫反射光谱的差异似乎还不如图 6-19 (a) 中一个样品两次装样所发生的差异。由此可见漫反射光谱由于制样、装样的随机差异可以引起光谱图上明显的差别，如果装样时不注意，或者不对光谱做后续数学处理，这种差异有可能掩盖由样品成分引起的光谱差异。

6.5.1 样品含水量对分析的影响

水分在近红外谱区有较强的吸收峰（约在 1450nm 和 1930nm 处），是近红外定量分析的主要干扰因素。有些样品，如牛奶、番茄、梨、豆腐、猪肉等含水量高达 70%~90%，其近红外光谱主要是水的吸收峰，参考文献 [5] 记录了梨在 1100~2500nm 范围内的近红外光谱，见图 6-20，其吸收带大约在水峰的位置，对该光谱求一阶导数后，样品在 1400~1410nm 处呈现出最强的吸收峰，此峰位置与纯水的吸收峰位置相同。干燥的样品如大豆、小麦、水稻中水的吸收峰与上述含水量高的样品有区别，其在 1410nm 处与自由水吸收相关的吸收峰逐渐消失，而代之以 1430nm 处结合水的吸收。

样品含水量的不同对近红外分析是有影响的，图 6-21 是一种蛋白质含量（12.4%）

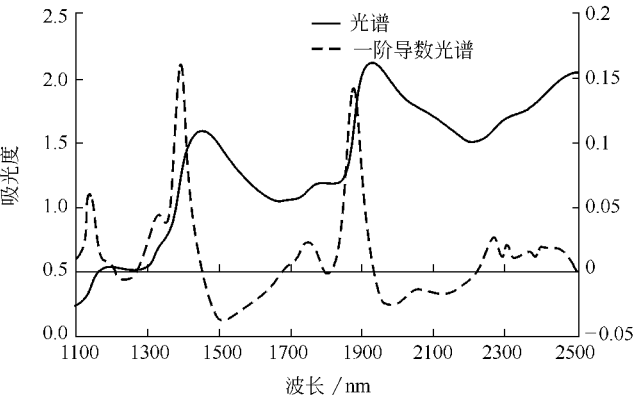


图 6-20 梨在 1100~2500nm 的近红外光谱

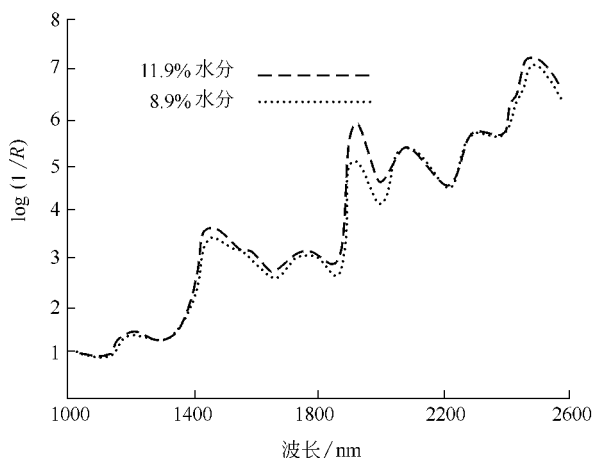


图 6-21 同一蛋白质含量水分含量不同小麦粉的近红外光谱

相同而水分含量不同的小麦粉末样品的近红外光谱^[12]。由图可见，同一蛋白含量水平的面粉由于含水量的差异在近红外仪上分析结果会稍有不同。

为了进一步说明样品含水量的影响，用 30 个谷子作为建模样品建立蛋白质含量数学模型，建模样品在室内经一周时间存放自然风干。另外两个待测样品，其蛋白质含量分别为 10.36% 和 14.95%，将这两个待测样品分成五份，分别放于不同的湿度环境下，使其含水量分别为 19.14%、16.97%、13.55%、12.02% 和 9.5%，其中含水量为 9.5% 是自然风干的结果，同时分别测出它们的近红外光谱后用建立的模型预测，结果见表 6-13。

表 6-13 同一样品含水量不同时的预测结果

		含水量 19.14%	含水量 16.97%	含水量 13.55%	含水量 12.02%	含水量 9.5% (风干)
样品 1	扣除水分变化后的实际蛋白质含量/%	9.5	9.6	9.9	10.1	10.36
	模型预测值/%	7.1	7.4	9.3	9.8	10.1
	绝对误差/%	2.4	2.2	0.6	0.3	0.23
样品 2	扣除水分变化后的实际蛋白质含量/%	13.6	13.9	14.3	14.6	14.95
	模型预测值/%	11.1	12.6	14.2	15.2	15
	绝对误差/%	2.5	1.3	0.1	-0.6	-0.05

由表可见，由低含水量样品（自然风干状态）建立的模型预测样品 1 和 2 的蛋白含量，随着样品含水量的升高，绝对误差总体趋势越来越大，所以由低含水量样品建立的模型一般很难适应高含水量的样品。

为了减小水分在分析中的影响，使模型具有稳定性能，可以采取如下措施：①在样品粉碎过程中避免粉碎机过热造成部分样品过多水分丢失；②粉碎后的样品不要马上在近红外仪上测定，要和定标样品（或称标样）放在同一环境中平衡 1~2 天或更长的时间，使样品中水分含量大体一致；③建模样品的含水背景应包括一定的变化范围，建立受水背景干扰影响较小、对背景水的变动有一定适应能力的模型，提高模型的稳定性。

6.5.2 样品温度对分析的影响

样品温度对近红外分析的影响已有很多报道。温度的变化会改变样品分子间的作用力，如温度变化，水分子中 O—H 的振动和转动会发生变化，从而改变近红外光谱吸收带的形状（如展宽）或发生漂移^[13]。Florian Wulfert 等^[10] 记录了将乙醇（ethanol）、水（water）和 2-丙醇（2-propanol）的混合物分别在 40℃、50℃、60℃ 和 70℃ 条件下测定的短波区近红外光谱，见图 6-22。由图可见，温度变化，970nm 处的水峰发生变化，但波长并未漂移。

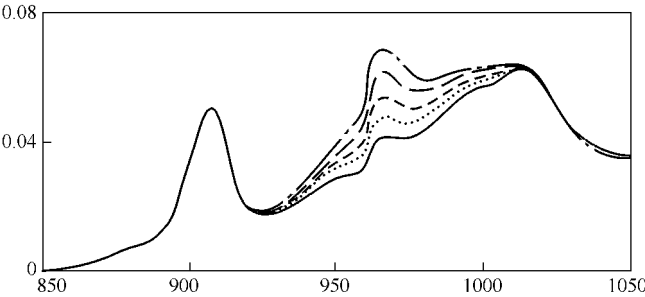


图 6-22 乙醇、水和 2-丙醇混合物在 40℃、50℃、60℃、70℃ 下的光谱

Mitsumoto, M. 等^[14] 在不同的温度下测定了牛肉的近红外光谱，光谱求二阶导数后也随着温度的变化而发生变化，见图 6-23，其中 5℃ 与 10℃ 间光谱的差异较小，但 10℃ 与 15℃ 以及 15℃ 与 20℃ 间的差异却要大多了，而且二阶导数光谱中还出现了波长的漂移（920nm 处）。

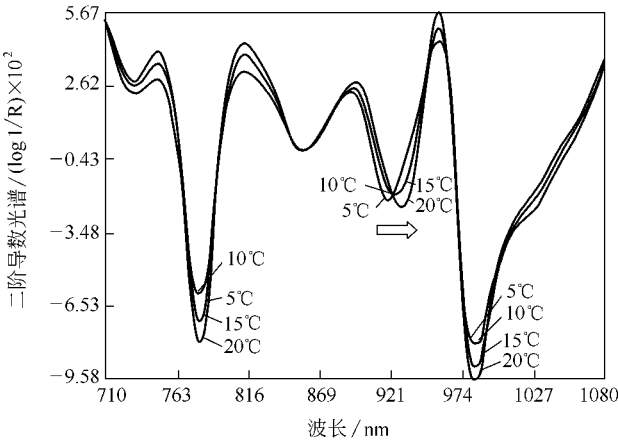


图 6-23 牛肉在 5℃、10℃、15℃、20℃ 下的近红外二阶导数光谱

温度的变化不仅可能逐日发生变化，甚至一天内都可能发生。由温度引起的波长的漂移和谱带的展宽属于非线性变化，这些非线性变化会使近红外数学模型变得复杂起来，因为通常我们的模型是用线性多元方法建立起来的^[10]。温度的变化影响模型从而影响近红外分析的结果，Williams 对小麦的研究发现，随着样品温度从 -30℃ 上升到 45℃，测定的小麦蛋白质含量呈下降趋势。当然，这种温度对模型的影响在一定程度上是可以降低

的，通常的方法是建模时把温度变化考虑进去，即建模样品的温度范围要包括通常分析条件下可以预期到的全部温度范围，这样建立的数学模型中对温度变化敏感的波长权重降低，提高了模型对温度变化的适应性，即模型对温度会有一定的自动校正作用。另外一种方法是在建立模型过程中将温度作为一个干扰项加入，即将温度作为一变量与光谱数据融合使用，形成一全局模型（Global Model）^[15]。

6.5.3 样品粒度对近红外分析的影响

在进行近红外分析时，为了尽量使样品均匀以减小误差，通常将样品磨碎后进行测量（有些样品采用透射技术除外），但在磨样时，虽然采取同样的磨，并过相同的筛，粉末的粒度大小仍会对近红外光谱和定量分析结果产生一定的影响。

由于样品的粒度影响光对样品的穿透特性和反射特性，粒度变化使光在样品粒子中的光程发生变化，从而影响样品对光的吸收系数和散射系数。其中样品的散射系数 S 和样品的粒度有关， S 随着粒度的增加而减少，粒度变化， S 变化，反射吸光度也随之变化，导致整个图谱发生变化，从而影响近红外数学模型的预测精度。另外，粒度的变化也影响光对粒子的衍射，且在高波数区和低波数区不同，产生与上述相同的结果。为比较样品不同粒度对数学模型的影响程度，下面以过 20 目筛和 40 目筛的 51 个烟草粉末样品的建模结果为例说明，见表 6-14。

表 6-14 样品粒度对近红外分析的影响				
成分名称	烟 碱		总 糖	
	R	$RMSEC$	R	$RMSEC$
粗粉(约 20 目)	0.9793	7.16	0.9514	7.25
细粉(约 40 目)	0.9832	5.47	0.9795	4.75

注： R 为相关系数 $RMSEC$ 为校正标准差

由表 6-14 可见，细粉样品的模型较粗粉样品的模型相关系数高，而且烟碱的校正标准差由 7.16 降到 5.47，总糖的校正标准差由 7.25 降到 4.75。这种结果可能是由于粒度大小及其分布影响了装样和样品的表面特征，产生了随机误差。粒度越小，在分析光范围内样品颗粒越多，从而产生的随机误差可能也就越小。

为了进一步说明样品粒度对近红外分析的影响，将同一样品分别过 40 目、60 目和 80 目筛，并用同一方法装样，扫描其近红外光谱。把过 40 目和过 60 目与过 80 目的样品光谱图进行比较，得到图 6-24 结果。图中曲线 1 是过 40 目与过 80 目的样品光谱的差谱，曲线 2 是过 60 目与过 80 目的样品光谱的差谱^[10]。

由图可见，由于粒度的变化，图谱基线发生了平移和旋转，同时差谱中在原图谱的峰谷处也出现了大小不

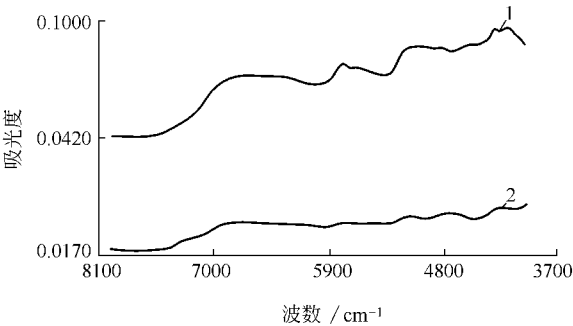


图 6-24 不同粒度测量结果的比较

同的峰谷，即表示存在非均匀变化以及其他随机变化。图谱中基线的平移是由于在漫反射过程中，样品的粒度影响光对样品的穿透特性和反射特性，粒度变化使光在样品粒子中的光程发生变化，从而影响样品对光的吸收系数与散射系数。其中样品的散射系数 S 和样品的粒度有关， S 随着粒度（直径 d ）的增加而减小。所以，当 d 变大时， S 变小， K/S 变大，吸光度 A 就会变大，整个图谱向上平移。另外样品粒度越小，在低波数区（长波区）由于对测试假定条件 $d \gg \lambda$ 满足程度不如在高波数区好，光对粒子的衍射现象的影响增大，从而使不同粒度的光谱基线发生旋转。粒度对样品随着波数的变化不是均匀变化的，但随着吸光度变化是均匀变化的，图上谷峰的高度变化又不一致，在差谱图上也就出现了非均匀变化，并且粒度越大，光程变化越大，所以在图 6-24 中曲线 1 中的这种非均匀变化比曲线 2 中这种非均匀变化更明显。

鉴于以上原因，在近红外分析测试时为提高分析的准确度，要么建模样品与分析样品粒度尽量保持一致；要么模型中包含粒度不同的样品，使模型的适应范围广，但这种方法的准确性可能会稍有下降。

6.5.4 样品的均匀性对近红外分析的影响

在近红外测试中，样品的均匀性是提高 NIR 分析准确性的一个重要因素。漫反射法测量的主要是浅层样品的某组分浓度，如果样品不均匀，某组分浓度与内部浓度不相等，这样漫反射测出的浓度就不能代表样品中这一组分的浓度。固体样品的均匀性主要受以下几个因素影响：①样品在脱壳时的彻底性；②磨样时样品粉末粒度的均匀性；③样品内水分、油分等影响均匀性的组分含量的高低；④样品表面平整性。液体样品的均匀性主要决定于样品的质地，有些液体样品具有分层现象，分析测试时要达到完全均匀有些困难。为尽量减小样品的不均匀性给分析结果带来的影响，测试过程中要特别注意如下：

① 在样品制备过程中将样品充分混匀：从粉碎机出来的粉末样品若是分层的，必须充分搅拌混匀；有些液体样品放置一段时间会有分层现象，也要摇匀后才能测定。

② 旋转样品池或光源：样品池旋转对减小不均匀性十分有效。表 6-15 数据是在旋转与不旋转情况下预测谷子蛋白质含量的标准差^[10]。其中“一次装样”是指同一样品池装一次样品后在旋转与不旋转两种情况下重复 10 次扫描光谱（每次间隔若干分钟，以保证温度平衡），然后分别用已经建立的模型计算出蛋白质含量，用标准差表示 10 次测定值之间的离散度，它表示了仪器噪声造成蛋白质测定值的不确定性。“多次装样”是指同一样品用同一样品池重复装样 10 次，用标准差表示 10 次重复装样测定值间的离散度，它表示了仪器噪声、样品的不均匀性、装样条件的差异等共同造成蛋白质测定值的不确定性。而且对“一次装样”实验与“多次装样”实验均重复 10 次以上，取各次实验标准差的最大值列于表 6-15 内。

表 6-15 样品池旋转消除样品不均匀性的影响

	一次装样的标准差	多次装样的标准差
不旋转	0.494	0.498
旋转	0.190	0.256

由表 6-15 可见，用标准差表示的不均匀性，由于样品池旋转降低了 1 倍多，大大提高了测量的精确度。样品池旋转对减小样品不均匀性的作用也可以通过表 6-16 的酒糟实验来说明。

表 6-16 样品池旋转与不旋转对酒糟中各成分预测性能的影响

成分	淀粉	水分	酸度	淀粉	水分	酸度
	同一样品装样 6 次、不旋转扫描的结果			同一样品装样 6 次、旋转扫描的结果		
1	10.4	58.7	2.1	9.9	59.5	2.1
2	9.9	59.8	2.3	9.8	60.2	2.1
3	9.5	59.1	2.1	10.1	60.6	2.1
4	11.0	60.5	2.4	10	60.4	2.0
5	9.5	61.7	1.8	10.1	60	1.8
6	11.3	60.4	2.4	10.2	60.5	1.9
标准差	0.77	1.08	0.23	0.15	0.4	0.13

酒糟样品的不均匀性较大，表 6-16 中同一样品重复装样 6 次，在样品池不旋转和旋转 2 种情况下得到的光谱用同一模型分析，结果发现酒糟中淀粉含量在样品池不旋转情况下 6 次测量的标准差为 0.77，旋转后降为 0.15，其他 2 种成分标准差在旋转情况下也显著下降，可见旋转样品池对降低样品的不均匀性是有好处的。

6.5.5 样品的松紧程度对近红外分析的影响

样品装入样品池时，样品的松紧程度随着装样人的不同而不同，随着装样的时间不同而不同，随着样品池的不同而不同，在实际测量过程要做到不同次装样松紧程度完全相同，确实不易，更何况还因人而异，下面举例讨论样品装样的松紧程度对近红外分析的影响^[10]。

同一样品取完全相同的样品量，装入 3 个相同的样品池，一个样品池不用压模压样，一个样品池压入 2mm，一个样品池压入 2.5mm，代表装样的不同松紧程度，在光谱仪上扫描其漫反射光图谱，为了比较，采用差谱技术，如图 6-25 所示，图中曲线 1 是不压样与压样 2.5mm 光谱的差谱；曲线 2 是压样 2mm 与压样 2.5mm 的光谱的差谱。

由图可见，图谱基线平移和差谱中的非均匀变化较明显，而基线的旋转不太明显，此外还有其他随机变化。原因可能是因为样品的松紧程度不同，使样品的散射系数 S 发生变化，从而使光谱的基线发生平移。另外，样品的松紧不同，入射光在样品中穿透程度和反射特性都不同，即光在样品中的光程不同。这样作用光中所携带的样品信息量就不同，所以其差谱中就出现了与反射吸光度有关的非均匀变化，而且样品越松这种差别就越大，差谱中的非均匀变化就越明显。

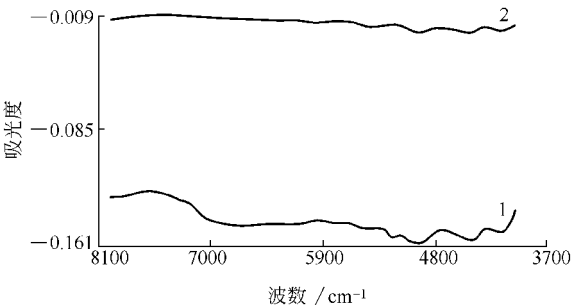


图 6-25 不同松紧程度的参比与样品的光谱差谱图

6.5.6 样品高度对近红外分析的影响

样品压样过程由于压力不同而造成样品高度不同是经常的事，对于普通的近红外漫反射分析，这就可能在漫反射图谱上产生较大的差异，从而影响定量分析的结果，为了分析

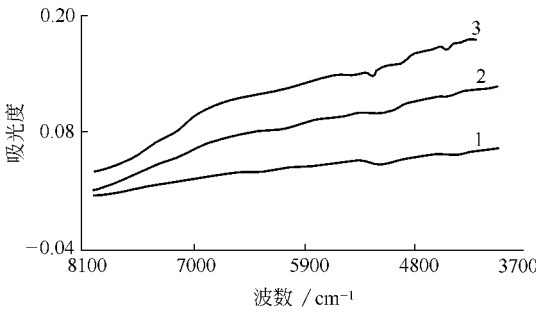


图 6-26 不同高度样品的光谱差谱图

1—垫 0.3mm 与不垫的差谱 2—垫 0.6mm 与不垫的差谱
3—垫 0.9mm 与不垫的差谱

样品高度产生的误差，严衍禄等设计了以下实验^[10]。

在同一样品池下垫不同高度的垫片以提高样品的高度，也可避免高度与松紧度同时变化。垫片的厚度分 4 个阶梯（不垫片，垫一层垫片，垫二层垫片，垫三层垫片，每层垫片约 0.3mm 厚），分别扫描 4 个高度样品，共得 4 个图谱。为了比较高度的影响，将垫一、二、三层垫片的图谱与不垫垫片的图谱做差谱处理，分别得图 6-26 中的曲线 1，2，3。

由图 6-26 可见，由于高度的不同，其图谱中主要产生了以下变化：

- ① 图谱基线平移；
 - ② 图谱的旋转；
 - ③ 在图谱的峰谷处差谱也出现了微小的波动，即存在非均匀变化与其他随机变化。
- 下面具体分析产生这些变化的原因。

无穷厚度的样品在受到入射光照射后发生吸收和漫反射。样品的高度发生变化后，其漫反射光线的方向和分布随之发生变化，如图 6-27，图中 1、2 为样品池的位置，其中 1 为正常位置，2 为垫高后的位置；3、4 为凹面反射镜，入射的分析光经反射镜 3 聚焦后，射于样品面上，样品的漫反射光经反射镜 4 反射，投向检测器。样品池位置发生变化，检测器接收光的多少和所携带的样品信息发生变化。其中光总能量的变化引起了图谱的基线平移。另外，由于在漫反射中散射系数 S 只是近似不变，而实际上的散射系数 S 随着波数的减小而变小，这样 K/S 就会变大，故： $A = \lg[1 + K/S + \sqrt{(K/S)^2 + 2(K/S)}]$ 也就会逐渐变大，随着样品高度的增加，在低波数区，不但散射光角度更大（位置 2），而且接受光的角度发生变化，这样低波数区光谱的吸光度值更大了，使光谱发生了旋转。由图 6-27 知光路的变化，加之光谱参数、吸光度和 K 、 S 的非线性关系，所以在差谱图上尤其是特征区（ $5600 \sim 4000\text{cm}^{-1}$ ）也出现了一些细小的峰谷，其大小与原光谱的吸光度大小有一定关系，此即图 6-26 中所示的非均匀变化。因此，为了减小因样品高度引起的以上变化，样品制备过程中需注意装样的高度。也有些漫反射仪器在设计时就保证了光源到样品面之间的距离不变，因此测试过程

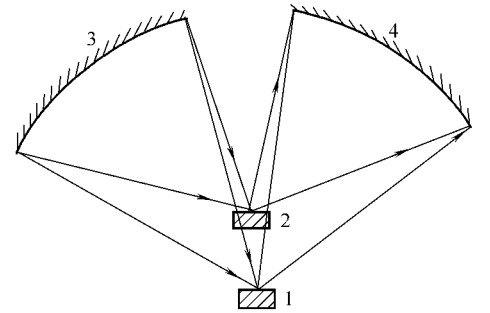


图 6-27 不同样品高度的漫反射光路图
1—正常位置 2—垫高后的位置 3、4—凹面反射镜

大，故： $A = \lg[1 + K/S + \sqrt{(K/S)^2 + 2(K/S)}]$ 也就会逐渐变大，随着样品高度的增加，在低波数区，不但散射光角度更大（位置 2），而且接受光的角度发生变化，这样低波数区光谱的吸光度值更大了，使光谱发生了旋转。由图 6-27 知光路的变化，加之光谱参数、吸光度和 K 、 S 的非线性关系，所以在差谱图上尤其是特征区（ $5600 \sim 4000\text{cm}^{-1}$ ）也出现了一些细小的峰谷，其大小与原光谱的吸光度大小有一定关系，此即图 6-26 中所示的非均匀变化。因此，为了减小因样品高度引起的以上变化，样品制备过程中需注意装样的高度。也有些漫反射仪器在设计时就保证了光源到样品面之间的距离不变，因此测试过程

中不再受这一因素的影响。

本章 6.4 和 6.5 两节内容着重讨论了近红外分析测试过程中样品光谱受仪器状态、扫描条件、样品状态和装样条件等因素的影响,以及由此对测量结果产生的影响。液体样品的透射分析法受装样条件和样品状态等因素影响较小,所以部分讨论着重以固体样品的漫反射测定为主介绍。

近红外分析测试过程中影响近红外光谱、造成光谱不确定性的因素主要来源于两个方面:一是与近红外仪器有关的各因素,如仪器波长的准确度、分辨率、信号能量,扫描参数的设置等;二是与样品有关的各因素,如样品的含水量、样品的温度、样品的粗细、样品的均匀性、样品的松紧度以及样品的高度等。这些因素对近红外光谱的影响可分解为下列四种独立影响的加和:①随机因素使样品光谱产生平移,这种结果表现在差谱是一条与横轴平行的直线;②随机因素对光谱值的影响随着波数点的变化而变化,如果这种影响的大小和波数成线性,则在差谱中存在一条有一定斜率的直线,斜率的大小表示影响随波数变化的程度,我们把这种对图谱影响称为光谱的基线旋转;③随机因素对光谱的影响随光谱值(反射吸光度值)的大小而变化,反应在差谱上是在原吸收峰谷的位置,差谱值有一定的变化,我们把这种对图谱的影响称为与吸光度有关的随波数的非均匀变化,简称非均匀变化;④随机因素对光谱的一些其他影响。分析测试过程中各因素对光谱造成的以上影响最终会影响到近红外分析结果的精确度与准确度,6.4 和 6.5 两节文中相应内容已提到减少这些影响的一些措施,下面将综合讨论减少各随机因素对光谱产生影响的方法。

6.6 校正各随机因素对近红外测定的影响

讨论减少各随机因素影响的措施方法是十分必要的,也是近红外技术能否广泛应用的重要内容。我们把校正随机因素对近红外光谱的影响分为两种,一种是用制样技术与仪器技术来校正,称为硬件控制措施,另一种是用数学方法来校正,称为软件控制措施。

6.6.1 硬件控制措施

要减少由随机因素带来的误差,采取硬件控制措施是首要的,也是行之有效的。

首先在样品制备过程中,有必要制定规范化操作规程。所有校正集的样品光谱都须在同一测量条件下逐一测量,这些测量条件也将在预测未知样品时使用。规范化操作规程包括对样品脱壳彻底性的控制;保持压模压力的一致,使样品松紧程度、高度一致;磨样过程中粒度的一致性;装样过程中样品的充分搅匀,使浓度分布和粒度分布均匀一致等都是十分重要的措施。为了减少装样不均匀引起的误差,建议每个样品装样 2~3 次,测定结果去掉异常数据后取平均值;有些厂家的仪器进样系统有旋转机构,尽量采用旋转以减小因混合和装样不均匀引起的误差,减少因这些方面的差异而引起的光谱的旋转和差谱中的非均匀变化,前已说明,旋转方法使重复测量间的标准差减少了 1 倍多。

其次,在扫描过程中,需控制分析光束截面积的大小保持不变和聚焦程度一致,这是很容易做到的措施。每次扫图前都首先调节仪器信号能量,尽量达到一致,而且每张图谱都扫描多次取平均值,也可大大减少光谱中基线的平移、旋转、差谱中的非均匀变化以及其他随机变化。扫图前对样品采取水分一致化处理,如在同一条件下自然平衡几天。扫图

时间要短，一般在几分钟内完成等措施，可大大减少水分的影响。这些硬件控制措施都是切实易行而又十分有效的，很大程度的保证了近红外分析的准确性。

当然，以上各随机因素对近红外分析产生的影响也可以在建立模型过程中通过建模样品集的广泛代表性而得以减小。例如若建模样品集中包含不同粒度的样品，则模型对不同粒度的样品都可分析，但其准确性一般会低于上述通过制定规范化操作规程而完全由同一粒度建立模型的结果。

6.6.2 软件措施——数学预处理方法

如果确实没有办法实现规范化操作，或者即使规定了规范化操作规程，但近红外测定过程仍不可避免的产生误差，这时可以进一步利用数学预处理方法消减光谱中受随机因素影响产生的误差部分，光谱的数学预处理是近红外分析测试过程中很重要的一环。

数学预处理方法有很多种，详见第 5 章内容。如一阶导数能明显消除光谱的基线平移，二阶导数可以消除光谱的基线旋转，样品粒度等原因引起的对光谱的影响可以利用附加散射校正（MSC）方法来校正，光谱中的噪声可以通过平滑、傅里叶变换法（FFT）、小波变换法（Wavelet Transform）等去除。下面以小波变换法为例说明数学预处理方法在光谱去噪的作用。选取 Mallat 正交小波基对小麦样品 4000~12000cm⁻¹ 的漫反射近红外光谱进行正交小波变换，根据实际情况取分解次数为 3，把第一阶小波系数置零反变换后得去噪后的光谱，其中 10000~12000cm⁻¹ 段信噪分离前后的对比见图 6-28，由图可见，去噪后的光谱与原始光谱相比，信噪比有了很大提高。

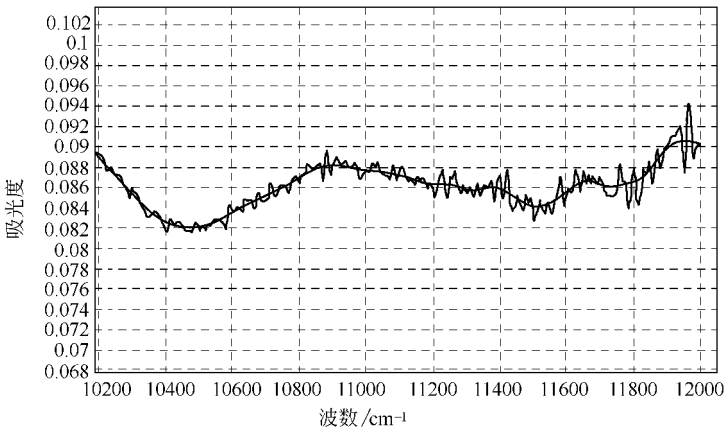


图 6-28 原始光谱和去噪后的光谱在 10000~12000cm⁻¹ 的对比

为了进一步观察小波变换的信噪分离作用，对 55 个小麦样品的近红外光谱只取 8000~12000cm⁻¹ 一段，此波段信息量少，实际计算通常不取此段谱区，但该谱区噪声强，更易看出信噪分离效果。未去噪前，此波段建立的模型效果见表 6-17，散点图见图 6-29（a）；利用小波变换去噪后，模型效果见表 6-17，散点图见图 6-29（b）。

表 6-17 信噪分离前后小麦蛋白的近红外数学模型对比（波长范围：8000~12000cm⁻¹）

	相关系数	校正标准差	相对标准差/%	平均绝对误差	平均相对误差/%
分离前	0.07	2.14	13.09	1.85	11.55
分离后	0.90	0.92	5.60	0.71	4.40

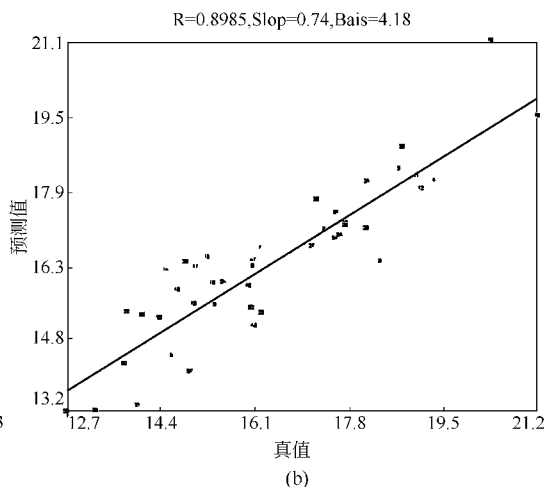
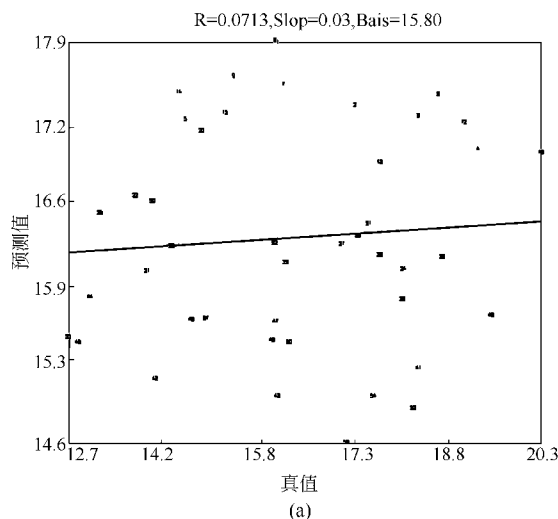


图 6-29 散点图

(a) 信噪分离前模型散点图 (b) 信噪分离后模型散点图

利用小波变换进行信噪分离前，模型的相关系数为 0.07，相对标准差为 13.09%，散点图也比较离散。信噪分离后，模型的相关系数变为 0.90，相对标准差为 5.60%，散点图也变得比较集中。可见小波变换确实具有较好的信噪分离作用，应用小波变换对近红外光谱做数学预处理对数学模型起到了很好的作用。

当然，没有哪种“神奇”的数学预处理算法能适合于所有物质的各种成分分析，所以在具体应用过程中应尽可能尝验多种数学方法，以便得到适合于该特定应用的最佳预处理方法。

根据近红外分析测试流程，光谱扫描后接下来要进行建立数学模型的工作，第 7 章“近红外光谱分析的数学模型及其建立”对此做了详述。优秀的近红外数学模型有三个基本标志，即模型的稳定性、可靠性及动态适应性。第 8 章“近红外光谱分析数学模型的优化”主要介绍得到可靠性、稳定性及动态适应性好的数学模型的一系列优化技术。建立一个优秀的近红外数学模型是一件费时、费力、耗资的繁重工作，在数学模型没有损失预测能力的前提下，人们希望能够长久而且不加变动的利用它，但是每个数学模型有一定的适配范围。第 9 章“近红外光谱分析数学模型的维护”讨论了数学模型的适配性检验方法，以及数学模型的维护问题，从而保证分析的准确度。以上都是近红外光谱分析测试技术中非常重要的环节，具体内容详见各章。

参 考 文 献

- [1] 严衍禄, 赵龙莲, 李军会等. 现代近红外光谱分析的信息处理技术. 光谱学与光谱分析, 2000, 120 (6): 777~780
- [2] 陆婉珍, 袁洪福等. 现代近红外光谱分析技术. 北京: 中国石化出版社, 2000
- [3] <http://www.bruker.com>
- [4] 江涛, 晓晨. 利用近红外分光法非破坏测定水果内部质量. 激光与光电子学进展, 2000 (8): 52~55
- [5] Hans Buning-Pfaue. Analysis of water in food by near infrared spectroscopy. Food Chemistry, 2003

- [6] 赵丽丽. 果品类内部品质近红外无损检测技术的研究. 硕士学位论文. 2003
- [7] 刘志华, 潘良文译. 应用近红外技术的实验方法. 现代商检科技, 1995: 59~65
- [8] 赵龙莲. 近红外光谱定量分析数学模型的转移. 硕士学位论文. 1999
- [9] 赵丽丽, 赵龙莲, 严衍禄. 傅里叶变换近红外光谱仪扫描条件对数学模型预测精度的影响. 光谱学与光谱分析. 2004 (1): 41~44
- [10] 严衍禄, 景茂等. 傅里叶变换近红外漫反射光谱分析测量误差的研究. 北京农业大学学报. 1990, 16 (增): 37~48
- [11] 吴瑾光主编. 近代傅里叶变换红外光谱技术及应用. 北京: 科学技术文献出版社, 1994
- [12] 王文真. 在近红外光谱定量分析中应注意的几个问题. 现代科学仪器, 1996 (1): 24
- [13] Florian Wulfert, Wim Th Kok. Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. Chemometrics and Intelligent Laboratory Systems, 2000, 51: 189~200
- [14] Mitsumoto M, Ozawa S, Mitsuhashi T. The influence of sample temperature on the determination of moisture and fat content in beef longissimus by near infrared spectroscopy. In: A M C Davies, R Giangiacomo, et al. Proceeding of the 9th International Conference, Near Infrared Spectroscopy. Chichester, UK: NIR Publications, 805~807
- [15] Robert N Feudale, Nathaniel A Woody, et al. Transfer of multivariate calibration models: a review. Chemometrics and intelligent laboratory systems, 2002 (64): 181~192

7. 近红外光谱分析的数学模型及其建立

7.1 概述

本书第2章已讨论了分子的近红外光谱主要包含分子振动倍频与合频的谱区。与中红外光谱类似，该谱区信息量丰富，特别是几乎包含全部含氢基团的信息，因此近红外光谱分析可对复杂样品进行多组分分析；与可见光谱类似，获得该谱区的光谱相对比较容易，而且分子对倍频与合频的吸收强度较低，近红外区穿透物质的能力相对较强，因此可进行复杂样品如生物样品的无损检测。

另一方面，在复杂物近红外光谱中每个波长点叠加了多种组分的信息，样品中每一种有机组分在近红外谱区的多个波段都有信息，属于复杂光谱。该谱区的光谱复杂、谱峰重叠、谱带较宽、光谱变动、干扰信息大，实行分析的难度较大。

复杂物的近红外光谱分析具有常规分析不同的一些困难，需采取一整套分析的特殊策略。

7.1.1 复杂物近红外光谱分析信息提取的策略

实现复杂样品近红外光谱分析的主要困难就是有效信息率（光谱中有效信息和总信息的比）低，这是指复杂物近红外光谱中能够直接用于分析的待测量信息强度低，而待测量以外的背景信息复杂、强度较高而且波动性大。

近红外光谱分析是从复杂、重叠、变动的背景中提取弱信息。这个过程可称为对光谱的校正。近红外光谱分析校正光谱的基本过程主要有两步：首先是建立用于校正光谱的校正样品光谱集，其次是运用算法利用校正样品光谱集建立光谱特征与待测量之间的数学关系，即数学模型，实际就是对光谱的校正。

校正样品光谱集是近红外光谱分析的信息源，是校正样品的基础，校正样品光谱集应当包括背景复杂、光谱重叠的信息（有效信息），并尽量降低干扰信息。建立校正样品集的策略是开辟与建立有效信息率高的分析信息源，可称之为“开源”的策略。

建立数学模型的策略是对校正样品集运用化学计量学算法，充分提取光谱中的有效信息，提高光谱信息的利用率，对光谱进行校正，把光谱表达为待测量，建立预测性能好的数学模型，可称之为“节流”的策略。

总之，校正光谱的基本策略是通过“开源”和“节流”的策略，解决光谱中有效信息率低的困难。

图7-1是近红外光谱分析解决分析的难点的技术策略。

图7-1中有关近红外光谱的特点及技术难点参见第2章。下面就近红外光谱分析的策略加以说明。

(1) 现代近红外光谱分析的“开源”策略 开源策略的目标是通过优化循环中的一系

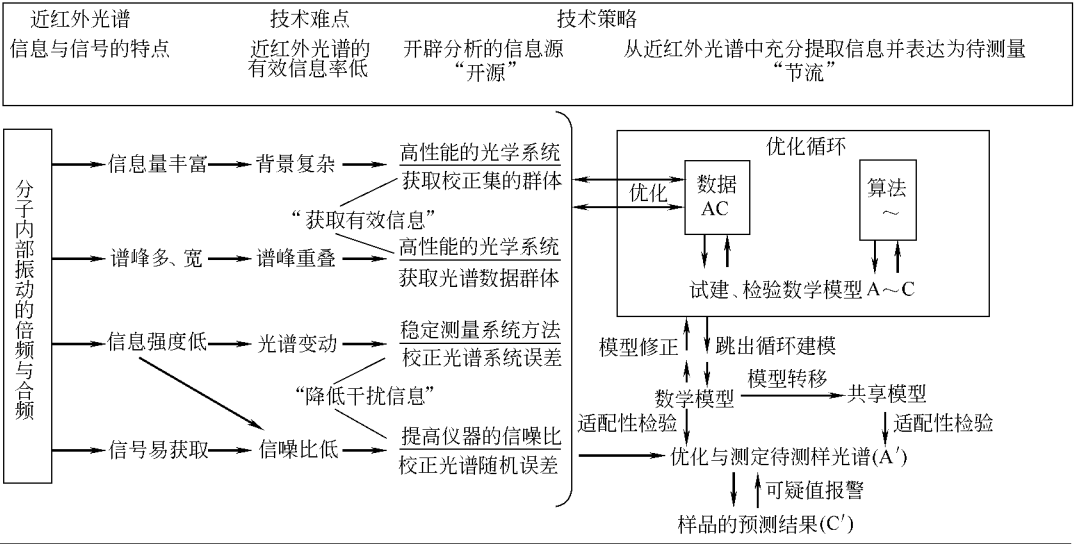


图 7-1 现代近红外光谱分析的技术策略

图中：连线与箭头（→）表示相关项；粗线框内为试建数学模型时对建模数据与建模算法的优化循环；“优化循环”、“适配性检验”、“模型转移”、“可疑值报警”等说明见后。

列优化技术，为近红外光谱分析开辟与建立有效信息率较高的信息源，即提高信息的有效率。

近红外光谱分析运用硬、软两个手段（图 7-1 的“开源”栏中分别列在横线的上、下方），通过图 7-1 中的“获取有效信息”与“降低干扰信息”两方面来实现“开源”。

开源策略的获取有效信息：即充分采集与获取完成近红外分析必需的有效信息。近红外光谱分析需要从样品复杂的光谱中提取两类信息：样品光谱中关于待测量的定性或定量信息，以及可以通过模型加以校正的背景信息；这两类信息都是近红外光谱分析需要的有效信息。近红外光谱分析主要通过两个数据群体，采集与获取光谱中的上述两类有效信息。这两个数据群体是：

① 用于光谱校正的样品数据集：为了取得有关复杂背景的信息，现代近红外光谱分析运用一批有代表性的、尽可能包含所有复杂背景信息的样品所组成的样品群体，该样品群体即建模所用的校正样品集。

② 用于光谱校正的光谱数据集：为了从重叠的光谱中取得特定的信息，现代近红外光谱分析必须运用校正样品集中各样品光谱的全部或部分特征波长点组成的光谱数据群体，以此获取样品中各种成分在光谱总体上的差异与特征。

近红外光谱分析这两个数据群体（集）的特征可以用校正样品集的二维光谱矩阵（图 7-1 中以矩阵 A 表示）来描述。实际为了建立样品光谱特征与待测量化学值之间的关系，还需测定校正样品集中各样品待测量的化学值（图 7-1 中以矩阵 C 表示）。因此矩阵 A、C 是通过“开源”策略建立的近红外分析所需的数据，是近红外分析的信息源。

开源策略的降低干扰信息：样品光谱中的干扰信息与无效信息主要是光谱测量的系统误差与随机误差。由图 7-1 可见，近红外光谱分析主要运用软件和硬件的滤波技术降低光谱中这两种误差，以降低光谱的不确定性并增加光谱的信噪比。

总之，近红外光谱分析的“开源”策略是通过“获取有效信息”与“降低干扰信息”中的四种主要措施，建立分析需要的信息源。从分析信息源中提取有效信息、确定待测量，则需要通过“节流”策略来完成。

(2) 现代近红外光谱分析的“节流”策略 采用节流策略的目标是通过优化循环中的一系列优化技术，从分析信息源中充分提取有效信息并表达为待测量，即提高有效信息的利用率。

近红外光谱分析通过优化与建立数学模型的方法，来实现“节流”：由于近红外分析的信息源是两个规模较大的二维光谱矩阵数据 A、C，不可能采用经典光谱分析方法从中提取需要的信息。由图 7-1 可见，近红外光谱分析只能运用化学计量学方法，优化与处理二维矩阵“A”与“C”，引入现代数学中的多元统计算法（图 7-1 中用“~”表示），来校正光谱，建立 A 与 C 之间的数学关系，（图 7-1 中用“A~C”表示的数学模型）。

分析待测样品时只要测定其近红外光谱（图 7-1 中用“A'”来表示），就可以通过数学模型，确定其待测量（图 7-1 中用“C'”来表示）。“节流”策略中关于“优化循环”、“适配性检验”、“模型转移”与“可疑值报警”等内容都在本章 7.5 与第 9 章中再做说明。

概括说，近红外光谱分析数学模型的性能决定于两个建模要素：建模数据和建模算法。这两个建模的要素需要通过“开源”与“节流”策略，在优化循环中反复优化才能建成优秀的数学模型。其中，按照“开源”的策略建立优秀的分析数据；按照“节流”的策略，运用算法，处理分析数据，建立优秀的数学模型；再运用数学模型，完成样品分析。

为了实现近红外分析的这些策略，需要采取一系列技术，主要有三项核心技术。

7.1.2 近红外光谱分析建立数学模型的核心技术

近红外光谱分析是从复杂的背景下提取弱信息。这种分析必须用足够大的样品集，才能采集样品的复杂背景信息。例如为了建立中国小麦等作物的数学模型，原则上应当取得国内不同地区、数以万计（达到 $10^4 \sim 10^6$ 左右）不同品种样品的作物光谱，建立中国作物样品光谱库，以此为基础建立数学模型。另外，近红外光谱分析必须用足够大的光谱数据点才能区别光谱中重叠的信息，例如：每个样品的傅里叶光谱通常有 10^3 个数据点。两个数据集的相乘使近红外光谱分析所采集的信息是海量的信息，如图 7-2 中达到 $10^7 \sim 10^9$ 数据规模的矩阵。但对某种具体的近红外光谱分析，有效的信息仅仅是样品某种成分或某一特征的微弱信息，如谷物中少量蛋白质含量的弱信息。近红外光谱分析待处理的弱信息叠加了各种干扰信号，包括系统的误差与随机的噪声。因此，近红外光谱分析最基本的困难是光谱信号的信息量大，有效信息率低。近红外光谱分析建立数学模型主要有三项核心技术：

(1) 压缩 近红外光谱分析的本质是海量数据的采集与处理，处理海量数据必先对数据加以压缩。所谓信号数据的压缩，是指保持信号中的有效信息量，降低表达信息的数据量，以便于对信号的处理。近红外光谱分析的信号压缩包括对样品集（样品数）的压缩和光谱数据点集的压缩两方面。压缩的算法主要有主成分分解、傅里叶变换、小波变换等。

(2) 恢复 近红外光谱有效信息的强度较低，属于弱信号。另外，每个光谱数据除了包含有效信息以外还包含测量误差等不确定的、难以校正的、干扰测定的无效信息或称干扰信息，这些干扰信息的强度一般较大。弱信号的处理需要放大，但是伴随弱信息的放

大，噪声也随着放大，所以处理弱信号首先要通过数学处理，降低叠加在弱信号上的噪声与系统干扰（误差），增强有效信息的相对强度，以恢复弱信号的原有特征，这就是弱信号的恢复。近红外光谱分析需要用两种校正技术，即消除或降低信号中系统误差和随机误差的干扰，来处理恢复近红外光谱的弱信号。恢复的算法主要包括平均、求导、矢量归一化、傅里叶变换、小波变换等。

弱信号恢复与信号压缩是从两个不同的层面来提高近红外光谱分析信号的有效信息率，都属于近红外光谱分析海量数据的前处理，是实现（近红外光谱分析）提取信息“开源”策略的两种核心技术。

（3）关联 经过前处理的光谱数据通过算法与待测量之间进行数学关联。所谓关联是指，建立两组量间的数学关系（即数学模型），通过该数学关系可从其中一个量得到另一个量。关联的方法可以通过解方程组，得到方程确定的解，也可通过统计拟合的方法（如回归方法）建立模型。“关联”是实现（近红外光谱分析）提取信息“节流”策略的一项核心技术。

近红外光谱分析三项核心技术的本质都是属于优化（建模数据与建模算法），目标是为了建立优秀的数学模型。三项核心技术是相互联系的，每项核心技术都不可能单独一次完成，近红外光谱分析必须反复通过一系列优化技术来完成分析；通过优化仪器参数与制样技术获取优良的近红外光谱；通过优化光谱与优化校正样品集来优化建模数据；通过优化建模数据与优化提取信息的算法来优化数学模型，通过优化的数学模型与待测样的光谱得到可靠的近红外分析结果。

图 7-2 中给出了三种核心技术（在图中的圆框内）之间的关系，以及为什么要采用这些技术的原因。

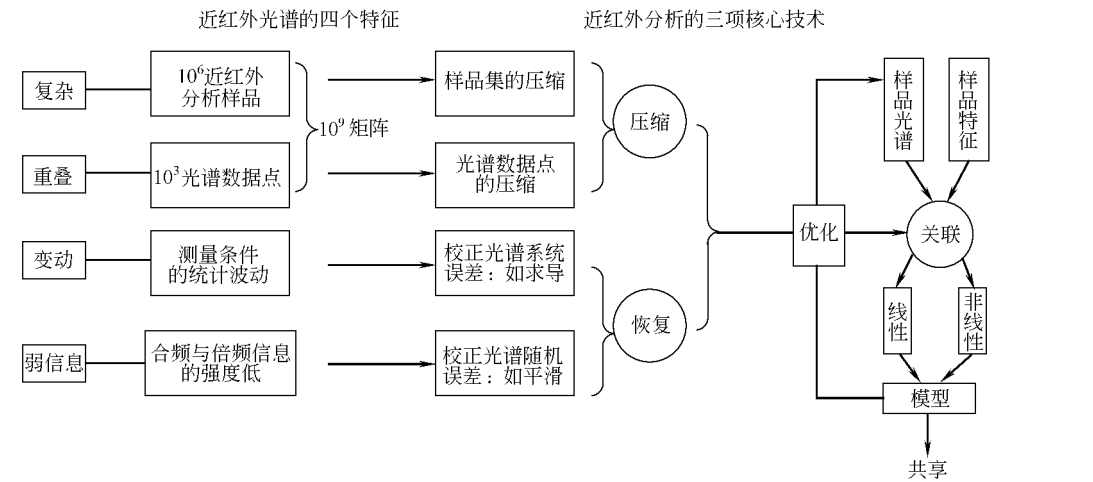


图 7-2 近红外光谱分析的三项核心技术

运用近红外光谱分析的三项核心技术要借助软件技术的帮助。

7.1.3 近红外光谱分析核心技术的运用原则

近红外光谱分析运用三项核心技术来优化建模数据与算法参数的原则是用最少的光谱

数据，最大限度包含分析的有效信息；简称“少而精”的原则。

图 7-3 表示数学模型引入的光谱数据 N 与数学模型预测误差 E 间的关系。建模过程应用的光谱数据越多，得到的有效信息就可能越多，这就使模型在不同时间与空间运用的稳定性得以提高，同时预测误差 E 减少、预测准确度也得以提高，见图中的曲线 A。另一方面，因为每个光谱数据除了包含有效信息以外还包含测量误差等不确定的、难以校正的、干扰测定的无效信息或称干扰信息，因此建模过程中每引入一个光谱数据的同时会带来影响提取有效信息的干扰信息，使模型的预测误差 E 增加、测定准确度下降，如图 7-3 中的曲线 B。曲线 C 是图中 A、B 的叠加。图中 A、B 两曲线的交点附近是建模所用光谱数据的最佳点。优化建模数据的目标就是确定或接近该最佳点，使数学模型的预测效果达到或接近最佳值。优秀的软件应能辅助确定数学模型的最佳参数。

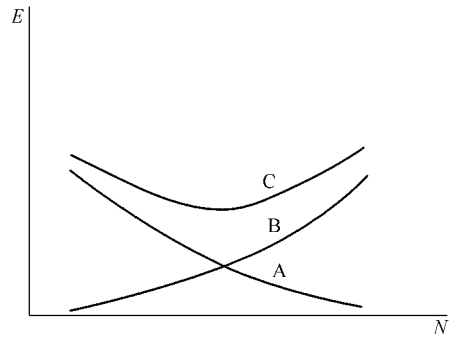


图 7-3 建模过程应用的光谱数据 (N) 与模型预测误差 (E) 的关系示意图

图中曲线 A 表示随着应用的光谱数据增加，由于有效信息增加，降低预测误差；曲线 B 表示随着应用的光谱数据增加，由于干扰信息增加，增加预测误差。

在这里所谓光谱数据是泛指近红外光谱分析的样品集或样品的光谱集，在建模过程的不同步骤，光谱数据所指不同。如：建模样品数的选择、各种处理数据的算法，分析谱区的选择、最佳阶数的确定、异常值的剔除、适配性的检验等等，这些特定光谱数据的处理有各自的技术，这些技术的本质都与提高分析信号中的有效信息率有关，运用这些技术都需要符合“少而精”的原则。

近红外光谱分析的算法是统计方法为主的多元校正，算法处理的对象——样品近红外光谱本身具有统计学的波动特性（见 7.4.2）。近红外光谱分析过程就是对一系列近红外光谱数据变量（或统计量）的优化处理。其中近红外光谱数据“少而精”处理原则的本质是指对近红外光谱控制有效信息量与有效信息率之间掌握合适的度。目前近红外光谱分析对这些数据变量优化处理的度主要靠经验或者通过软件实际测试来确定，对此还缺少精确的理论分析与定量的确定。这种优化处理尺度的掌握贯穿于近红外分析的全过程，如选定模型预测样品的适配范围以及主成分空间的大小（见 7.2.2）、确定算法的参数（见 7.3.2）、确定近红外光谱分析对变动背景的自校正与外校正（见 7.4.4）、模型预测结果的准确度与精确度（见 7.4.1）等等近红外光谱数据的处理都需要掌握合适的度。

运用近红外光谱分析“少而精”的原则也要借助软件技术的帮助。

本章的后面几节将分别讨论近红外光谱分析的几种核心技术及其应用。其中 7.2 节着重讨论“恢复、压缩”的应用；7.3 节用简单关联模型为例，讨论“关联”、“优化”的应用；7.4 节以简单关联模型为代表，讨论模型的结构与模型预测性能的关系；7.5 节主要讨论综合运用这些技术实际建立与优化数学模型的过程，还论述了近红外光谱分析资源共享的思想。在这些讨论中将介绍一些实例。

7.2 近红外光谱分析数据的前处理

近红外光谱（NIR）分析的关键是建立预测效果优秀的数学模型。数学模型预测样品的效果决定于建模所用数据，以及（用算法）对建模数据中信息的充分提取。NIR 分析大致有一半的误差来自于建模数据。因此在运用关联算法建立数学模型前，还需要对光谱数据进行前处理，以优化建模数据。近红外光谱数据前处理的主要技术是前面已讨论过的“恢复、压缩”。

建模数据也就是建立数学模型所用校正样品集。校正样品集包括直接用于建立模型的建模样品集与检验模型的检验样品集。现代 NIR 分析包括一系列优化校正样品集光谱的技术，如：建模集与检验集的分割，优化校正样品集总体的样品组成，优化各样品光谱，优化选择用于建立数学模型的谱区以及优化选择各种 NIR 定量分析算法的最佳参数等等多种多样的处理技术，由上节可知这些前处理技术的本质都是压缩和恢复，都应符合“少而精”的原则，目标都是提高建模数据的有效信息率。

近红外光谱分析建模数据的各种前处理技术，以及这些技术的目标见图 7-4。

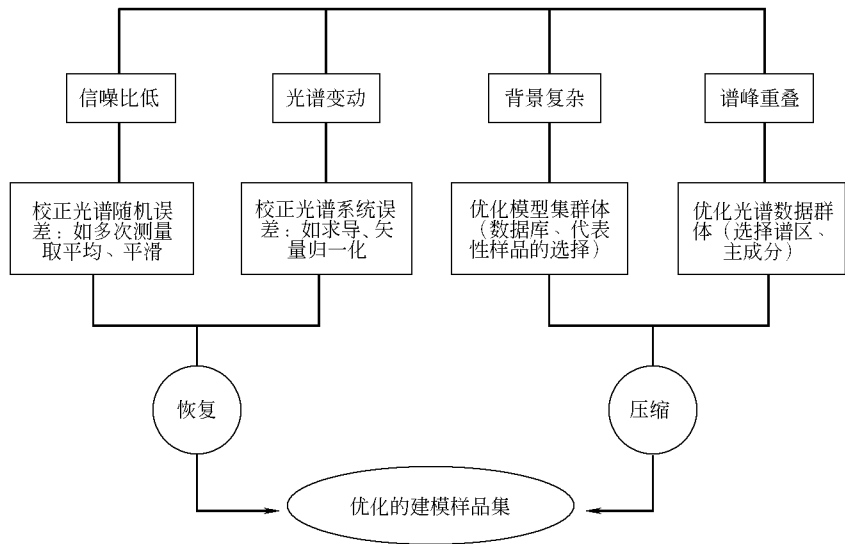


图 7-4 近红外光谱信息前处理

7.2.1 近红外光谱分析数据的弱信号恢复

7.2.1.1 降低近红外光谱信号中的系统误差

NIR 光谱的误差主要有两种：随机误差与系统误差。样品集光谱的前处理就是为了降低这两类误差。可以通过多次测定取平均、平滑、傅里叶滤波等方法降低随机误差；用导数光谱等方法降低或消除某些系统误差。

由泰勒级数得到任意函数可用不同幕次的幂函数之和来逼近：

$$f(t) = f(0) + f'(0)t + \frac{1}{2!}f''(0)t^2 + \frac{1}{3!}f^{(3)}(0)t^3 + \dots \tag{7-1}$$

函数展开式中的零次项相当于函数对应曲线的平移，一次幂项相当于对应曲线的旋转。通过对光谱求导数，可以降低对应函数的幂次，以消除对光谱的某些干扰。对光谱求一次导，可消除对应函数展开式中的零次项，即可消除光谱的平移误差，求二次导可消除展开式中的一次项，即可消除光谱的旋转误差。

傅里叶变换近红外光谱仪的信噪比一般较高，通常能达到 10^4 以上，其光谱误差主要是系统误差；此类失真在建模样品集光谱的前处理时可以用导数方法给以校正。导数处理近红外光谱分析建模光谱集提高所建模型的预测效果见表 7-1。

表 7-1 导数处理噪声较低的光谱对数学模型预测效果的影响

前处理方法	仪器、建模样品 与评价指标	A-FT 光谱仪 83 个烟草样品		B-FT 光谱仪 34 个玉米样品		C-FT 光谱仪 32 个烟草样品	
		R^2	SECV	R	SECV	R	SECV
不处理		95.22	0.162	0.957	1.119	0.991	0.513
一阶导数		97.33	0.0957	0.963	0.973	0.995	0.401
二阶导数		97.39	0.094	0.966	0.948	0.999	0.214

注： R^2 为决定系数， R 为相关系数，SECV 为内部交叉证实标准差。

表 7-1 内三类样品分别在 A、B、C 三台高性能傅里叶变换（FT）近红外光谱仪上测定 NIR 光谱；这些光谱分别经平滑处理后再用导数法处理，用 PLS 法建立数学模型、用 SECV 表示前处理对预测效果影响。

由表内数据可见，各种条件下对不同的 FT 光谱仪扫描的漫反射光谱，用导数处理均对数学模型的预测效果有较好的作用。这是因为扫描样品的漫反射光谱时，由于仪器状态、样品状态与测量条件的差异，容易造成 NIR 光谱的平移或旋转。光谱的平移或旋转对模型的预测效果有一定影响，但可以通过一阶导数或二阶导数给以校正，从而使模型预测的决定系数得以提高，SECV 得以下降。

表 7-1 数据是对低噪声光谱导数处理的结果。对于噪声较高的光谱，导数处理可能反而降低模型的预测效果。表 7-2 是 32 个烟草样品用色散型 NIR 光谱仪实验样机（信噪比约 10^2 ）所作的光谱，用导数处理后对模型预测效果的影响（表内数据为 PLS 法建立的数学模型预测 16 个检验样品的结果）。

表 7-2 导数处理噪声较高的光谱对数学模型预测效果的影响

前处理方法	前处理方法与 评价指标	原 光 谱		五点三次平滑	
		R	SECV	R	SECV
不处理		0.994	0.065	0.994	0.065
一阶导数		0.976	0.098	0.984	0.094
二阶导数		0.762	0.290	0.757	0.284

表内数据可见：导数处理噪声较高的原光谱后，反而降低了数学模型的预测效果，而且导数的阶数越高，影响越大；五点三次平滑对一阶导数的预测效果略有改善，但不明显；对二阶导数则无效。这是因为导数处理提高了噪声值。

7.2.1.2 降低近红外光谱信号中的随机误差

提高近红外光谱的信噪比可以应用多种的硬件与软件技术，最常用的算法是各种求平均、平滑的方法，其中通用的算法是利用傅里叶技术和小波变换。

由傅里叶级数可知：任意一个周期性函数可用不同频率的余弦函数之和来逼近，由此可得到该函数频谱：

$$f(t) = a_0 + \sum_{i=1}^{\infty} a_i \cos(2\pi v_i t) \quad (7-2)$$

频谱中不同的频率分量代表了时域中原函数的不同（频率）信息成分。原谱中的噪声部分在频谱中主要出现在高频部分，因此，对原谱进行傅里叶变换得到频谱后滤去高频分量，再反傅里叶变换得到的新谱就相当于降低噪声的原谱。有关内容详见第 8 章。

7.2.2 近红外光谱分析数据的压缩

7.2.2.1 近红外光谱的主成分及其在近红外分析数据压缩中的应用

近红外光谱分析海量数据的压缩包括两项，即对近红外样品集的压缩与对近红外光谱数据集的压缩；压缩这两个集的原则都是“少而精”。矩阵主成分分解法在近红外光谱分析样品集的压缩、光谱数据集的压缩以及异常样品的选取等多方面都有重要的应用。

主成分分解即：将 n 个样品，每个样品 m 个波长点处的光谱信息组成的矩阵 A_{nm} ，压缩为由 n 个样品，每个样品光谱信息用少数几个（如 d 个）主成分向量构成的矩阵，而每个主成分为 m 个波长点处光谱信息的线性组合，且各主成分向量之间相互正交。见式 (7-3)：

$$A_{nm} = T_{nd} P_{dm} + E_{nm} \quad (7-3)$$

式中 E_{nm} 为残差矩阵，表示将 n 个样品 m 个波长点（如 2000 个点）处的光谱信息压缩为 d 个（如 3 个）主成分后，所丢失的信息。有关主成分分解的数学方法详见第 3 章，本章只简单讨论主成分分析在近红外光谱数据压缩中的应用。

因为光谱的每个主成分都分别大体表征光谱中某一定类型的信息，因此可以通过样品集中各样品主成分得分（SCORE）的空间分布图（简称样品的主成分空间分布），来了解样品的总体包含的有效信息与背景信息范围，即通常所说的样品代表性；而正是这个信息范围决定了由该样品集建立的数学模型在预测样品时能够适应与配合的范围——称为适配范围。一般认为数学模型对于空间分布范围以外的样品或“空洞”（指样品集的主成分空间中有某些缺乏校正样品的区域）内样品的预测准确度较低。因此稳定性高的数学模型应具有相对较大的适配范围，其建模样品集中样品的空间分布应有足够的范围，且不存在大范围的“空洞”。

用主成分分解把光谱压缩成 d 个主成分后，在 d 个主成分组成的空间中，每个样品用一个点来表达其特征。利用样品光谱在主成分空间中的分布，可以了解每个样品光谱特征在建模样品集中的代表性，因此可以用于选择代表性样品。用主成分方法在样品集的主成分空间内分散而且均匀选择若干样品，使这些样品可以大体覆盖原主成分空间，可以从成千上万个样品中挑选出少量（数十个）具有代表性的样品，将建模样品压缩上百倍。用这些代表性样品建立的数学模型覆盖样品范围与大批样品建立的模型相似，如果按少而精的原

则选择代表性样品建立的数学模型，其预测效果甚至比用大量样品直接建模预测效果更好。

PLS 算法把光谱主成分与待测量进行关联运算，使 PLS 主成分具有更丰富的待测量信息。PLS 的数学方法详见第 3 章。

近年来发展起来的正交信号校正法（ORTHOGONAL SIGNAL CORRECTION, OSC）可以滤除与待测量信息无关的干扰信息。OSC 用于 PLS 算法的前处理，可在近红外光谱分析建立数学模型时使主成分数进一步下降（一般 OSC-PLS 的主成分数目用 3~4 个即可），从而极大地压缩近红外光谱分析的数据量，进一步提高模型的稳定性。

7.2.2.2 近红外光谱分析对样品集的压缩

有关用主成分分解压缩建模样品集的应用实例见第 8 章。本节只简单介绍关联算法中对样品集的压缩——局部回归法与压缩样品集的平衡——异常样品的取舍。

局部回归或称局部校正（Local Calibration）是指在分析样品时对每个待分析样品都分别从校正样品集中选取一部分与其接近的样品作为建模样品集。而舍去其他与其差别大的样品。由关联算法试建一系列模型，通过比较选出最佳的建模样品数与主成分阶数。用这些优化选取的样品与建模参数来建立数学模型，并用模型预测待分析样。局部回归法是关联中包含压缩的一种算法，是试建选优与压缩结合的建模方法，因此对每个待分析样都用不同的少量经过压缩最优化的样品建模对其进行分析，从而提高分析的准确度。

在近红外光谱分析中用局部回归法有可能建立复合样品（多种不同类型样品组成的样品集称为复合样品）的数学模型，该数学模型可以用于预测复合样内各类样品的指定成分。局部回归的应用实例见第 8 章。

样品集的压缩是在样品集的有效信息量与有效信息率间的平衡，因为每个样品既包含有效信息，又有干扰信息，在样品集中加入某个样品，一方面能够增加样品集的有效信息，扩大模型的适配范围，另一方面能增加干扰信息，降低了模型的分析精度，两者要权衡利弊，决定取舍，这就是压缩样品集的平衡。建模样品集中异常样品的选取就是样品集压缩平衡。

NIR 分析中通常需要将校正样品集中的某些异常样品剔除。异常样品中有些是因为其光谱或化学值误差较大，这些异常样品的有效信息率较低，应加以剔除。有些异常样品的光谱或化学值是正确的，只是由于其类型（背景）特殊；这些异常样品有的保留在校正样品集中虽然有可能降低预测准确度，但却可以提高所建模型的适配范围。

如用某种（如大麦）样品建立的数学模型，对另一种样品（如大豆）是“异常”样品，该数学模型不能预测大豆样品。实验证明在大麦样品建立的数学模型中加入少数几个“异常”的大豆样品，则原大麦样品建立的数学模型的适配范围可以扩大，以至也可以预测大麦、大豆混合样品。

7.2.2.3 近红外光谱分析光谱数据集的压缩

近红外光谱分析建模时对光谱数据的压缩包括对光谱谱区的压缩（选择较窄的谱区进行建模）以及在关联算法中选择算法的参数达到对数据压缩的目的。用主成分分解对光谱数据进行压缩见 7.2.2.1

（1）选择样品光谱的建模谱区

表 7-3 是 83 个烟草样品用不同谱区的光谱数据，用 PLS 法建立数学模型的预测效果（表 7-3 内数据为内部交叉证实结果，主成分数取 10 阶、13 点平滑后取一阶导数）。

表 7-3 选择建模光谱的谱区对数学模型预测效果的影响

起点终点波数/cm ⁻¹	谱区范围	R ²	RMSECV
4000~8000	4000	97.33	0.0957
4500~8500	4000	96.99	0.128
5000~9000	4000	96.22	0.144
6000~10000	4000	87.65	0.26
7000~11000	4000	63.11	0.449
8000~12000	4000	32.95	0.606
4000~12000	8000	97.13	0.125
4000~11000	7000	97.85	0.108
4000~10000	6000	97.3	0.0966
4000~9000	5000	97.3	0.0963
4000~8000	4000	97.33	0.0957
4000~7000	3000	97.28	0.0971
4000~6000	2000	97.27	0.0972

由表 7-3 内数据可见，在建模光谱中谱区范围同样选取等宽度 4000cm⁻¹ 作为建模谱区时，选取不同的谱区对建模效果有极大的影响，这说明在不同谱区内包含的指定待测量的信息很不相同。必须选择有效信息率最高的谱区用于分析，才能提高模型的预测效果。

另外，由表内还可看到建模谱区不是越宽越好，12000~4000cm⁻¹ 虽然包含了 8000~4000cm⁻¹，但前者建模的预测效果反而不如后者。这显然是因为 8000~4000cm⁻¹ 谱区的有效信息率高于 12000~4000cm⁻¹ 的有效信息率。

表 7-3 说明由选择建模谱区来提高数学模型预测性能的本质是：通过对光谱数据集的压缩选择部分有效信息量高的建模谱区，减少直接用于建立数学模型的光谱数据点，提高建模光谱数据的有效信息率。这也是优化建模光谱数据的“少而精”原则。

(2) 关联算法中对光谱数据的压缩及压缩的优化

近红外光谱分析数据的关联算法本质上分为两步，首先是对光谱数据进行压缩，然后对压缩的数据通过线性或非线性的算法与待测量关联建成数学模型。现代近红外关联算法通常包含了对光谱数据上百倍的压缩。其中简单关联的算法如逐步回归、所有可能回归等对原光谱中上千个光谱波长数据点，通过“逐步”的比较，只选择其几个波长点用于建立定标方程，达到压缩的目标。全谱区的关联算法（PCR、PLS）用主成分分解法，对原光谱中上千个光谱波长数据点最终压缩只选取几个主成分，再通过线性回归或非线性回归与待测量关联建立数学模型。关联算法对光谱数据压缩详见第 3 章。

在选择 NIR 定量分析算法的参数时也需考虑有效信息率，如逐步回归算法选择波长点时有一个最佳值，偏最小二乘法选择主成分数时也有一个最佳的阶数，都是为了提高建模光谱数据有效信息率，本质上是一种对压缩的优化过程，详见 7.3 节。超过该最佳值称为“过拟合”，过拟合的数学模型其预测效果下降。

图 7-5 是对 83 个烟草样品用 PLS 法建立数学模型时，主要成分的阶数与内部交叉证实 RMSECV 的关系。

由图 7-5 可见用 PLS 法建立数学模型时，主成分的阶数与内部交叉证实 RMSECV 的关系有一个最低点，图中主成分数取 10 阶时模型的交叉证实标准差最小。超过 10 阶的过拟合数学模型预测效果反而有下降趋势。

准确选择各种 NIR 定量分析算法参数的本质也是为了提高建模光谱数据的有效信息

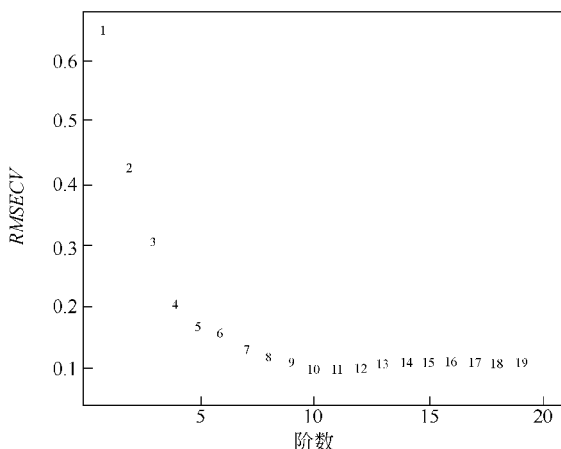


图 7-5 PLS 法建立数学模型主要成分的阶数与 $RMSECV$

率。也体现了建模光谱数据“少而精”的原则。

7.3 近红外光谱分析数据的简单关联模型

用近红外光谱分析某种待测量前必须先建立相应的数学模型，才能依据样品的光谱通过数学模型确定待测量。建模数据的前处理是便于通过关联算法建立优秀数学模型。

建立数学模型的过程就是关联样品光谱与待测量之间的数学关系。数学模型大体上有两种类型：简单的数学模型就是以通过压缩得到的少数几个波长点光谱参数（如吸光度）为自变量，待测量为因变量的某种多元方程，可以称为浓度计算方程或浓度定标方程。简单的数学模型运用简单的关联算法来建立。

较复杂的数学模型需要用特殊的算法如主要成分回归（PCR）、以及偏最小二乘（PLS）算法，对全谱区的光谱压缩得到的少数光谱参数，再与待测量做关联，建立更复杂形式的数学模型。有关问题已在第 3 章讨论。

为了更清楚的了解近红外光谱分析数学模型建立的微观过程、模型的结构及其表达信息的形式、预测性能等基本问题，本节只讨论用简单的数学关联方法，建立较简单形式的数学模型——定标方程。

7.3.1 近红外光谱分析的简单数学模型——定标方程

若把下标 i 记为近红外光谱中的波长序号，把下标 j 记为复杂样品中各组分的序号， F_i 记为复杂物近红外光谱的某一线性函数（如吸光度）在波长 i 的值，若一共选择了 m 个波长处的 F_i 值来计算复杂样品中某 j 组分的含量，则这个组分的含量定标方程为

$$C_j = K_0 + \sum_{i=1}^m K_{ij} F_i \quad (7-4)$$

式中 C_j 为样品中第 j 个待测成分的浓度， F_i 是样品近红外光谱经线性变化后的第 i 波长的某一光谱线性函数值， K_0 、 K_{ij} 为浓度定标方程中作为校正系数的常数与系数。实际的浓度定标方程还应加上一项误差项 E ，包括由于线性变化的近似性以及测量 F_i 所产生的随机误差与系统误差。方程（7-4）可改为

$$C_j = K_0 + \sum_{i=1}^m K_{ij} F_i + E \quad (7-5)$$

式 (7-4) 就是近红外光谱分析测定样品中某一组分含量的浓度定标方程或称为浓度计算方程, 即近红外光谱分析的数学模型。

由式 (7-4) 可以把近红外光谱分析确定样品中某一成分含量定标方程 (即建立数学模型的过程) 的结构要素有 3 个, 即:

- ① 选择适当形式的光谱函数;
- ② 确定定标方程的项数;
- ③ 选择最佳波长组合, 并确定相应的方程中常数项与各项的系数。

7.3.2 近红外光谱分析定标方程的结构

7.3.2.1 近红外光谱分析定标方程中的光谱线性函数

近红外光谱分析定标方程中必须选择适当形式的光谱函数 (主要是线性函数) F , 以降低误差 E 。对透射光谱, 函数 F 的基本形式为吸光度 A ; 对固体样品漫反射光谱分析法, 根据 Kubelka-Munk 定律 (见第 1 章) 光谱函数主要是反射吸光度 A 。

$$A = -\log(1 + K/S - [(K/S)^2 + 2(K/S)]^{1/2}) \quad (7-6)$$

式 (7-6) 中 K 为样品的吸收系数, S 为样品的反射系数。

浓度范围较窄时可以近似为

$$A = a + b (K/S)$$

样品的物理状态一致, S 为一定时

$$A = a + b \times C$$

由此可见, 对于近红外漫反射光谱分析, 当样品组分含量范围较窄时, 样品的紧实度等物理条件不变的条件下, 漫反射吸光度 A 或者 A 的一次、二次微分 dA , d^2A 与组分浓度成正比, 可以作为定标方程中的光谱线性函数。对光谱求导可以校正光谱平移与光谱转动的作用。因此, 一般以光谱的微分、二次微分形式建立的定标方程预测精度较高 (见 7.2.1)。

7.3.2.2 近红外光谱分析浓度定标方程项数及方程最佳波长组合

近红外光谱原有的波长点达到 10^3 , 通过对光谱压缩, 从中选出几个波长点建立数学模型。确定选用波长点的数目, 即式 (7-5) 的项数。如果选的波长数目少, 即计算方程中的项数太少, 则由于覆盖面不够, 就不能把样品中的各种影响测定待测成分的背景因素包含在内, 定标方程对待测成分预测的浓度也就不可靠了。若波长选得过多, 也会使误差 E 增大, 而且计算量过大。二者关系成 U 型的曲线 (见图 7-3) 中间有一个测定误差最小的最佳波长点数。一般定标方程的项数是 3~4 项为宜。

近红外光谱分析为了建立定标方程需要在近红外谱区内, 通过对光谱压缩, 选择一组波长 λ_i , 确定相应的定标方程中的常数与系数 K_0 、 K_{ij} , 使方程在预测未知样品浓度时精度最高, 这样的波长组合称为最佳波长组合。上述步骤是在近红外光谱分析定标方程的建立过程中完成的。

7.3.3 近红外光谱分析定标方程的建立

建立定标方程主要有两种方法: 解联立方程组的方法与统计拟合的方法。

7.3.3.1 解联立方程组建立定标方程

原则上可用解联立方程组的办法建立定标方程。解联立方程组建立定标方程分成压缩与关联两部分：信息压缩是对代表性样品（已经压缩的样品集）人工选择几个特征波长。为了建立蛋白质、脂肪和水分的定标方程由光谱理论知道 4200cm^{-1} 、 4600cm^{-1} 、 5200cm^{-1} 、分别是样品中组分，面粉（用“面”来表示）、脂肪（用 F 来表示）、蛋白质（用 P 来表示）与水分（用 W 来表示）的特征吸收波数。选择这几个波数点，按照 Berr 定律可以建立方程组：

$$\begin{aligned}A^{\text{面}}_{4200} &= A^{\text{F}}_{4200} + A^{\text{P}}_{4200} + A^{\text{W}}_{4200} + A^{\text{B}}_{4200} \\&= K^{\text{F}}_{4200} C_{\text{F}} + K^{\text{P}}_{4200} C_{\text{P}} + K^{\text{W}}_{4200} C_{\text{W}} + K^{\text{B}}_{4200} C_{\text{B}} \\A^{\text{面}}_{4600} &= A^{\text{F}}_{4600} + A^{\text{P}}_{4600} + A^{\text{W}}_{4600} + A^{\text{B}}_{4600} \\&= K^{\text{F}}_{4600} C_{\text{F}} + K^{\text{P}}_{4600} C_{\text{P}} + K^{\text{W}}_{4600} C_{\text{W}} + K^{\text{B}}_{4600} C_{\text{B}} \\A^{\text{面}}_{5200} &= A^{\text{F}}_{5200} + A^{\text{P}}_{5200} + A^{\text{W}}_{5200} + A^{\text{B}}_{5200} \\&= K^{\text{F}}_{5200} C_{\text{F}} + K^{\text{P}}_{5200} C_{\text{P}} + K^{\text{W}}_{5200} C_{\text{W}} + K^{\text{B}}_{5200} C_{\text{B}}\end{aligned}$$

方程中的下标为波数点，上标为组分名称，其中 B 表示背景。

解联立方程组关联光谱与待测量；方程的解具有惟一性就是定标方程。得定标方程：

$$\begin{aligned}C_{\text{F}} &= M^{\text{F-B}_0} + M^{\text{F}_1} A^{\text{面}}_{4200} + M^{\text{F}_2} A^{\text{面}}_{4600} + M^{\text{F}_3} A^{\text{面}}_{5200} \\C_{\text{P}} &= M^{\text{P-B}_0} + M^{\text{P}_1} A^{\text{面}}_{4200} + M^{\text{P}_2} A^{\text{面}}_{4600} + M^{\text{P}_3} A^{\text{面}}_{5200} \\C_{\text{W}} &= M^{\text{W-B}_0} + M^{\text{W}_1} A^{\text{面}}_{4200} + M^{\text{W}_2} A^{\text{面}}_{4600} + M^{\text{W}_3} A^{\text{面}}_{5200}\end{aligned}$$

但实际由于样品的组分含量以及各组分单独的光谱特性（即方程中的 K 值）无法预先知道，因此近红外光谱分析样品组分浓度的预测定标方程实际不能用解联立方程的方法，只能用多元统计拟合的方法得到。在 NIR 全谱区范围内选择最佳波数点建立定标方程，这些波数点包含待测成分的信息量最大，受背景干扰最小，这些波数点的组合可以校正背景的变化。这些方法就是数学上的自变量选择（Variable Selection）。常用的自变量选择有逐步线性回归（Stepwise Linear Regression）、所有可能的回归（All Possible Regression）、最大相关系数法（Rmax）等。为了减少建模过程的信息丢失，20 世纪 80 年代以来逐步发展了一些不选择最佳波长的全谱数据处理，如主成分回归（PCR）与偏差最小二乘法（PLS）等，这些方法都属于多元线性回归。有关近红外光谱的数学处理详见第 3 章。本节重点讨论选择最佳波长点的方法建立定标方程作为数学模型。

7.3.3.2 用多元统计建立多元浓度定标方程

只选一个波数点，建立的浓度一元方程对于背景的变动没有任何校正作用。选几个波数点，在方程中增添了背景的信息，建立的浓度多元方程对于背景的变动就有校正作用，多元定标方程中所选的多个波数点中的各个波数点单独的相关系数并不一定高，一般这些波数中有 1、2 个是处在待测成分的特征波数区，而其余波数点不在待测成分的特征区，以便包含更多背景成分的信息，在最佳波数组合中起校正背景变动的作用。建立多元定标方程的方法主要有所有可能回归与逐步回归。

(1) 所有可能回归

若原光谱有 1000 个波长数据点，用统计拟合方法建立一个 3 个波长的自变量来计算浓度的三元定标方程，所有可能回归的是在 1000 个光谱点中任意取 3 个波长点组合数中，挑选出计算的结果和实际值相关系数最大的组合作为定标方程。但所有可能回归的计算量

太大，像上面的定标过程需要进行 10^9 次回归计算，一般计算机耗时过长难以接受，因此对数据点多的光谱常用逐步回归与改进的逐步回归法，所有可能的回归可用于波长点少的离散型的近红外光谱仪。

(2) 逐步回归

用一批组分含量已知的样品的近红外光谱，通过多元线性回归的方法来得到该组分浓度定标方程。其一般原则如下：选取一批 n 个样品，其中待测组分的含量预先用标准方法提出，设各样品中待测组分含量分别为 $C_1, C_2, \dots, C_n$ ，同时测出他们的近红外光谱。每个样品光谱有 m 个波长 (λ) 点，用吸光度 A 矩阵表示他们的光谱见表 7-4。

表 7-4 近红外光谱分析用逐步回归法建立数学模型

n 个标准样品的化学值与光谱	C	λ_1	λ_2		λ_m
第 1 个样品的化学值 C 与光谱各波长的 A	C_1	$A_{1.1}$	$A_{1.2}$		$A_{1.m}$
第 2 个样品的化学值 C 与光谱各波长的 A	C_2	$A_{2.1}$	$A_{2.2}$		$A_{2.m}$
第 n 个样品的化学值 C 与光谱各波长的 A	C_n	$A_{n.1}$	$A_{n.2}$		$A_{n.m}$

注： C 与各波长点 A 间的相关系数 R 与标差 σ $R_1 \sigma_1 R_2 \sigma_2 R_m \sigma_m$ 。

① 化学值 C 分别与各波长点的吸光度作线性回归，得到信息的待测量表达式 $C = Aa + b$ ，并得到相关系数 R 与标准差 σ ；

提取有效信息：

② 选择相关系数 R 最大，标准差最小的 λ_{i1} 点，即某组分近红外光谱中信息量最大的波长点，得 $C = K_0 + K_1 A_{i1}$ ；

③ 在光谱 m 个波长点中除 i_1 点外剩下的 $m - 1$ 个波长点中选择信息量次大的波长点 λ_{i2} 得 $C = K_0 + K_1 A_{i1} + K_2 A_{i2}$ ；

④ 在剩下的 $m - 2$ 个波长点中继续选择信息量较大的波长点；

⑤ 选出的 λ_{ip} 点，组成线性方程组（即数学模型）用于计算该组分的含量 $C = K_0 + K_1 A(\lambda_{i1}) + K_2 A(\lambda_{i2}) + \dots + K_p A(\lambda_{ip})$ 表达为待测量 C 。

逐步回归的算法的程序流程见图 7-6 中的算法 1。

(3) 改进的逐步回归法

逐步回归所选出的波长组合不一定是最佳的，如首先选出了波长点 i 的吸光度 A_i ，再选第 2 个波长点 A_j ， A_i 和 A_j 的组合不一定是最佳的，于是保留 A_j 继续选第 2 个波长点为 A_r ，若 A_r 不等于 A_i ，说明 A_i 和 A_j 的组合不是最佳的，于是保留 A_j 继续再选其他第 2 个波长，反复进行上述过程，直到达到需要为止，算法的流程框见图 7-6 中的算法 2。

7.3.4 近红外光谱分析定标方程的信息表达

7.3.4.1 一元定标方程与组分信息在谱区的分布

一元浓度定标方程是指用近红外光谱一个波数点处的反射吸光度值来计算样品中某一组分的浓度，一元定标方程对于样品中待测成分的背景变化没有校正作用。不同波数点含有组分浓度的信息不同，在样品的近红外光谱中选取不同波数点的反射吸光度 A 和组分浓度 C 做一元线性回归时，其相关系数与标准差不同。可以在波长范围内逐点通过回归建立各波长点的吸光度与待测量之间的回归方程，从中选出信息最大的波长点，称为逐步回归法。

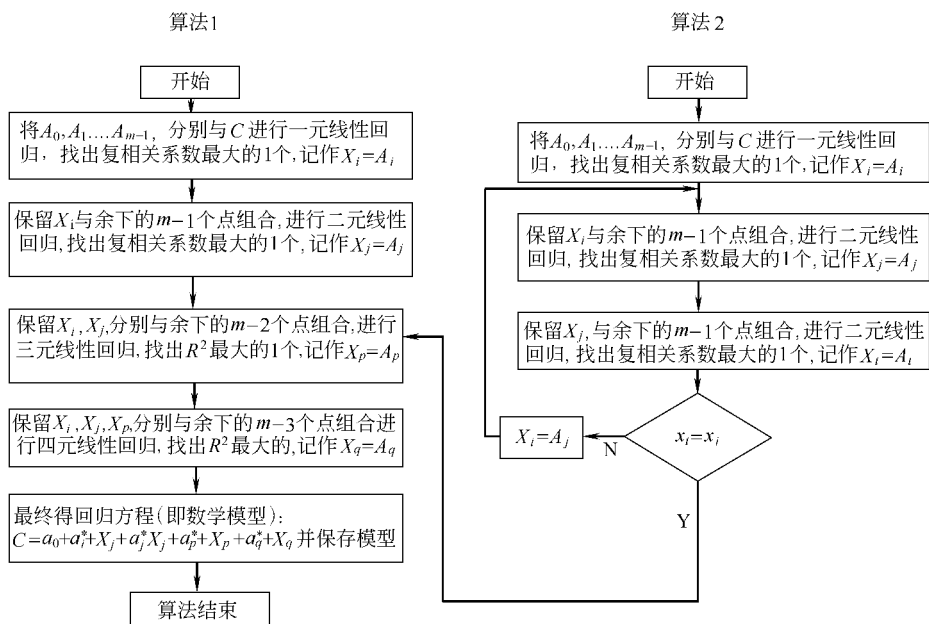


图 7-6 逐步回归法与改进的逐步回归法

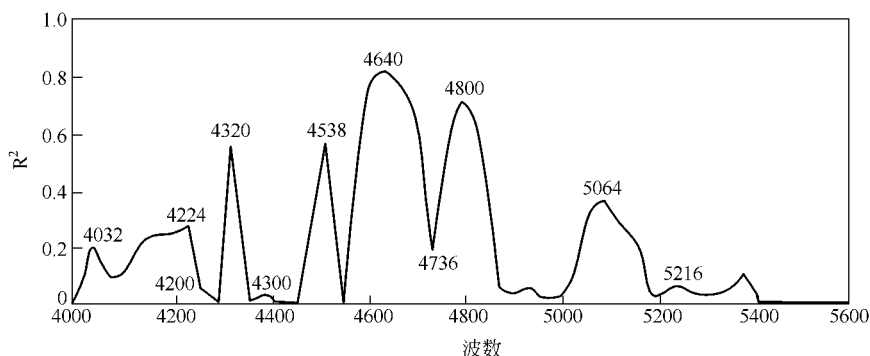


图 7-7 谷子 NIRDRS 各波数点 A~C 的相关指数 R^2 与波数 WN 间的关系

图 7-7 是谷子近红外光谱（各波数点处反射吸光度 A 与蛋白质含量 C 间单相关）的相关指数 R^2 与波数 WN 间的关系图。由图可见 R^2 值的极高点（即图 7-7 中的峰值点）都接近蛋白质的吸收峰值，而且与主要背景成分的吸收峰值相差较远。 R^2 值的极低点（即图 7-7 中的谷区），都接近样品中主要背景成分的吸收区。极高点与极低点 R^2 值相差千倍以上。

由第 2 章图 2-5 可见，蛋白质组分在 4596cm^{-1} 、 4858cm^{-1} 有较宽的吸收峰，而以 4600cm^{-1} 为主；淀粉组分在 4750cm^{-1} 、 5169cm^{-1} 有较宽的吸收峰，而以 4750cm^{-1} 为主；水在 5200cm^{-1} 有较强的吸收峰； $4100\sim 4300\text{cm}^{-1}$ 的谱区是 CH 的吸收峰，各种有机分子在此谱区都有吸收。根据上述各组分的特征吸收可以对图 7-7 中各相关指数最低点与最高点（图中的谷与峰）做出合理的解释：

图 7-7 中与蛋白质组分相关指数高的谱区是 $4500\sim 4700\text{cm}^{-1}$ ，在该谱区中虽然相关

指数有所波动，但总体相关指数较高，而且稳定，这是与蛋白质光谱的特征相符合的。图 7-7 中 $4200\sim 4400\text{cm}^{-1}$ 、 $4700\sim 4800\text{cm}^{-1}$ 、 $4900\sim 5050\text{cm}^{-1}$ 、 $5100\sim 5300\text{cm}^{-1}$ 是蛋白质的低相关区，这是与样品中作为背景成分的脂肪、淀粉、水的吸收峰有关。例如 4288cm^{-1} 、 4352cm^{-1} 波数的吸光度 A_{4288} 、 A_{4352} 与蛋白质含量相关数低的原因显然与作为背景的脂肪在 4283cm^{-1} 与 4355cm^{-1} 的吸收峰（见图 2-5）有关。而图 7-7 中 4736cm^{-1} 与 5216cm^{-1} 处 R 的极低点与样品中作为待测成分背景的淀粉与水在 4750cm^{-1} 、 5169cm^{-1} 、 5182cm^{-1} 吸收峰（见图 2-5）有关。图 7-7 中相关指数最高的一些波数点 4512cm^{-1} 、 4640cm^{-1} 、 4800cm^{-1} 都在蛋白质 NIR 吸收峰附近，尤其 4640cm^{-1} 该波数点位于蛋白质分子中 N—H 合频吸收带附近，是蛋白质含量信息量最大的波数点之一，又与变化背景（水分、淀粉、脂肪）的吸收峰有较远的偏离。可称这些波数点附近的谱区为蛋白质的特征区。

7.3.4.2 多元定标方程的信息表达

定标方程中所选用的波长决定了方程中信息表达的特征。表 7-5 是用改进逐步回归最大相关系数法对最佳波数组合选择的一个实例。

表 7-5用逐步回归法建立模型组合波长的选择过程

选择次序	选中的波数组合/ cm^{-1}	相关系数
1	4640 4416	0.948
2	4416 4640 4512	0.961
3	4640 4512 4608	0.967
4	4512 4608 4448	0.974
5	4608 4448 4480	0.977
6	4448 4480 4608	0.988

由表可见，第 1 次选择的 2 个最佳波长为 4640cm^{-1} 与 4416cm^{-1} ；第 2 次选择时保留此 2 波长，选择第 3 个波长组合为 4416cm^{-1} 、 4640cm^{-1} 与 4512cm^{-1} ；此 3 个波长不一定是最佳的波长组合，于是第 3 次选择是保留 4640cm^{-1} 与 4512cm^{-1} 由此选出了 4608cm^{-1} ，此组合与第 2 次选择的 3 波长组合不同，因此进行第 4、5 次选择；直到第 5、6 次选择相同的波长，以第 6 次选择的波长作为数学模型的定标方程。

由该选择的过程可见，第 2、3、4 次所选的 3 个波点中都有 2 个以上在蛋白质特征吸收区内，但定标方程计算结果的相关系数不是最高的。最后第 4、5 次选出的相关系数最高的浓度定标方程中，只有一个波数点 4608cm^{-1} 属于蛋白质的特征吸收区，从图 2-5 中可看到，其他 2 个波数点虽然也在蛋白质 NIR 吸收区内，但这 2 个波数点受背景的影响变化都较大，单相关系数 R 很小，说明这 2 个波数点具有较强的背景信息，它们与校正背景的变化关系更大。这 2 类数学模型虽然在建模时都有较高的相关系数，但预计它们对样品背景的变化能力有所不同。

总之，优秀的数学模型不但应包含待测组分的信息，而且应该有丰富的背景信息，使模型不但在建模时有较大的相关系数，较小的标准差，而且应当对背景的变化有较强的适应能力，这就提高了模型的稳定性。

本节只讨论了用简单的关联建立较简单的数学模型（定标方程）等有关问题，定标方程只用了少数几个波数点，因此会丢失一定的有效信息，影响模型的预测稳定性与准确度。近年来，更多的用全谱区的信息，如主成分的回归来建立数学模型（PLS、PCR

法)，有关问题参看第 3 章。

7.4 近红外光谱分析数学模型的结构与预测性能

7.4.1 近红外光谱分析数学模型的一般预测性能

近红外光谱分析通过数学模型预测样品的待测量，因此该项技术的分析精确度等特征决定于数学模型的预测性能。近红外光谱分析数学模型的预测性能是指用模型预测样品待测量的精确度、预测最低极限、预测速度以及模型的预测可靠性（准确度）和预测稳定性等。其中前 3 种预测性能主要决定于近红外光谱的特征，后 2 种预测性能主要决定于建立近红外光谱分析数学模型样品集及其化学值。这些参数也就是近红外光谱分析技术的对应参数。

7.4.1.1 近红外光谱分析的分析精确度

近红外光谱分析的分析精确度是指用近红外法对同一样品重复测定结果之间离散的大小。分析的精确定于分析过程操作的离散性，在分析过程中每一步都有离散，都可能带来各种误差，这些误差的累加降低了分析的精确度，因此一种分析的流程越长，分析的精确度可能越低。通常化学分析的流程较长，有可能影响化学分析的精确度；近红外光谱分析由于不需要样品前处理，操作步骤简单，因此近红外光谱分析的精确定较高，甚至比分析流程较长的化学分析的精确度更高。表 7-6 是某生物样品中总糖等五种成分近红外光谱分析的精确定与化学值分析精确定的比较。

表 7-6 近红外光谱分析精确定与化学分析的精确度比较

化学值分析的标准差				近红外光谱分析 相对标准差/%
成分	分析员	各分析员测定的 相对标准差	分析员之间的 相对标准差	
总糖	1	2.8	4.19	3.5
	2	3.6		
	3	4.7		
还原糖	1	3.6	4.96	3.3
	2	5.0		
	3	6.3		
生物碱	1	17.8	6.57	3.8
	2	15.7		
	3	10.2		
总氮	1	5.4	9.17	4.4
	2	6.1		
	3	3.4		
蛋白质	1	7.8	12.08	5.2
	2	7.2		
	3	4.4		

表中五种成分化学值的分析由 1、2、3 三名有经验的分析员来完成，每个样品每个分析员重复做 10 次，由此计算各个分析员分析每一种成分的精确度（由分析的相对标准差来表示），同时还可计算出每一种化学成分分别由三个分析员分析结果的离散（也用分析

的相对标准差来表示)。这些样品的近红外光谱分别与相应的化学值平均值关联建立数学模型,用近红外数学模型预测样品结果的相对标准差均小于相应的分析员内部或分析员之间的分析标准差,由此可见,近红外光谱分析的精确度可以高于化学值分析的精确度。近红外光谱分析的精确度决定于光谱测定的精确度,因此,溶液样品透射光谱的分析精度一般高于漫反射样品的分析精度,为了提高近红外光谱分析的精确度必须控制近红外光谱分析的条件,提高光谱的测量精度。

7.4.1.2 近红外光谱分析的检测极限

近红外光谱分析的检测极限决定于光谱中样品信号和噪声比,即信噪比,常规仪器分析的检测极限定位在3倍噪声的信号所对应样品的成分含量。在近红外光谱分析中由于应用了多种弱信号恢复的信息技术,即使对低于噪声强度的弱信号也能够通过信息处理提取其中的信息。因此,近红外光谱分析的检测极限大体上可以定为与噪声相同数量级的信号所对应的待测分量。近红外光谱分析仪器的信噪比大体可做到 10^3 、 10^4 ,近红外光谱分析的检测限可定为 10^{-3} 、 10^{-4} 吸光度所对应的待测量。虽然由于一般成分的近红外吸光系数很低,但考虑到近红外光谱分析的光程较长,上述吸光度对应的待测量含量大约相当于 10^{-3} 、 10^{-4} 左右。

7.4.1.3 近红外光谱分析的速度

用已有的模型完成近红外光谱分析的速度决定于三个条件:即样品的处理、扫描光谱与数据处理。近红外光谱分析的样品处理比一般化学分析简单得多,甚至可以分析整粒谷物样品,不需要化学或物理的处理。近红外图谱的扫描也比较简单,为了提高光谱的信噪比,一般需要多次扫描取平均,以降低噪声,即使如此,每个样品的扫描大体需要1~2min,甚至更快。近红外光谱分析数学模型的建立因为需要反复的优化,比较费时,但是模型建立后用模型来计算待测值几乎不需要花费时间,现代近红外光谱分析可以通过网络取得数学模型可以节省建模的时间。此外,近红外光谱分析还可以实现扫描一个样品光谱,完成多个成分的分析。因此,从总体来讲,近红外光谱分析的速度比常规分析速度快得多。建立近红外数学模型的有关问题详见7.5节。

7.4.1.4 近红外光谱分析的分析准确度及模型的可靠性

近红外光谱分析的分析准确度是指分析的预测值与样品真值的偏差。近红外光谱分析的准确度与多种因素有关。若所用仪器的性能良好,光谱的测定准确,则近红外光谱分析结果的准确度决定于建模样品的代表性与化学值测定的精度。这是由于近红外光谱分析的预测值是依靠数学模型计算得到的,而数学模型是依据化学值和光谱依靠算法所建立的,如果光谱测定准确,则(依据数学模型)计算所得到的近红外预测值的准确度决定于建模样品化学值测定的准确度。近红外光谱分析的准确度可以逼近化学值测定的准确度,因此化学值测定的准确度是决定近红外光谱分析关键,参见本书第12章。对一定的建模样品,近红外光谱分析的准确度决定于数学模型,也称为模型的分析可靠性,关于数学模型的可靠性以及表示方法见第8章。

7.4.1.5 近红外光谱分析数学模型的适配范围及模型的稳定性

近红外光谱分析依靠数学模型来完成,每个数学模型有一定的适配范围,即每种数学模型只能应用于一定范围的样品,该应用的范围可以在建立近红外光谱分析数学模型样品集的主成分空间中用一定的空间区域来表示[见7.2.2.1中(1)]。

模型的稳定性和可靠性是近红外光谱分析数学模型的主要两种预测性能。模型的稳定性和可靠性之间是对立的，一般适配范围较宽的数学模型预测的精度较低，反之适配范围较窄的数学模型分析精度较高。

模型的可靠性主要决定于建立样品集的化学值；模型的稳定性决定于数学模型适配范围，它不但与建模样品集有关，还与建立模型全过程的优化有关，是本书重点讨论的内容，详见本章 7.4.2。

7.4.2 近红外分析光谱与数学模型结构的波动性

近红外光谱分析数学模型的预测性能决定于模型的结构。对于复杂的数学模型两者的关系不易看清，简单的数学模型结构比较简单，模型的性能与结构的关系也比较清晰，我们以定标方程（简单的数学模型）讨论模型的性能与结构的关系。

近红外光谱特别是固态或粉末样品近红外漫反射光谱由于测量条件难以严格控制，光谱具有一定的随机波动，可称为光谱的波动性。这是因为固态或粉末样品不能像真溶液那样达到均匀，由于测量条件（尤其是散射性系数 S）不易保持严格一致而造成近红外光谱较差的重现性。另一方面分子在 NIR 谱区吸收弱谱带重叠，谷物的某组分发生变化时，只在 NIR 谱上产生不大的变化。这种 NIRDRS 对测量条件差异的敏感性高，而对样品间差异不敏感的特点造成了样品 NIRDRS 的方程波动性，这种光谱波动性可见图 2-10。

近红外光谱分析的数学模型由近红外光谱通过关联算法建成，近红外光谱的波动性决定了所建模型的波动性，本节以较简单的数学模型——定标方程为例进行讨论。用同一批 30 个谷子标样品，分两个条件，在不同时间扫描四个光谱，用相同的逐步回归程序建立四个数学模型——浓度定标方程：方程 1、2 的标准样品是经过低温烘烤，水分含量比较一致的扫描光谱；而建立方程 3、4 的标准样品是在室内经 1 周时间存放自然风干，样品中水分含量有所不同的扫描光谱。尽管这些方程都是用同一批样品建立而成的，而且，预测样品蛋白质含量都有较小的标准差（0.4% 以下）与较高的相关系数（0.98 以上），但这些方程中被选的波数点与方程的常数、系数均有很大的差异。见表 7-7。

表 7-7 同一组标样不同图谱所建立的浓度定标方程

方程号	蛋白质浓度定标方程	备 注
1	$C_1 = 13.24 + 3696(A_{4640} - A_{4656}) - 2308(A_{5126} - A_{5142}) - 833(A_{4288} - A_{4304})$	建模样品
2	$C_2 = 12.42 + 3384(A_{4640} - A_{4656}) - 341(A_{5248} - A_{5264}) + 390(A_{4096} - A_{4102})$	含水量一致
3	$C_3 = 15.64 + 2595(A_{4640} - A_{4656}) - 1263(A_{4864} - A_{4880}) - 1004(A_{4480} - A_{4496})$	建模样品
4	$C_4 = 9.42 + 4400(A_{4672} - A_{4688}) - 4349(A_{4288} - A_{4304}) + 1598(A_{4032} - A_{4048})$	含水量不一致

由表 7-7 可见各浓度定标方程中被选定的最佳波数点与常数、系数不完全相同，其中方程 1、2 中包含水的吸收峰（5200cm⁻¹），而方程 3、4 不包含水吸收峰。这种同一批样品用不同的光谱建立的模型，在结构上的差别可以称为近红外光谱分析浓度定标方程结构的波动性，这种统计波动性决定于光谱的波动性。但这些方程都有一些共同点：所有的浓度定标方程都有一个较大值的常数项。定标方程所计算的浓度是以这个常数为基础，由三个波数点的反射吸光度差乘以数值较大的系数，给常数以校正而取得的。此外各方程中都包含一项蛋白质特征吸收（4600~4700cm⁻¹）的吸光度，而且该波数点吸光度乘的系数也是最大的，即在方程中该项所起的作用最大。另外两项分别与修正待测组分蛋白质的背

景成分有关。

近红外光谱数学模型结构的波动性说明近红外样品和数学模型不是惟一对应的：同一批样品由于扫描条件的不同以及统计变动的因素，扫描的光谱有所不同；即使同一批光谱由于建模算法的不同或优化方法的不同，所建模型的结构也有所不同，由此产生模型的不同预测效果。因此建立一个优秀的数学模型必需在建模的全过程经过反复的优化才能实现。

7.4.3 近红外光谱分析数学模型结构波动性对模型预测性能的影响

不同结构的数学模型用于预测未知样品时，预测性能不同。选取两个待测样品其蛋白质含量分别为 10.36%和 14.95%，来检验模型的预测精度。将这两个待测样品分成 5 份，分别放于不同的湿度环境，使其含水量分别为 19%～19.2%、16.7%～17.0%、13.5%～13.7%、11.7%～12%和 9.0%～10%（与表 7-8 内 A、B、C、D、E 相对应）同时分别测出它们的近红外光谱，见图 7-8。

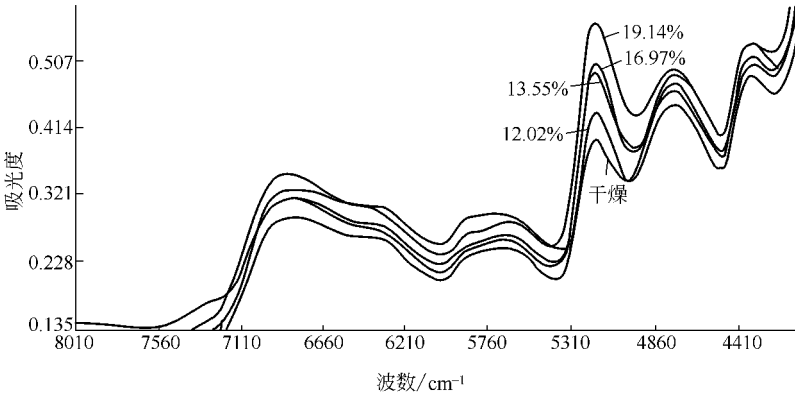


图 7-8 含水量不同的谷子样品的近红外漫反射光谱

由图 7-8 可见，同一样品含水量不同的光谱在水的近红外吸收峰 5200cm^{-1} 附近有明显的差别：样品中水分含量越高， 5200cm^{-1} 的水峰也就越高。用表 7-7 的 4 个浓度定标方程，分别计算上述 2 个含水分不同样品的蛋白质含量，所得结果见表 7-8。

表 7-8 不同定标方程预测含水量不同样品中蛋白质含量

方程号	样品 1 蛋白质含量 10.36%					样品 2 蛋白质含量 14.95%				
	A	B	C	D	E	A	B	C	D	E
1	14.8	14.8	14.9	13.3	11.8	19.1	17.6	17.7	17.21	16.2
2	17.1	16.4	15.9	14.3	13.3	21.7	20.4	19.9	19.1	17.8
3	7.1	7.4	9.3	9.8	10.1	11.1	12.6	14.2	15.2	15
4	5.1	6.5	8	7.7	10.7	7.4	10.3	12.4	13.2	14.4

注：表内数据均为用定标方程计算的样品中蛋白质百分含量，表中各方程参见表 7-7。

表中 A、B、C、D、E 分别为样品含水量 19.0%～19.2%、16.7%～17.0%、13.5%～13.7%、11.7%～12.0%、9.0%～10.0%。

表 7-8 说明，由方程 1、2 预测样品中蛋白质含量时，样品中水分含量对预测结果的影响较大，这两个数学模型的稳定性差；而用方程 3、4 预测蛋白质含量时，在一定的范

围内计算结果受样品中水分含量变化的影响相对较小，那两个数学模型的稳定性相对较好。表 7-8 中的数据还可以看出方程 3、4 对高含水样品 A、B 预测的结果仍然不理想，说明每个数学模型有一定的适用范围。

另外，方程 3、4 的建模样品是相同的，但预测效果不同，3 号方程预测结果（平均值）比 4 号方程更接近真实值，这是因为这两个方程结构不同，所表达的信息不同，因此说明模型需要优化，才能够获得稳定性、可靠性均好的数学模型。

近红外光谱分析数学模型的预测性能是模型的一种宏观性质，而其数学结构是模型的一种微观性质；模型的微观结构决定了模型的宏观性能。它们的关系是：建模样品集的组成与建模的技术决定了模型的数学结构；而模型的结构，即模型规定的（光谱与待测量间）数学关系决定了该模型表达的信息，进而决定了模型的预测性能。

近红外光谱分析的定标方程是一种简单的数学模型，从方程的结构可以直观的看出定标方程（数学模型）所表达的信息；但是，复杂的数学模型（如 PCR、PLS 法建立的数学模型）用相关主成分的得分（SCORE）阵来表达，而各个主成分分别用全光谱的载荷（LOADING）阵才能表达。这样的模型结构十分复杂，因此直接从模型结构来分析所表达的信息较为困难。但是仍然符合上述的规律；建立数学模型的样品集与算法决定模型的结构，后者决定了表达的信息，进而决定了模型的预测性能。不管是简单模型还是复杂模型，近红外光谱分析的建模数据对所建立的数学模型的预测性能有重要的作用。

7.4.4 近红外分析数学模型的稳定性与建模过程的自校正

为了提高数学模型的稳定性，建模标样中应包括尽可能多影响测量的背景信息，由此建立的数学模型受背景的变化影响较小。例如，为了校正样品中水分背景对测定样品中某成分的影响，用多元线性回归建立该成分的定标方程时，建模样品的含水背景应包括一定的变化范围，这样在建模过程中多元校正的算法会自动给（需要避免的）水吸收谱区的光谱参数以适当的权重（如表 7-7 中方程 3、4，避免水的吸收谱区），建立受水背景干扰影响较小、对背景水的变动有一定的适应能力的模型。但当背景水的变动过大，超出适应范围后模型就难以实现自校正，仍然需要用硬件或软件直接校正变动背景。同样道理，为了提高模型对温度变化适应性，建模样品扫描光谱时，样品温度应有一定的变化范围，这样建立的数学模型才能适应一定的温度范围，在此范围内，对温度有一定的自动校正作用，但是温度变化过大还是需要硬或软件的校正。

这种多元校正法建立数学模型的过程中在一定范围内对变动的背景具有自我校正（或自动校正、内部校正）的作用，简称建模过程的自校正或内校正；对于需要用硬件或软件直接校正变动背景的方法称为外校正。数学模型对变动背景的自校正作用是多元校正区别于一元校正的一种重要特点，这种作用具有相当的广泛性。利用建模过程的自校正不但可以建立对背景水或温度的变化有一定自动校正的模型，还可对样品的密度、厚度等样品参数的变动以及积分时间等测量参数的变动等均可以在一定范围的自校正。

建模过程的自校正作用的方式与强度主要决定于建模的算法；相反，外校正作用的方式需要由近红外光谱分析技术人员来掌握。

数学模型对变动背景具有自校正作用的原因可能是：建模样品背景或状态一定范围的变动，也就是建模样品具有代表性，将影响或改变建模样品主成分空间的范围，从而影响

或改变模型的适配空间，进而对提高模型的稳定性产生一定作用。

在近红外光谱分析技术发展过程的早期，十分强调建模样品资源背景条件以及测量条件的严格一致性。但随着多元校正方法的改进，现代近红外光谱分析不再过分强调这些条件的严格一致性，而更加注重改进建模的算法；另一方面对于是像样品含水量等背景因素，反而要求其有一定的离散，以便通过建模的自动校正作用来减少这些背景的影响。但是对于校正某些具体的变动背景是用自校正还是外校正，以及对模型自校正作用的范围等问题，目前主要依靠经验或实际检验、测试与比较。

7.5 建立近红外光谱分析数学模型一般过程与资源共享

前面 7.3.2 节 讨论的是一种特定数学模型的建立，本节讨论一般数学模型的建立。

7.5.1 近红外光谱分析的基本流程

近红外光谱分析的流程大体包括 3 个层次：分析样品的分析层次、建立数学模型的建模层次与优化模型的优化层次。每个层次分别都包含若干步骤。运用近红外光谱技术对样品中某化学（或物理）待测量进行定量分析的基本流程见图 7-9。

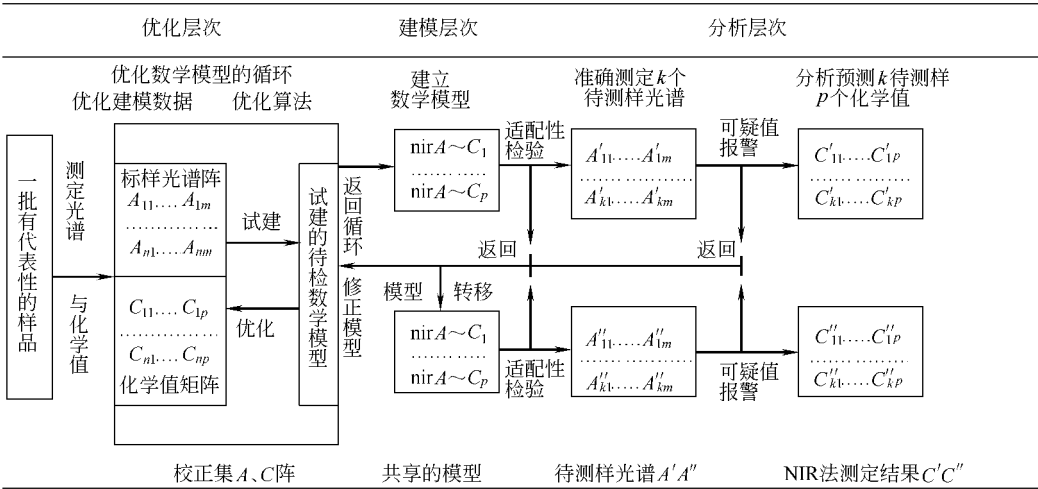


图 7-9 近红外光谱定量法分析 k 个样品中 p 种待测量的基本流程

由图 7-9 可见，用近红外光谱技术测定样品中某待测量必须首先建立相应的数学模型，需要选择一批（图 7-9 中为 n 个）有代表性的样品作为校准样品集，分别准确测定其近红外光谱（图中光谱由 m 个数据点组成）与各种（图中为 p 种）待测量（称为化学值），得到二维光谱矩阵（图中以 n 行、 m 列的矩阵 A 表示）与对应的二维化学值矩阵（图中以 n 行、 p 列的矩阵 C 表示）。然后以此作为建模数据，运用化学计量学方法，分别通过优化层次与建模层次，选择最佳参数建立数学模型，图中表示共建立了 p 个数学模型（以“ $nirA \sim C$ ”表示）。对于待测样（ k 个），只要测定它们各自的光谱，就可以分别用相应的数学模型，预测每一个待测样的各（ p ）种待测值。

下面分别简单介绍近红外光谱分析过程的分析层次、建模层次优化层次。

优化层次：在近红外光谱分析过程如果尚无合适的数学模型，则进行样品分析前需先建立合适的数学模型。在建模过程中建模数据和建模算法属于建模参数，是决定数学模型性能的两个要素。

优化并确定最佳建模参数的过程在“优化层次”、通过“优化循环”反复优化才能完成。优化循环中按照“开源节流”的技术策略，通过试建与检验模型，反复、循环优化各建模参数，以确定最佳建模参数。有关数学模型优化的详细讨论参见第 8 章。

建模层次：在建模层次，利用（通过优化层次确定的）最佳建模参数，将内部交叉检验的建模样品集以及外部检验的检验样品集加以合并，建成能够实际用于分析未知样品的数学模型。在建模层次也可以通过模型转移产生共享的模型。

每一种数学模型只能适用一定的时间、空间与样品范围，该适用范围决定于建模样品所覆盖的（主成分）空间范围，因此在运用数学模型预测未知样品之前应当先对待测样与模型之间是进行“适配性检验”。常见的适配性检验是用主成分空间或马氏距离，来检验待测样是否在建模样品集覆盖范围内。不适配模型则通过模型转移形成的适配模型，或者返回循环修正模型，即选择新的参数或者添加一些新的样品，修正原来的数学模型。已建好的数学模型也需定期进行检验，特别是对于新类型的样品，必须首先检验其是否与原有的模型相适配，这些过程统称为对数学模型的维护，因此可认为模型需要“终身优化”。有关适配性检验、模型转移与模型的修正详见本书第 9 章。

分析层次：在近红外分析的过程中如果已经建有合适的数学模型，则可以直接进入分析层次，只要准确测定待测样品的近红外光谱，运用数学模型分析样品，进行近红外光谱分析。分析层次直接决定了分析结果。分析结果的好坏决定于数学模型的性能与样品光谱的质量，因此分析层次的两个要素是数学模型与待测样品光谱。

近红外光谱测定是一种间接测定方法，由于样品光谱可能存在的一些失真以及模型与样品之间的适配程度，因此近红外测定的结果只能在一定的统计误差范围内保证测定结果的可靠性，对于测定结果可疑的样品软件应当给出报警。可以用某些数学方法来检验数学模型预测结果的可靠性。所谓“可疑值报警”就是用数学方法指出预测结果可靠性较差的某些样品。对于少数可疑样品，如有必要需用经典方法进行化学值测定；如果分析结果中可疑样品较多时，需要通过检验，另外选用适配的数学模型再做分析。

7.5.2 近红外光谱分析建立数学模型的主要步骤

本章 7.1 节已经讨论了近红外光谱分析校正光谱的基本过程主要有两步：建立校正样品集与通过校正样品集运用算法建立数学模型。校正光谱的目标是建立优秀的数学模型，其过程大体包括 11 个主要步骤：

① **近红外光谱库的建立：**为了全面地收集近红外光谱分析样品的复杂背景，需要大量收集近红外光谱分析样品，并在规范条件下测定这些样品的近红外光谱，光谱的测试技术详见第 6 章。近红外光谱信息量巨大，因此需要通过数据库技术来进行管理。样品光谱应当区分与记录仪器的类型，样品的品质、来源不同状态（籽粒或者粉末），并用高精度、高分辨的仪器与规范一致的测试条件，用尽可能宽的谱区来扫描样品光谱。有关图谱库的建立和管理详见第 13 章。

② 近红外光谱分析代表性样品的选择：根据待测样品的不同类型（不同的作物、不同的地区、不同的分析条件等）运用算法从图谱库中选择代表性样品，以此作为建模样品集。有关代表性样品的选择详见第 8 章。通过代表性样品的选择可以使图谱库样品的海量信息压缩上百倍。以代表性样品来建立数学模型可以大大提高建模样品中的有效信息率。

③ 近红外光谱分析代表性样品化学值与光谱的分析扫描：由于化学值的分析需要花费大量资源，应用已经压缩的代表性样品建立数学模型可以减少化学分析的工作量。代表性样品的化学值与它们对应的近红外光谱（光谱可以重新扫描或者利用图谱库中原有的光谱）由两个矩阵即样品的化学值矩阵与光谱矩阵来表示。这两个矩阵是建模的基本数据，它们的准确度将决定分析模型的预测准确度。建模数据中的光谱矩阵由两个群体（集）组成，即样品集中的样品组成与每个样品的光谱数据集。在利用这些数据建立数学模型以前，需要预先对这两个数学集进行优化，称为建模数据的前处理。由于建模数据的信息量仍然太大，有效信息率太低，因此前处理包括对建模数据中的样品群体和光谱数据群体做进一步的压缩以及对光谱数据弱信号的恢复。

④ 近红外光谱数据集弱信号的恢复：该项处理是光谱数据前处理的一部分，弱信号的恢复主要是降低噪声与降低或消除系统误差。降低噪声最常用的算法是平均、平滑、积分、傅里叶变换、小波变换等等，有关这些算法的原理参见第 3 章。消除近红外光谱中系统误差最常用的算法是求导，也可以应用傅里叶变换、小波变换等算法。所有这些弱信号恢复的方法，目的都是提高光谱中的有效信息率。现代近红外定量分析软件一般都具有上述降低噪声与减少系统误差算法的功能。

⑤ 近红外光谱分析建模谱区的选择：该项处理也是光谱前处理的一种。对于波长连续的近红外光谱，为了保证对样品光谱的分辨，光谱的数据一般有数百到数千个数据点，但谱区中各个数据点包含的信息是不同的，为了压缩光谱数据可以运用算法在全光谱中选择部分谱区用于建立数学模型。谱区的选择也是对光谱数据集压缩的一种方法。分析谱区选择过宽会增加无效信息，降低有效信息率；分析谱区过窄，则可能丢掉有效信息降低数学模型分析的准确度，因此谱区的选择要根据分析对象的不同进行反复的优化。

⑥ 近红外光谱分析数学模型的试建：经过光谱数据点压缩与弱信号恢复以及对样品集压缩等前处理后，建模数据可以用算法来关联光谱和待测量，建立数学模型。现代的建模关联算法如：逐步回归，主成分回归等需要确定数学模型的阶数或波长数，建立的数学模型通常不用建模光谱的全部光谱点而选择少数几个信息量最大的波长点（所有可能回归法或逐步回归法）或选择几个主成分（PCR 或 PLS 法，称为阶数），从而达到进一步压缩建模数据的目的。当建模样品数少于 100 时，所选的阶数大体上不超过样品数的 $1/60$ 到 $1/10$ 。阶数过小称为模型的欠拟合，即未能将光谱中的信息充分提取；过高的阶数所建的模型称为过拟合，过拟合的模型虽然对模型内部的样品预测效果可能较好，但预测其他样品时误差较大，详见后。

⑦ 近红外光谱分析数学模型的优化：优秀的数学模型一般不可能一次建成，而需要通过试建以后，用不同的样品进行检验，检验模型的稳定性与可靠性，然后根据试样的结果进一步剔除建模样品中的异常样品，修改建模的谱区与其他参数，以提高模型的稳定性与可靠性，这些过程称为对模型的优化。剔除模型中的异常值样品是优化数学模型的一项技术，建模过程中某些样品加入模型后所建的模型的预测精度会大大降低，这些样品的化

学值或光谱值称为异常值。异常值有两类，一类是这些异常值样品的化学值分析或光谱的扫描有较大的误差，在模型的样品集中剔除这些样品可以提高建模数据的有效信息率，从而提高所建模型的预测性能。另一类异常值的样品是样品的化学值与光谱值数据都是准确的，但这些样品和建模的大部分样品有较大的差异，其化学值或光谱与其他样品相比有特殊性。在模型中加入这些样品后虽然会降低模型对一般样品的预测精度，但这些样品加入模型后提高了模型的有效信息，因此可以提高模型的适配范围。异常样品的处理详见第 11 章。模型优化也包括返回并重复进行前面②～⑥步，有关模型优化详见第 8 章。

某些新的建模算法能将模型试建与优化的部分功能的结合在一起，如局部回归的算法，这种算法能在原有的建模集中找出和待测样品比较接近的局部样品，用该局部的样品集建立模型来预测待测样。局部样品集减小了原有样品集中的样品数，减少了建模样品集中的异常样品，相当于建模过程中优化了样品集，从而提高了有效信息率，符合“少”而“精”原则，使预测效果更加准确。

⑧ 近红外光谱分析数学模型的试用与适配性检验：用数学模型预测一批新样品时（如用 2002 年建立的模型预测 2003 年的样品时）需要检验模型的适配性。适配性检验可以用统计略方法来检验待测样与模型的适配性，也可以通过对数学模型的试用，即用几个化学值已知的新样品来检验是否能被原有的模型准确预测；如果达不到预测的准确度，说明原有模型的有效信息不能覆盖新的样品。如果原模型预测原有样品的效果良好，而且新样品的光谱和化学值测试准确，但是原模型不能准确测试新样品，其原因可能有三种：

- a. 原建模样品和新样品测试的条件不同，如样品的粒度和松紧度的差别，由此造成两批光谱间的平移与旋转；
- b. 原建模样品和新样品测试仪器不同，包括两台仪器间横坐标波长之间的差异与纵坐标吸光度的偏差；
- c. 新样品的内部信息超出了建模集的样品信息。

上述 a、b 两种情况可以通过模型转移，使原模型与新样品相适配，上述 c 情况需要通过模型修正，形成新的模型使之与新的样品相适配（见第 9 章）。

⑨ 近红外光谱分析数学模型的转移：为了建立一个稳定、准确预测样品的数学模型，除了需要有优良的仪器硬件与软件条件外，用户还应拥有丰富的样品资源、准确测定化学值的设备，并由有经验的分析人员运用定量分析软件，经过对数学模型的反复优化，才能建立比较理想的数学模型；而且数学模型使用过程还需不断修正、维护或转移，因此优秀的数学模型是近红外光谱分析中的一个重要资源。但由于数学模型包含了测量条件等信息，一个数学模型只能在一定的时间与空间中应用。实现数学模型在不同仪器间的共享是近红外光谱分析技术的重要课题。用某一台仪器、在一定的操作条件（仪器参数、制样条件、分析环境、样品的背景等）下测量的光谱所建立的数学模型，应用到其他仪器、操作条件或样品类型的过程称为“模型转移”；通过模型转移修改原模型，可以使数学模型扩大共享范围，使之能够预测新样品的待测量。如果新样品与原模型适配不好的原因是两批样品间的平移与转移造成的，则用斜率——截距法来转移模型，若不适配的原因是仪器状态的不同，这需要通过一批标样分别在原建模测试的仪器，与新样品测试的仪器上测定，然后根据同一批样品在同两种仪器测出的光谱在横坐标与纵坐标的差别，用算法进行校正，使之一致，由此实现模型的转移。模型转移的算法使得原有模型增加新的信息，使之

覆盖新的样品，从而使原有模型能准确预测新样品。

⑩ 近红外光谱分析数学模型的修正：在原有的模型中增加几个新样品，使原有模型增添新的信息，形成新的修正的模型，能覆盖新样品，从而能准确预测新样品（见第 9 章）。

⑪ 近红外光谱分析数学模型的应用：用经过检验的模型预测未知样品的待测量，并给出报告，完成未知样品的分析过程。此外，运用算法可对测定结果可疑的样品给出报警。

上述 11 个步骤概括了近红外光谱分析从信息采集到提取，以完成分析的主要步骤。

这些步骤中包括了多种建模技术，这些技术的形式虽然不同，但本质都属于“恢复、压缩、关联、优化、共享”等五种近红外光谱分析的核心技术。

7.5.3 运用软件技术建立近红外光谱分析数学模型

近红外分析的主要过程（建立数学模型）依靠专门的分析软件来完成。近红外分析软件除了能够采用一定的化学计量学算法建立并运用数学模型进行分析外还需精心设计软件界面。由于近红外分析过程比较复杂，优秀近红外分析软件还应按照上述分析的基本流程，通过软件界面，使程序具有指导分析的功能。这样的软件可以帮助操作人员准确完成近红外光谱分析的各个步骤与环节，提高所建模型的稳定性与预测准确度，得到准确的分析结果。

一般定量分析软件主界面应有四个功能钮：“建立模型”、“检验模型”、“转移模型”、“分析样品”。进行分析的操作人员在下列的任一情况下都能直接由主界面进入分析流程。

如果分析人员已经有相应的模型，而且确认能够准确测定当前需要分析的待测样品，则可以启动“分析样品”钮进入分析层次，按照图 7-9 的分析流程进行样品分析。

如果分析人员尚未建立模型，则可以启动主界面的“建立模型”钮，进入“建模层次”，正式建立可以用于分析样品的实用数学模型。

如果分析人员虽然已经有相应的模型，但是不能确认是否可以用来准确测定当前需要分析的一批样品，则启动主界面的“检验样品”钮，用化学值已知的一些当前样品来检验指定的模型是否能准确测定当前样品。如果检验的结果确认能准确分析当前样品，则启动“分析样品”钮进入分析样品。否则需将新的样品加入建模样品集，启动“建立模型”钮，重新修正建立数学模型后再进入分析层次，完成分析样品；另一种选择是按图 7-9 进行模型转移。

如果分析人员虽然已经有相应的模型，但是不能与当前待分析的样品相适配，则可以启动主界面的“转移模型”钮，通过模型转移，使模型能适应当前待分析的样品，然后再进入分析样品层次。

总之定量软件可以帮助进行近红外分析的操作人员按照图 7-9 的近红外分析流程，在各种情况下都通过主界面，直接切入分析流程的各位置，进入分析程序，完成分析。

7.5.4 近红外光谱分析的资源共享

目前的近红外光谱仪一般需要用户自行建立数学模型。为了运用近红外光谱分析技术，用户除了需要具备仪器外，还需要 4 个条件，才能自行建立 NIR 分析的数学模型：

① 具有一批代表性样品，可以用于作为建立数学模型标样；

② 具有按照标准的方法正确分析样品待测量的条件，以便能正确建模样品集中样品的各待测量；

③ 具有建立数学模型的分析软件。NIR 分析软件应具有建立和优化校正样品集的功能。应能通过剔除异常的样品与光谱相同或过分相似的样品，优化校正样品集的组成；通过选择谱区以及预处理方法优化校正样品集的光谱；总之能提高建模数据的有效信息率。NIR 分析软件应具有优秀的化学计量学算法，并能优化算法的有关参数（如 PLS 算法的阶数，逐步回归算法选择的波长数）。NIR 分析软件应能进行严格的统计学测试，选择最佳参数的数学模型；

④ 具有有经验的近红外分析人员，能制定规范的分析条件，建模所用分析条件与以后分析样品的分析条件必须一致。

因为近红外光谱仪的一般用户不一定同时都具备上述各项建立优秀数学模型的条件，这是当前影响近红外光谱技术应用的一个重要原因。在我国 20 世纪 70 年代与 80 年代曾有过两次较大规模引进近红外分析技术，但许多引进单位均未能掌握与运用该技术，主要原因是建模的困难：一般不具备标样或缺乏有经验的近红外分析人员，单独的“一家一户”难以建立优秀的近红外分析数学模型；通常只有掌握代表性样品并具有精确分析化学值的条件，才能由专业分析人员来建成优秀的模型；而且任何一个数学模型都不可能永久使用，随着样品、仪器、测量条件的变化，数学模型也需不断修正、维护、优化或转移，这些对于一般用户都有一定的困难。近红外光谱分析的数学模型是一种重要的资源，需要研究这种资源共享的“产业化”技术。

20 世纪 80~90 年代，某些原来难以操作或者必须具有高学历、通过专门训练人员才能使用的仪器，由于越来越多的利用了高技术，使得操作应用越来越简便并且能够被广大的“普通”人员使用，这就是所谓仪器的“傻瓜化”。这种趋势在 1999 年的第 50 届匹茨堡会议上被称为是分析仪器（信息）的大众化（Democratization）。近红外光谱技术由于建模等操作步骤的复杂，致使该技术难以被一般人员掌握，为了使这种重要的技术能在全中国农业生产和科研中广泛使用，我们认为必须要研究通过运用共享技术实现近红外光谱技术的“大众化”。

表面上看来近红外技术具有“产业化”特点是该技术难以“大众化”的原因，但是我们认为如果在近红外仪器的“硬”、“软”两方面充分运用当代的信息技术，将近红外技术的“产业化”和“大众化”结合起来。其中硬的方面指发展性能优良、操作方便的仪器；软的方面就包括设计、开发与发展使用方便的智能化软件，加强使用人员的培训等，但最好的方法是利用数据库技术和网络技术：利用数据库技术可以按照近红外技术“产业化”的特点，由专门机构在一定的或全国范围内收集样品，通过优秀的软件由专业人员选取校正样品集，再建立、维护、修正与优化各种数学模型，通过网络技术将最佳的模型直接提供给近红外的最终用户，实现资源共享。用户可以通过远程分析与模型下载等方式享受有关资源，进行“全傻瓜”的近红外分析，并保证分析结果的稳定、可靠，从而实现近红外光谱技术的“大众化”。

我国政府部门在“十五”期间已经在国家级的科研项目中安排了多个与近红外分析技术有关的研究项目，特别是在“十五”科技攻关重大项目《科学仪器研制与开发》中设立

了“建立与完善我国部分农产品近红外光谱库”的研究专题，到 2003 年末，已经建立包括多种农作物总数约 10000 条光谱的近红外基础数据库，该光谱库的建立将对我国在农业领域应用近红外光谱技术起积极的作用。

参 考 文 献

- [1] 陆婉珍, 袁洪福等. 现代近红外光谱分析技术. 北京: 中国石化出版社, 2000
- [2] 严衍禄主编. 现代仪器分析. 北京: 北京农业大学出版社, 1995
- [3] 吴瑾光主编. 近代傅里叶变换红外光谱技术及应用. 北京: 科学技术文献出版社, 1994
- [4] 林水, 吴平等主编. 实用傅里叶变换红外光谱学. 北京: 中国环境科学出版社, 1991, 318~319
- [5] Donald A Burns, Emil W Ciurzak. Handbook of Near-infrared analysis. MARCEL DEKKER INC, 1992
- [6] 尾崎幸洋, 河田聪 编. 近赤外分光法. 日本分光学会测定法シリ-ズ32, 学会出版センター, 1995
- [7] Osborne B G, Fearn T. Near-Infrared Spectroscopy in Food Analysis. Longman Scientific and Technical, New York, 1986
- [8] Weler J G, Appl Spectroscopy Riv, 1985, 21 (1, 2): 1~43
- [9] Stark E, Luchter K, Appl Spectroscopy Riv, 1986, 22 (4): 335~399
- [10] K A Martin. Appl Spectroscopy Riv, 1992, 27 (4): 325~383
- [11] Kathryn A, Bunding Lee. Appl Spectroscopy Riv, 1993, 28 (3): 231~284
- [12] Jerome J Workman, JR. Appl Spectroscopy Riv, 1996, 31 (3): 251~320
- [13] Jerome J Workman, JR. Appl Spectroscopy Riv, 1996, 31 (1, 2): 73~124
- [14] Paul R Mobley, Bruce R Kowalski. Appl Spectroscopy Riv, 1996, 31 (4): 334~368
- [15] Jason M Brenchley, Uwe Horchner and John H. Kalivas. Applied Spectroscopy, 1997, 51 (5): 689~699
- [16] Shenk J S M O Westerhaus. Crop Sci, 1991 (31): 469~474
- [17] “10th International Conference on Near-Infrared Spectroscopy” Abstracts and Programs (2001)
- [18] 严衍禄, 赵龙莲等. 现代近红外光谱分析的信息处理技术. 光谱学与光谱分析, 2000, 20 (6): 777~780
- [19] 李军会, 赵龙莲等. 农业近红外光谱分析技术软件及网络系统的研制. 现代仪器, 2000, 6

8. 近红外光谱分析数学模型的优化

近红外光谱的信息来源于分子内部振动的倍频与合频，其集中体现在复杂自然物的近红外光谱分析是属于从复杂、重叠、变动的光谱中来提取弱信息的技术。所谓复杂是指背景信息复杂，其背景信息既包括由样品的粒度、松紧程度等引起的物理背景信息，又包括除待测量以外其他化学成分引起的化学背景信息；重叠是指谱峰重叠，譬如 C—H 伸缩振动与弯曲振动的合频可能与 N—H、O—H 振动的合频吸收重叠在一起，而 C—H 倍频的合频可能与 C—H 二倍频或 O—H 倍频吸收重叠在一起等等；变动是指近红外光谱受仪器状态、外部环境等因素的影响较大；弱信息是指其吸收强度较低。近红外光谱分析是从复杂、重叠、变动的光谱中提取微弱信息，其分析难度相当大，实现近红外光谱分析的前提是采用化学计量学方法建立光谱特征与待测量之间的关系——称为数学模型；因此建立分析性能稳定、结果可靠的数学模型是实现近红外光谱分析的前提和关键。近红外光谱分析的物理、化学基础以及建立模型的数学算法基础详见第 1~3 章，本章主要介绍建立可靠性、稳定性及动态适应性好的一系列建模优化技术。

8.1 概述

现代近红外光谱分析需要通过一系列的优化技术来实际建立优秀的数学模型，对数学模型的优化是建立优秀数学模型的关键。

8.1.1 优化近红外光谱分析数学模型的基本方式

与常规的光谱分析方法不同，近红外分析的光谱和数学模型都具有统计性质，近红外光谱分析的样品集与光谱集之间不是惟一对应的，而且光谱集和数学模型之间也不是惟一对应的。同一批样品由于扫描条件的不同或随机变动的因素，扫描的光谱集有所不同（如图 8-1 有 n 个光谱集）；即使同一批光谱由于建模算法的不同或优化方法的不同，所建模型的结构也有所不同（如图 8-1 有 m 个模型集），由此产生不同的预测效果。这就是近红外光谱分析中的“一对多”现象。“一对多”是近红外光谱和数学模型统计性质（详细介绍请参见第 7 章）的一种表现。

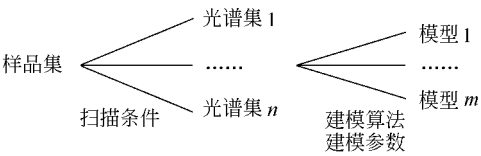


图 8-1 近红外光谱分析的“一对多”关系

由于存在“一对多”关系，近红外光谱分析在建模的过程中必须从多种光谱、多种可能的建模参数或多种试建的模型中选择相应的最优条件，才能建成优秀的数学模型，这个优化的过程称为“多中选优”。

优秀的数学模型不但在建模时有较高的相关系数和较低的预测标准差，而且还要检验对样品背景变化的稳定性。为了提高数学模型的稳定性，不但要尽量选择适配范围宽、代

表性好的样品集，还需要优化测量仪器参数与测量条件，扫描好的样品光谱；不但要有好的光谱，还需要在一定的范围反复优化各建模参数。例如：近红外漫反射分析时要求标样与待测样品的散射系数 S 值控制在一定的范围 [由自校正作用能够加以校正的范围（本章中如“自校正”、“外校正”、“少而精”等术语的详细介绍请参见第 7 章）]，因此要控制光谱的测定条件包括：装样条件、仪器参数等尽量控制在一定的范围。所谓掌握“一定的范围”意味着在建模过程中分析人员需要掌握各建模参数范围的“度”，以达到优化数学模型的目标。这种优化的方式可以称为“以度取优”。

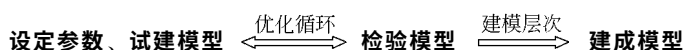
“多中选优”与“以度取优”是近红外光谱分析优化数学模型的两种基本方式，它们都来自于近红外分析光谱和数学模型的统计性质，实施这两种方式的原则都是“少而精”，一般都需要由分析人员根据经验或在计算机辅助下进行。

8.1.2 近红外光谱分析建模过程的“全程优化”与“循环优化”

从建模标样的选择、光谱的测定到模型的优化，在建模的全过程中每一环节都可能影响模型的预测性能，因此，需要从建模样品的收集、光谱的扫描到建立模型的全过程通过优化层次进行优化，以建立最佳的数学模型。这个过程可以称为数学模型的“全程优化”。数学模型的“全程优化”包括通过数学模型对变动的背景的自校正与分析人员用硬件或软件方法校正变动背景的外校正。建模过程的自校正作用的方式与大小主要决定于建模的算法，外校正作用的方式、大小与程度需要由近红外光谱分析技术人员来掌握。

由于优化的各建模参数之间有相互影响，因此对模型的优化不可能一次成功，需要建模人员反复地试建、测试、检验、比较，对试建的模型进行试用，然后才能判断和选定需要的参数，因此近红外光谱分析的优化过程是一个反复、循环优化的过程，这个过程称为“循环优化”。

近红外分析为了建成优秀的数学模型，对于“建模数据”与“建模算法”这两个建模要素都需要反复优化；这个优化过程是在优化层次中通过优化循环（本处“循环”的含义与软件程序中“循环”所指严格的重复不完全一致）来完成的。优化循环是按照前述近红外光谱分析的压缩、恢复、关联、优化四项核心技术策略，反复优化建模数据与建模算法，以至最终确定最佳建模要素的过程。优化循环可以简单示意为如下的模式。



其中，按照设定的建模数据与算法，建立一个临时的待检模型，称为“试建”模型；再对此待检模型进行严格的检验，试建模型与检验模型的有关步骤组成模型的优化循环。模型通过检验，确定试建的模型是否达到分析样品的要求，以决定是跳出优化循环正式建成实用模型，还是继续在循环内，优化建模参数后再试建与检验。

图 8-2 是上述“循环优化”与“全程优化”模式较详细的图解。

SEP ：外部证实的预测标准差 SEC ：内部交叉证实的校正标准差 SES ：校正集的标准样品化学值测定的误差 $SEP \rightarrow SEC \rightarrow SES$ ； SEP 逼近 SEC 、 SEC 逼近 SES 。

8.1.3 近红外光谱分析优化数学模型的过程

由图 8-2 可见近红外光谱分析优化数学模型的过程是在优化循环中实现的。图中粗线

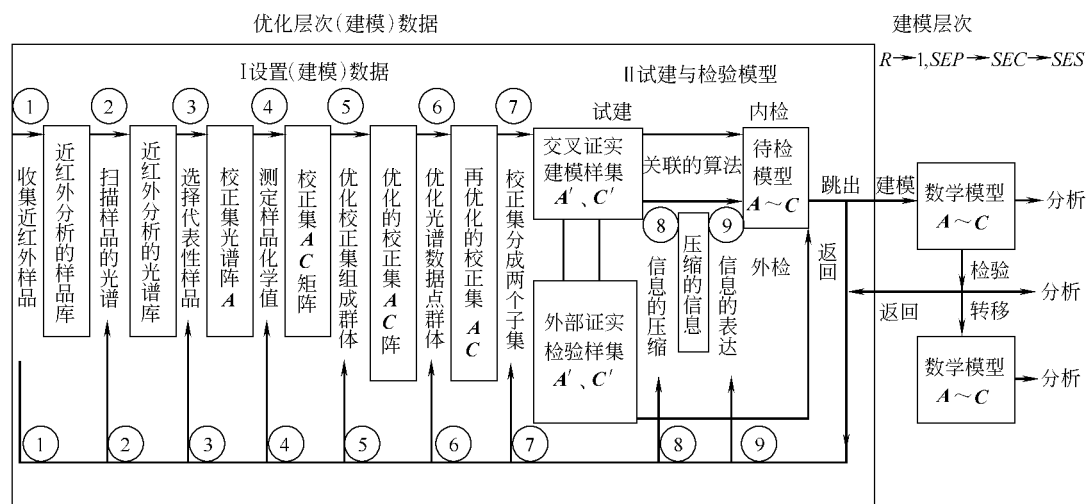


图 8-2 现代近红外光谱分析运用软件技术建立数学模型的优化层次与建模层次

图中粗线框内为优化数学模型的循环，由 I、II 两个环节与①~⑨共 9 个优化节点组成

框内为优化循环的图解。图中表示：循环一般包括三个步骤：设置建模数据、运用算法试建数学模型与检验模型以确定试建的模型是否达到要求、是否需要继续循环优化。图 8-2 中优化循环的后两个步骤，即试建数学模型与检验模型在一些软件中可以合并进行；这样，建立数学模型的全过程优化包括两个环节，即环节 I（设置与优化建模数据）与环节 II（优化算法并试建、检验模型），在图 8-2 中还指出了①~⑨共 9 个优化数学模型节点。优化数学模型是通过试建模型，反复循环优化①~⑨各个优化节点，来实现对模型的优化，最后确定最佳建模参数，跳出优化循环，建成模型。

8.1.3.1 设置与优化建模数据

建模数据指用于建立数学模型的校正集光谱矩阵与化学值矩阵，它们是建立数学模型的原信息。建模数据决定了由此数据建立的数学模型的预测准确度，也与稳定性有重大关系，因此建模数据是建立数学模型的基础。建模数据应包含分析样品的全部组分与背景的特征信息（与模型稳定性有关），而且光谱与化学值的数据可靠、干扰背景较少（与预测的准确度有关）。优化建模数据大体通过 7 个优化节点来实现：

优化节点①：优化样品信息来源。更加广泛收集近红外分析的样品，建立近红外样品库。

优化节点②：初步优化样品光谱。按规范的分析条件准确扫描样品光谱，建立样品近红外光谱库。

优化节点③：初步优化建模样品集。按照“压缩”的策略，根据样品集内样品的主成分空间分布，从样品库中选取有代表性的样品，建立校正样品集的光谱矩阵，提高有效信息率；校正样品集的范围对所建模型的适用范围，即模型的稳定性有重大影响。

优化节点④：优化样品化学值。准确测定这些选出样品的各待测量（称为化学值），建立校正集的化学值矩阵。化学值测定的精度决定了模型预测结果的准确度。

优化节点⑤：再次优化建模样品集。按照“压缩”的策略，通过统计检验方法，从校

正集中剔除光谱或化学值异常的一些样品，优化校正集样品组成群体，提高有效信息率。

优化节点⑥：再次优化建模样品光谱。按照“恢复”的策略，通过对样品光谱的各种数学预处理，降低光谱的随机误差与系统误差，优化校正集样品的的光谱，提高有效信息率。

优化节点⑦：将校正集的样品分成两个有代表性的子集（建模样品集与用于外部证实的检验样品集）。建立数学模型的过程需要反复的设定参数试建模型与检验模型，因此，检验样品集与建模样品集一样需要有代表性。校正集的样品需要通过一定的算法（如样品在主成分空间分布的方法）分两个有代表性的样品：建模样品集与用于外部证实的检验样品集。

8.1.3.2 试建、检验模型与优化建模算法

数学模型的性能不但与建模数据有关，还与建模的算法有关；用算法试建数学模型包括两方面内容，即压缩特定信息与关联表达信息，二者都影响所建模型的预测性能，因此分别列为优化循环中的⑧、⑨优化节点。

优化节点⑧：优化压缩信息的算法。初步优化算法，是在保持特定的结构和成分信息的前提下，将原来较大规模的二维光谱矩阵，分别压缩成少数选定的几个波长点、几个主成分或几个潜变量。经过如此处理后的光谱数据才容易通过关联将其表达为待测量。

优化节点⑨：优化表达信息的算法。是再次优化算法，表达信息是指对已压缩的信息，选择不同的关联算法，通过数学拟合（主要是回归方法）将其表达为待测量（如浓度 C ），以便建立预测性能优良的数学模型 $A \sim C$ 。

现有的一些软件为了便于将试建模型与检验模型结合在一起，也为了对试建模型进行客观检验，经常把校正集样品分为两个子集：作内部交叉检验的建模样品集以及外部检验的检验样品集；用建模样品集试建模型，用检验样品集检验试建的模型。试建的待检模型首先通过内部交叉检验（在图 8-2 中简称为“内检”），然后再用检验样品集对待检模型作进一步的外部检验（在图 8-2 中简称为“外检”），来评价模型对样品预测的性能。根据用数学模型对样品预测的结果与样品的实际值之间相吻合的程度，用相关系数 R （或决定系数 R^2 ）以及校正标准差 SEC 和预测标准差 SEP 来评价待检模型；若预测的结果已经符合分析要求，则跳出优化循环，用最佳参数正式建立模型；若尚不能达到分析要求，则可以通过检验的统计图、表或统计参数了解影响模型性能的因素，以便继续在优化循环内，进一步优化校正样品集与算法后再试建数学模型。

各个优化节点的优化原则是“少而精”，具体实施中对于“少”和“精”度的掌握，与校正变动背景是用自校正还是外校正一样，目前主要依靠经验或实际的检验与测试。可以说，近红外光谱分析建立数学模型的过程是优化建模参数的过程，优化的过程是近红外光谱数据处理掌握度的过程。

8.2 评价近红外数学模型优劣的标准

本节主要介绍近红外数学模型优劣的标志和判断的标准，包括模型自身相关性的优劣以及实际预测能力优劣的标准。

8.2.1 评价近红外数学模型自身相关性能及测量偏差的数学标准

评价近红外数学模型的相关性除通常采用相关系数外，还经常采用决定系数；评价近

红外数学模型的测量偏差除通常采用的平均绝对误差和相对误差外，还经常采用标准差、相对标准差、预测残差平方和等。

(1) 决定系数 R^2

$$R^2 = \left[1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_m)^2} \right] \times 100\%$$

n 为样本数； $\hat{y}_i$ 为第 i 个样品的预测值； y_i 为第 i 个样品的参考值（真值）； y_m 为样本真值的平均值。

R^2 接近 100% 表示预测浓度值接近真值，若 $R^2 = 1$ 则说明存在完全拟合。当预测残差平方和大于总平方和（真值的方差）， R^2 成为负值，表示拟合效果极差。

(2) 校正标准差（RMSEC）和预测标准差（RMSEP）

$$RMSEC = \sqrt{\frac{1}{N-P-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (P \text{ 为回归因子数目})$$

$$RMSEP = \sqrt{\frac{1}{NN} \sum_{i=1}^{NN} (y_{Ni} - \hat{y}_{Ni})^2}$$

标准差是衡量模型好坏的一个重要指标。

(3) 相对标准差（RSD%）

校正相对标准差： $RSD\% = RMSEC / y_{mc} \times 100\%$

预测相对标准差： $RSD\% = RMSEP / y_{mp} \times 100\%$

(4) 预测残差平方和（PRESS 值）

$$PRESS = E_1^2 + E_2^2 + \cdots + E_m^2$$

E_i 就是模型预测第 i 个样品的预测误差。

8.2.2 评价近红外数学模型实际分析预测能力优劣的标准

评价近红外数学模型实际分析预测能力的优劣有 3 个基本标志，即模型稳定性、可靠性及动态适应性。

(1) 模型稳定性 是指数学模型对样品的类型或其背景信息的适应性

模型的稳定性决定于建模样品集所覆盖范围大小，这需从大量的样品中挑选有代表性的样品才能实现。所谓代表性就是建模样品集包含的有效信息与背景信息的范围要足够宽。提高模型稳定性的措施、方法与实例详见本书 8.3.2 部分内容。

(2) 模型可靠性 是指数学模型的预测准确度

要得到准确可靠的分析结果，建立数学模型必须遵循“少而精”的原则。所谓“少”是指建模样品数要尽量少，所谓“精”是指待分析样品的特征要与建模样品集的特征相匹配。只有实现“少而精”的原则，才能提高数学模型的预测准确度。我们可以看出近红外数学模型的稳定性和可靠性是一对矛盾，如何解决这一矛盾是建立优秀近红外数学模型的重大问题。提高模型可靠性的措施、方法与实例详见本书 8.3.3 部分内容。

(3) 模型动态适应性（模型进化） 动态适应性是指数学模型受仪器性能的变化、分析时间、条件或样品状态等差别的影响

通过光谱的预处理方法可以消除部分干扰因素的影响。但随着时间、空间的变化，这些干扰因素不断增加时，必须通过模型转移或者需要在模型中将新增的干扰因素加入，这就需要对建模样品集要不断丰富。模型的动态适应性是模型长时间稳定性的体现。提高模型动态适应性的措施、方法与实例本章只做简单介绍，详见本书第 9 章“近红外光谱分析数学模型的维护”。

8.3 优化近红外数学模型的措施与方法

8.3.1 提高近红外数学模型相关性及相关可靠性的措施与方法

8.3.1.1 提高近红外数学模型相关性的措施与方法

相关性高、偏差低是优秀数学模型的标志之一，也是评价一个近红外模型是否优秀的前提和基础，但其相关性又不是由过拟合造成的。在近红外数学模型建立的过程中除使仪器得到尽可能高信噪比的光谱外，通常使用以下几种措施和方法来提高其相关性。

① 使用光谱预处理的方法：譬如中心化、归一化、散射校正、导数、傅里叶滤波等等（这些算法的详细介绍详见本书第 5 章“近红外光谱分析的信号处理”），以减少样品物理性质和其他干扰因素对光谱测量的干扰，提高光谱的有效信息量。

② 选择合适的定量分析谱区范围：对分离波长点的近红外光谱数据，通常采用逐步回归、所有可能回归等算法挑选出相关性最好的波长点组合来建立多元线性回归模型；对连续谱区的近红外光谱数据，虽然可以应用基于全谱数据分析的 PLS 或 PCR 等算法建立数学模型，但选择合适的谱区范围也是必要的。表 8-1 是 66 个小麦样品选择不同谱区范围建模时的相关情况（此模型为蛋白含量模型，样品的化学标定由中国农科院品资所提供，光谱采集使用的是德国布鲁克公司的 VECTER22 傅里叶近红外仪器）。

表 8-1 选择不同谱区范围的建模结果

所选谱区范围/ cm^{-1}	4000~6000	4000~8000	4000~10000	7000~10000
决定系数	98.1	98.5	97.5	69.5

在建立数学模型的过程中应选择哪些谱段来建立数学模型，一般是根据经验反复实验以确定最佳谱区范围，一般的近红外建模软件都提供该选择的功能，对大多数农产品而言 4000~8000 cm^{-1} 一般是较理想的谱区范围；现在有些商品化的近红外建模软件也有采用将整个谱区按照各个基团的合频、不同倍频吸收划分为不同的谱段，然后让计算机自动处理以挑选出最佳的谱段组合的方法。

③ 在建模型过程中对异常样品的处理：出现异常样品是不可避免的，而且异常样品对模型也是很重要的。如果这些样品对模型可能有坏作用，应该很仔细地将这些异常样品改正或去除，以提高模型自身的相关性。如果可能的话，应该尽量鉴别样品产生的原因。

校正过程中异常样品的检测方法最直观也是最简单的方法就是用真实值/拟合值 (True/Fit) 散点图进行查看，对于严重偏离拟合线或其偏差大于 2 倍或 3 倍平均偏差以上的样品通常认为是异常样品，为进一步分析是否为光谱异常、样品异常或化学值异常等，通常用以下几种方法进行进一步检测：是否有高的光谱残差值、是否有异常高的杠杆

利率 (leverage) 值、用主成分得分值 (Score) 散点图查看; 在预测过程中通常用预测样品主成分向量与校正集平均主成分向量之间的马氏距离值或光谱残差值来进行检测 (有关异常样品检测的更详细介绍详见本书第 11 章“异常样品的分析”)。

异常样品有好的异常样品和坏的异常样品两种, 在建模校正过程中我们只能剔除坏的异常样品, 而对模型有重要作用的好的异常样品则应该保留。如果异常样品的待分析组分含量处在最高或最低范围内, 那么这些样品是能提供有用信息的, 并认为是好的异常样品, 而且不应该去除; 坏的异常样品: 检查是否将这个样品的组分含量值输错了; 重新测量一些异常样品的 NIR 光谱; 重新测量异常样品的参考化学值 (组分含量值), 鉴别异常样品可能会非常费时。如果模型中好的样品很多而异常样品很少, 而且这些异常样品的待分析组分含量不处在最高或最低范围内, 那么这些异常样品可以放心地去除, 但是当校正样品的数目很少时, 过量的剔除异常样品应该避免。

以下介绍 2 种类型异常样品检测的实例。

图 8-3 是 90 个玉米样品其傅里叶光谱的第 1、2 主成分的空间分布图, 从图中明显的可以看出分布在右上方的 2 个样品为异常, 在建立其蛋白校正模型中这 2 个样品的相关性也处在较低的水平上, 此蛋白模型的平均相对偏差为 2.4%, 这两个样品的相对偏差分别为 4.0% 和 5.9%, 但这两个样品却并非坏的异常样品, 而是好的异常样品, 在建模校正过程中应保留而不应剔除, 因为这两个样品为高蛋白和高赖氨酸样品, 其蛋白和赖氨酸含量分别为 16.862%、17.627%; 0.72%、0.72%。

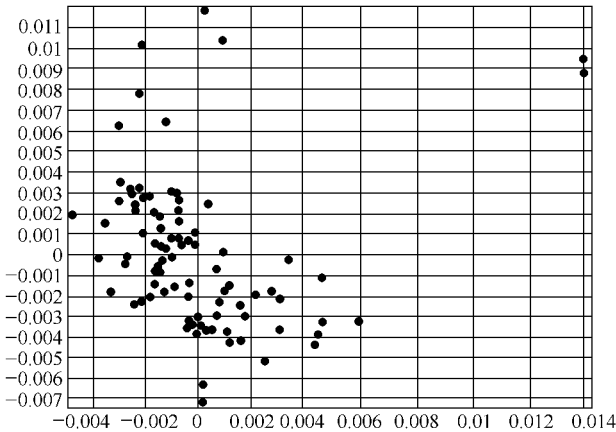


图 8-3 90 个玉米样品的第 1、2 主成分分布图

从图 8-4 中可以明显的看出 60 号样品严重的偏离了拟合线, 而该样品的化学值含量又不在最高、最低范围之内, 因此可以把这个样品看作是坏的异常样品, 可以放心地从模型中将该样品剔除, 该样品剔除后模型的相关性有明显的提高, 其相关系数由原来的 0.968 提高到 0.987。

8.3.1.2 提高近红外数学模型相关可靠性的措施与方法

在建立回归拟合模型的时候, 通常会产生以下几种情况:

① 欠拟合: 如果 PLS 所用的因子数太少, 那么光谱中一些有用的信息就没有包含在模型中, 那么校正的结果和预测的结果都不会很好。

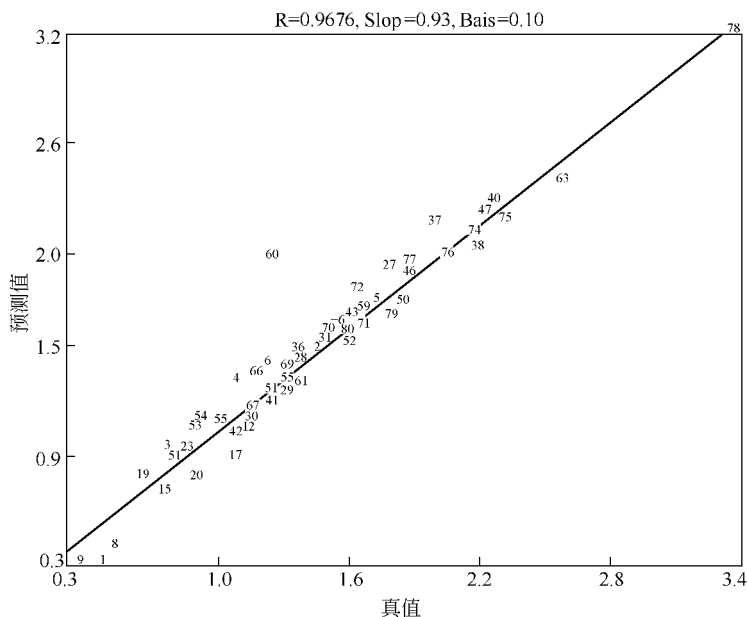


图 8-4 80 个青烟样品尼古丁含量模型真实值与拟合值的散点图

② 最佳拟合：应该使校正的结果和预测的结果都很好。

③ 过拟合：如果 PLS 所用的因子数太多，那么虽然对校正集样品的校正结果很好，但是对没有在校正集中样品的预测结果会很差。

在近红外数学模型的建立过程中，得到相关性好、偏差小的模型固然重要，但也要防止过拟合现象的产生，譬如在基于单波长点的多元线性回归模型中所选择的波长点数过多或在 PLS、PCR 回归中所用的主成分因子过多等都可能造成过拟合。下面以 PLS 方法为例说明在建模过程中如何防止过拟合现象的产生，这些方法同样适用于 PCR 方法，也同样适用于多元线性回归，只不过 PLS、PCR 等所确定的是最佳主成分因子数，而多元线性回归所确定的是最佳点数。

能够有效的运用 PCR 和 PLS 方法的关键就是为特定模型选择最佳的因子数目。这可以通过交叉证实和（或者）外部检验过程来确定。为了确定最佳的因子数通常采用两种检验方法：外部检验法、内部交叉证实法。

(1) 外部检验法

将标准样品集被分为两组：校正集和检验集。

用校正集样品建立一组 PLS 模型

用 1 个 PLS 因子建立 PLS 模型（阶数为 1），并且计算检验集的 $RMSEP$ （预测均方差）。

用 2 个 PLS 因子建立 PLS 模型（阶数为 2），并且计算检验集的 $RMSEP$ 。

用 3 个 PLS 因子建立 PLS 模型（阶数为 3），并且计算检验集的 $RMSEP$ 。

•
•
•

用 $h_{\max}$ 个 PLS 因子建立 PLS 模型（阶数为 $h_{\max}$ ），并计算检验集的 $RMSEP$ 。

通常选择检验集具有最小 $RMSEP$ 的阶数来作为最佳 PLS 因子数。

应用外部检验法来确定主成分因子数是比较客观的方法，能防止过拟合现象的产生，但应用该方法的关键是如何来确定哪些样品作校正集，哪些样品来作检验集，如果这两部分样品是人为的随机挑选，很容易造成检验样品集不具有普遍性，随机挑选的检验样品可能其特征会很相似，譬如其背景信息基本相同，或者待测组分的含量分布范围较小或不合理等，其检验的结果也是不可靠的。因此采用按照光谱（或光谱主成分）特征来自动分离建模、检验样品集，可以提高检验集样品的普遍性，保证检验结果的客观性，分离过程由软件辅助完成，称为自动分离，自动分离时按照建模、检验样品的光谱（或光谱主成分）特征都尽可能宽的原则来进行自动分离。我们以 180 个玉米样品建立蛋白含量模型时按 4 : 1 的比例来分离建模、检验样品集来验证其分离效果。自动分离的方法采用“主成分重心法”，即利用样品光谱的主成分特征向量，与

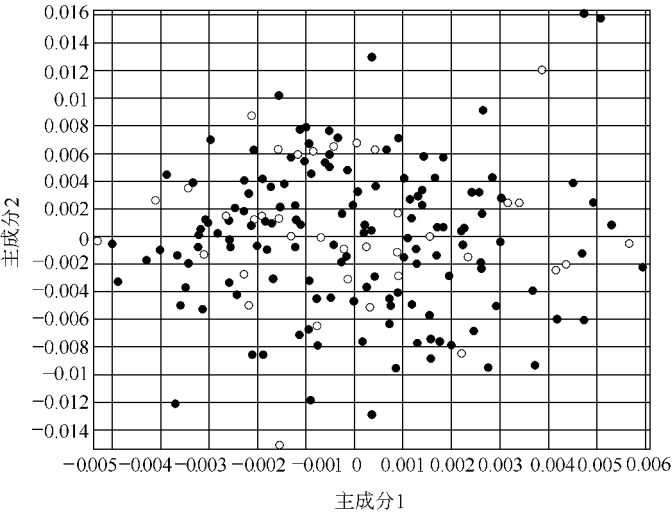


图 8-5 建模、检验样品光谱的第 1、2 主成分特征在空间的分布
(实心表示建模集，空心表示检验集)

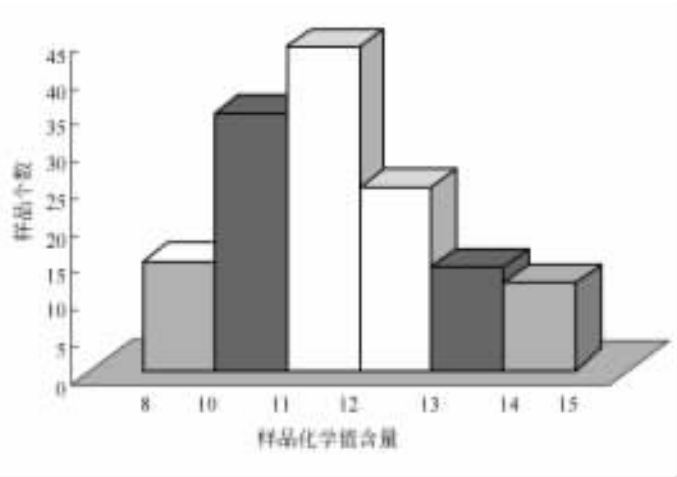


图 8-6 建模样品集蛋白含量的化学值分布

样品平均光谱的主成分特征向量之间的距离（马氏距离或欧氏距离）值，按从小到大排序，并按一定间隔来挑选样品的方法。图 8-5 表示建模、检验样品光谱的第 1、2 主成分特征在空间的分布情况；图 8-6、图 8-7 分别表示建模、检验样品化学值蛋白含量的分布范围图。

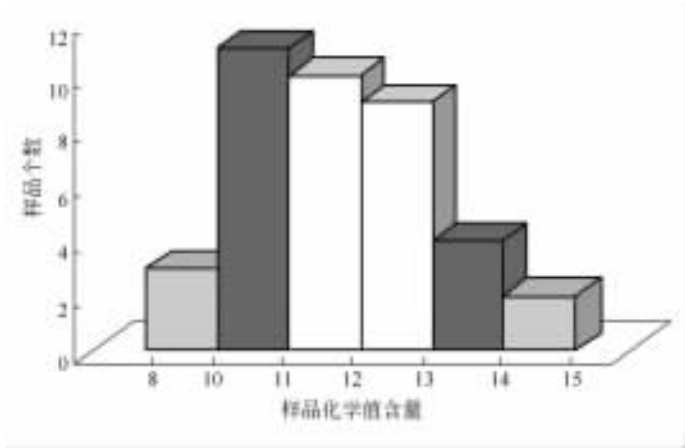


图 8-7 检验样品集蛋白含量的化学值分布

从图 8-5 中我们可以看出检验样品集基本上能包含所有样品的光谱特征信息，因为其在整个样品中的分布范围较大，分布较合理。

从图 8-6 和图 8-7 中我们可以看出检验集样品化学值蛋白含量的分布范围基本涵盖了建模集样品化学值蛋白含量的分布范围，因此无论从光谱的特征信息分布还是从蛋白含量的分布范围来看，检验样品集基本涵盖了建模样品集全部的背景信息和特征信息。因此这使得检验样品集具有普遍性，使得检验的结果具有客观性。

(2) 交叉证实法

交叉证实的过程可用图 8-8 表示。

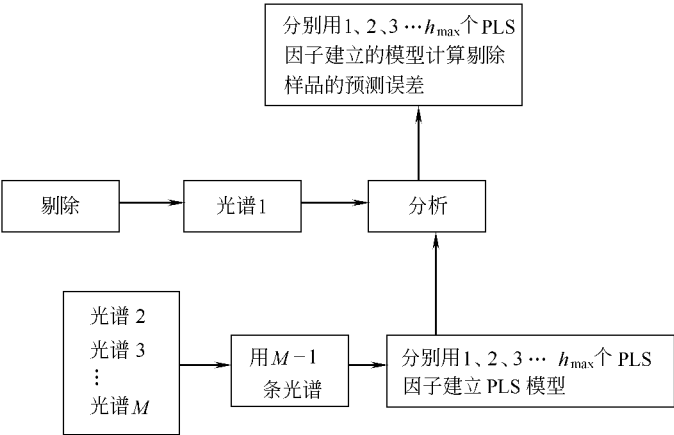


图 8-8 $h_{\max}$ 用户指定的最大因子数

当计算完“光谱 1”的预测误差后，再将其恢复到校正集中，然后再剔除“光谱 2”，分别用 $1 \sim h_{\max}$ 个 PLS 因子建立的模型预测“光谱 2”，并且计算其预测误差。对校

正集中的每 1 条光谱都重复进行这一过程，直到最后 1 个 PLS 模型对校正集中每一条光谱都有 1 个总体预测结果。我们可以看出通常选择交叉证实所得到的校正结果会比较校正所得的结果更客观一些，因为交叉证实所得的每一个样品的预测结果都没有包含在模型自身中，因此通常选择交叉证实最小 PRESS 值所在的因子数为最佳因子数。但用最小 PRESS 值的因子数也可能会导致过拟合，最佳阶数是经常通过 F 检验来确定，每一阶的 PRESS 值都会同前一阶进行比较，如果较高的阶数在统计上对较低阶数有显著的改变，则就会推荐较高的阶数，否则则使用较低的阶数，以防止过拟合现象的产生，同时，也可以通过查看 PLS 因子的残差光谱图形，那些只包含噪声的高阶因子应该不用。

8.3.2 提高近红外数学模型稳定性的措施与方法

定标模型的稳定性决定于建模样品集所覆盖范围大小，在经典光谱分析中的定标工作（外标法）很简单、很便宜，但是在应用 NIRS 对于复杂背景的农产品、食品等天然产物进行定量分析时，定标就比较复杂，这需要有代表性的数据和适合的定标方法。在建立一个优秀的定标模型中，要花费一定的时间和经费投入。建立定标模型是要在 1 个小样本的基础上获得可适应于大群样品或整群样品的模型。从下面这个用不同的小麦蛋白模型来分析预测来自河南小麦样品的实验数据，我们可以充分认识到建模样品集的选择对模型预测性能的重要性。

模型一是用 1998 年收集的部分全国范围内 66 个小麦样品建立的，其建模样品光谱的主成分空间分布特征见图 8-9，模型的预测结果见表 8-2；模型二是从几年内在全国范围里收集的 2000 多份样品中挑选出有代表性的 120 个样品建立的模型，其建模样品光谱的主成分空间分布特征见图 8-10，模型的预测结果见表 8-3；模型三是从 2000 多份样品中挑选出有代表性的 120 个样品加上 1998、1999 年在全国范围内收集的部分样品共 270 个，其建模样品光谱的主成分空间分布特征见图 8-11，模型的预测结果见表 8-4。

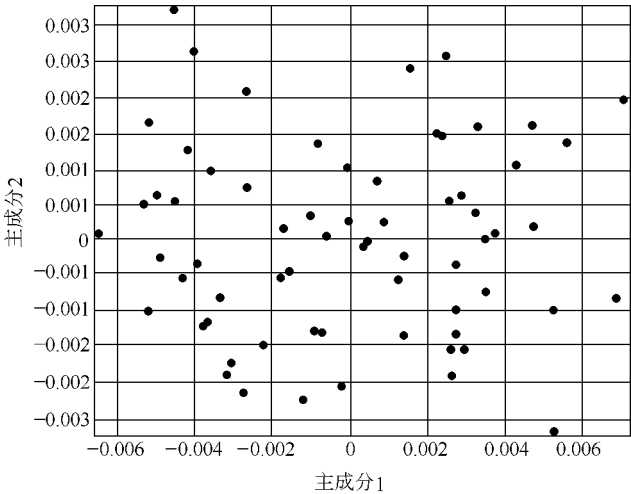


图 8-9 66 个小麦建模样品第 1、2 维主成分空间分布图

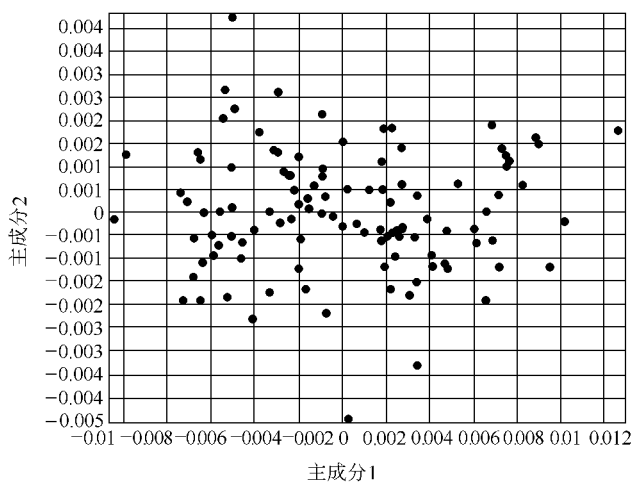


图 8-10 120 个小麦建模样品第 1、2 维主成分空间分布图

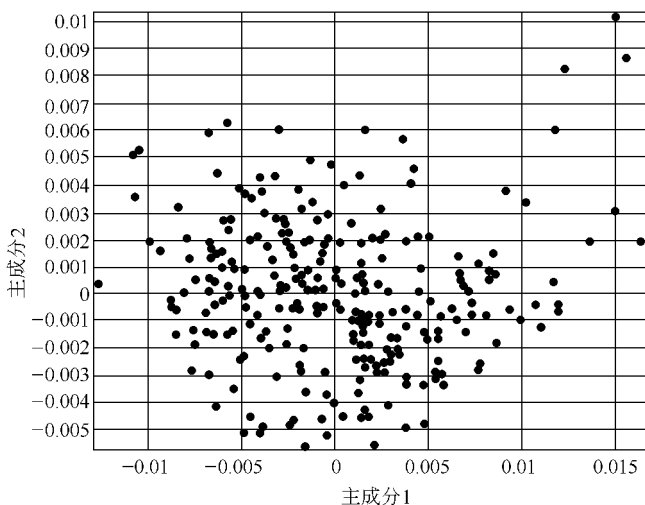


图 8-11 270 个小麦建模样品第 1、2 维主成分空间分布图

表 8-2

模型一的预测结果

样品号	真实值	预测值	绝对误差	相对误差/%
1	13.84	15.48	1.64	11.85
2	13.77	15.71	1.94	14.06
3	14.01	15.40	1.39	9.92
4	14.23	15.82	1.59	11.15
5	14.17	15.93	1.76	12.44
6	13.57	15.21	1.64	12.11
7	13.72	15.37	1.65	12.00
平均误差			1.66	11.93

表 8-3		模型二的预测结果		
样品号	真实值	预测值	绝对误差	相对误差/%
1	13.84	13.78	0.06	0.46
2	13.77	14.26	0.49	3.56
3	14.01	13.88	-0.13	0.90
4	14.23	14.11	-0.12	0.84
5	14.17	14.17	0.00	0.00
6	13.57	13.46	-0.11	0.84
7	13.72	13.42	-0.30	2.16
平均误差			0.17	1.25

表 8-4		模型三的预测结果		
样品号	真实值	预测值	绝对误差	相对误差/%
1	13.84	13.64	-0.20	1.47
2	13.77	14.18	0.41	2.95
3	14.01	13.70	-0.31	2.24
4	14.23	14.04	-0.19	1.34
5	14.17	14.43	0.26	1.81
6	13.57	13.33	-0.24	1.74
7	13.72	13.53	-0.19	1.41
平均误差			0.26	1.85

从上表中我们可以看出模型二和模型三的预测结果明显好于模型一，主要原因是模型二和模型三的建模样品集是从大量样品中挑选出来的有代表性样品，从建模样品的光谱主成分特征分布图中也可以看出。图 8-9、图 8-10 中，我们可以看出模型二的建模样品的光谱第一、第二主成分的分布范围比模型一的建模样品的光谱第一、第二主成分的分布范围宽，约扩大了 90％和 30％，因此模型二的稳定性就好多了，即对不同类型样品的适应性强，所以预测性能也就好。因此挑选有代表性样品建模有重要意义。

从图 8-11 中我们可以看出其建模样品的光谱第一、第二主成分的分布范围比模型二的建模样品的光谱第一、第二主成分的分布范围还要宽，但预测性能却有所下降，这是因为我们建立、优化模型必须按照“少而精”的原则，当建模样品数太多时，由于干扰信息的增强，会影响模型的性能。至于需要多少样品来建模型，是一个需要不断进行实践验证的过程，但首先应该考虑模型的稳定性问题。

因此，建立稳定性好的数学模型的关键之一是如何得到一个好的、有代表性的定标用数据，这需要从大量的样品中挑选有代表性的样品才能实现。所谓代表性就是建模样品集包含的有效信息与背景信息的范围要足够宽。

8.3.2.1 挑选有代表性的建模样品的方法

① 根据样品的可知因子参数进行人工挑选。这些可知因子参数对农产品而言，譬如其产地、品种、年份、颜色等等，对一些在线的生产产品，譬如其生产周期内不同时刻的样品、工艺流程的变动、溶剂的变化等等，在挑选建模样品时应尽可能的增大这些因子的

变异范围，以得到代表性尽可能好的建模样品集；当然还有许多未知的因子参数，对农产品而言譬如其各种内在的化学品质等，如果用人工挑选代表性样品，只能采用随机挑选的方法。

② 对大量样品知道其某些待建模型的化学值的情况下，可按照这些组分的化学值进行挑选。在挑选过程中应尽可能的包含那些分布在两端即最高、最低范围内的样品；对建模样品集化学值什么样的分布最好，现在还没有统一的观点，有人认为呈均匀分布好，即高、中、低含量的样品数量大体相当，但我们认为建模样品集的化学值范围呈正态分布好一些，因为将来用这个模型分析大量其他样品时，大多数自然产物化学值的分布也是呈正态的，那么这个模型对大多数样品的适应性会更好，当然对一些工业在线产品的化学值分布应另当别论；如果在条件允许的情况下，最好不要只根据一种化学组分的含量来挑选代表性样品，譬如即使蛋白含量完全相同的两个样品，它们的淀粉或其他组分的含量差异可能会很大，如果在挑选时只选择了其中一个样品那也不是很好的。

③ 用数学方法根据大量样品的近红外光谱特征来进行挑选。如果在收集大量样品时也兼顾那些可知因子参数来收集，再配合这些挑选方法，将是挑选有代表性样品的最理想方法。这些根据样品光谱特征挑选有代表性样品的数学方法主要有：

a. 聚类方法：通过对光谱的聚类分析，可保证所选择的样品，其光谱与其他样品光谱不是很一致。对光谱的聚类分析方法不止一种，有系统聚类法、分层方法（Hierarchical Procedure）、模糊分聚类方法（Fuzzy Clustering）等。挪威科学家 Nas 总结了聚类分析的不同方法，其共同的规则为：

- 选择覆盖整个变异的样品，避免选择光谱相似的样品；
- 选择接近跨度边缘的样品；
- 从每一个聚类光谱中，仅选择一个或较少的样品。

b. 主成分因子法：这种方法是利用群体光谱主成分分析的主成分因子对样品进行挑选，其基本原理是通过主成分代表因子，以反映样品光谱的异质性，即样品的异质性。这种挑选方法比只利用样品的原始近红外光谱来进行挑选更具有优势，因为通过主成分因子更能体现样品之间的异质性，避免了原始光谱所带来的其他众多干扰因素。

利用主成分因子法进行代表性样品挑选的方法也很多，一种最简单的方法就是对每一阶主成分因子都根据其值（即得分值大小）从大到小来进行筛选，直到最佳的主成分因子数为止；还有一些数学家设计出了根据多维主成分的空间分布情况来进行自动挑选的数学方法，譬如由中国农业大学张录达教授提出的网格法等，其基本思路在只有 2 个主成分时可用图 8-12 来说明。

其基本思路即是把二维主成分的空间分布按所设定的网格来划分，使每一个样品的主成分向量都落在所有的网格之内，然后根据每个网格内所包含的样品数量来进行挑选，挑选的基本原则是所有有样品的网格内都应该取到，然后再根据每个网格内样品的数量进行按比例筛选，在多维的主成分空间内其原理基本类似，只不过主成分不是二维向量，而是多维向量，网格不是二维网格，而是多维网格。

8.3.2.2 挑选代表性样品的最佳数量

到底挑选多少有代表性的样品来建模才算最好，不能一概而论，要根据建模对象的复

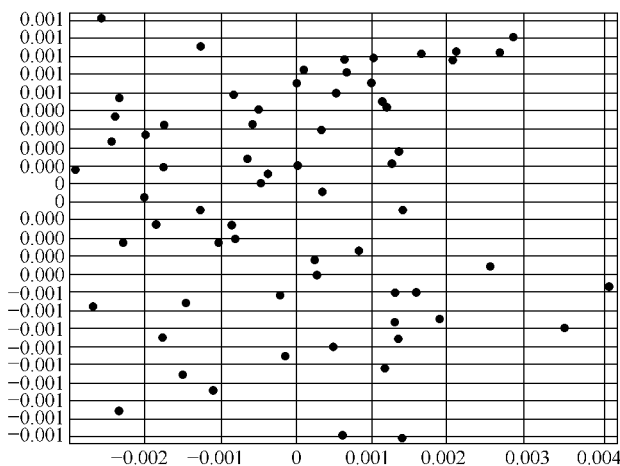


图 8-12 二维网格方法说明示意图

杂程度而具体问题具体分析，但对于大多数对象一般来说所挑选的建模样品数应不少于 50 个，但也不能过多，样品数过多就会带来越多的干扰因素，因为模型中每增加一个样品时在增加信息量的同时也增加了干扰信息，当建模样品的数量太多时干扰信息会掩盖大部分有用信息，使所建模型性能大大下降，对模型也是不利的。一般来说建模样品数不宜大于 300 个。

8.3.2.3 用数学方法挑选代表性样品建模实现模型稳定性的实例

前面我们提到优秀数学模型的标志之一是稳定性要好，即数学模型对将来待分析样品的类型或其背景信息的适应性要好，这需要从大量的样品中挑选有代表性的样品才能实现。譬如我们用蛋白质含量在 14%~22% 范围内的小麦样品建立的模型，那么这个模型对蛋白质含量在 12% 以下或 24% 以上的样品会不适应，对这些样品的预测性能会显著降低；如果我们用栽培品种大豆的样品建立的数学模型来分析野生大豆，那么这个模型对野生大豆样品可能不适应。因此我们要建立稳定性好的模型，必须从大量样品中来挑选有代表性的样品，以尽可能大的包含样品的背景信息。

要挑选代表性样品建模的原因是我们需要按照“少而精”的原则来建立、优化模型，使模型中尽可能避免背景信息基本相同的样品出现，使建模样品尽可能的“精”即有代表性。当然建模样品数目的“少”是有一定限度的，这需要根据具体情况来确定建模样品数目，但建模样品数目一般不能过多，譬如超过 300 个，因为模型中每增加一个样品时在增加信息量的同时也增加了干扰信息，当建模样品的数量太多时干扰信息会掩盖大部分有用信息使所建模型性能大大下降。

下面介绍我们采用的在主成分多维空间内运用“网格法”为云南烟草科学研究院从 1800 个烟草样品中挑选 300 个代表性样品的实验数据。

我们首先来看一下 300 个代表性样品在 1800 个样品中，其第一、二维主成分空间分布的情况（图 8-13）。

从图 8-13 我们可以看出挑选样品在二维主成分空间范围较大，且分布较合理，也就是说这些样品从光谱特征上来看是有代表性的，所挑选 300 个样品的光谱主成分在全部

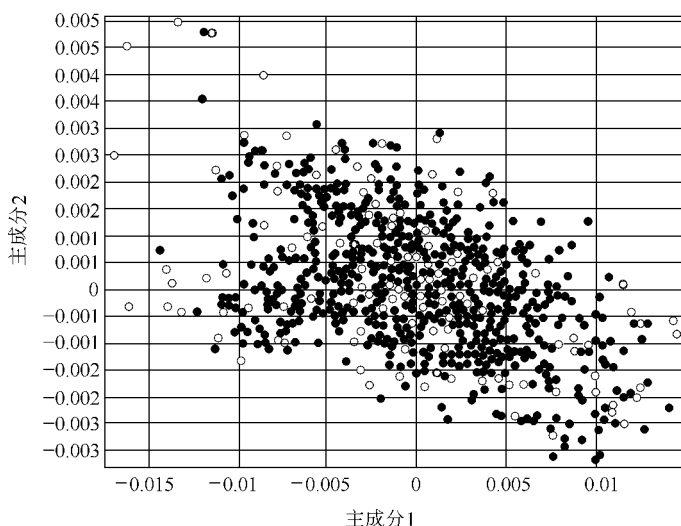


图 8-13 1800 个烟草样品第一、二维主成分空间分布图
(空心表示挑选样品，实心表示未挑选样品)

1800 个样品的光谱主成分中的空间分布比较合理，基本上能包含所有样品的光谱特征信息。

另外从这些样品各组分的含量范围我们也可以看出这些样品的代表性，见表 8-5。

表 8-5 云南烟草科学研究院挑选的 300 个代表性样品各组分的含量范围

组分名称	含量范围	组分名称	含量范围
总糖	4.71%~38.31%	尼古丁	0.47%~4.04%
还原糖	4.75%~34.01%	总氮	0.973%~4.531%

以上除尼古丁含量上限值与文献报道近红外建模数据相当外，尼古丁含量下限值与其他各组分上、下限值均比国内外文献报道的近红外建模数据的范围要宽许多，因此用这些样品建立的模型其适应性会更宽、更好。

所挑选样品各组分的含量值范围不但宽（已经把 1800 个样品中各组分含量的最大值和最小值样品全部挑选出来了），而且化学值含量的分布范围都非常合理，仅以还原糖含量的分布范围为例进行说明。如图 8-14 所示，我们可以看出建模样品的化学值含量基本符合正态分布，这样不但能保证所建模型的稳定性，而且同时能保证模型的预测准确度即模型的可靠性。

8.3.2.4 挑选代表性样品以建立稳定数学模型的基础是建立近红外光谱库

该部分的详细介绍请参见本书第 13 章“近红外光谱分析的光谱库与网络技术”。

8.3.3 提高近红外数学模型可靠性的措施与方法

近红外数学模型的优劣即其预测能力是不能只靠模型的稳定性来衡量的，也就是说并不是建模样品的代表性越好、模型的适应范围越宽、模型的稳定性越好就意味着模型的预测能力也越好。定标模型中样品的变异范围越宽，则模型中所遇到的非线性或异质性问题

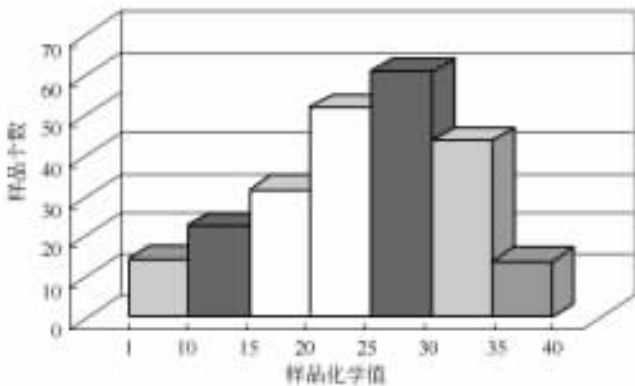


图 8-14 代表性样品总糖含量的分布图

越严重，干扰因素也越多，对某些预测样品而言其预测能力会下降，如果不知道今后要预测样品的类型，就评价一个定标模型的预测能力是很不可靠的，举一个极端的例子：如果用许多农作物譬如小麦、玉米等样品的蛋白含量模型用来预测小麦样品的蛋白含量，那么它的预测性能肯定不如只用小麦样品建立的模型；但如果用这个混合模型来分析预测玉米样品的蛋白含量，那么其预测性能肯定比用小麦样品建立的模型来预测玉米样品的预测性能要好。再举一个实际的例子：我们用蛋白质含量在 14%~22% 范围内的小麦样品建立的模型，虽然其稳定性，即预测 12% 以下或 24% 以上的样品会不好，但如果要预测蛋白质含量在 14%~22% 范围内的样品，其预测准确度会比用含量在 10%~30% 范围内样品建立的模型要好。

要得到准确可靠的分析结果，除建立模型的标样化学值要准确外，建立数学模型的过程中必须遵循“少而精”的原则。所谓“少”是指建模样品数要尽量少，所谓“精”是指待分析样品的特征要与建模样品集的特征相匹配。从大量样品中挑选有代表性的样品可以体现“少”的原则，但“代表性”即样品的变异范围要尽可能宽与“精”的原则即待分析样品的特征要与建模样品集的特征相匹配却是一对矛盾，因此，在设计一个定标模型中，我们必须在模型的稳定性（宽的变异范围）和可靠性（窄的变异范围）之间进行妥协，找到最佳契合点；模型稳定性的预测能力我们可以称为模型的相对预测能力，模型可靠性的预测能力我们可以称为绝对预测能力，因此，评价一个好的定标模型时，要综合绝对预测能力和相对预测能力这两个指标来评价。

那么我们在建立实用的分析模型时应采取哪些措施既要保证模型的一定稳定性又能提高它的可靠性呢，即如何缓和这一矛盾呢？

8.3.3.1 在保证模型一定稳定性的前提下，通过建立变异范围相对小的数学模型来提高模型的可靠性

当定标模型所覆盖的范围比较小时，由于在一个窄的范围内线性更好，因此，可靠性（绝对预测能力）提高了。但是，由于窄范围内，样品变异小，所以，稳定性（相对预测能力）会下降，但如果我们对将来待分析的大多数样品的一些特征是已知的，那么我们在建立定标模型时可以不考虑一些样品的变异因素。

这些方法和措施有很多，可以具体问题具体分析，譬如如果我们要建立一个供河南

地区使用的小麦模型，就没有必要收集全国甚至全世界范围内的小麦样品来建立这个模型；如果我们只建立一个烤烟烟草样品的模型，那么青烟等其他类型的烟草样品就没有必要收集；如果我们要建立一个无铅汽油的模型，那么含铅汽油的样品就没有必要收集等等。

如果对待分析样品的许多其他特性无法预知，那么我们可以利用一部分待分析样品的光谱从光谱库中挑选一些与其特征相似的样品来建立模型；但我们也不能认为模型的变异范围越窄会越好，应该是在某些变异因素不考虑的前提下，其他的一些变异因素也应该尽可能的宽，因此即使是建立变异范围相对窄的数学模型时，挑选代表性样品来建模的方法也同样适用的。

8.3.3.2 用局部回归数学算法来缓解模型的稳定性和可靠性这一矛盾

局部回归算法的核心思想是：在提取群体样品的特征因子（主成分因子）时利用的是大量的有代表性的样品光谱来提取，但在分析预测每一个未知样品时，都会根据待分析样品光谱的主成分因子特征，从建模样品集中挑选出部分与其最匹配的样品来为每一个待分析样品建立一个与其相适应的局部模型，并对其进行预测；寻找匹配性样品的原则可以根据待分析样品光谱的主成分因子与建模样品集中样品光谱的主成分因子之间马氏或欧氏距离的大小来进行匹配性的选择。如果把主成分因子简化为一个只含有两个元素的向量，采用的是欧氏距离的选择方法，那么这种算法的原理可以用图 8-15 来描述。

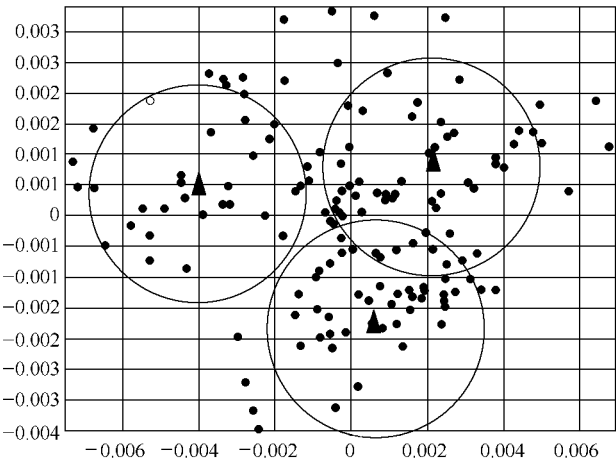


图 8-15 局部回归算法原理说明示意图

图 8-15 中的各个圆点表示建模样品光谱的二维主成分在空间的分布，3 个三角形表示 3 个待分析样品光谱的二维主成分在空间的分布，那么用局部回归算法分析这 3 个样品时，并不是利用所有样品建立的模型来分析这 3 个样品，而是根据每一个待分析样品的主成分特征挑选与其最相似的部分样品来为这个待分析样品单独建立一个局部的模型，并对其进行分析预测，所挑选的局部相似样品可以用各自样品大圆圈范围内的样品来表示。

采用这种数学算法可以很好地提高模型的可靠性，提高分析样品的预测准确度，譬如我们用小麦蛋白含量模型来预测天津农科院通过网络传送过来的 4 个小麦样品的例子，其具体结果见表 8-6。

序 号	真 实 值	不采用局部回归预测结果	采用局部回归预测结果
1	15. 25	14. 52	15. 25
2	14. 80	15. 13	15. 09
3	15. 83	16. 29	16. 09
4	16. 63	16. 04	16. 69

其中不采用局部回归的预测平均相对误差为 3. 37%，而用局部回归的预测平均相对误差为 1. 32%。

同时采用这种数学算法也可以很好地缓解模型的稳定性和可靠性这一矛盾，表 8-7 中的数据是上海烟草集团公司技术中心用使用了该方法的中国农业大学近红外定量分析软件，所建立的不同地区模型来预测不同地区样品的结果，其中所建的贵州地区模型共有建模样品 80 个，所建的山东地区模型共有建模样品 180 个，混合模型的建模样品为两地区样品量的总和，共 260 个。预测检验样品是上海烟草集团公司技术人员从贵州、山东样品中随机挑选的各 10 个样品，所有预测检验样品都不在建模样品集内。

表 8-7 用地区模型与综合模型分析
 预测山东样品的结果

成分名称	地区模型(山东)预测 平均相对偏差/%	综合模型预测平均 相对偏差/%
尼古丁	2. 45	2. 38
钾	10. 54	10. 64
还原糖	2. 47	2. 25
总糖	3. 43	2. 72
总氮	10. 25	10. 46

表 8-8 用地区模型与综合模型分析
 预测贵州样品的结果

成分名称	地区模型(贵州)预测 平均相对偏差/%	综合模型预测平均 相对偏差/%
尼古丁	5. 20	2. 07
钾	5. 76	8. 81
还原糖	2. 51	2. 24
总糖	2. 39	3. 19
总氮	5. 42	6. 08

从表 8-7、表 8-8 中的分析预测结果我们可以看出：用混合模型预测无论是山东样品还是贵州样品，其预测性能（预测相对平均偏差）与用各自地区模型来预测本地区样品的预测性能基本相当。这充分说明运用局部回归算法在加大模型适用范围的同时又保证了其预测准确度，在保证模型稳定性的同时又保证了其可靠性，保证所建模型预测准确度的情况下，提高了模型的稳定性（既模型的适应能力）。

8.3.4 提高近红外数学模型动态适应性的措施与方法

模型动态适应性是指数学模型受仪器性能的变化、分析时间、条件或样品状态等差别的影响较小。通过光谱的预处理方法可以消除部分干扰因素的影响。但随着时间、空间的变化，这些干扰因素不断增加时，必须通过模型校正或者需要在模型中将新增的干扰因素加入等措施维护模型的性能。所谓模型的转移是指通过算法将原来只能适用于某一范围的数学模型校正使之能适用另一范围，范围的变化主要体现在光谱特征的变化上，这主要体现在不同仪器之间的差别、仪器性能的变化、分析时间、条件或样品状态、类型的差别等。

随着时间、空间的变化模型动态适应性会降低的原因是由于近红外光谱具有变动的特征，而且数学模型本身也会受到各种条件变化的影响。所以提高所建模型的动态适应性以实现近红外数学模型资源的充分共享是推广应用近红外技术的重大课题，有关本节的一些具体介绍详见本书第 9 章“近红外光谱分析数学模型的维护”。

8.4 近红外光谱分析数学模型优化技术的发展趋势

近红外光谱分析的“多中选优”、“以度取优”、“全程优化”、“循环优化”和“终身优化”等说明该项技术在很大程度上需要依靠分析人员的经验，这是该项技术应用的困难所在之处。因此近红外光谱分析技术当前发展趋势是：

① 发展免建模的专用仪器以及运用网络技术实现资源共享的免建模近红外光谱仪器，这类仪器建立和优化数学模型的工作由专业人员完成，避免应用人员建模过程的困难；

② 发展新的算法，扩大建模过程中算法对变动背景的自校正作用，方便分析人员用于建立优秀的数学模型；

③ 发展智能化软件，通过计算机运算实现参数的最优化，减少需要依靠分析人员的经验来判断最优参数的困难。

参 考 文 献

- [1] 严衍禄，赵龙莲，李军会. 现代近红外光谱分析的信息处理技术. 光谱学与光谱分析，2000，20 (5)
- [2] Balbine Garcia-Criado, Antonia Garcia-Ciudad. Application of Near-Infrared Reflectance Spectroscopy to Protein Analysis of Grassland Herbages Samples. JSci Food Agric, 1990 (50): 479~484
- [3] M Ferroari, D A Wilson, et al. Noninvasive Determination of Hemoglobin Saturation in Dogs by Derivative Near-Infrared Spectroscopy. Am J Physiol, 1989 (256): 1493~1499
- [4] E Peuchant, C Salles, R Jensen. Determination of Serum Cholesterol by Near-Infrared Reflectance Spectroscopy. Anal Chem, 1987 (59): 1816~1819
- [5] J W Pyper. The Determination of Moisture in Solids, A Selected Review. Anal Chem Acta, 1985 (170): 159~175
- [6] P Dubios, J R Martimes, et al. Determination of Five Components in a Pharmaceutical Formulation Using Near-Infrared Reflectance Spectrometry. Analyst, 1987 (112): 1675~1679
- [7] E W Ciurczab, R P Tormi. Analysis of Solid and Liquid Dosage Forms Using Near-Infrared Reflectance Spectroscopy. Spectroscopy, 1987 (2): 41~43
- [8] P L Walling, J M Dabney. Application of Near-Infrared Reflectance Spectroscopy to the Quality Assurance of Surfactants. J Soc Cosmet Chem, 1986 (37): 445~459
- [9] J G Montalve Jr. A Comparison Study of NIR Diffuse Reflectance of Cottons Grouped According to Fiber Cross- Sectional Dimensions Part I: Fundamentals. Appl Spectrosc, 1991 (45): 779~789
- [10] C E Miller. Near-Infrared Spectroscopy of Synthetic Polymers. Appl Spectrosc Rev, 1991 (26): 277~339
- [11] C S Jacobsen. Trends in the Infrared and Near Infrared Properties of Organic Conductors. J Phys C, 1986 (19): 5643~5654

- [12] E Nishio, N Ikuta, et al. Fourier Transform Near- Infrared Attenuated Total Reflectance Spectroscopy of Silance Coupling Agents on Glass IRE. Appl Spectrosc, 1990 (44): 614~617
- [13] L P McDermott. Approaches to Applying Near-Infrared Spectroscopy On-line. Adv Instrum. Control, 1990 (45): 669~677
- [14] M Forina, C Armanino, R. Giangiacomo. In: Hildrum, Isaksson, Naes, Tandbery, et al. Near Infrared Spectroscopy. Ellis Horwood, Chichester, 1992
- [15] M Forina, G drava, C Armanino, et al. Ontell Lab Syst, in press
- [16] K A Martin. Recent Advanced in Near- Infrared Spectroscopy On-line. Appllied Spectroscopy Reviews, 1992, 27 (4): 325~383
- [17] B R kowalski. Chemometrics. Anal Chem, 1980, 52 (5): 112~122
- [18] Massie D R, et al. The spectral Refleectance and Properties of Grain in the Visblie and Near Infra-red. Trans Am Soc Eng, 1965 (8): 598
- [19] **严衍禄等. 北京农业大学学报, 1990, 16 (增)**
- [20] Yan Yanlu, et al. Determination of Cereal Quality by Fourier Transform Near Infrared Diffuse Reflectance Spectroscopy. Proc 3rd Beijing Conf And Exhib On Instrum Analysis, 1989
- [21] **严衍禄等. 傅里叶变换近红外光谱法测定谷物的品质. 光谱学与光谱分析, 1989 (10)**
- [22] **严国光. 严衍禄等. 仪器分析原理及其在农业中的应用. 北京: 科学出版社, 1982**
- [23] **吴瑾光. 近代傅里叶变换近红外光谱技术及应用 (上卷). 北京: 科学技术文献出版社, 1994: 12**
- [24] **严衍禄, 张录达. 傅里叶变换近红外漫反射光谱分析应用基础的研究. 北京农业大学学报, 1990, 16 (增)**
- [25] **严衍禄, 劳彩莲, 李军会等. 傅里叶红外谱仪技术及应用论文集: 分析中的建模样品集总体组成与模型的预测稳定性. 1999**
- [26] **严衍禄, 李军会, 劳彩莲等. 傅里叶红外谱仪技术及应用论文集: 分析中的建模样品集的光谱优化与数学模型的预测性能. 1999**
- [27] Tomas Isaksson, el al. Selection of samples for calibration using spectral information. NIR News, 1995, 6 (1)
- [28] Tormod Naes, Tomas Isaksson. Locally Weighted Regression Scatter Correction for Near-Infrared Reflectance Data. Anal Chem, 1990 (62): 664~673
- [29] Edward V Thomas. A Primer on Multivariate Calibration. Anal Chem, 1994, 66 (15)

9. 近红外光谱分析数学模型的维护

近红外光谱分析的前提是数学模型，建立一个长期稳健（Robust）、可靠（Reliable）、准确（Accurate）的数学模型是相当复杂的过程，它需要有优秀的近红外分析软件，有经验的近红外专业人员，并掌握一大批样品资源，还要进行大量困难而复杂的开发工作，为此，就要耗费大量的时间、人力、物力和财力。但因为数学模型包含了样品信息、装样条件和仪器状态等的信息，所以在使用前必须对模型是否适配于新样品的光谱，即对模型的正确预测能力做出评估，这就是数学模型的适配性检验技术。

根据近红外分析的流程，如果模型适配性检验的结果满足需求，则可以直接进入分析层次，运用数学模型分析待测样品。如果模型不适配，则应返回进行模型的维护工作。模型的维护又分为两类，一类称为模型的修正，另一类称为模型转移（Calibration Transfer）。如果待测样品本身的物化信息超出了原有模型的范围，模型的维护工作主要是扩充建模样品所覆盖的（主成分）空间范围，则称这种维护工作为模型的修正。模型的修正通常是通过在模型中添加原模型中不包含的新信息完成的，原建模样品与新添加样品合并到一起重新进入模型的优化循环，并重新得到最佳模型，即为修正后的模型。利用修正后的模型在预测分析未知样品前仍需进行模型的适配性检验工作。如果待测样品本身的信息与建模样品信息并无区别，或者说主成分空间是一致的，但待测光谱因为装样条件、仪器状态等与建模时不一致，使得直接用原模型预测待测光谱结果有较大偏差，希望通过模型的维护，使同一待测样品在两种情况下得到一致的结果，我们通常称这种模型维护工作为模型转移，模型转移通常发生在不同仪器之间、同一仪器的不同附件之间或不同测量时间等。

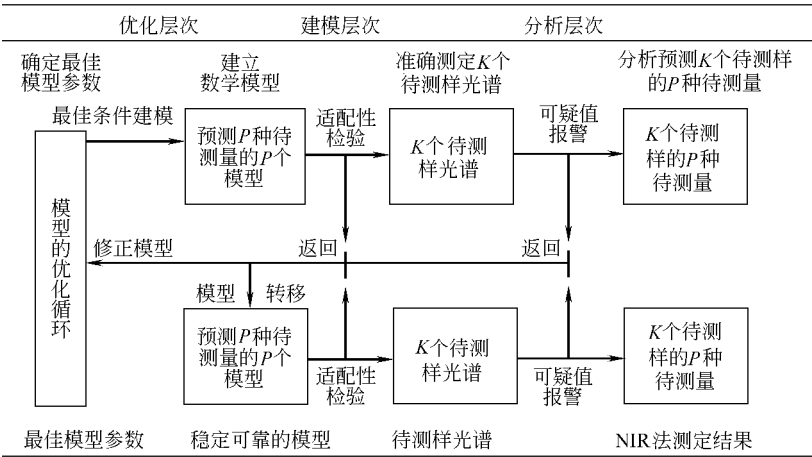


图 9-1 模型的适配性检验，模型的修正和模型的转移流程图

注：如有问题本章 9.3.2 节内容请与杨曙明联系，其他内容与赵龙莲联系。

模型得到维护后进入样品分析层次，分析结果的好坏可以通过“可疑值报警”给出标识，所谓“可疑值报警”就是用数学方法指出预测结果可靠性较差的某些样品。如果“可疑值报警”样品数太多，则模型仍需返回重新进行模型的维护工作。模型的适配性检验，模型的修正及模型的转移这一系列工作可以用图 9-1 所示流程描述。

因为模型的修正在添加样品后实际进入模型的优化循环，本书第 8 章内容对此有详细阐述，因此本章将主要介绍模型维护工作中模型的转移。

成功的实现数学模型的修正与转移，是应用近红外技术的基础，也是实现近红外资源共享，促进近红外光谱分析技术产业化发展的核心技术之一。

9.1 数学模型的适配性检验

如前所述，近红外光谱分析的基础是数学模型，在近红外数学模型没有损失预测能力的前提下，人们希望能够长久而且不加变动地利用它。但是每个数学模型有一定的适配范围，即每种数学模型只能适用于一定范围的样品、一定的装样条件和一定的仪器状态。因此，用数学模型预测一批新样品时首先需要检验模型的适配性。

进行模型适配性检验常用的一种方法是选择一化学值已知且其物化性质不随时间变化的标准物 (Standard Material)，在每次进行测量前，先扫描此标准物的图谱，用已知模型预测此图谱，看预测值是否满足用户测试需求。如能满足，则可利用此模型预测分析未知样品。否则，就需要进行对模型维护，以保证分析样品的准确度。BRAN+LUBEE 19 个滤光片仪器 IA450 就是采用这种方法，其标样用塑料袋或锡箔密封起来，放在冷冻室内保存以防化学组分的变化，需用时随时取出，和待测样品在同一环境中平衡 1~2d 即可使用。

检验模型常用的另外一种方法是利用模式识别的方法，如马氏距离法、主成分空间分布法、SIMCA 法、判别分析法、模糊聚类法等对未知待测样品做出判断，从而决策出模型是否需要维护。下面分别就几种方法在模型适配性检验方面的应用做简单介绍，各方法的原理在本书第 3、4 章有详细阐述。

9.1.1 马氏距离法

未知样品光谱与建模样品光谱之间马氏距离的大小可以作为衡量模型是否适配于未知样品光谱的方法之一。

在利用已有模型预测未知样品前，首先计算该未知样品与建模样品范围之间的马氏距离，如果距离超过设定阈值，则给出报警，说明模型的有效信息不能充分覆盖样品，该样品预测的效果可能较差，测定值的准确性可疑。如果报警样品的数量过多，应对数学模型进行相应的维护。

例：建立 1998 年 66 个小麦样品蛋白含量的数学模型，用此模型直接预测 1999 年的 12 个小麦样品，结果见表 9-1，由表可见，12 个待测样品的光谱与建模样品光谱的马氏距离全部大于阈值 3，近红外预测值全部报警（用 * 表示），这说明 1998 年建立的模型不适配于 1999 年的小麦光谱，或者说 1999 年小麦样品光谱不在 1998 年小麦模型所涵盖的范围内，模型需要维护。

编 号	样品光谱名称	化学值	近红外预测值	马氏距离	是否报警(＊)
1	99WT01. DX	16. 10	17. 86	6. 28	*
2	99WT02. DX	17. 34	18. 41	5. 58	*
3	99WT03. DX	15. 39	16. 94	3. 86	*
4	99WT04. DX	19. 32	19. 88	8. 35	*
5	99WT05. DX	14. 62	16. 48	4. 43	*
6	99WT06. DX	16. 06	18. 14	6. 13	*
7	99WT07. DX	16. 19	17. 77	6. 99	*
8	99WT08. DX	18. 35	19. 00	5. 56	*
9	99WT09. DX	18. 68	19. 44	6. 36	*
10	99WT10. DX	21. 17	21. 94	9. 59	*
11	99WT11. DX	19. 01	19. 92	7. 98	*
12	99WT12. DX	19. 09	20. 38	6. 70	*

9. 1. 2 主成分空间分布法

光谱主成分的得分构筑成主成分空间，利用样品的主成分在空间的分布也可以对模型进行检验。仍用上例表示，66 个 1998 年小麦建模集样品集的第一主成分与第二主成分在主成分空间的分布见图 9-2 中右上方圆点表示的部分，而 1999 年 12 个小麦样品的主成分在主成分空间的分布见图 9-2 中左下方三角形表示的部分。

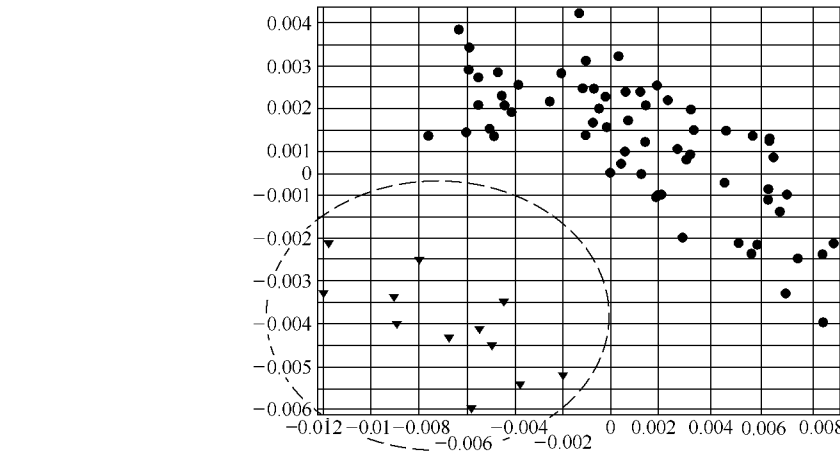


图 9-2 66 个 1998 年小麦与 12 个 1999 年小麦的第一主成分与第二主成分空间分布图

由图 9-2 可见，12 个 1999 年小麦与 66 个 1998 年小麦在主成分空间中分成了两个区域，说明用 1998 年 66 个小麦建立的数学模型不适合于直接预测 1999 年的 12 个小麦，检验结果与马氏距离法相同。

用马氏距离法对模型进行检验时，如果采用全谱计算，计算工作量极大，且可能由于共线性的问题导致矩阵运算不稳定，因此实际应用时通常将主成分（PC）与马氏距离法

相结合。首先用 PCA 对校正集样本光谱进行处理，然后再用各光谱得到的主成分阵计算马氏距离。这样既利用了 PCA 对光谱降维处理不丢失信息的特点，又发挥了马氏距离法对离群值有效识别的优点。

9.2 模型的修正

前已提过，近红外光谱分析的基础是数学模型，为提高近红外光谱分析的精确度与准确度，一方面建立模型时，需要通过建模样品集的认真全面挑选、谱区的合理选择以及求导、多元散射校正、傅里叶变换、小波变换、标准正交变量变换等预处理方法的应用来提高模型的稳定性。但另一方面，每个数学模型仍有自己的适配范围，如果待分析样品本身的物化信息超出了模型所能适应的范围（如待测样的蛋白含量超出了模型的蛋白含量范围），就需要对模型进行修正。模型的修正通常是扩充建模样品集所覆盖的（主成分）空间范围，使待测样品的信息能够涵盖在建模样品集内，因此常用的方法是在原有模型中添加几个包含新信息的新样品。此外，人们习惯上将调整原有数学模型的斜率和截距也看作模型的修正，这种方法实际并没有主成分空间的扩展，但按照习惯也将其列为模型修正方法的一种。下面分别对两类方法作具体介绍。

9.2.1 添加新样品

既然新样品包含了原数学模型中没有的新信息，就可以在原有的模型中增加几个包含新信息的样品，即通过丰富建模样品集，扩大主成分空间，使原有模型增添与新样品相适配的新信息，并形成新的模型。修正后的新模型能够覆盖新样品，因而就能准确测定出其待测含量。例：如果用 1998 年 66 个小麦样品建立的蛋白数学模型直接预测 1999 年的 12 个小麦样品，本章 9.1.1 中已做说明，12 个样品的近红外预测值全部报警，模型需要修正。在 1998 年的小麦蛋白模型中加入另外 10 个 1999 年的小麦光谱，并重新经过模型的优化循环得到修正后的模型，再用修正后的模型重新预测 12 个 1999 年的小麦样品，预测结果和马氏距离以及是否有可疑值报警列于表 9-2（模型修正后的预测结果）。为比较模型修正前后的结果，将模型修正前的预测结果也列于表 9-2。

表 9-2 模型修正（添加新样品）前、后预测 12 个 1999 年小麦的结果								
编号	样品光谱名称	化学值	模型修正前的预测结果			模型修正后的预测结果		
			近红外预测值	马氏距离	是否报警（*）	近红外预测值	马氏距离	是否报警（*）
1	99WT01.DX	16.10	17.86	6.28	*	16.18	2.41	无
2	99WT02.DX	17.34	18.41	5.58	*	17.26	1.82	无
3	99WT03.DX	15.39	16.94	3.86	*	15.78	1.76	无
4	99WT04.DX	19.32	19.88	8.35	*	18.87	2.21	无
5	99WT05.DX	14.62	16.48	4.43	*	15.12	1.54	无
6	99WT06.DX	16.06	18.14	6.13	*	16.64	2.36	无
7	99WT07.DX	16.19	17.77	6.99	*	16.32	2.08	无

续表

编号	样品光谱名称	化学值	模型修正前的预测结果			模型修正后的预测结果		
			近红外预测值	马氏距离	是否报警 (*)	近红外预测值	马氏距离	是否报警 (*)
8	99WT08. DX	18. 35	19. 00	5. 56	*	18. 12	1. 39	无
9	99WT09. DX	18. 68	19. 44	6. 36	*	18. 38	1. 42	无
10	99WT10. DX	21. 17	21. 94	9. 59	*	21. 33	2. 04	无
11	99WT11. DX	19. 01	19. 92	7. 98	*	18. 88	1. 84	无
12	99WT12. DX	19. 09	20. 38	6. 70	*	19. 22	1. 71	无

由表 9-2 可见，通过添加新样品信息的方法对模型修正后，12 个 1999 年的小麦样品预测值全部接近真实值，马氏距离全部小于阈值 3，没有一个样品报警，说明该方法对模型的修正是合适的。

9.2.2 斜率/截距修正法

人们习惯上将斜率/截距修正法也看作模型修正的一种方法。

斜率/截距修正的原理是：在欲分析的包含新信息的样品中选择几个样品（样品选择的方法根据第 8 章相关内容进行），其化学值已知，用原有模型预测这几个新样品，并将预测值与真实值用一条直线拟合，记录该直线的斜率和截距，作为原有模型的修正参数。分析未知样品时，将预测数据用上述参数修正后作为其真正的预测值。斜率/截距修正的具体算法在本章 9.4.2.1 模型转移方法中会做详细介绍。

仍用上述实例对该方法做解释说明。1998 年小麦蛋白模型未修正时分析 12 个 1999 年小麦的预测值，以及真实值与预测值的绝对误差列于表 9-3 中，分析误差超过了允许范围。为了对该模型进行修正，选取 8 个 1999 年的小麦样品，用此模型预测其蛋白含量，将其预测值与真实值用一条直线拟合，得到用以修正模型的斜率和截距并保存下来。将 12 个 1999 年小麦用 1998 年模型预测的结果用以上两个参数修正后作为最终结果，见表 9-3，表中同时还列出了它与真实值的绝对误差。

表 9-3 模型修正（斜率/截距法）前、后预测 12 个 1999 年小麦的结果

编 号	样品光谱名称	化 学 值	模型修正前的预测结果		模型修正后的预测结果	
			近红外预测值	绝对误差	近红外预测值	绝对误差
1	99WT01. DX	16. 10	17. 86	－1. 76	16. 59	－0. 49
2	99WT02. DX	17. 34	18. 41	－1. 07	17. 22	0. 12
3	99WT03. DX	15. 39	16. 94	－1. 55	15. 53	－0. 14
4	99WT04. DX	19. 32	19. 88	－0. 56	18. 89	0. 43
5	99WT05. DX	14. 62	16. 48	－1. 86	15. 01	－0. 39
6	99WT06. DX	16. 06	18. 14	－2. 08	16. 91	－0. 85
7	99WT07. DX	16. 19	17. 77	－1. 58	16. 49	－0. 30
8	99WT08. DX	18. 35	19. 00	－0. 65	17. 89	0. 46

续表

编 号	样品光谱名称	化 学 值	模型修正前的预测结果		模型修正后的预测结果	
			近红外预测值	绝对误差	近红外预测值	绝对误差
9	99WT09. DX	18. 68	19. 44	－0. 76	18. 39	0. 29
10	99WT10. DX	21. 17	21. 94	－0. 77	21. 24	－0. 07
11	99WT11. DX	19. 01	19. 92	－0. 91	18. 93	0. 08
12	99WT12. DX	19. 09	20. 38	－1. 29	19. 46	－0. 37
平均绝对误差(误差取绝对值)			1. 24		0. 29	

由表 9-3 可见，12 个 1999 年的小麦，直接用 1998 年小麦建立的模型预测平均绝对误差（误差取绝对值）为 1. 24，而用斜率/截距法对模型进行修正后再来预测这 12 个样品，平均绝对误差降为 0. 29，可见预测效果有明显提高。但这种模型修正方法的效果没有 9. 2. 1 的方法好，因为这种方法并没有在原模型中增添新的信息，这从模型修正后近红外预测值与化学值之间的绝对误差也可以看出，9. 2. 1 的方法对模型修正后，绝对误差要小于该方法的绝对误差。

一般情况下，若待测样品本身的物化信息超出了模型范围，单纯通过斜率/截距法进行模型的修正效果不会很理想，需要通过添加新样品的方法才合适。

9. 3 模型转移

9. 3. 1 概述

近红外数学模型包含了样品状态、装样条件、仪器性能和分析时间等的信息。在一定时间、一定条件下，数学模型受仪器性能变化、分析时间、条件或样品状态等差别的影响较小，通过光谱的预处理方法等可以消除部分干扰因素的影响。但随着时间、空间、条件的变化，这些干扰因素不断增加，模型不能适应新的变化，预测未知样品的能力会下降。

虽然近红外仪器厂商一直致力于完善仪器加工工艺，提高仪器加工的标准化，力求仪器的稳定性和重现性。但因为每一种近红外数学模型只能适应一定的样品、时间与空间范围，在运用 NIR 定量分析数学模型的过程中，仍可能会遇到各种各样的问题，使得不能直接运用原来的数学模型预测分析未知样品的某成分含量，否则，就可能产生错误的预测结果。这几种情况是：

① 假设模型建立在“主仪器”（Master Instrument）上，而现在拥有的是另外一台仪器，称为“从仪器”（Slave Instrument）。因为是两台不同的仪器，所得到的光谱即便形状相似，但光谱强度一般会发生变化，即从“从仪器”与“主仪器”上得到的光谱不同，如果要用“主仪器”上已有的模型来预测“从仪器”上得到的光谱，则预测结果通常是不理想甚至是错误的。即便是同一台仪器，若构成其系统的元器件更换（如光源、光学系统及检测器等），或者元器件随着时间老化，则其响应也必定会发生或多或少的变化。那么，用原来的模型预测仪器变化后采集的光谱，也可能导致预测结果的不理想。

② 样品的某些性质如样品的黏度、颗粒大小、表面的质地等发生变化，会使原来的

模型不适合预测新样品。通常不同批次的样品或处于不同时间、不同背景条件下的样品会有这种情况发生。

③ 仪器的工作环境如温度或湿度等发生变化，会使光谱发生漂移、谱峰展宽和吸光度非线性变化等现象，也使得原有模型不适合分析新样品。此外，不同操作者、不同操作条件下得到的样品光谱，通常也需要首先检验其是否与原有模型相适配，如果模型不适配，则需要对模型进行维护。

在近红外定量分析过程中遇到的以上几种情况，虽然产生变异的来源不同，但结果却是相同的，即都使得已经建立的数学模型不能直接预测未知样品，否则就可能产生不理想甚至错误的结果。解决这一问题最直接的方法就是在新的仪器、新的环境、新的测试条件下重新扫描全部建模样品，并重新建立数学模型。可是这也一定不是最好的方法，因为如果建模样品数很多，或者其物化性质不稳定，又或者样品有毒，那么重扫样品并重建模型就是一种很不经济也很不实用的方法。我们转向另外一种解决这一问题的方法，这就是数学模型的转移 (Calibration Transfer)，也称为仪器的标准化 (Instrument Standardization)。对以上提到的三种情况，虽然变异可能并不一定发生在两台仪器间，我们仍把建立模型所对应的仪器、工作环境、测量条件等称为“主仪器”，分析新样品时所对应的仪器、工作环境、测量条件等称为“从仪器”，仪器的标准化或者说模型转移就是在“主仪器”和“从仪器”间进行的。

9.3.2 仪器间标准化的途径

对于同一样品，不同仪器得到的近红外光谱不尽相同，来自于不同仪器生产厂家的仪器如此，即使相同厂家生产的同一型号不同批次间的仪器也是如此，只是光谱的差异较小。仪器的标准化就是为了避免重建模型而用数学的手段，尽量消除近红外仪器间对同一样品获得不同光谱和不同分析结果的差异。

实现近红外光谱分析标准化有两条途径：①调整已经建立模型的斜率和截距，使不同仪器对同一样品的分析获得相同结果。这是一种简单的标准化处理，这种技术仅适应于比较相似的仪器之间，所达到的仪器间标准化程度较低。②通过数学方法，减小因仪器的波长准确性、带宽、工作环境、操作条件等给“从仪器”上光谱带来的影响，使不同仪器对于同一样品，尽可能获得相同的光谱信号。这种标准化技术本质上是将“从仪器”上的光谱标准化成与“主仪器”上光谱一致的光谱，仪器间经过这种处理后，无需再对定标模型的斜率和截距进行调整。在进行这种标准化处理时，标准化样品分别在主、从仪器上扫描。如果所有标样光谱在主、从仪器间差异的形式一致，则可用单个标样进行标准化处理 (Single Sample Standardization, SSS)；如果差异的形式不一致，则需要多个标样进行标准化处理 (Multiple Sample Standardization, MSS)。其中途径①在 9.4.2.1 有详细描述，下面着重介绍途径②时仪器的标准化。

9.3.2.1 单个样品的标准化

单个样品的标准化 (Single Sample Standardization, SSS) 是用单个标准化样品在仪器间做偏移量的校正。如果仪器间的差别主要是偏移量的不同，而其他类型的差别相对较小时，选用 SSS 法进行仪器的标准化是恰当的。如果仪器间的差别不以偏移量为主，其他类型的差别也占有相当大的比重，那么，SSS 校正就是不恰当的。根据经验，标准化时

如果所用的标准化样品不是实测样品，而是纯水（无论是固体还是液体）或任何其他惰性的或人工的样品，则仪器经标准化后，结果并不令人满意。而用单一某实测样品作为标准样品对仪器进行标准化处理，对解决大多数 NIR 分析问题是恰当的，分析结果也是可靠的。

单个标准化样品的挑选最好是在建模样品集中挑选出与平均光谱接近的样品。这可以通过对建模样品集内光谱的主成分分析、比较光谱的得分阵来完成。例如，如果想用一个样品作为所有谷物或所有饲料样品的标准化样品，那么该标准化样品的选择就应该是从所有谷物或所用饲料 NIR 光谱库中挑选出来的。如果建模样品库中所拥有的 NIR 光谱量足够多，而挑选出来的样品 NIR 光谱就是所有光谱中的平均光谱，这是最完美的。通过该标准化样品对仪器的标准化，在“从仪器”上预测样品时预测误差和系统误差都应该较小。在确定一个标准化样品前应进行检验，以确认用这个标准样品对仪器进行标准化处理后标准化结果是可以被接受的。当作为标准化的样品确定后，应将其放置于密封的进样槽内，防止其变质，使其可以长期保存。

当前，利用单一标准化样品对近红外光谱分析网络系统中各近红外仪器进行标准化的有澳大利亚谷粒工业中心的 ToroTurfgrass 近红外光谱分析网络和北美 ISI 的饲料网络等。表 9-4 是用 SSS 法对 16 台近红外仪器进行标准化的结果，标准化样品为实测青贮玉米，用标准化后的仪器预测分析玉米样品中的不同成分，每台仪器的分析结果以及 16 台仪器间结果的标准差列于表 9-4 中。

表 9-4 16 台仪器以青贮玉米为标样进行 SSS 标准化后对玉米几种主要成分的预测结果

仪 器	粗蛋白(CP)	酸性洗涤纤维(ADF)	干物质(DM)	中性洗涤纤维(NDF)
1	13.093	20.032	90.285	44.644
2	12.958	20.114	90.295	44.717
3	13.014	19.971	90.383	44.585
4	13.102	20.054	90.254	44.595
5	13.048	20.058	90.208	44.564
6	13.080	19.986	90.266	44.501
7	13.114	19.998	90.303	44.515
8	13.064	20.034	90.245	44.601
9	13.096	19.993	90.293	44.468
10	13.169	20.009	90.160	44.504
11	13.026	20.102	90.289	44.686
12	13.058	20.103	90.197	44.724
13	13.061	19.991	90.272	44.484
14	13.154	19.970	90.268	44.503
15	13.043	19.854	90.507	44.505
16	12.878	20.281	90.228	44.033
标准差 SD	0.071	0.092	0.079	0.142

由表可见，16 台仪器经过 SSS 标准化后分析同一样品，分析结果之间的标准差很小，说明仪器经标准化后非常相似，达到了标准化的目的。

9.3.2.2 多样品的标准化

用多个标准化样品对仪器的标准化（Multiple Sample Standardization, MSS）原理与单一样品标准化的原理相同。因为近红外光谱分析仪型号种类越来越多，不同制造厂生产的仪器不一样，甚至同一制造厂生产的仪器也不一样。对于所有这些仪器，除了要进行吸光度的校正之外，还要进行波长校正。仅靠仪器制造厂对波长的校正是不充分的，因此，有必要用外部样品对仪器间波长和吸光度一起校正。要完成这两个任务，需要以多个样品完成影响近红外数学模型各参数的标准化。

与 SSS 一样，MSS 标准化中标准化样品通常不是经典光谱分析意义上的标准品，而是实际分析中的样品如玉米、大豆等，只不过是通过主成分或其他方法挑选出来的。标准化样品可能包括许多类型，每个类型代表要分析的一类，这些样品都要放在被密封的进样槽内。MSS 可以对于仪器间的线性变化进行标准化处理。当仪器间波长变异属于非线性模型时，也可以用 MSS 进行标准化处理。对于一些含水量非常高的样品，由于其表面反射率很高，NIR 光谱峰谷多、变化复杂，因此，仪器间的变异往往属于非线性模型，应该用多个样品对仪器进行标准化处理。

表 9-5 是用 MSS 法对仪器进行标准化处理前后预测分析相同样品标准差的变化情况。由表可见，仪器标准化前预测分析四类（玉米、干草、粗饲料、青贮玉米）各 8 个样品四种成分（粗蛋白、酸性洗涤纤维、干物质、中性洗涤纤维）的总体平均标准差为 1.17。经标准化 1 后，总体平均标准差变为 0.56，经标准化 2 后，总体平均标准差变为 0.27，其中标准化 1 和标准化 2 都采用 MSS 法，但标准化样品并不是实测样品。而用实测青贮玉米作为标准化样品对仪器进行标准化后预测样品的总体平均标准差变为 0.14，可见用实测样品作为标准化样品进行 MSS 效果是最好的。

表 9-5 MSS 标准化处理前后分析 8 个样品的标准差的变化				
	粗蛋白(CP)	酸性洗涤纤维(ADF)	干物质(DM)	中性洗涤纤维(NDF)
标准化前				
玉米	0.975	0.474	0.351	1.678
干草	1.481	1.144	0.322	2.564
粗饲料	1.371	1.173	0.803	1.351
青贮玉米	0.695	0.316	0.488	1.139
总体平均标准差:1.17				
标准化 1 后				
玉米	0.346	1.058	0.184	1.106
干草	0.320	0.417	0.229	0.904
粗饲料	0.290	0.535	0.313	0.638
青贮玉米	0.293	0.300	0.356	0.471
总体平均标准差:0.56				
标准化 2 后				
玉米	0.226	0.249	0.047	0.282
干草	0.272	0.259	0.039	0.392
粗饲料	0.237	0.317	0.087	0.309
青贮玉米	0.246	0.200	0.113	0.382

续表				
	粗蛋白(CP)	酸性洗涤纤维(ADF)	干物质(DM)	中性洗涤纤维(NDF)
总体平均标准差:0.27				
以青贮玉米为标样进行标准化后				
玉米	0.073	0.147	0.038	0.131
干草	0.118	0.190	0.045	0.292
粗饲料	0.098	0.188	0.052	0.208
青贮玉米	0.071	0.092	0.079	0.142
总体平均标准差:0.14				

只有在仪器间非常相似的情况下，单个标准化样品的方法才是可行的。一般情况标准化样品都会多于一个，即都选择用 MSS 法进行仪器的标准化。以下介绍模型转移的数学实现时标准化样品也都是选择了多个。

9.3.2.3 透射光谱的标准化

上面谈到的 SSS 和 MSS 方法一般是针对近红外漫反射光谱仪器间的标准化，近红外透射仪器间的标准化方法有所不同。透射仪器间的误差主要是由于带宽的变化引起的光谱基线漂移和换算系数不同所导致的，对于近红外透射仪器，波长位移的校正并不是最重要的。标准化样品一般使用固体的整粒样品，例如整粒小麦。在透射模式中，由于仪器设计方面的限制，将标准样品密闭是不可能的。为解决这个问题，一般是将标准样品在扫描时重复 3 次，取其平均值。一般需要 3~10 个样品来进行仪器间的偏差和斜率校正。在不同仪器间扫描这些样品时，最好使用相同的样品和光程。

需要注意的是：仪器间的标准化不仅要使得用同一个定标模型对同一个样品在不同仪器上测定的结果是相同或近似的；而且仪器间对奇异样品的敏感性也应该是一致或近似的，即对于相同样品，不同仪器间应获得相同或近似的光谱，这是实现近红外数学模型在仪器间转移的基础。如果仪器进行了标准化后，对相同样品的预测误差不一致，通常将这些样品保留下来，增加到建模样品集中，重新进入模型的优化循环，进行模型的修正。如果仪器间对相同样品的预测误差一致或近似，但是 NIR 光谱仍有一定差别，仪器间对奇异样品的敏感性也有一定误差，解决这个问题通常的方法也是将这些 NIR 光谱有差异的样品保留下来，增加到建模样品集中进行模型修正，这样将减少模型对未知样品作为奇异点的敏感性。

9.4 模型转移的数学实现

9.4.1 概述

模型转移的数学实现是指通过算法来校正光谱的失真，或修正模型使其适用于变化了的图谱。转移的方法有两类：一类是直接修正已经建立的数学模型，使它适于预测未知样品的图谱。另一类是校正“从仪器”上得到的光谱，使之能够利用原来的模型进行预测。第一类方法有斜率/截距校正法（Slope/Bias Correction Method）、传统数学模型标准化法（Standardization with the Classical Calibration Model）、反向数学模型标准化法（Stand-

ardization with the Inverse Calibration Model) 等。第二类方法是一种更广泛的模型转移的方法，它是将“从仪器”上获得的光谱转移到“主仪器”上，使得这些光谱就好像是在“主仪器”扫描的一样，应用“主仪器”上已建立的模型预测这些修正后的光谱，即可得到未知样品某成分的含量。这一类方法是通过转移使两台仪器的响应 (Response) 相匹配，具体的方法有：Shenk 算法、PDS 法和 FIR 法等。

国外从 20 世纪 80 年代起，就开始有关于近红外数学模型转移的研究报道。为了利用面粉中蛋白质和水分的数学模型，Osborne 和 Fearn 第一次提出了一种模型转移的方法，这是一种最容易想到的方法。“主仪器”上已建有蛋白质和水分的一元或多元线性回归模型，20 个面粉样品在 9 台不同的“从仪器”上扫描。当“从仪器”上获得的光谱用“主仪器”上建立的模型预测时，为了校正预测值，用一斜率截距校正因子来修正模型。Jones 等人也运用了相同的方法转移了测定冷冻注射产品中水含量的数学模型。而 E. Bouversse、D. L. Massart 则详尽的报道了用 Slope/Bias 法进行的数学模型转移，报道中共用了三类不同的样品集，第一类为 60 个大豆种子，样品分别在两台不同的 NIR 仪器上扫描，扫描区间为 1100~2500nm，建立第一台仪器上大豆种子中水、蛋白质和油分三个参数的模型，用 Slope/Bias 法成功的实现了两台仪器间数学模型转移。第二类样品为 30 个汽油，也分别在两台仪器上采集 800~1600nm 范围内的透射光谱；第三类样品为 39 个不同批号的药片，在 1100~2500nm 扫描其反射光谱；同样，用 Slope/Bias 法成功的实现了两台仪器间数学模型的转移。

为了处理一些复杂的情形，相继有许多文献描述了另外一类模型转移的方法。1989 年，J. S. Shenk 提出了模型转移的一种专利算法，该算法将“从仪器”上的光谱通过波长校正和吸光度校正两步转移成与“主仪器”上光谱相一致的光谱。E. Bouversse 等用 Shenk 专利算法将 NIRSystem 6500 上 482 个玉米样品建立的测定蛋白质含量的数学模型转移到 NIRsystem 5000 上，模型校正后，预测标准差从校正前的 0.40 减小到 0.12。此外，Wang Yongdong 等人提出了 PDS (Piecewise Direct Standardization) 的模型转移方法，为了求得两台仪器间的转移参数，它也需要从建模样品中选择一些有代表性的标准化样品 (Standardization Samples) 分别在“主、从”两台仪器上扫描，但对波长的校正是在较小的窗口内进行的。不同温度对 NIR 光谱是有影响的，Su-chin Lo 等人将 23.0℃ 下 NaCl 水溶液的模型用 PDS 法转移到 28.5℃ 下，光谱标准化后其预测标准差 (SEP) 为 0.008M，若模型不校正，则 SEP 为 0.048M。可见模型经 PDS 法校正后，提高了预测的准确度。为了消除杂散光 (Stray Light) 等对光谱的影响，Kowalski 等人在 PDS 的基础上提出了附加背景校正法 (Additive Background Correction)，称为 PDSB，实验结果表明 PDSB 法比 PDS 法在预测效果上有一定水平的提高。以上几种模型转移的方法都需要部分标准化样品分别在“主、从”2 台仪器上同时测量，Steven D. Brown 等人则提出了另外一种无需标准样品的光谱校正法：FIR 法 (Finite Impulse Response Filtering)。

下面着重介绍几种常用的实现近红外数学模型转移的方法原理，并以实例验证。

9.4.2 模型转移的数学实现

先做以下符号约定：

- 大写黑斜体字母表示矩阵

- 小写黑斜体字母表示向量
- 普通字母表示标量
- $^m A$ 表示“主仪器”上校正样品集的吸光度矩阵
- $^s A$ 表示“从仪器”上校正样品集的吸光度矩阵
- $^m T$ 表示“主仪器”上标准化样品集的吸光度矩阵（由校正样品集选则得到）
- $^s T$ 表示“从仪器”上标准化样品集的吸光度矩阵（由校正样品集选则得到）
- $^s X$ 表示“从仪器”上检验样品集的吸光度矩阵
- y 表示校正样品集的浓度向量
- y_{un} 表示预测得到的检验样品集的浓度向量
- X_{ji} 表示矩阵第 j 行第 i 列的元素（对应于第 j 个样品第 i 个波长处的吸光度值）
- N_w ：表示波长点数
- N_s ：表示校正集样品数
- N_t ：表示标准化集样品数

因本节介绍的几种模型转移方法中部分实例采用中国农业大学赵龙莲 1999 年硕士论文中的数据，为避免重复，故先对其中仪器设备、实验材料与方法做总体介绍如下：

模型转移是在三台不同厂家生产的傅里叶变换近红外仪器上进行的，三台仪器分别是：Bruker Vector 22/N 傅里叶变换近红外光谱仪，由德国 Bruker 公司生产；Bio-Rad FTS165 傅里叶变换近红外光谱仪，由美国 Bio-Rad 公司生产；WQF-400N 傅里叶变换近红外光谱仪，由中国北京第二光学仪器厂生产。

样品为小麦样品 66 个，粉末状（过 60 目筛），并由化学法测定蛋白含量。分别用上述三种傅里叶变换光谱仪采集小麦样品的近红外漫反射光谱，扫描的参数皆为：

扫描区间：4000~9000 cm^{-1}

扫描分辨率：8 cm^{-1}

数据点数：1297

扫描次数：64 次取平均

共得四套图谱集：

第 1 套：1998 年 11 月 19 日，用 Bruker Vector 22/N 傅里叶变换近红外光谱仪扫描，样品池为积分球，得到的光谱简称光谱集“Bruker1”，此光谱集用于建立数学模型，它对应的仪器及相应条件称为“主仪器”。

第 2 套：1999 年 5 月 9 日，用 Bruker Vector 22/N 傅里叶变换近红外光谱仪扫描，样品池为积分球，得到的光谱简称光谱集“Bruker2”。

第 3 套：1998 年 10 月 3 日，用 Bio-Rad FTS165 傅里叶变换近红外光谱仪扫描，样品池为普通漫反射样品池，得到的光谱简称光谱集“Bio-Rad”。

第 4 套：1999 年 5 月 9 日，用 WQF-400N 傅里叶变换近红外光谱仪扫描，样品池为积分球，得到的光谱简称光谱集“Berg”。

第 2、3、4 套光谱集所对应的仪器及相应条件称为“从仪器”。4 套图谱集分别见图 9-3。由图可见，同一组样品在不同仪器和不同测试条件下，得到的图谱有较大差别。

为实现数学模型转移需要有三组样品集：校正样品集，标准化样品集和检验样品集。校正样品集用于在“主仪器”上建立模型，它的数目可能很大。标准化样品集用于计算模

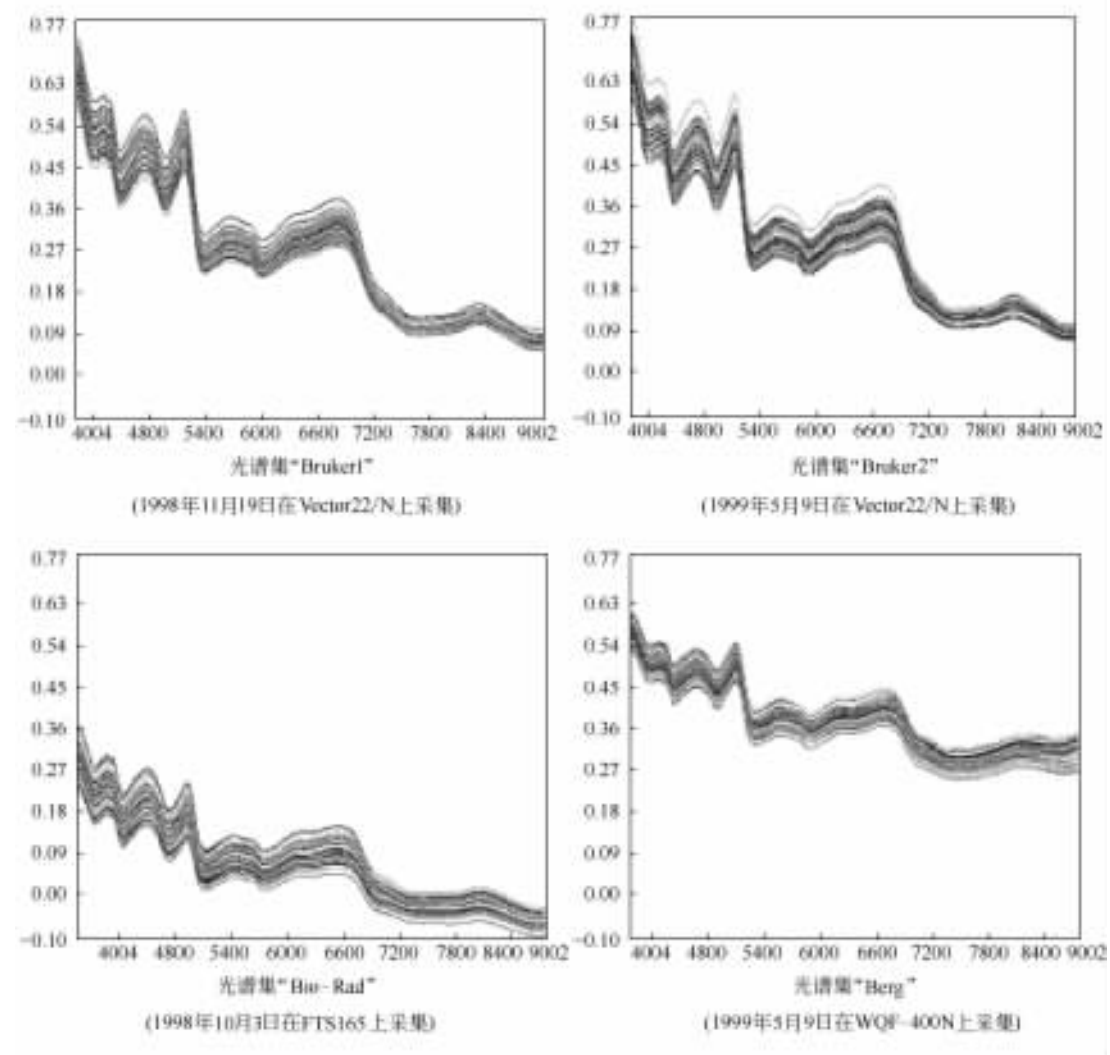


图 9-3 不同仪器、不同时间采集的 4 套小麦样品图谱集

型转移参数，是通过一定的方法由校正样品集中选择出来的，需分别在“主、从仪器”上扫描其光谱。检验样品集用于检验模型转移的效果。实例中 66 个小麦样品通过聚类法分成两部分：44 个校正样品集和 22 个检验样品集，并将全部校正集样品选为标准化样品集，当然，也可以从校正样品集中选择部分样品作为标准化样品集。

应用偏最小二乘法（PLS）对光谱集“Bruker1”中的 44 个校正样品建立数学模型，模型结果见表 9-6，这一模型就是后面模型转移的基准模型，称之为“主模型”。利用此“主模型”分别预测 4 套图谱集中的 22 个检验样品集，预测结果见表 9-7。

表 9-6 “主仪器”上光谱集“Bruker1”的建模结果

成分	维数	校正样品集(44 个)		
		R^2	SEC	RSD%
蛋白质	4	98.31	0.24	1.51

表 9-7 用“主仪器”上模型分别预测分析 4 套图谱集中 22 个检验集样品的结果

	样品集“Bruker1”	样品集“Bruker2”	样品集“Bio-rad”	样品集“Berg”
R^2	97.53	31.02	1.95	-158.66
SEP	0.29	1.51	1.80	2.92
RSD%	1.85	9.78	11.66	18.94

由表 9-7 可见，利用“主仪器”上建立的模型预测“主仪器”上光谱集“Bruker1”中的 22 个检验样品， R^2 为 97.53，SEP 为 0.29，与模型的结果近似；但如果直接用此模型分析光谱集“Bruker2”、“Bio-rad”和“Berg”中的 22 个检验样品， R^2 分别降为 31.02、1.95 和 -158.66，SEP 则分别变为 1.51、1.80 和 2.92。分析原因如下：光谱集“Bruker2”与光谱集“Bruker1”相比，二者虽是在同一台仪器（皆为 Bruker Vector 22/N）上测量的，但前者的测量时间是 1999 年 5 月 9 日，而后者的测量时间为 1998 年 11 月 19 日，两次测量的环境不同（如温度、湿度等），而且在这么长的一段时间内，仪器的整机状态也可能会发生或多或少的变化（如信号能量的变化等），这些变化都会对图谱产生影响，从而导致用“主模型”预测光谱集“Bruker2”中的检验样品结果不理想。光谱集“Bio-Rad”和“Berg”与“主仪器”上的“Bruker1”相比，光谱是在不同厂家生产的近红外光谱仪上得到的，图谱之间差异较大，因此直接用“主仪器”上的模型预测分析结果也不理想。由上可见，不能直接用“主仪器”上的模型预测分析“从仪器”上的图谱，模型必须经过维护或者“从仪器”上的图谱必须经过校正后才能得到正确的预测结果，即实现模型转移，才可以实现不同仪器间的资源共享。

下面具体介绍常用的几种模型转移的数学方法，并以上述 3 套“从仪器”光谱集为例验证不同方法模型转移的效果。

9.4.2.1 斜率/截距校正法 (Slope/Bias Correction Method)

这一方法的基本思想是：“主仪器”上用 N_s 个样品建立模型，这 N_s 个建模样品的光谱在“主仪器”上采集，产生响应矩阵^mA，浓度向量 y 通过一参考方法得到。^mA 和 y 的回归系数矩阵为 b 。分别在“主、从仪器”上扫描 N_t 个标准化样品（Standardization Samples）的光谱，用^mT、^sT 表示，则

$$y^{(mt)} = {}^mT \times b$$
$$y^{(st)} = {}^sT \times b$$

其中， $y^{(mt)}$ 是 N_t 个标准化样品由光谱矩阵和“主仪器”上建立的模型相乘得到的浓度值。 $y^{(st)}$ 是同一组标准化样品由“从仪器”上的光谱与回归系数矩阵 b 相乘得到的浓度值，并未将两台仪器的差别考虑在内。

用一元线性回归方程来拟合 $y^{(mt)}$ 和 $y^{(st)}$ ，求得此线性方程的普通最小二乘解（Ordinary Least Squares），即为这一线性模型的斜率和截距：

$$\text{slope} = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4(S_{xy})^2}}{2S_{xy}}$$
$$\text{bias} = \bar{y}^{(mT)} - \text{slope} \times \bar{y}^{(sT)}$$

其中：

$$S_{xx} = \sum (y_j^{(sT)} - \overline{y}^{(sT)})^2$$

$$S_{yy} = \sum (y_j^{(mT)} - \overline{y}^{(mT)})^2$$

$$S_{xy} = \sum (y^{(sT)} - \overline{y}^{(sT)}) (y^{(mT)} - \overline{y}^{(mT)})$$

有了回归系数矩阵 b 和 Slope/Bias 校正项，我们就可以直接由“从仪器”上得到的光谱计算其浓度值。设“从仪器”上得到的光谱矩阵为 ${}^sX$ ，则它的浓度矩阵为

$$y^{(sX)} \text{ std} = \text{bias} + \text{slope} \times ({}^sX \times b)$$

此过程可用图 9-4 示意。

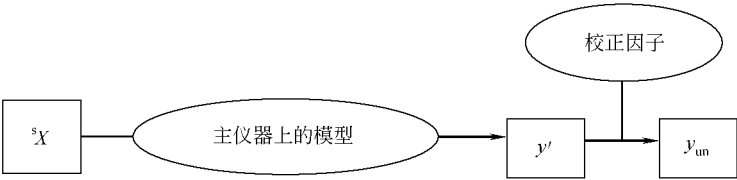


图 9-4 斜率/截距校正法示意图

参考文献 [34] 利用 Slope/Bias 法将“主仪器”上的小麦蛋白模型转移给“从仪器”。方法如下：用一元线性回归方程拟合上述提到的 44 个标准化样品在“主仪器”和“从仪器”上的预测分析结果，拟合得到的斜率和截距作为 Slope/Bias 法修正“主仪器”上模型的校正因子。利用修正后的模型预测分析 22 个检验集样品，该预测结果作为模型转移后的结果见表 9-8。为了比较 Slope/Bias 法模型转移前后的预测效果，模型转移前直接用“主模型”预测分析同一批样品的结果也列于表 9-8 中。

表 9-8 数学模型转移前、后 (Slope/Bias 法) 预测分析 3 套检验集样品的结果比较

	“主仪器”上的 预测结果	光谱集“Bruker2”		光谱集“Bio-rad”		光谱集“Berg”	
		转移前	转移后	转移前	转移后	转移前	转移后
$R^2/\%$	97.53	31.02	97.55	1.95	95.10	-158.66	94.95
SEP	0.29	1.51	0.28	1.80	0.40	2.92	0.41
RSD%	1.85	9.78	1.85	11.66	2.61	18.94	2.65

由表 9-8 可见，模型经 Slope/Bias 法转移后，“Bruker2”中检验样品集的决定系数 (R^2) 由为 31.02 变为 97.55，预测标准差 (SEP) 由 1.51 变为 0.28，这与“主仪器”上的分析结果基本一致，说明模型转移是成功的。同样，光谱集“Bio-Rad”的 R^2 由转移前的 1.95 变为 95.10，预测标准差 (SEP) 由 1.80 变为 0.40；光谱集“Berg”的 R^2 由转移前的 -158.66 变为 94.95，预测标准差 (SEP) 由 2.92 变为 0.41。在 $\alpha = 0.05$ 水平上，对模型转移后的数据与“主仪器”上的分析结果进行 F 检验，两者差异皆不显著，说明 Slope/Bias 算法能够将“主仪器”上的模型转移到以上三种不同的“从仪器”上，实现了不同仪器间的标准化。

9.4.2.2 Shenk 算法

Shenk 算法 (Shenk’s Algorithm) 是一种一元全光谱标准化方法，它是通过对“从

仪器”上图谱每一光谱点的校正来实现的。

用 Shenk 算法进行仪器标准化（模型转移）共分两步。第一步计算转移参数，第二步利用得到的转移参数对未知光谱进行标准化。

假设用于计算转移参数的标准化样品在“主、从仪器”上的吸光度矩阵分别记为： ${}^mT, {}^sT$ ，则具体算法如下：

第一步：计算转移参数

计算波长校正系数：

计算波长校正系数又分三步进行：

① 首先，为了找出“从仪器”上与“主仪器”第 i 个波长点相关最大的值，也就是为了找出“主仪器”第 i 个波长在“从仪器”上漂到了哪一个波长点，我们开设一个小窗口（ w ），假设“从仪器”第 i 个波长点漂移值在相对于“主仪器”第 i 个波长点 $\pm w$ （ w 为正整数）内，分别计算 T_{mi} 与 T_{sk} （ $k=i-w \sim i+w$ ）的相关系数 R ，如窗口大小为 w ，则可计算出 $2 \cdot w + 1$ 个相关系数，找到其中最大的相关系数所对应的波长 p ，就找到了“从仪器”上与“主仪器”相关最大的波长点。

② 为了更精确地估计出“从仪器”上与“主仪器”第 i 个波长点最相关的波长点 i' ，我们令 x 值为波长点， y 值为相关系数，用相关系数最大的波长点 p 及两个临近波长点 $p-1, p+1$ 拟合一个二次模型

$$\text{Correlation} = a + bi + ci^2 \quad (y = a + bx + cx^2)$$

由此等式可得到相关系数最大处的波长。

③ “主仪器”每个波长 i 都可计算出一个二次模型，由此模型就可计算出相关系数最大的精确波长值 i' ，对“主仪器”原波长 i 及“从仪器”上校正后的波长 i' 再做一次拟合

$$i' = A + Bi + Ci^2$$

（其中 A, B, C 就是所要计算的波长校正系数）

移动窗口，计算所有波长点的波长校正系数。

计算吸光度校正系数：

其过程也分三小步：

① 对测量的“从仪器”光谱，用第二步中计算出的校正波长 i' 进行内插（也就是计算出“从仪器”上第 j 个样品第 i' 个波长处的吸光度值）

$$\text{for}(i=1 \text{ to } N_w)$$

$$T_{si\#} = T_{si} \quad (T_{si\#} \text{ 表示 } T_{si} \text{ 内插后的结果})$$

② 用校正后的“从仪器”光谱的第 i 列 $T_{si\#}$ ，以及“主仪器”光谱 T_{mi} 相应的第 i 列进行线性拟合，计算出吸光度校正系数。

$$\text{for}(i=1 \text{ to } N_w)$$

$$T_{mi} = a[i] + b[i] \times T_{si\#}$$

（对于每个波长 i 都要计算 a, b ）

③ 再用 $X_{s\#}$ 及 a 、 b ，计算出标准化后的光谱 X_{std}

$$\text{for}(i=1\text{to } N_w)$$

$$X_{stdi} = a[i] + b[i] \times X_{si\#}$$

保存转移参数：

将波长校正系数 A 、 B 、 C 及吸光度校正系数 $a[i]$ 、 $b[i]$ ($i=1\text{To } N_w$) 保存在标准化文件中。

第二步：利用转移参数对未知光谱进行标准化

① 读入标准化参数及“从仪器”上的原始光谱数据 X_s ；

② 利用等式 (2-5) 计算出校正后的“从仪器”的波长 i' ；

③ 用光谱矩阵 X_s 及校正后的波长 i' 进行内插，得到内插后的光谱矩阵 $X_{s\#}$ ；

④ 从波长点 1 到波长点 N_w 用 $X_{s\#}$ 及 $a[i]$ 、 $b[i]$ ($i=1\text{To } N_w$)，计算出标准矩阵 X_{std} 。此过程用图 9-5 表示。

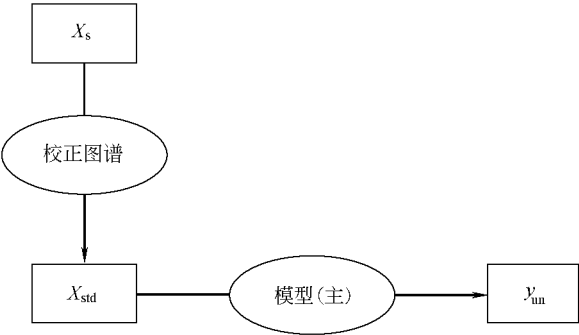


图 9-5 Shenk 算法示意图

参考文献 [34] 根据 44 个标准化样品在“主仪器”和“从仪器”上的光谱，用“Shenk”算法计算出“主、从仪器”间光谱的校正参数，并用得到的校正参数分别修正从仪器光谱集“Bruker2”、“Bio-Rad”和“Berg”中的 22 条检验集光谱，使其与“主仪器”上的光谱匹配，然后用“主仪器”上的模型预测分析修正后的光谱，结果见表 9-9。

表 9-9 模型转移前、后（“Shenk 算法”）预测分析 3 套检验集样品的结果比较

	“主仪器”上的 预测结果	光谱集“Bruker2”		光谱集“Bio-rad”		光谱集“Berg”	
		转移前	转移后	转移前	转移后	转移前	转移后
R^2	97.53	31.02	98.16	1.95	91.19	-158.66	90.94
SEP	0.29	1.51	0.28	1.80	0.54	2.92	0.55
RSD%	1.85	9.78	1.60	11.66	3.49	18.94	3.54

由表 9-9 可见，模型转移后，3 套检验集样品的决定系数 (R^2)、预测标准差 (SEP) 和平均相对误差 (RSD) 均优于模型转移前的结果，也接近“主仪器”上的分析结果。在 $\alpha=0.05$ 水平上，对模型转移后的数据与“主仪器”上的分析结果进行 F 检验，差异皆不显著，说明“Shenk”算法能够将“从仪器”上的光谱修正成适应于“主仪器”上模型的光谱，“主、从仪器”对同一样品的响应接近一致，实现了不同仪器间的标准化。

9.4.2.3 PDS 算法

PDS 算法 (Piecewise Direct Standardization) 是一种多元全光谱标准化方法。它的基本思想是：“主仪器”上的光谱矩阵 T_m ，“从仪器”上的光谱矩阵 T_s ，二者因为在不同的条件下测定，所以存在着差异，通过一转移矩阵 F 将 T_s 转换成 T_m ，使二者相匹配。其中 F 描述了“主、从”两种测定条件下的差异。

$$T_m = T_s F + E \text{ (} E \text{ 为残差矩阵), 则}$$

$$F = T_s^{-} T_m, \text{ 其中 } T_s^{-} \text{ 为 } T_s \text{ 的广义逆}$$

考虑到这样一个事实：对于实际的光谱数据，波长点的漂移通常只局限于一个小的区域。因此，“主仪器”上的每个波长点只是与“从仪器”上相应波长点附近的波长关联大，而并非与全谱区的光谱点都关联。设这个小的区域为一窗口 Z_i ，为了重构“主仪器”上的每一点，令 Z_i 表示“从仪器”上从 $i-k$ 到 $i+k$ 共 $2k+1$ 个点的吸光度矩阵。

$$Z_i = [T_{s,i-k}, T_{s,i-k+1}, \cdots, T_{s,i+k-1}, T_{s,i+k}]$$

对“主仪器”上的每一向量 X_i ，构造一如下多元回归模型：

$$X_{m,i} = Z_i b_i + e_i$$

此方程由 PLS 方法求解。将所有的回归系数 b_i 置于 F 的主对角线上，并将其他元素置为 0。这样可得到一带状矩阵 F

$$F = \text{diag}(b_1^T, b_2^T, \cdots, b_i^T, \cdots, b_p^T)$$

有了转移矩阵 F ，就可以将“从仪器”得到的未知样品的光谱 X_{su} 转化成与“主仪器”光谱相匹配的光谱 X_{st} 了。用公式表述为

$$X_{st} = F X_{su}$$

运用“主仪器”上的模型预测校正后的光谱即可得到未知样品某成分的含量了。这一过程用图 9-6 表示为：

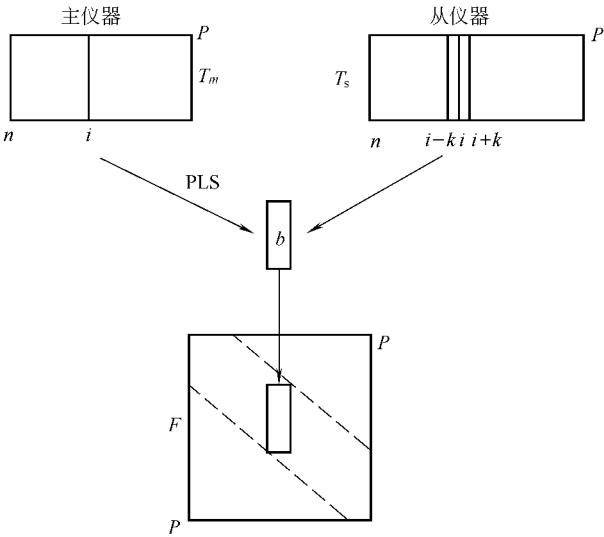


图 9-6 PDS 算法示意图

参考文献 [34] 根据 44 个标准化样品在“主仪器”和“从仪器”上的光谱, 用 PDS 算法计算出“主、从仪器”间光谱的校正参数, 并用得到的校正参数修正“Bruker2”、“Bio-Rad”和“Berg”中的 22 条检验集光谱, 使其与“主仪器”上的光谱匹配, 然后用“主仪器”上的模型预测分析修正后的光谱, 结果见表 9-10。

表 9-10 模型转移前、后 (“PDS 算法”) 预测分析 3 套检验集样品的结果比较

	“主仪器”上的 预测结果	光谱集“Bruker2”		光谱集“Bio-rad”		光谱集“Berg”	
		转移前	转移后	转移前	转移后	转移前	转移后
R^2	97.53	31.02	97.13	1.95	93.41	-158.66	90.52
SEP	0.29	1.51	0.31	1.80	0.47	2.92	0.56
RSD%	1.85	9.78	1.99	11.66	3.02	18.94	3.63

由表 9-10 可见, 模型转移后, 决定系数 (R^2)、预测标准差 (SEP) 和平均相对误差 (RSD) 均优于模型转移前的结果, 也接近“主仪器”上的分析结果。在 $\alpha=0.05$ 水平上, 对模型转移后的数据与“主仪器”上的分析结果进行 F 检验, 差异皆不显著, 说明“PDS”法能够将“从仪器”上的光谱修正成适应于“主仪器”上模型的光谱, “主、从仪

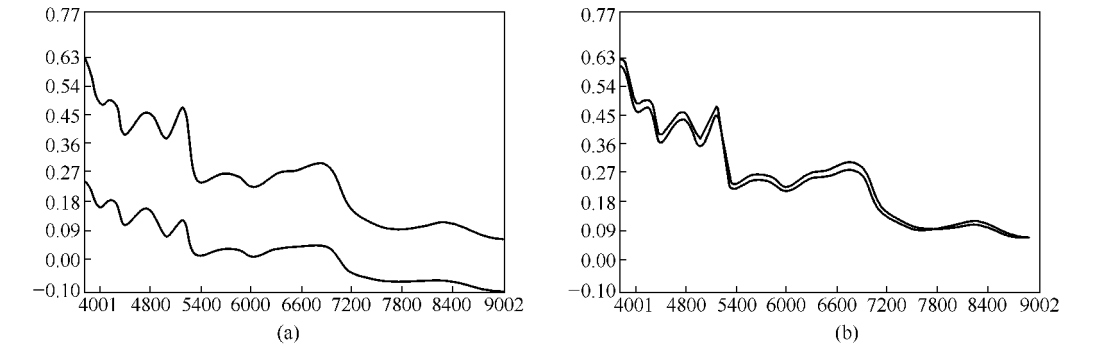


图 9-7 模型转移前后光谱的比较 (一)

(a) 样品在主仪器上的光谱 (上) 及样品在“Bio-Rad”上的光谱 (下)
(b) 主仪器上的光谱及“Bio-Rad”光谱经 Shenk 和 PDS 算法校正后的光谱

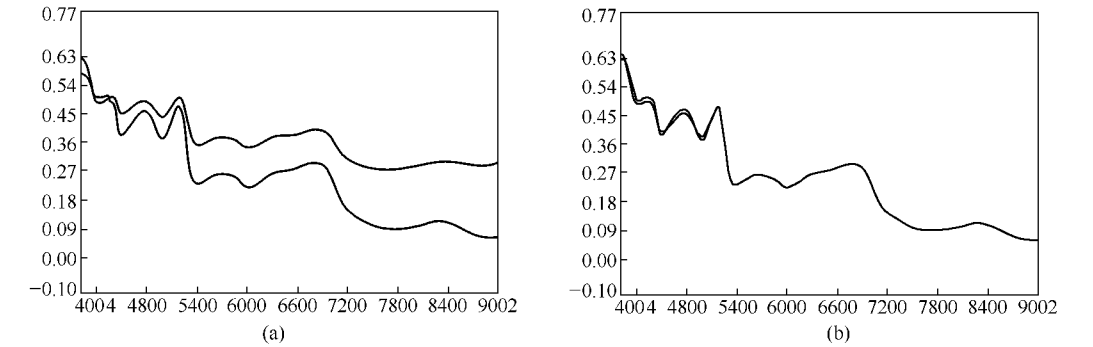


图 9-8 模型转移前后光谱的比较 (二)

(a) 2 号样原图 (下) 以及 2 号“Berg”图谱 (上)
(b) 2 号样原图、2 号“Berg”经 Shenk 和 PDS 校正后的图谱 (三者几乎重叠)

器”对同一样品的响应接近一致，实现了不同仪器间的标准化。

Shenk 算法和 PDS 算法模型转移的效果也可以从模型转移前后光谱的对照上看出来，图 9-7(a) 是同一样品在“主仪器”上和“从仪器 Bio-Rad”上的光谱，二者至少在纵轴偏移量上有较大区别，经过 Shenk 算法和 PDS 算法校正后的图谱见图 9-7(b)，由图可见，校正后的“从仪器”光谱与主仪器上的原光谱三者几乎重叠在一起。

同样，图 9-8(a) 是同一样品在“主仪器”和“从仪器 Berg”上光谱的对照，二者之间也存在较大区别，纵轴存在平移和基线旋转，经过 Shenk 算法和 PDS 算法两种方法校正后，“从仪器”光谱与“主仪器”上的原光谱三者也几乎重叠在一起见图 9-8(b)。

图 9-7 和图 9-8 说明模型转移方法确实能将“从仪器”上的光谱转移成与“主仪器”上光谱相类似的光谱，达到模型转移的目的，实现不同近红外仪器间的标准化。

以上介绍了几种典型的模型转移数学方法的原理与应用，其中：

① Slope/Bias 法是一种简单的对预测值的一元校正法，得到的这一简单线性模型和原校正模型合二为一给出了转移后的模型，用这一转移后的模型预测“从仪器”上得到的光谱就可得到正确的结果。这种方法只有在比较相似的仪器上才有较好的效果，此时使用简单的斜率/截距法既有效又方便。如果使用的“主、从两台仪器”差别很大，那么在两台仪器上测量同一样品时，光谱间的差异不仅很大也会很复杂，这时，用这种简单的拟合直线来预测没有标准化的光谱，通常不会获得满意的结果。

② Shenk 算法是一种一元全光谱校正法，它包括波长的校正和吸光度的校正两步。因此，在有波长漂移的色散型仪器间数学模型的转移问题上，这种算法的校正效果可能比较好。这种方法的一元特征是其主要缺点，因为它不能处理波长漂移和吸光度变化之间的复杂相互作用。

③ PDS 算法是一种多元全光谱校正法，这种算法的提出是为了处理一些更加复杂的情况。国外已有报道说明这种方法比其他几种方法的转移效果好^[27]。它的缺点一是计算强度大，大约要进行 N_w 次多元回归，二是在对每个波长点进行 PLS 回归的时候，要选择局部最佳维数，这对计算模型转移参数并用此参数校正未知光谱来说是极其重要的，同时也是很难决定的。

模型转移方法的选择通常可根据以下原则：如果两台近红外仪器差别很大（如光学系统不同，信噪比不同），或者由于测试条件不同（如不同的温度），则同一样品在两台仪器上测得的光谱差异会比较大而且复杂（如：非常量波长漂移、光谱强度的较大变化，不同类型的噪声）。在这种情况下，可用 PDS 或其他复杂的方法来计算模型转移参数^[30]。如果两台仪器相似，给出的光谱响应的差异非常简单（如无显著波长漂移，光谱强度仅有常量偏移，或者两台仪器的信噪比近似），这时，仅用简单又快速的 Slope/Bias 法就可以得到满意的结果，若用比较复杂的 PDS 法则可能是多余的，PDS 法需要较多多元回归模型的计算，而且 Slope/Bias 法只需要基本的统计知识，一般人都应用。

9.4.2.4 有限脉冲响应法 (FIR)

以上三种模型转移的数学方法 Slope/Bias 法、Shenk 算法和 PDS 算法有一共同点是：为了找到“主、从”两台仪器间的映射关系，需要选择部分标准化样品，并分别在“主、从”两台仪器上同时测量。由 Steven D. Brown 等人提出了另外一种多元模型转移的方

法；FIR 法（Finite Impulse Response Filtering）。FIR 法是一种无需标准化样品的光谱校正法，它根据有限脉冲滤波的基本原则降低仪器变动带来的影响，其基本思想如下：“主仪器”上的光谱用建模样品集光谱阵的平均值 r_m 代替，“从仪器”上测得未知样品光谱 S ，对应于 S_i （第 i 个波长点的光谱值）。

① 首先在 i 波长点左右选择窗口大小为 $(2p+1)$ 的光谱段，将 S_i 和 r_m 按下式在 $2p+1$ 窗口内进行均值中心化处理

$$s_i = [s_{i-p} - \text{mean}(s_i), s_{i-p+1} - \text{mean}(s_i), \dots, s_{i+p} - \text{mean}(s_i)]$$

$$r_{m,i} = [r_{m,i-p} - \text{mean}(r_{m,i}), r_{m,i-p+1} - \text{mean}(r_{m,i}), \dots, r_{m,i+p} - \text{mean}(r_{m,i})]$$

$$\text{mean}(x_i) = \sum_{k=i-p}^{i+p} \frac{x_k}{2p+1}$$

② 由 s_i 和 $r_{m,i}$ 计算第 i 波长点的回归系数： $b_i = (r_{m,i}^T \cdot r_{m,i})^{-1} \cdot r_{m,i}^T \cdot s_i$

③ 将“主仪器”上的光谱矩阵 S 转换成从仪器上的光谱矩阵 S^* ： $S_i^* = s_i/b_i + \text{mean}(r_{m,i})$

Thomas B. Blank 等人 1996 年的实验是在 NIR5000 近红外光谱仪上利用 PLS-2 法建立由阿司匹林、咖啡因和微晶纤维素按不同比例混合得到的 16 个样品的数学模型，并利用直接转移法（DS）、FIR 法和 PDS 法等结合不同的光谱预处理方法（无预处理、二阶导数加平滑处理）将此模型转移到 NIR6500 上，然后利用转移后的模型预测分析 15 个 NIR6500 上的检验集样品，不同转移方法的结果列于表 9-11。

表 9-11 NIR5000 的数学模型转移到 NIR6500 上的结果

方 法	窗口 /nm	数据预处理方法	主成分个数	预测标准差 /mg		
				阿司匹林	咖啡因	微晶纤维素
直接法		无	6	107.2	58.9	48.5
直接法		二阶导数 7 点平滑	6	22.7	9.5	14.0
FIR 无标样	220	无	6	8.5	4.1	5.4
PDS 7 个标样	14	无	6	6.9	5.9	3.3
PDS 7 个标样	10	无	6	7.0	6.0	3.2
PDS 7 个标样	8	无	6	7.0	6.0	3.3
PDS 7 个标样	12	无	6	6.9	6.1	3.4
PDS 7 个标样	12	转移后二阶导数 11 点平滑	6	7.5	7.6	2.1
PDS 7 个标样	12	转移后二阶导数 11 点平滑	6	8.6	9.0	3.1

FIR 法在计算过程中可能会导致光谱中出现奇异噪声（Artifacts）。图 9-9 中光谱在 1600nm 处的毛刺就是模型转移过程中产生的，原因可能是局部回归过程中回归系数太小所致。

因为 FIR 算法模型转移过程中不需标准化样品的测定，这对标准化样品无法长时间保存或样品有毒等的情况比较合适，但该方法应用文章较少，其主要原因是运算中常常出现无法解释的噪声（Artifacts）以及窗口大小的选择等问题，致使其应用受到限制。

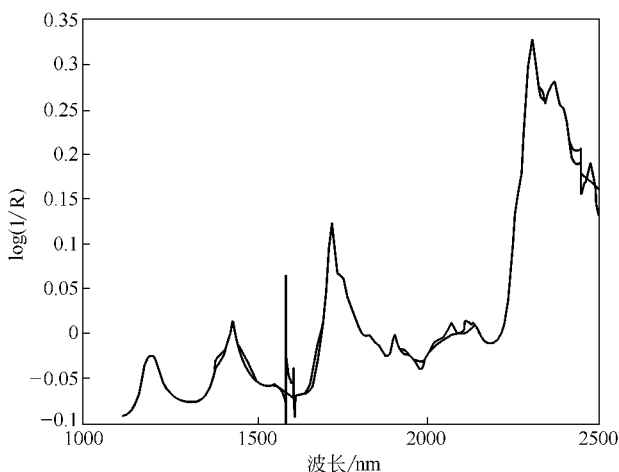


图 9-9 “从仪器”光谱经 FIR 法校正后的结果，1600nm 处出现 Artifacts

9.4.2.5 模型转移的其他方法

(1) 小波变换域中的模型转移 (Standardization in the Wavelet Domain, SWD)

以上介绍的模型转移的实现方法都是在光谱所在域内进行的，B. Walczak 尝试在光谱的小波变换域中进行模型转移。过程如下：分别对“主仪器”和“从仪器”上得到的标准化样品集的光谱 ^mS 、 ^sS 进行小波变换，得到小波域中的小波系数 ^mW 、 ^sW ，利用线性拟合法求 ^mW 、 ^sW 的转移参数矩阵 F ，此参数阵就是“主、从仪器”间的映射矩阵，如图 9-10 (a)。分析样品时，对“从仪器”上得到的预测样品集光谱 $^sX_{new}$ 利用相同的小波变换法转化成小波系数，并根据上述求得的转移参数矩阵 F 将 $^sW_{new}$ 转换成 $^mW_{new}^*$ ，再将 $^mW_{new}^*$ 进行小波逆变换得“从仪器”上标准化后得光谱 $^sX_{news}^*$ ，见图 9-10(b)，然后用“主仪器”上的模型预测分析 $^sX_{news}^*$ 就可得到正确的分析结果。

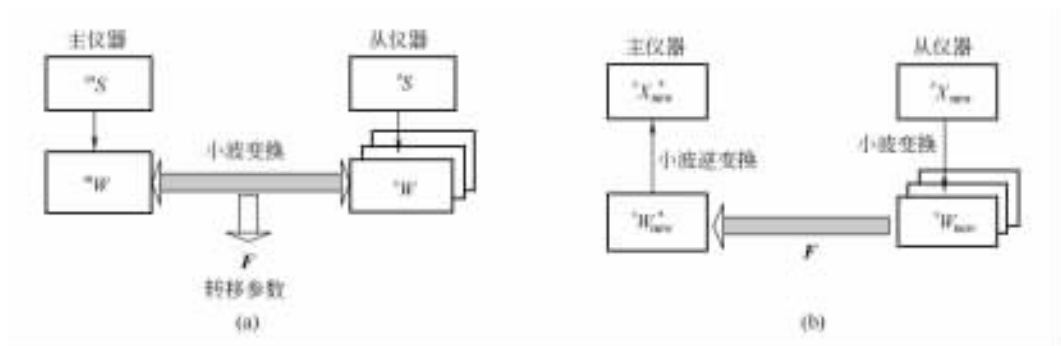


图 9-10 模型转移

(a) 小波变化法计算模型转移参数 (b) 利用转移参数校正“从仪器”光谱

图 9-11(a)、(b)、(c) 显示的是 30 个汽油样品在“主、从仪器”上的光谱图及二者的差谱。图 9-12(a) 是用“主仪器”上建立的模型直接预测分析 30 条图 9-11 (a) 汽油光谱五种组分结果的散点图，横坐标是化学实测值，纵坐标为近红外预测值，图中将五种组分的结果同时显示出来。图 9-12(b) 是用“主仪器”上的模型直接预测分析 30 条“从仪器”光谱 [图 9-11(b)] 的结果。由散点图可见，如果模型不转移，“从仪器”上预测结

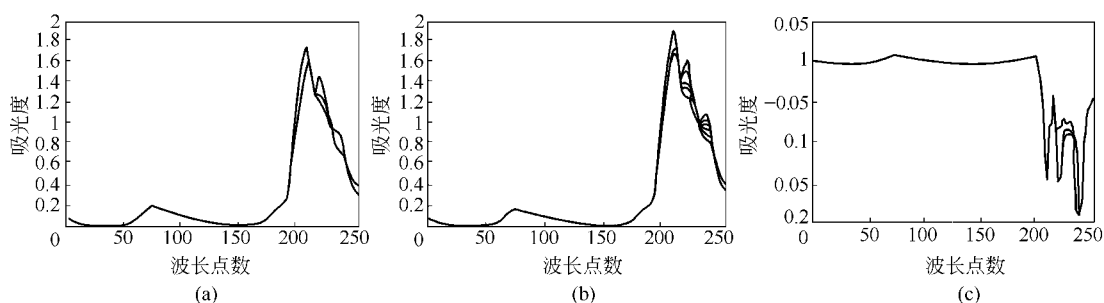


图 9-11 样品光谱

(a) 30 个汽油样品在“主仪器”上的光谱图 (b) 30 个汽油样品在“从仪器”上的光谱图
(c) 30 个汽油样品在“主、从仪器”上的差谱

果与“主仪器”上预测结果会有较大差异，说明“主、从仪器”存在差别。利用 SWD 法将“从仪器”上的光谱标准化成与“主仪器”相符合的光谱后，再用原模型重新预测标准化后的光谱，结果见图 9-12 (c)，此结果已经与在“主仪器”上的分析结果基本一致了。

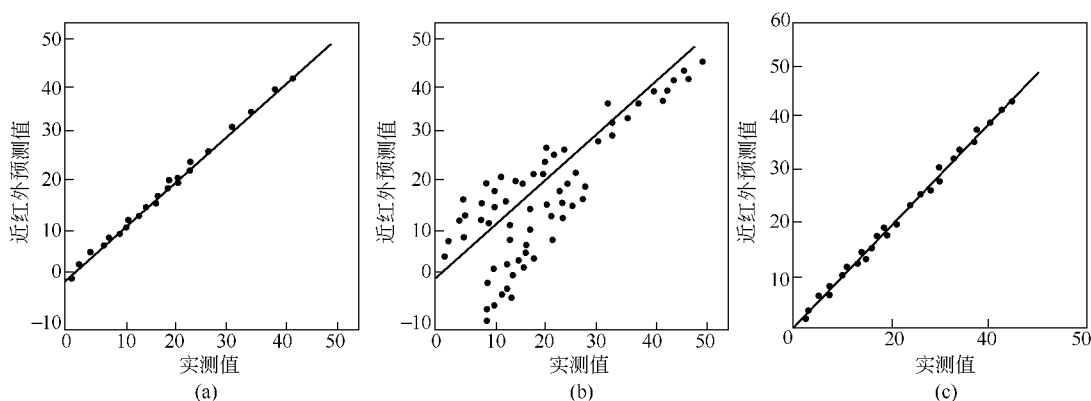


图 9-12 散点图

(a) “主仪器”上预测结果的散点图 (b) 模型转移前“从仪器”上预测结果的散点图
(c) 模型转移后“从仪器”上预测结果的散点图

与 PDS 算法相比，SWD 通常需要较少的标准化样品数即可获得与 PDS 相当的结果。虽然 SWD 法与 PDS 法二者都具有局部 (Local) 和多元 (Multivariate) 特性，但 PDS 在转移过程中易受到局部维数 (Local Rank) 值的影响，而 SWD 不会。

(2) 其他方法

最近一些新的化学计量学算法如 ANN、最大似然主成分分析 (MLPCA, Maximum Likelihood Principal Analysis)、正矩阵因子分析 (PMF, Positive Matrix Factorization) 等相继应用于模型转移，在此不一一而足。Yulong Xie 等将模型传递看作遗失数据重建问题 (Missing Data Reconstruction)，用 PMF 对近红外光谱测定汽油五种组成的校正模型进行了转移。与 PMF 原理相同，加拿大 Daren T Andrews 等用 MLPCA 对近红外光谱测定甲苯、氯苯和庚烷的校正模型进行了转移。Frederic Despagne 等基于人工神经网络 (ANN) 提出了一种解决模型转移的算法，与 PDS 不同，它对所有窗口光谱只用一个变换模型，避免了奇异噪声 (Artifacts) 的出现，也不必对背景进行预处理，如果标准化样

品集选择合适，可以得到比 PDS 更好的结果。英国 R Goodacre 等也用 ANN 对热解质谱定量测定溶菌酶含量以及定性判断真细菌类属的校正模型在三台热解质谱间实现传递。

9.5 不同类型近红外光谱仪器间的模型转移

对相同类型的近红外光谱仪器，如傅里叶变换仪与傅里叶变换仪之间以及光栅型仪器与光栅型仪器之间，一般情况下都可以采用以上介绍的方法进行模型转移，但转移过程中需要注意相同波长的数据要对齐。滤光片型仪器之间因为波长离散，且间隔很大，Shenk 算法和 PDS 算法、FIR 算法等基本不适合，但可以应用斜率/截距算法。

对不同类型的近红外光谱仪器，尤其是连续型光谱仪与离散型光谱仪之间，模型转移因为涉及到数据量的多少以及波长对齐的问题，需要单独考虑。为了在两类仪器间进行模型转移，首先要使两组光谱数据匹配，就是使它们的波段范围、波长位置及数据间隔一一对应。匹配的方法一般是先将含信息量高的仪器以含信息量低的仪器为准，生成一个“虚拟仪器”，该“虚拟仪器”的波段范围、波长位置及数据间隔与含信息量低的仪器一致。完成该“虚拟仪器”后，再实现仪器间的标准化问题，在生成“虚拟仪器”的过程中可能会降低原仪器的信息量。以天津大学高荣强与范世福等的研究为例说明：小麦蛋白的近红外数学模型建立在傅里叶变换仪器上，为了将该模型转移到光栅型仪器上，以光栅型仪器数据为准，对傅里叶变换仪器进行抽取和插值，从而使新得到的傅里叶变换光谱与光栅光谱匹配（对同样样品，光谱数据矩阵具有相同的维数，波数也一一对应），即得到一“虚拟傅里叶变换仪器”，对应光谱称为“虚拟光谱”。为了检验针对光栅光谱数据作匹配的虚拟傅里叶变换光谱是否有效地保留了原傅里叶变换光谱仪中的有效信息，仍用主成分回归算法对虚拟光谱建立数学模型，结果见表 9-12 所示。

表 9-12 傅里叶变换光谱仪与虚拟傅里叶变换仪器的模型结果

	模型结果		
	R^2	SEC	RSD(%)
原始傅里叶变换仪器	98.45	0.27	1.73
虚拟傅里叶变换仪器	97.00	0.35	2.24

由表可见，与原始傅里叶变换光谱的建模结果相比，虚拟傅里叶变换仪器 R^2 值有所降低，SEC 和 RSD 值有所增大，说明原傅里叶变换光谱经过与光栅光谱匹配处理后有效光谱信息有所损失，但仍然保留了绝大部分有用信息。

虽然在生成“虚拟仪器”的过程中可能会降低原仪器的信息量，但该方法不需要对样品进行重新扫描并重新建模，为不同类型 NIR 仪器间实现资源共享提供了一个方法。下面以实例说明不同类型近红外光谱仪间的模型转移。

(1) 傅里叶变换近红外光谱仪与光栅型仪器间数学模型的转移

仍以天津大学高荣强与范世福等的研究为例说明：55 个小麦样品分别由傅里叶变换型近红光谱仪和光栅型近红外农产品品质分析仪扫描所得的近红外光谱图如图 9-13 和图 9-14 所示，二者波段范围、采样间隔等均不同，且光栅仪器在 $4000\sim5000\text{cm}^{-1}$ 信噪比较差。

模型建立在精度和稳定性很好的傅里叶变换型近红外光谱仪上，为了能在光栅型近红

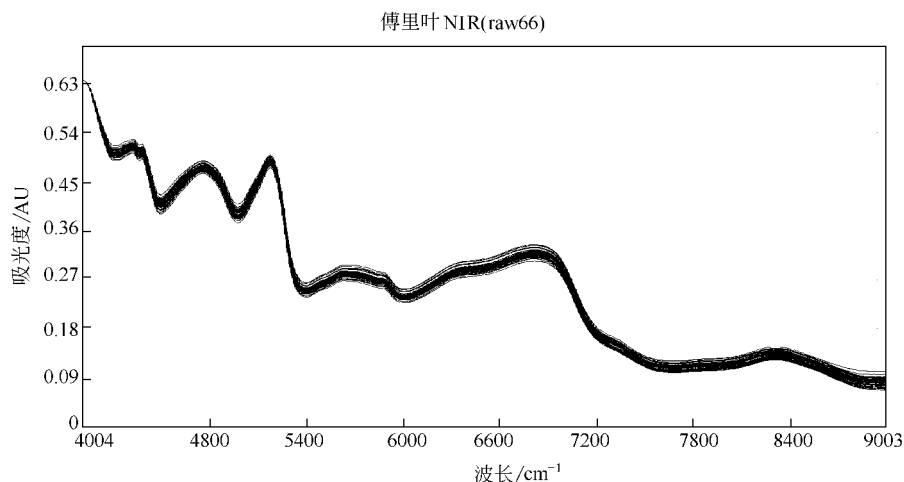


图 9-13 55 个小麦样品的傅里叶变换近红外光谱，范围：4004~9003 cm^{-1} ，
间隔 3.875 cm^{-1} ，共 1297 个点

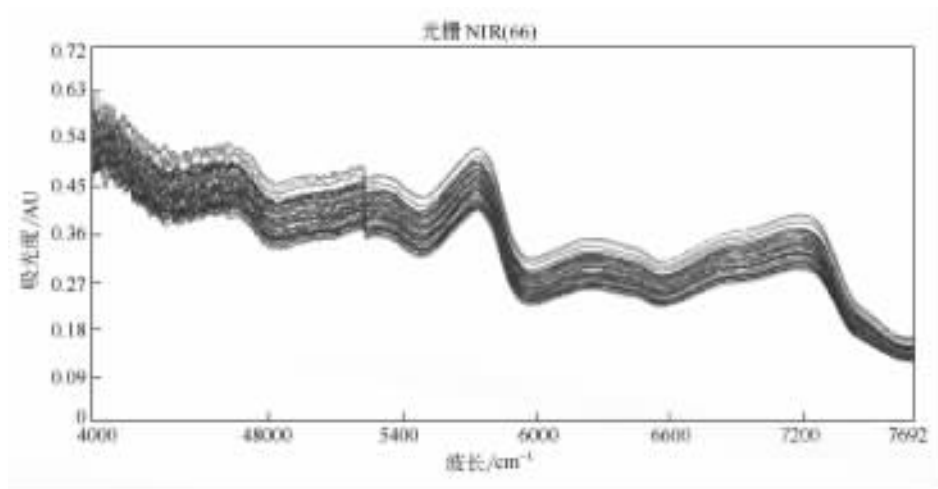


图 9-14 55 个小麦样品的光栅近红外光谱，范围：4000~7692 cm^{-1} ，
间隔 12.308 cm^{-1} ，共 301 个点

外光谱仪上应用此模型，首先生成一与光栅仪器具有相同波段范围、波数位置及采样间隔的虚拟傅里叶变换仪，作为模型转移的“主仪器”。选取 55 个样品中的 36 个建立模型，19 个作为检验样品，该主仪器上的模型结果见表 9-12。分别采用 Slope/Bias（斜率/截距法）和 Shenk 算法将虚拟主仪器上的模型转移到光栅型仪器上，模型转移前、后预测分析 19 个检验集光谱的结果见表 9-13。

表 9-13 傅里叶变换仪与光栅型仪器间数学模型转移前后预测集的结果比较

	转移前	转移后	
		Slope/Bias	Shenk
$R^2/\%$	— 22.64	12.42	63.86
SEP	2.08		1.13

表 9-13 说明，采用斜率/截距算法将傅里叶变换仪器上的模型转移到光栅光谱仪上后预测分析检验集样品， R^2 仅为 12.42，达不到实际应用需要。采用 Shenk 算法， R^2 则为 63.86，虽与实际应用仍有一定距离，但该方法已逼近光栅光谱仪自身模型的结果，且与斜率/截距算法相比有很大改进，说明 Shenk 算法可以应用在“主、从仪器”差别较大情况下的模型转移，而斜率/截距算法只能对比较相似的仪器才会有较好的效果。

(2) 傅里叶变换近红外光谱仪与滤光片型仪器间数学模型的转移

傅里叶变换近红外光谱仪是连续光谱仪，而滤光片型仪器只有几个固定波长点，各个波长间隔较大（滤光片的带宽一般可达 10nm），上述提到的模型转移的方法，如 Shenk 专利算法、PDS、FIR 等不再适应这两种类型仪器间的模型转移。因为没有方法使滤光片仪器直接与傅里叶变换仪器比较，一般先将傅里叶变换仪器生成一个虚拟的滤光片型“主仪器”。完成该虚拟的主仪器后，再将各滤光型仪器对该虚拟仪器进行标准化。下面以中国农业大学近红外课题组与中国农科院品资所的研究为例说明。

傅里叶变换近红外光谱仪上建有小麦粉蛋白模型和玉米粉蛋白模型，为了将该模型转移到 Bran+Luebbe IA450 型滤光片式近红外光谱仪上，首先以滤光片型仪器（“从仪器”）为准，将傅里叶变换仪器生成一个虚拟的滤光片型“主仪器”并建立“主模型”。然后分别在两台仪器上扫描相同的一组蛋白质含量已知的小麦、玉米样品光谱，并用“主模型”预测分析，分析结果用 Slope/Bias 法求得“主、从仪器”间小麦模型和玉米模型的转移系数，原“主模型”与转移系数合二为一给出了 IA450 型滤光片型仪器上的小麦模型和玉米模型。利用该模型直接预测 IA450 滤光片仪器上的光谱即可得到正确结果。

为了检验此模型，在 Bran+Luebbe IA450 型光谱仪上分别扫描了 17 个小麦样品和 17 个玉米样品，并用 IA450 上自己建立的小麦和玉米的模型（简称“自身模型”）分析测定这两组样品的蛋白质含量。在完全相同的条件下，用傅里叶光谱仪转移后的小麦和玉米的模型（简称“转移后模型”）也来分析这 2 组样品，具体结果见表 9-14 和表 9-15。

表 9-14 用“自身模型”和“转移后模型”分析 17 个小麦样品蛋白质含量结果比较

编 号	小麦真值①	自身模型②	转移后模型③	真值_自身模型④	真值_转移后模型⑤
1	17.34	17.32	17.30	0.02	0.04
2	16.06	15.58	16.14	0.48	−0.08
3	18.35	17.78	17.89	0.57	0.46
4	13.81	14.42	14.73	−0.61	−0.92
5	18.12	18.78	18.48	−0.66	−0.36
6	18.10	17.84	17.64	0.26	0.46
7	18.74	20.08	19.61	−1.34	−0.87
8	14.08	15.16	15.50	−1.08	−1.42
9	16.05	16.02	16.73	0.03	−0.68
10	15.55	16.11	16.39	−0.56	−0.84
11	17.64	18.42	18.06	−0.78	−0.42
12	14.39	14.37	14.72	0.02	−0.33
13	19.53	20.46	20.05	−0.93	−0.52

续表					
编 号	小麦真值 ^①	自身模型 ^②	转移后模型 ^③	真值-自身模型 ^④	真值-转移后模型 ^⑤
14	17.72	17.57	17.31	0.15	0.41
15	14.67	13.69	13.38	0.98	1.29
16	17.28	16.73	16.42	0.55	0.86
17	26.64	26.43	24.69	0.21	1.95
绝对标准差				0.469	0.767
相对标准差				2.714	4.436
相关系数				0.975	0.961

注：① 代表用凯氏定氮法测得的小麦或玉米的蛋白质含量（称为真值）。
 ② 代表用转移后模型分析得到的 18 个小麦和 17 个玉米样品的蛋白质含量。
 ③ 代表用自身模型分析得到的 18 个小麦和 17 个玉米样品的蛋白质含量。
 ④ 代表真值和转移后模型预测值的差值以及二者之间的关系。
 ⑤ 代表真值和自身模型预测值的差值以及二者之间的关系。

表 9-15 用“自身模型”和“转移后模型”分析 17 个玉米样品蛋白质含量结果比较

编 号	玉米真值 ^①	自身模型 ^②	转移后模型 ^③	真值-自身模型 ^④	真值-转移后模型 ^⑤
1	13.37	14.12	14.25	－0.75	－0.88
2	12.46	12.74	13.23	－0.28	－0.77
3	12.4	11.89	11.93	0.51	0.47
4	11.98	11.95	11.76	0.03	0.22
5	11.86	11.79	12.19	0.07	－0.33
6	11.72	11.74	11.81	－0.02	－0.09
7	11.70	11.66	11.69	0.04	0.01
8	11.65	11.79	12.2	－0.14	－0.55
9	11.23	11.16	10.98	0.07	0.25
10	11.05	11.5	11.43	－0.45	－0.38
11	10.67	10.55	11.2	0.12	－0.53
12	10.63	11.11	11.02	－0.48	－0.39
13	10.55	10.86	11.01	－0.31	－0.46
14	10.51	10.54	10.38	－0.03	0.13
15	10.24	10.12	10.4	0.12	－0.16
16	10.14	9.87	9.75	0.27	0.39
17	9.32	9.23	9.3	0.09	0.02
绝对标准差				0.314	0.440
相对标准差				2.78	3.91
相关系数				0.966	0.948

注：①②③④⑤同表 9-14 注解。

分析结果表明，对实验的小麦样品和玉米样品，用傅里叶光谱仪转移到滤光片上的模型和用 Bran+Luebbe 滤光片型近红外分析仪自己建立的模型来预测分析滤光片仪器上的

光谱, 其分析结果是相当的。这对不同厂家、不同类型的近红外光谱仪建立的数学模型实现资源共享, 以及不同用户之间进行交流具有实际意义。

9.6 模型转移中的其他问题

9.6.1 标准化样品集的选取

9.6.1.1 不同类型的标准化样品集

模型转移过程中,用于计算“主、从仪器”间映射参数的标准化样品集的选取对模型转移效果有重要影响,标准化样品集必须具有稳定性(Stability)和代表性(Representativity),稳定性是指标准化样品集的物理化学性质稳定,代表性是指标准化样品集要能代表预测样品集。标准化样品集的选择通常有两大类方法。

一类方法是从建模样品集或预测样品集中选择实测样品作为标准化样品。另一类是利用普通标样（Generic Standardization Samples）进行模型转移，普通标样又分两种，一种是基本无光谱特性但吸光度要涵盖一定范围的普通标样；另一种是密封保存的物理化学性质不变的样品（如纯水、其他惰性的或人工纯品）。Yongdong Wang 等的实验表明利用普通标准物质对仪器进行标准化，结果并不令人满意，普通标样在不同仪器上测量的差异必须要反映实际测试样品的信号变化，否则会出现较大预测偏差。

目前，一般采用第一种方法进行模型转移，因为选择实测样品作为标准化样品可以同时考虑温度等对样品的影响。以 E. Bouveresse 1994 年的研究为例说明：在 NIR6500 近红外光谱仪（“主仪器”）上建立饲草、玉米、菜籽的近红外数学模型，模型结果见表 9-16。

表 9-16 **NIR6500 近红外光谱仪的数学模型**

样 品	品质参数	建模样品数	SEC	R^2
饲草 (SE)	蛋白	564	0.85	0.98
	纤维	563	1.27	0.96
玉米 (MA)	蛋白	482	0.43	0.90
	纤维	489	1.04	0.93
菜籽 (CO)	蛋白	382	0.56	0.94
	脂肪	354	0.73	0.91

要在 NIR5000 近红外光谱仪上应用这些模型，因为建模样品数很多，且仪器不在同一地点，不可能通过重扫样品并重建模型来实现，E. Bouveresse 采用三种类型的标准化样品集研究了主仪器上数学模型的转移，三种标准化样品集的描述见表 9-17，其近红外光谱图见图 9-15，三种标准化样品集的选取遵从上述介绍的选取方法，其中 STD1 为实测样品，其光谱与建模样品类似；STD2 是不具光谱特征的普通标样，吸光度范围为 0~1.6，与建模样品集吸光度范围一致；STD3 是由 12 个有机或无机纯品组成的。

表 9-17

三种标准化样品集

标准化样品集	描 述
STD1	30 个农产品,包括饲草、玉米、小麦、菜籽等,光谱与预测集样品的光谱相似,吸光度范围一致
STD2	6 个普通标样(Generic standards),光谱很平,但吸光度范围较宽(0~1.6)
STD3	12 个有机或无机纯品标样,包括尿素、甘油酯、丁酸、聚乙烯、碳酸铵、淀粉、萘、乳糖、蔗糖、硫酸钙、柠檬酸、草酸

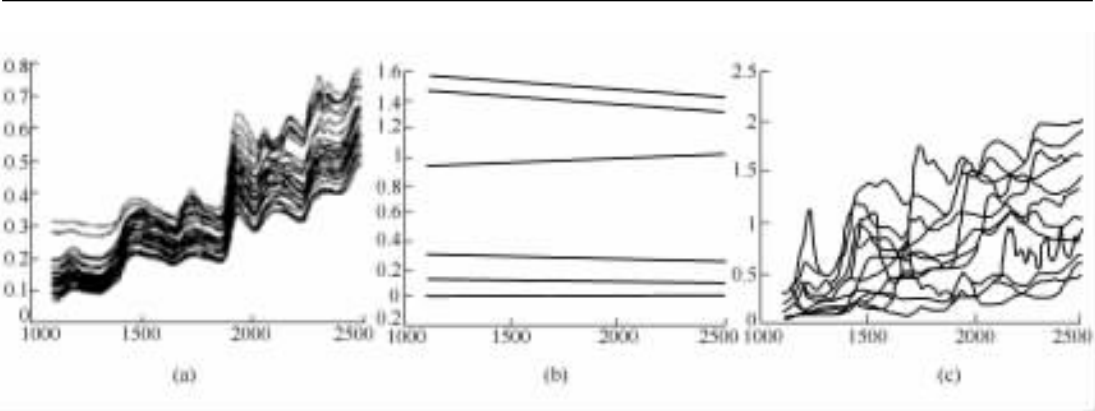


图 9-15 三种标准化样品集的近红外光谱

分别用这三类标准化样品集标准化位于其他地点的“从仪器”NIR5000 近红外光谱仪，即对“主仪器”进行模型转移，转移的方法为 Shenk 算法。为了检验模型转移的效果，用转移后的模型分别预测分析 6 个新的饲草样品和 6 个玉米样品。表 9-18 列出了模型转移前的预测结果，以及分别用 STD1、STD2、STD3 将光谱校正后再行预测的结果。

表 9-18 用三类标准化样品集对数学模型转移后预测分析的结果比较

样 品	品质参数	模型的 SEC	转移前的 SEP	用 STD1 标准化后的 SEP	用 STD2 标准化后的 SEP	用 STD3 标准化后的 SEP
饲草	蛋白	0.85	0.66	0.21	0.35	1.18
	纤维	1.27	0.45	0.21	0.93	7.07
玉米	蛋白	0.43	0.40	0.12	0.95	1.24
	纤维	1.04	0.47	0.30	0.38	2.89

由表 9-18 可见，四个近红外数学模型，用 30 个农产品（即 STD1）对“从仪器”上的光谱标准化后的 SEP 效果最好，均优于转移前的结果，STD2 次之，而 STD3 的结果很差。这是因为标准化样品集 STD1 的性质与待分析样品（饲草和玉米）相似，而且光谱的吸光度范围相同，而 STD2 和 STD3 与饲草样品（SE）和玉米样品（MA）的光谱差异较大。

9.6.1.2 选取标准化样品集的数学方法

从建模样品集中选择标准化样品有多种方法，最初用的数学方法是从建模样品中选高杠杆值（High Leverages）样品，但这种方法对样品不同组分的浓度分布比较敏感，当预测样品集超出标准化样品集的空间时，会出现奇异噪声（Artifacts）。比较常用的方法是

KS (Kennard-Stone) 算法，它利用样品间的欧氏距离（可以是原光谱之间的距离，也可以是光谱主成分之间的距离）来选择标准化样品。还有另外一种方法 MSID (Maximizes The Smallest Inter-point Distance)，但这种方法一般计算量大，计算时间长。各种挑选标准化样品的数学方法文献有详尽描述，在此不再列出。

9.6.2 标准化样品个数的多少

标准化样品集中样品个数的多少对模型转移效果也有影响。若标准化样品个数太少，会导致求出的模型转移参数信息不充分，从而导致模型转移效果不理想；若标准化样品数太多，实际应用时不方便，且标准化样品数太多时可能引入冗余信息，反倒对模型转移没有好处。一般标准化样品个数的多少没有固定数学算法来确定，通常是根据检验集样品预测偏差 (SEP) 的大小来确定的。

B. Walczak 研究了标准化样品个数对模型转移效果的影响。用“主仪器”上建立的数学模型直接分析“从仪器”上 60 个大豆样品的预测标准差是 8.9194。分别用 PDS、SWD 和 Slope/bias 法对“主仪器”上的模型进行转移，再用转移后的模型分析 60 个大豆样品，预测标准差列于表 9-19 中，由表可见，模型转移后的标准差皆远小于模型转移前的标准差。表中同时还列出了不同标准化样品个数对模型转移效果的影响。

标准化样品个数	模型转移(不同方法)后预测 60 个样品的标准差		
	PDS	SWD	Slope/Bias
3	1.7904	1.5670	2.3405
4	1.7353	1.5000	1.9494
5	1.7951	1.5000	1.6842
6	1.7889	1.5851	1.5510
7	1.8379	1.4259	1.5882
8	1.7654	1.3944	1.5986

随着标准化样品个数的增加，三种方法模型转移的预测标准差都在开始阶段呈下降趋势，到达一定数目后预测标准差降低的效果不再明显，甚至会有些微增加，说明并不是标准化样品个数越多模型转移效果越好。同时，表中结果还说明不同模型转移的方法所要求的最佳标准化样品个数并不完全相同。

褚小立等用普鲁克分析 (Procrustes Analysis, PA) 将“从仪器”上汽油样品的光谱标准化时也得到了类似结果。检验集样品 C8 芳烃的预测标准差 (SEP) 随标准化样品个数的变化趋势见图 9-16。

由图 9-16 可以看出，随标准化样品数的增多，SEP 逐渐下降，当标样数增至 6 时，SEP 变化幅度趋于平稳，因此，

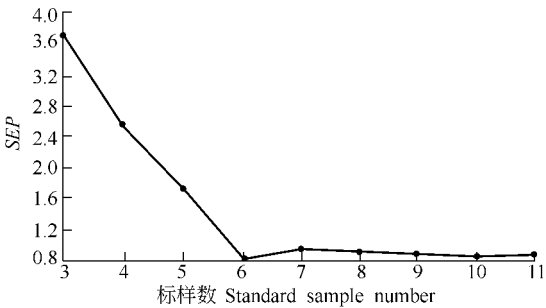


图 9-16 检验集的预测标准偏差 (SEP) 随标样数的变化

选取标准化样品数为 6。

结束语:

本章内容主要讨论数学模型的维护问题。

近红外分析的前提是数学模型, 数学模型的维护是应用近红外技术的基础, 也是实现近红外资源共享, 推广应用近红外光谱分析技术的核心技术之一。本章将模型的维护分为两类, 一类称为模型的修正, 另一类称为模型转移。如果待测样品本身的物化信息超出了原有模型的范围, 模型的维护工作主要是扩充建模样品所覆盖的(主成分)空间范围, 则称这种维护工作为模型的修正, 通常通过在原建模样品集中添加新样品完成。如果待测样品因为测试条件(如装样条件、仪器状态等)与建模时不一致, 希望通过模型的维护使同一待测样品在两种情况下得到一致的结果, 通常称这种模型维护工作为模型转移, 模型转移通常发生在不同仪器之间, 同一仪器的不同附件之间, 或不同测量时间等。本章仅就比较熟悉的几个内容做了讨论, 其他问题有待更深入探讨。对数学模型进行维护的方法很多, 但目前有些方法研究得还远不够深入, 同一方法处理不同的问题, 其结果有时相差较大, 解决实际普遍存在的问题也尚有一定距离。希望随着人们对近红外技术的日益重视、近红外技术本身的发展, 以及相关理论和方法的进一步完善, 数学模型的维护问题得到切实解决, 真正做到实用。

参 考 文 献

- [1] E Bouversse, D L Massart. Standardization of Near-Infrared Spectrometric Instruments. Anal Chem, 1996, 68 (6): 982~99
- [2] E Bouversse, D L Massart. calibration transfer across near-infrared spectrometric instruments using Shenk' s algorithm; effects of different standardization samples. Analytica Chimica Acta, 1994 (297): 405~416
- [3] E Bouversse, D L Massart. Chemom Intell Lab Syst, 1996 (32): 201~213
- [4] E Bouversse, D L Massart. Standardization of near-infrared spectrometric instruments: a review. vibrational Spectroscopy, 1996 (11): 3~15
- [5] B Walczak, E Bouversse, D L Massart. Standardization of near-infrared spectra in the wavelet domain. Chemom Intell Lab Syst, 1997 (36): 41~51
- [6] E Bouversse, D L Massart. Improvement of the piecewise direct standardization procedure for the transfer of NIR spectra for multivariate calibration. Chemometrics and Intelligent Laboratory Systems, 1996 (32) : 201~213
- [7] M Daszykowski, B Walczak, D L Massart. Representative subset selection. Analytica Chimica Acta, 2002 (468): 91~103
- [8] M Forina a, G Drava. Transfer of calibration function in near-infrared spectroscopy. Chemometrics and Intelligent Laboratory Systems, 1995 (27) : 189~203
- [9] Robert N Feudale a, Nathaniel A. Transfer of multivariate calibration models: a review. Chemometrics and Intelligent Laboratory Systems, 2002 (64): 181~192
- [10] J S Shenk, M O Westerhaus. U. S. Pat. , 4866644, Sept. 12, 1989
- [11] Yongdong Wang, David J Velramp, Bruce R Kowalski. Multivariate Instrument Standardization. Anal Chem, 1991 (63): 2750~2756

- [12] Yongdong Wang, Bruce R Kowalske. *App Spectrosc*, 1992, 46 (5): 764
- [13] Y D Wang, D J Veltkamp, B R Kowalske. *Anal Chem*, 1991 (63): 2750~2756
- [14] Thomas B Blank, Steven D Brown. Transfer of Near-Infrared Multivariate Calibrations without Standards. *Anal Chem*, 1996, 68 (17): 2987~2995
- [15] JieLin, Su-chin Lo. Near-IR Calibration Transfer Between Different Measurement Temperature. **匹茨堡会议论文摘要**, 1997
- [16] Onno E DeNoord. Multivariate Calibration Standardization. *Chemom Intell Lab Syst*, 1994 (25): 85~97
- [17] Ziyi Wang, Bruce R Kowalski. Additive Background Correction in Multivariate Instrument Standardization. *Anal Chem*, 1995 (67): 2379~2385
- [18] H Swierenga, W G Haanstra, A P de Weijer, et al. Comparison of Two Different Approaches Toward Model Transferability in NIR Spectroscopy. *Applied Spectroscopy*, 1998, 52 (1): 7~16
- [19] Tomas Isaksson, Bruce Kowalski. Piece-Wise Multiplicative Scatter Correction Applied to Near-Infrared Diffuse Transmittance Data from Meat Products. *Applied Spectroscopy*, 1993, 47 (6): 702~709
- [20] Jonas Sjoblom, Olof Svensson, Mats Josefson, et al. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 1998 (44): 229~244
- [21] Jeongah Yoon a, Byungwoo Lee b, Chonghun Han. Calibration transfer of near-infrared spectra based on compression of wavelet coefficients. *Chemometrics and Intelligent Laboratory Systems*, 2002 (64): 1~14
- [22] Paul Geladi, Haans Barring, Eigil Dabakk. *J Near Infrared Spectrosc*, 1999 (7): 251
- [23] Yulong Xie, Philip K Hopke. *Analytica Chimica Acta*, 1999 (384): 193~205
- [24] Daren T Andrews, Peter D Wentzell. *Analytica Chimica Acta*, 1997 (350): 341
- [25] Frederic Despagne, Beata Walczak, Desire-luc Massart. *Appl Spectrosc*, 1998, 52 (5): 732
- [26] R Goodacre, E M Timmins, A Jones. *Analytica Chimica Acta*, 1997 (348): 511
- [27] R W Kennard, L A Stone. *Technometrics*, 1969 (11): 137~148
- [28] J J Workman, P R Mobery. *Applied Spectroscopy Reviews*, 1996, 31 (1, 2): 73~124
- [29] C E Miller. Near-Infrared Spectroscopy of Synthetic Polymers. *Appl Spectrosc Rev*, 1991 (26): 277~339
- [30] Osborne B C, Fearen T J. *Food Technol*, 1983 (18): 453~460
- [31] Jones J A, Last I R, MacDonald B F. *J Pharm Biomed Anal*, 1993 (11): 1227~1231
- [32] 徐广通, 袁洪福, 陆婉珍. 近红外光谱定量校正模型适用性研究. *光谱学与光谱分析*, 2001, 21 (4): 459~463
- [33] 褚小立, 袁洪福, 陆婉珍. 普鲁克分析用于近红外光谱仪的分析模型传递. *分析化学*, 2002, 30 (1): 114~119
- [34] 赵龙莲. 近红外光谱定量分析数学模型的转移. [硕士学位论文], 北京: 中国农业大学, 1999
- [35] 高荣强. 现代近红外光谱仪器定标和模型传递. [硕士学位论文], 天津: 天津大学, 2003
- [36] 王文真. 在近红外光谱定量分析中应注意的几个问题. *现代科学仪器*, 1996 (1): 24
- [37] 褚小立, 袁洪福, 陆婉珍. 光谱多元校正中的模型传递, *光谱学与光谱分析*, 2001, 21 (6): 881~885
- [38] 赵龙莲, 劳彩莲, 严衍禄. 不同漫反射样品池间近红外数学模型的转移. *中国农业大学学报*, 1998 (增刊): 91~96

10. 近红外光谱定性判别分析

10.1 概述

近红外 (NIR) 光谱含有丰富的物质信息, 光谱与物质本身的组成及含量是密切相关的, 通过对光谱特征的分析, 可以获得有关物质结构与组成的信息; 绝大多数光谱定性分类的方法是直接地依靠光谱的分析研究, 如研究吸收峰的吸收强度、吸收频率、带宽和峰形来确定样品的特性和归属。由于近红外光谱主要是物质的倍频和组合频, 信号较弱, 且谱带较宽, 重叠严重, 尽管现代分析仪器的信噪比很高, 对这样的光谱进行类似中红外光谱的结构剖析仍有较大困难, 因此目前近红外光谱定性分析主要用于物质的定性判别分析上, 即通过比较未知样与已知样或标准样品的光谱来确定未知样的归属。由于近红外光谱的吸光度差异通常很小, 而且近红外光谱具有复杂性与多变性, 通过人眼直接对照的可靠性大受影响, 在近红外光谱定性分析中多依靠计算机和算法来进行光谱的比较和识别, 利用算法来分离提取近红外光谱信息特征。模式识别是最常用的一种借助于数学方法进行光谱模式的区分和识别的一种综合方法。

简单的光谱比较可以采用差谱法、光谱相关系数法或计算光谱间的距离 (如欧氏距离) 来进行直接比较, 也可以采用光谱转换 (压缩) 后的特征数据来进行比较, 如依靠主成分分析来进行比较。而对大量复杂的样本集来说, 模式识别显示出其强大的判别能力。常用的方法有两大类, 一是有监督的模式识别, 如判别分析 (Fisher 判别、Bayes 判别、逐步判别)、线性学习机 (LMD)、KNN、SIMCA、DPLS, 神经网络等; 二是无监督的模式识别方法, 如聚类分析、最小生成树等。另外, 分类图、线性映射、非线性映射技术和一些联用算法也可用于分类研究中。

光谱比较的主要用途是确定某未知样品的归属, 确定未知样品是否与某标样一致。如果只确定是否是某一特定物质, 可以采用简单的判别; 而对于大量复杂的样品, 确定未知样品的归属 (属于某一类) 则需要用模式识别来进行。从根本上说, 判别分析就是利用样本的光谱所提供的最大信息差异或相似性来进行识别。

近红外光谱的定性判别分析目前广泛应用于质量控制中, 如原料及中间体及最终产品的监控, 化工生产工艺优化, 农产品产地、品质、风味等的分析研究, 在制药业、石油化工、环保、轻工、食品加工等领域正日益受到重视。

10.2 近红外光谱定性判别分析的一般程序

近红外定性判别分析依实际应用需要其过程有所差异, 一般可以分为以下几个步骤: ①标准 (参考) 样品光谱集的建立; ②光谱校正及预处理; ③光谱特征的提取; ④定性判别分析模型的建立和评价; ⑤未知样品定性判别。

10.2.1 标准（参考）样品光谱集的建立

光谱定性分析就是对标准光谱与样品光谱进行比较，因此建立稳定可靠的标样光谱集是非常重要的环节。需要测量每种物质的标准光谱；一般同一样品需多次重复测量，不同批号的样品也需重复测量，以平均光谱近似作为该样品标准光谱。

如果是以模式识别来进行判别分析，则需将已知样品的光谱分成学习集和检验集两部分，划分的依据是学习集和检验集中的类别种类应相同，具有广泛的代表性，这一点与定量分析很类似。另外还需对各种类样品依先验知识进行赋初值（类指标的数值化），以便于识别；具体指标常见有百分制（专家评分）、二进制编码等。

10.2.2 光谱的校正与预处理

与定量分析类似，光谱校正的作用是光谱图的规范化、抵消背景干扰及提高光谱的质量。采用何种校正方法要依光谱的质量及干扰的情况来选择，预处理也可以把原来隐藏的信号差异放大出来，提高光谱的分辨率，使定性判别更加直观、可靠。常用的光谱预处理方法有中心化、导数、归一化等，有关预处理方法的详细介绍请参见本书第5章“近红外光谱分析的信号处理”。

10.2.3 光谱特征的提取

定性判别分析的重要依据是光谱特征的相似性（或差异性），因此正确获取光谱特征是定性分析的关键一环；目前，在模式识别中应用广泛的特征筛选方法有偏差权重法、K-W（Kruskal-Wallis testing）检验、主成分分析（Principal Component Analysis, PCA）、偏最小二乘法（Partial Least Square, PLS）等。

在常规定性分析中，人们通常依据某几个峰值或其相对比值就可以确定物质的归属和纯度，如峰值法等，这种方法在近红外光谱中仍然适用，只是可靠性差些，这是由近红外光谱多变性、背景的复杂性决定的。因此，在近红外光谱中，定性判别分析更多依靠若干个峰组（频率段）甚至全光谱来进行定性判别，如PCA、DPLS、SIMCA、LLM、Fisher判别、KNN、ANN等以提高定性分析结果的可靠性。另外，分类学的研究常用聚类分析方法。

对大量光谱数据来计算光谱距离是非常复杂的，尤其是在模式识别过程中，如果用如此多的原始光谱数据直接计算几乎是不可能的，也是没有必要的。实际上只需提取原光谱特征，然后依据这些特征间差异来进行模式识别就可以了，无需采用全部信息。常见的数据压缩变换方法有FFT、PCA、小波变换等方法，具体算法见相关文献。

采用主成分进行定性判别分析是最常见的一种光谱特征提取方法，其主要优点是可利用主成分间新描述的原光谱特征间的差异对样品进行分类鉴定，可选定以各主成分构成的多维空间来描述原样品，利用不同类的样品的某些特定主成分的得分不同而把样品分类鉴定。如图10-1就显示了20个烤烟和20个青烟样品光谱在第1、2主成分空间的分布，其中位于左半部分的实心点表示烤烟样品，位于右半部分的空心点表示青烟样品。

从混合物NIR光谱提取定性信息可用许多方法来完成，其中大部分方法都是试图从

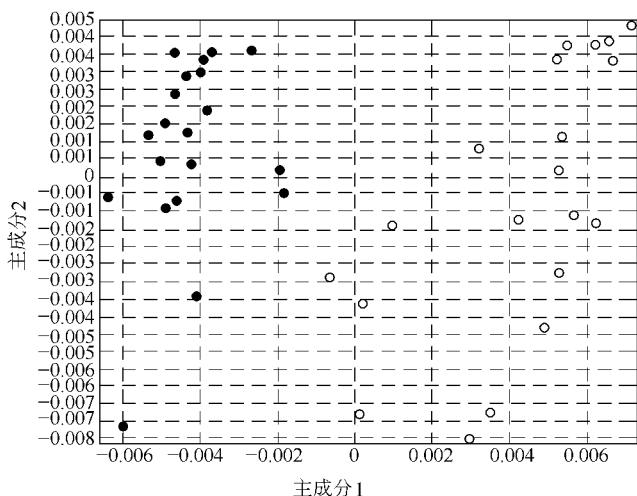


图 10-1 烤烟、青烟样品光谱的第 1、2 主成分空间分布图

原混合物的光谱中重构纯组分的光谱。光谱分解 (Spectral Rationing) 需要含不同配比纯组分的一系列混合物的光谱, 分解法强调纯组分某特定的吸收峰, 而且其他组分无干扰, 这种方法只限于简单混合物, 人们已注意到对于大多数 NIR 光谱而言, 宽且重叠的吸收峰将使上述方法变得不可靠。

可以获得混合物中的组成信息, 尽管存在分子间作用和散射效应, 每一波长点的光谱强度与组分浓度的相关系数图 (单相关图), 最终光谱经常与纯组分的非常相似。散射校正 (如 MSC) 将有利于相关图的解释。对于散射校正数据, 与纯组分相关的吸收带不是被假定与光谱重叠或内相关, 而是由于参比数据的误差引起吸收带变形。举一例子, 当样品中其他成分呈相近关系时, 饲料样品中的某些组分间 (消化度和 ADF-淀粉) 呈现出直接的反相关关系。

基于交叉相关的组分光谱重构已见报道, 所需知道的只是一系列已经知道感兴趣组分含量的样品集的光谱; 如果存在氢键或组分的内相关性, 最后重构光谱与纯组分的光谱将不相同。必须说明的是这方法应用于小麦时, 需要有足够大的和有代表性的样品集, 并需对颗粒度进行校正 (如果存在的话)。

基于 Bootstrap Method 来评价置信区间的从光谱-浓度相关性衍生出来的一种减少光谱重构误差的方法已有报道。相比之下, 曲线拟合需要知道纯组分的光谱并且没有类似氢键等的相互作用。值得注意的是 Hacschkgan Norris 提出的标准最小二乘曲线拟合过程的关键问题, 他们使用已知的小麦样品光谱来补偿组分间相互作用。

Cowe 等人通过主成分图解释组分的信息, 另外, 回归系数和组分权重的组合可用来获得光谱, 这光谱与感兴趣成分相似。尽管光谱重构的主要贡献来源于感兴趣的成分, 它们也受其他成分的内相关性的影响, 然而散射效应不会太多干扰这种重构方法, 这方法的主要缺点是尽管 PCA 的主成分与浓度误差无关, 它们与感兴趣成分的关系也不密切, 优点是这种方法不仅可以应用于单一组分 (如蛋白质), 而且可应用于复杂组分, 如感光特性 (Sensory Attributes), 因此这些化学方法多少有些非化学测量的味道。

NIR 判别光谱模式是由因子判别分析的 PCA 得分值来决定的, 最终光谱模式强调定

义样品群间的差异。一个小麦样品集经三种不同湿度处理后，判别光谱模式表现出随机和键合水（Bound Water）有关。在随后的研究中，这种方法被用来鉴定并消除水分的干扰。

10.3 近红外定性判别分析的方法——有监督模式识别算法

常见的定性判别分析之一是根据测得的光谱数据来确认未知试样是否属于某特定若干种类化合物之一，这属于有监督模式识别方式。在这种识别方式中，要求事先知道获得这些种（类）化合物的某些代表性样品（参照物，Reference）的光谱，目的是利用已有信息来确认某未知物到底属于哪一种。在物料确证、产品质量鉴定、过程控制等领域绝大多数问题可采用这类方法来解决。下面具体介绍几种常用的模式识别方法，并以最近邻法为例详细说明定性判别分析过程中的注意事项和问题。

10.3.1 最近邻法

最近邻法（K-NN）可分为建库和样品的鉴定两个步骤。在建库阶段，必须收集各种各样的能够鉴别的物质来扫描它们的光谱。对于库里的每一个物质可以通过和库中参考光谱进行统计分析而计算出每一种物质所对应的阈值，这个阈值就可以作为鉴定新未知样品的标准。通过收集光谱，计算每一种物质相应的阈值并且对库进行检验后，建库阶段就完成了。检验过程实际上就是确认库里的每一种物质都只能被惟一的鉴定，而不能与其他物质相混淆。在鉴定阶段，新物质的光谱得到后就可以对其进行鉴定了。未知物质样品的光谱可以和库中每一种物质的参考光谱进行比较。如果新光谱和库中参考光谱距离（后面会详细讨论）小于参考光谱对应的阈值，那么这个新样品可能和库中的这种物质属于同一种物质。但是对这个新样品最后鉴定结果必须将它同库中所有的物质比较后才能得出。如果新样品的光谱在库中只在一种物质阈值之内，那么就认为这个样品被惟一鉴定为库中某一种物质。如果新样品光谱在库中多种物质阈值之内，那么就认为这个样品被这些物质混淆了，如果新样品的光谱在库中所有物质阈值之外，那么就认为这个样品不能用这个库来鉴定。

关于如何获取库中光谱，产生阈值，检验库，并且分析新样品光谱过程将在以下几个部分进行详细说明。

10.3.1.1 获取参考光谱

获取参考光谱时所采用的测量方法，必须同获取新未知样品光谱时所采用的测量方法保持一致，库中的参考光谱和新未知样品光谱之间的变异来源应该尽可能减少，只有这样新样品光谱才能和库中光谱尽可能接近。比如采样方法（探测器、样品池等）、分辨率、扫描间隔或扫描时间、前置放大器增益、孔径或滤光设置等这些测量参数，应该尽可能的保持一致。

虽然在建立 Ident 库时对每一种物质只使用一条光谱也是可以的，但通常的做法是对每一种物质都收集几条光谱，并且用这几条光谱来产生一条这种物质的平均光谱，然后利用平均光谱来建库。这种取平均的方法有许多的好处。

在许多的情况下，被鉴定的物质如果来自不同批号，会表现出一些细微的光谱差异。因此一个理想的库应该允许这种批号间差异，只有这样才能使来自不同批号的相同物质被鉴定为同一种物质。同样，由于测量技术变化影响，同一种物质的光谱也可能会出现细微的差别，这些变化可能出现在比如采样探测器同样品或样品池之间的相对位置变化上。这些与重复性相关的变化通常应用矢量归一化（Vector Normalization）这一预处理方法来大大减小。

一方面，希望阈值尽可能高，能使已知物质光谱都能够在此阈值之内，只有这样才能进行正确鉴定。但另一方面，也希望阈值能够尽可能低，只有这样才能将不属于同一种物质而又具有相似光谱特征的物质区别开来。因此在设置阈值时，必须在忽略物质本身之间差异和测量过程中误差能够产生误判之间获得一个平衡点，如果在库中对每一种物质只收集了一条参考光谱，将无法得到每种物质的光谱的内部变异信息，于是也就没有办法为每一种物质来设置 1 个阈值。相反，如果对每一种物质都收集了多条单独的光谱，那么就许多种算法来帮助我们优化阈值，这些算法都是通过每一条单独参考光谱和平均参考光谱之间的关系进行统计分析的。

通常获得库中参考光谱的过程包括得到同一种物质不同批号样品的多条光谱，其中包括对每一批号的多次测量，然后用这些单独光谱来产生一条平均光谱。平均光谱是通过单独光谱的每一个波长点的吸光度值取平均得到的。这个平均光谱就可以用来作为库中这种物质的参考光谱。通常平均光谱比任何用来产生平均光谱的单独光谱都具有代表性，因为它包含了批号间及每次测量中带来的变异信息，用平均光谱作为参考光谱比用任何单独参考光谱都好。

10.3.1.2 欧氏距离

前面已经提到，产生平均光谱的另外一个好处就是能为库中每一种物质能够得到一个统计优化的阈值提供根据。理解这种分析的关键在于建立欧氏距离的概念。欧氏距离是两条光谱差异程度的一个尺度，也就是说欧氏距离越大，光谱的差异也就越大，而两条完全相同光谱的欧氏距离为零。1 条样品光谱同库中 1 条光谱的距离（ D_s ）被定义为

$$D_s = \sqrt{\sum_k (A_{sk} - A_{lk})^2}$$

式中 A_{sk} 为样品光谱在波长点 K 处吸光度值， A_{lk} 为库中光谱在波长点 K 处吸光度值。

欧氏距离的本质是两条光谱在所选波长范围内 Y 轴差异的总和。所以对两条完全相同的光谱，它们的吸光度是完全相同的，差异的总和也就为零，而且随着吸光度值差异的增加，欧氏距离也会相应的增大。

10.3.1.3 设置阈值

欧氏距离可用来比较给定物质的单独参考光谱和平均参考光谱之间的关系，阈值实际上就是对库中每一条平均参考光谱所定义的一个欧氏距离限度。设置这个限度的目的是为了选取合适的阈值，使同一种物质中所有单独参考光谱同平均参考光谱的欧氏距离都小于这个限度，而库中所有其他物质光谱同这个平均参考光谱的欧氏距离都大于这个限度。也就是说阈值给平均参考光谱定义了一个所谓的置信区间。如果代表物质光谱的点落在给

定物质的置信区间内，那么就认为这些物质与给定物质是相同的，而落在给定物质置信区间之外的物质，就认为与该给定物质是不同的。需要反复地为库中每一种物质来设置阈值，直到库中每一种物质都有一个平均参考光谱和一个阈值为止。

前面已经提到，对每一种物质收集多条单独的参考光谱并产生相应的平均参考光谱，可以使用统计方法去优化这些阈值。为了达到这个目的。需要计算每一条单独参考光谱与相应平均参考光谱之间的欧氏距离 D_i 。在计算完这一组单独的欧氏距离 D_i 后，就可以得到两个参数，即平均距离 (D_m) 和单独欧氏距离的标准差 (S_0)。

$$D_m = \sum_i \frac{D_i}{n}; \quad s_0 = \sqrt{\frac{\sum_i D_i^2}{n-1}}$$

式中 n 为对于给定物质单独参考光谱的数目。

平均欧氏距离 D_m 和欧氏距离标准差 S_0 提供了与库中物质相关的大量信息。具有比较小的平均欧氏距离和标准差的物质，也就是说它们的单独参考光谱和平均参考光谱非常接近，它们与平均参考光谱的欧氏距离变异也很小。这种情况就表明物质的测量重复性好，而且批号间变异范围也非常小。因此对这种物质阈值的设置也应该很小，但是也必须使其所有的单独光谱在此阈值之内。在这种情况下，一个小的阈值就足以能够惟一鉴别出所有这种物质的单独样品光谱，而且能够使其与其他物质很好地分离开来。相反，具有比较大的平均欧氏距离和标准差的物质，其每条单独光谱到平均光谱的距离比较大，并且距离分布范围也很大。因此为了能包含所有的单独参考光谱。必须为这种物质设置一个相对比较大的阈值。这种大的阈值虽然能够将所有单独样品光谱惟一鉴别开来，但是也会有与其他相邻物质发生重叠的危险，也就是存在将一种物质混淆为另一种物质的可能。

基于单独参考光谱和其对应平均参考之间的欧氏距离，我们可以多种算法优化这个阈值，其中最常用的方法是目标固定算法 (Fixed Algorithm)，它定义阈值 D_t 为：

$$D_t = D_{\max} + \frac{S_0}{4}$$

其中， $D_{\max}$ 为对于给定物质其所有单独参考光谱到平均参考光谱欧氏距离的最大值。

Fixed 算法能够确保给定物质的所有单独参考光谱能被绝对地鉴定出来，因为阈值是由比所有单独光谱的最大欧氏距离还要大来定义的。其阈值中比最大欧氏距离还要大的部分是通过加入一部分单独参考光谱距离的标准差来得到的。加入这个附加因素的目的是为了将来测未知样品光谱时能够允许存在一定误差范围。因为这个附加因素与标准差有关，所以对单独参考光谱具有较大变异范围的物质，其给定的允许误差范围也就较大，这样就能适应将来测这种物质样品时同样也会出现较大变异时的情况。单独参考光谱具有较小变异范围的物质，其给定的允许误差范围也就较小，这样就能适应将来测这种物质样品时其与平均参考光谱的变异范围也很小的情况。

除上面的 Fixed 算法，还有另外一种优化阈值的算法，它被称为常量置信水平方法 (Constant Confidence Level method)。当应用这种方法时，用户必须输入一个优化阈值的置信水平。置信水平可以理解为单独参考光谱在某一置信区间范围内的比率。如果输入的值为 100%，那么阈值将会设为大到能够将所有单独参考光谱都落在这个置信区间之内，

也就是说相当于用最大欧氏距离方法 (Maximum Euclidean Distance) 来设置该物质的阈值。如果将这个水平值设为 90%，那么阈值将会减小，直到有 10% 的单独参考光谱落在此置信区间范围之外。为什么不能简单地将每种物质的置信水平都设为 100% 呢？因为问题出在库中，与我们感兴趣物质相邻的其他物质的置信区间可能与我们感兴趣物质的置信区间发生重叠。当这些区域发生重叠时，任何未知样品光谱如果落在这个重叠区内，也就是说落在了多于一种物质的置信区间内，那么对这一未知样品就不能进行惟一的鉴定了。

10.3.1.4 检验库

当得到平均光谱并且为库中每一种物质都设置了一个阈值之后，必须对这个库进行检验。检验就是检查库的参数是否是优化的，检验通常有两种方法。第一种是检查平均参考光谱的方法，通过检查库中所有平均参考光谱，并且显示出那些置信区间发生重叠的物质。第二种是检查原始参考光谱的方法，通过检查每一条单独的参考光谱来观察每一条光谱是否被惟一鉴定，还是被其他参考光谱相混淆，或者没有被鉴定出来。如果发现重叠或没有被鉴定出来，可以调整相应的阈值以去除重叠或没有被鉴定出来的情况，这可能需要调整优化阈值的方法，甚至会需要对单独参考光谱进行重新评价。检验结果感到满意后，这个库就可以用来分析新样品了。

10.3.1.5 鉴定样品

鉴定的测试结果会有三种情况，这三种情况将在下面用二维空间图方法进行说明。在这个二维空间图中，平均参考光谱和其他光谱都将用二维空间中点来表示，其中两个坐标轴的值是仅有两个波长点光谱的吸光度值。在这个二维空间中的例子中，一种物质的阈值 (D_i) 用一个圆形区域的半径来表示，而这个区域的圆心就代表这种物质的平均参考光谱。这个圆形区域就被称为是一个置信区域，所有与中心点距离（平均参考光谱）小于阈值的点都必须在这个区域范围之内。任一样品与参考物质的欧氏距离 (D_s) 用样品光谱点到参考物质光谱点之间的向量来表示。如果一个新样品只同两种参考物质（A 和 B）做比较，那么可能出现的鉴定结果会有以下三种情况。

第一种情况：

新样品光谱到参考光谱的距离只小于一种参考物质的阈值，比如说是 A，那么未知样品能被惟一地鉴定为物质 A。在这种情况下，如果选择了验证新样品是否为 A 物质，那么 Ident 检验测试结果将会是“可以鉴定为是物质 A”，如图 10-2。

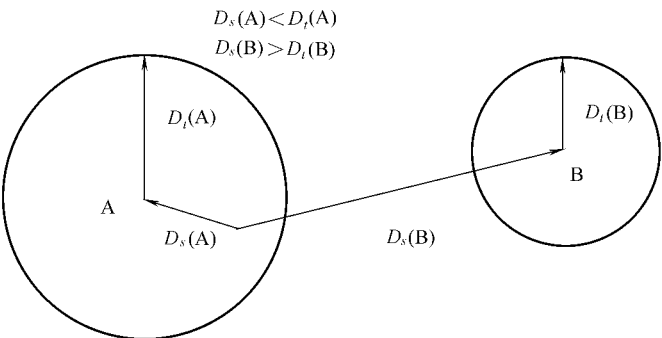


图 10-2 鉴定结果为第一种情况示意图

第二种情况：

新样品光谱到所有参考光谱的欧氏距离小于两种或更多参考物质的阈值，比如说是 A 和 B，那么新样品不能被惟一地鉴定为 A 或是 B。在这种情况下，如果选择了验证新样品是否为物质 A，那么 Ident 鉴定测试结果将会是“与 B 物质相混淆”，如图 10-3。

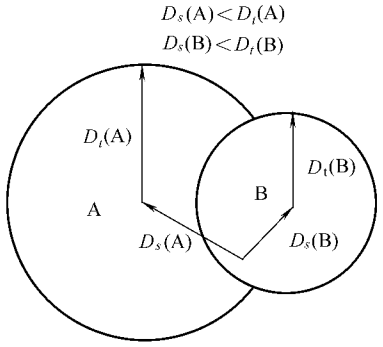


图 10-3 鉴定结果为第二种情况示意图

第三种情况：

新样品光谱到所有参考光谱欧氏距离大于所有物质阈值，在本例中就是大于 A 和 B，那么就认为这个新样品不能被这个库来鉴定。在这种情况下，如果选择了验证新样品是否为物质 A，那么 Ident 鉴定测试结果将会是“不能鉴定为物质 A”，如图 10-4。

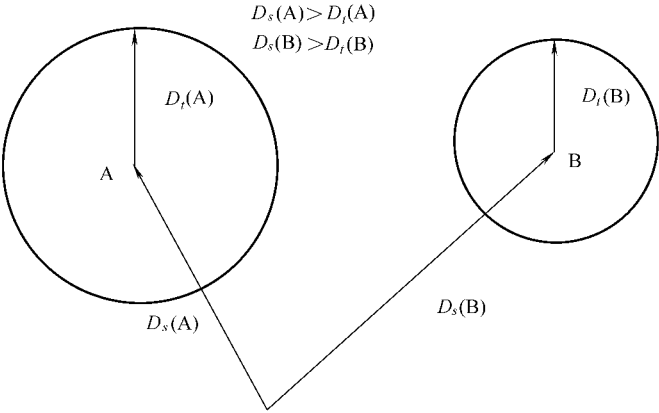


图 10-4 鉴定结果为第三种情况示意图

10.3.2 SIMCA 方法

SIMCA (Soft Independent Modelling of Class Analogy) 法利用样品在主成分空间中不同类别的样品之间的类距离来对类的归属进行判别的方法。设有 q 类样本数据阵 $\{^qX_{mi}\}$ (光谱变量序数 $k=1, 2, \dots, m$; 样本序数 $i=1, 2, \dots, n$; 类序数 $h=1, 2, \dots, q$)，对样品光谱 a ，在类 q 中第 k 个变量被表达成

$$^qx_{ik} = ^qa_k + \sum ^qp_{kj}^qt_{ji} + ^qe_{ik} \quad (\text{主成分序数 } j=1, 2, \dots, A_q)$$

其中： ^qa_k 是类 q 变量 k 的均值； $^qp_{kj}$ 是 A_q 个主成分数； $^qt_{ji}$ 是样本 i 的光谱得分； $^qe_{ik}$ 是样本 i 的第 k 个变量的光谱残差。用主成分空间表征未知样本 ω ，有：

$$Z_k = x_{\omega k} - ^qa_k = \sum t_{\omega j}^qp_{kj} + ^qe_{\omega k} \quad (j=1, 2, \dots, A_q)$$

类 q 的光谱总剩余方差 $^qS_0^2$ 和未知样本 ω 光谱残差 $e_{\omega k}$ 的方差 $^qS^2\omega$ 见下式：

$$^qS_0^2 = \sum \sum (^qe_{ik})^2 / [(n_q - A_q - 1)(m - A_q)]$$

$${}^qS_{\omega}^2 = \sum (e_{ok})^2 / (m - A_q)$$

q 类样本的光谱残差比 (Residual Ratio) 和模为

$${}^qR_a^{\text{spectral}} = {}^qS_{\omega} / {}^qS_0 \quad \text{和} \quad {}^qR_a^{\text{model}} = \sqrt{\sum_i \frac{(t_{ij} - t_j^*)^2}{\sum_{i=1}^{n_q} \frac{(t_{ij} - \bar{t}_j)^2}{n_q}}}$$

$\bar{t}_j$ 为校正集样品光谱第 j 个主成分的平均得分。对所有主成分 i , 当 t_{aj} 大于上限 $t_{j, \text{upper}}$ 时,

$$t_{j, \text{upper}} = t_{j, \text{max}} + \sqrt{\sum_{i=1}^{n_q} \frac{(t_{ij} - \bar{t}_j)^2}{n_q}} \quad t_{j, \text{lower}} = t_{j, \text{min}} - \sqrt{\sum_{i=1}^{n_q} \frac{(t_{ij} - \bar{t}_j)^2}{n_q}}$$

其中 $t_{j, \text{max}}$ 为校正集样品光谱第 j 个主成分得分的最大值, $t_{j, \text{min}}$ 为相应得分的最小值。

组合残差距离 ${}^qR_a^{\text{combined}}$ 为: $({}^qR_a^{\text{combined}})^2 = ({}^qR_a^{\text{spectral}})^2 + ({}^qR_a^{\text{model}})^2$

当样品的组合残差小于临界值时, 可判别该样品归属 q 类; 否则不属于 q 类。

分类研究中一般有两类错误。第一类错误 (type I error) 是拒绝用其他标准可以接受的样品, 把该归入该类的样品拒于该类之外; 第二类错误 (type II error) 是接受用其他标准不该接受的样品, 把不该归入该类的样品归于其中。SIMCA 法分类模型的好坏可用识别率、误判率等指标来评价。

下面介绍一个应用该方法鉴别不同来源的布洛芬原料的实例。

NIR 漫反射光谱法测定布洛芬原料, 以标准陶瓷片为参比。扫描范围 $4000 \sim 10000\text{cm}^{-1}$, 分辨率 8cm^{-1} , 其中湖北类 62 个样品光谱中选取不同批号 20 个样品光谱作为训练集, 获得湖北类布洛芬原料主成分分析结果, 光谱预处理采用一阶导数、5 点平滑和 SNV, 采用留一法的全交叉验证法确定取 5 个主成分, 其累积贡献率为 90.01%。从山东类 88 个样品光谱选取不同批号 28 个样品光谱作为训练集, 获得山东类布洛芬主成分分析结果, 处理方法同上, 取 5 个主成分, 其累积贡献率为 83.15%。再将上述两个结果用 SIMCA 建立分类方法。剩余光谱全部作为验证集, 以检验方法的可靠性。

用 SIMCA 所建立分类模型 (临界概率水平为 0.05), 并对训练集共 48 个样品进行判别, SIMCA 分类模型正确率为 100%。湖北类与山东类类间距离 (interclass distance) 为 5.22。湖北类临界距离 (critical distance) 为 1.31, 山东类临界距离为 1.25。类间距离 5.22 远大于两类的各自类内距离 1.31 和 1.25, 两类有明显的类界限, 因此可以完全区别。

湖北类对自身 20 个训练集样品的识别率 (Recognition Rate) 为 100%, 把山东类 28 个训练集样品全部拒绝, 拒绝率 (Rejection Rate) 为 100%; 山东类对自身 28 个训练集样品全部正确识别, 把湖北类 20 个训练集样品全部拒绝, 具体结果见图 10-5(a)。

采用建立的 SIMCA 分类模型对检验集样品进行分类判别分析, 结果见图 10-5 (b)。湖北类检验集样品 42 个, 犯第一类错误仅为 2 个 (4.8%), 无第二类错误; 山东类检验集 60 个样品, 犯第一类错误仅为 1 个 (1.7%), 第二类错误仅为 1 个 (2.4%), 判别正确率平均为 98%。

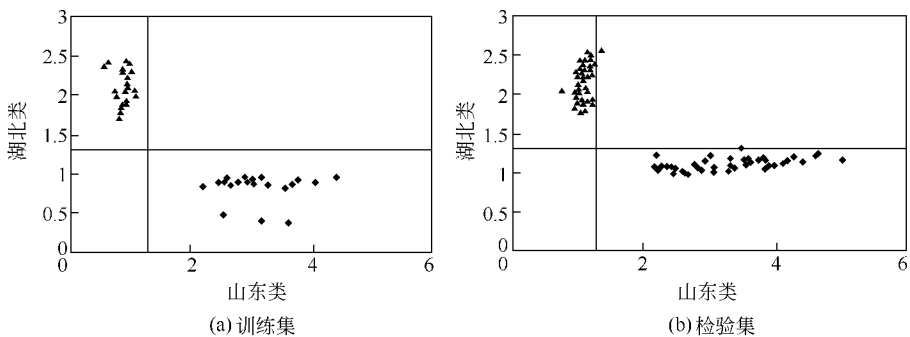


图 10-5 SIMCA 法分类结果

10.3.3 DPLS

DPLS (Discriminant PLS) 是基于判别分析基础上的 PLS 算法，并且以 Y 变量是二进制度变量（类别变量）来取代浓度变量；事实上，DPLS 是用来计算光谱向量 X 与类别向量 Y 的相关关系，取得 X 与 Y 的最大协方差 $\text{Cov}(X, Y)$ ，因此，DPLS 成分可以看作是 PCA 的旋转，并且使得所预测的类贡献值的正交性最大。为了决定混合物中某种物质的类归属， Y 矩阵必须能描述特定种类的样品，混合物中以“1”表示属于这类，“0”表示属于其他类。一般可以设定一个临界值，来判定归属，如有 A、B、C、D、E、F 共 6 类样品，则 Y 值可设定为：

A	1	0	0	0	0	0
B	0	1	0	0	0	0
C	0	0	1	0	0	0
D	0	0	0	1	0	0
E	0	0	0	0	1	0
F	0	0	0	0	0	1

如果待分析鉴别的只有两类，分别用 1, 0 表示即可。

应用 DPLS 进行定性判别分析时，作为样品光谱的输入变量矩阵和二进制度变量（类别变量）即输出变量矩阵之间的关系可用图 10-6 描述。其中 N 表示建模时的样品数， K 表示样品光谱的吸光度点数， M 表示类别的数目。

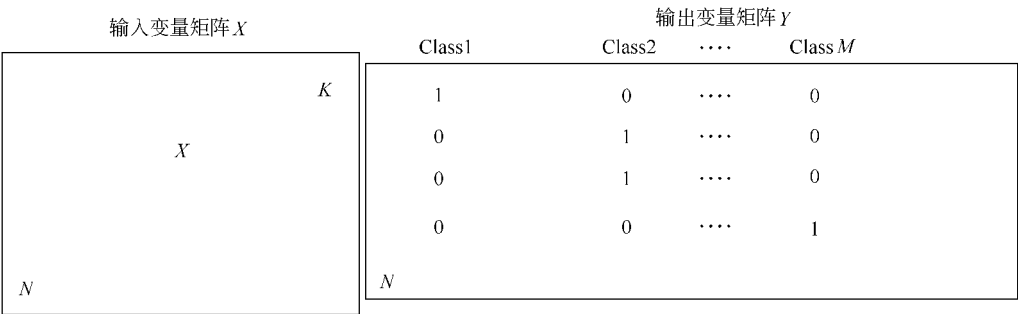


图 10-6 输入变量矩阵和输出变量矩阵（二进制度变量）之间关系示意图

DPLS 法不但可以确定未知纯样的类归属，而且还可以确定混合样主要成分的类归属。

10.3.4 人工神经网络

人工神经网络（ANN）可用来解决许多化学问题，包括非线性多元回归和模式识别。ANN 分类的主要目标是广泛的，在 ANN 的众多方法中，多层向前（Feed-Forward）传递网络形成一个包含有高度连接和互相作用单元的动态系统，适用建立一个非线性计算模型，最常用的多层网络的训练方法是 BP 算法，这种算法基于误差的梯度下降法思想。

由于用来作为人工神经网络计算的输入层数据个数不宜太多，一般不用原光谱数据直接计算。因此，如何有效压缩和提取数据特征显得异常重要，通常可以特征波长的吸收（或吸收比），PCA、FFT、小波变换等方法来提取特征，具体请参见本书第 5 章“近红外光谱分析的信号处理”。

BP（Back Propagation）模型是一种典型的人工神经网络模型，也是比较常用的一种。图 10-7 为该种模型的基本结构，它由三部分组成：输入层，隐含层和输出层。图中圆圈表示神经元，数据由输入层输入，并施以权重传递到第二层，即隐含层；隐含层进行输入的权重加和、转换，然后输到输出层；输出层给出神经网络的预测值或模式判别结果。神经网络通过对简单的非线性（或线性）函数进行数次复合，可近似复杂函数。

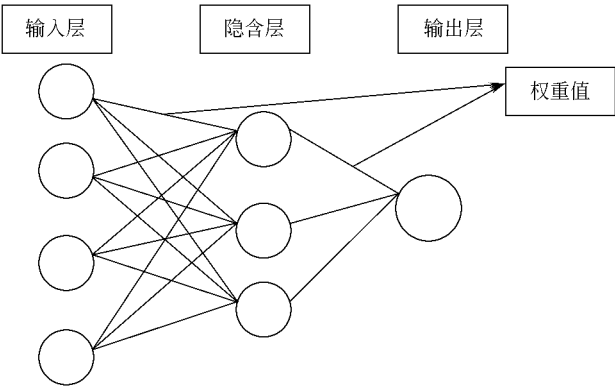


图 10-7 BP-ANN 算法原理示意图

误差反传算法 BP（Back Propagation）为 3 层 ANN 系统，它包括一定数量节点（神经元），图中圆圈表示神经元，而且上一层的每一个节点都同下层的每一个节点连接。同外界的联系通过输入、输出层的节点进行，隐含于内部的中间层决定系统的主要计算能力。在 ANN 的每条连接线上是权重因子，代表系统记忆能力的重要部分，当网络学习时，权重的数值可随着正在网络中流通的新信息而改变，一般应用时先用某种类型的一组已知数据输入 ANN 模型，经过训练后 ANN 就能用学过的已知数据推测出新输入的未知数据。

与 DPLS 类似，ANN 的输出层数据可以是二进制的，如果未知样为包含不同类的混合样，而每个类输出值可能常落在 0~1 之间。在这种情况下，输出值可被认为是可能属于某类的概率，即代表该类在混合样的可能比例，可以选择一个阈值（如 0.1~0.2）来

确定混合物中是否含有某类样品。如果某类的输出值不满足阈值条件，则可认为该样是少数样或未归类的样品。有关人工神经网络的详细算法过程参见本书第 3 章“近红外光谱分析的数学基础”。

下面介绍一个应用 BP-ANN 方法鉴别中药大黄正伪的实例。

大黄是一种比较常用的中草药，按生物形态学划分有很多种类，但从药理学角度又分为两种：正品大黄（药用价值高）和伪品大黄（药用价值很低），而且在晒干粉碎后很难用人眼区分。

NIR 漫反射光谱法测定大黄样品 103 个，扫描范围 $4000 \sim 10000\text{cm}^{-1}$ ，分辨率 4cm^{-1} 。建模样品共有 33 个，其中正品大黄 14 个，训练值用“1”表示；伪品大黄 19 个，训练值用“0”表示。模型的参数如下：

网络的输入节点数：采用光谱的前 8 个主成分数据作为网络的 8 个输入节点；

网络的输出节点数：1 个，即 1 或 0（正品或伪品）；

网络的隐含节点数：4 个；

网络的训练次数：50000 次；

学习效率：0.6；

调整步长：0.1。

训练学习的结果见表 10-1。

表 10-1 中药大黄正、伪判别模型的学习结果

序号	样品名	正伪品 训练值 (0,1)	学习值	序号	样品名	正伪品 训练值 (0,1)	学习值	序号	样品名	正伪品 训练值 (0,1)	学习值
1	1	1	1	12	4	1	0.8	23	70	0	0
2	13	1	0.9	13	40	0	-0.1	24	74	0	0
3	14	1	0.8	14	42	0	-0.1	25	77	0	-0.1
4	17	1	1	15	47	0	0	26	79	0	-0.1
5	18_1	1	0.9	16	5	1	1	27	8	1	1
6	20	1	0.9	17	55	0	-0.1	28	82	0	0
7	23	1	0.8	18	59	0	-0.1	29	85	0	0
8	26	1	1	19	63	0	0	30	88	0	0
9	27	1	0.9	20	65	0	-0.1	31	92	0	0
10	3	1	0.7	21	67	0	-0.1	32	96	0	0
11	30	1	0.9	22	69	0	0.1	33	97	0	0

如果对学习输出值的阈值设定为 0.2，则判断为正品大黄的下限值为 0.8，上限值为 1.2；判断为伪品大黄的下限值为 -0.2，上限值为 0.2。从表 10-1 中数据可以看出只有第 10 号样品为不能归类的样品，如果对学习输出值的阈值设定为 0.3，则所有样品都能正确归类。

应用该模型对剩余的 70 个样品的判别分析结果见表 10-2，判别分析的效果请读者自己分析。

表 10-2

中药大黄正、伪判别模型的分析预测结果

序号	样品名	正、伪 品属性	分析值	序号	样品名	正、伪 品属性	分析值	序号	样品名	正、伪 品属性	分析值
1	10	正品	0.8	25	41	伪品	0	49	72	伪品	−0.1
2	11	正品	1.1	26	43	伪品	−0.1	50	73	伪品	0.2
3	12	正品	0.9	27	44	伪品	−0.1	51	75	伪品	0.2
4	15	正品	1.1	28	45	伪品	−0.2	52	76	伪品	0.3
5	16	正品	0.7	29	46	伪品	0.1	53	78	伪品	−0.1
6	18_2	正品	0.9	30	48	伪品	−0.2	54	80	伪品	−0.1
7	18_3	正品	0.9	31	49	伪品	−0.1	55	81	伪品	0
8	19	正品	0.9	32	50	伪品	−0.1	56	83	伪品	0
9	2	正品	0.9	33	51	伪品	0	57	84	伪品	−0.5
10	21	正品	0.7	34	52	伪品	0.1	58	86	伪品	−0.2
11	22	正品	0.8	35	53	伪品	0.1	59	87	伪品	0.3
12	24	正品	0.7	36	54	伪品	0.2	60	89	伪品	0
13	25	正品	0.9	37	56	伪品	−0.2	61	9	正品	0.8
14	28	正品	0.8	38	57	伪品	−0.1	62	90	伪品	−0.2
15	29	正品	0.8	39	58	伪品	0.1	63	91	伪品	0
16	31	正品	0.9	40	6	正品	1	64	93	伪品	0.1
17	32	正品	0.9	41	60	伪品	−0.1	65	94	伪品	0
18	33	正品	1	42	61	伪品	−0.2	66	95	伪品	0
19	34	有争议	0.6	43	62	伪品	−0.2	67	98	伪品	0.1
20	35	有争议	0.6	44	64	伪品	0	68	99	伪品	0.9
21	36	有争议	0.8	45	66	伪品	−0.1	69	100	伪品	0.7
22	37	有争议	0.6	46	68	伪品	−0.1	70	101	伪品	0.7
23	38	有争议	0.6	47	7	伪品	0.7				
24	39	伪品	−0.1	48	71	正品	0				

10.3.5 判别分析

在近红外定性判别分析中，尤其是工业分析中经常遇到这样的情况，即要判断这批原料的真假或与目前的原料是否相同，中间体或产品的纯度与均匀性问题等，这就需要利用测得光谱来对样品进行归属指定，即要决定该样本是否归属某一个总体，这就是差距分析的应用范围。在多元统计分析中，常用的有马氏距离判别法、贝叶斯判别法、费歇判别法、模糊判别法、逐步判别法等，这些判别法的原理及应用请参见本书第 3 章“近红外光谱分析的数学基础”及相关文献，本章中不做详细介绍。

10.4 近红外分类学研究——无监督模式识别方法

有监督模式识别方法是基于一组先验知识来做的判别，但在自然科学和社会科学研究中，面对世界不断涌现的新知识、新数据和新信息，更大量的情况是在无先验知识条件，依一定规律对已有的信息进行分类研究处理并加以利用，这种分类方法称无监督模式识

别。用这类方法对大量的样品进行分类，而且事先并不知道可能分成几类和每一类的特征，只是采用一些数学方法使得同一类样品性质相似，而类与类之间各不相同。常见的方法有系统聚类、最小先成树、映射图等。

对于表面上毫无规律的大量数据来进行分类研究而言，聚类分析是一种建立分类模式的有效方法，它基于的主要思想是“物以类聚”，即同类样本的性质（特征）是相似的，在多维空间中，距离相对小些；相反，不同类的样本的（特征）性质不同，则距离相对远些。其中系统聚类主要过程为：① 计算样本之间和类与类之间的距离；② 在各自成类的样本中将距离最近的两类合并，重新计算新类与其他类间的距离，并按最小距离归类；③ 重复过程②，每次减少一类，直至所有的样本成为一类为止。选择不同的变量（特征）或采用不同的类距离计算方法将得到不同的聚类结果。有关聚类方法的原理及应用请参见本书第 3 章“近红外光谱分析的数学基础”及相关文献，本章中不做详细介绍。

10.5 定性判别分析的结果评价

近红外定性判别分析的基本思想是应用获得的光谱来对样品进行分类，与定量分析有所不同的是定性判别分析的结果应先判别样本的类归属，类归属的判别正确率、错误率等都可以用来评价定性判别方法的优劣。

参 考 文 献

- [1] Stark E, Luchter K, Margoshes M. Appl Spectrosc Rev, 1986, 22 (4): 335
- [2] Sabdra E Kays, W R Windham, Franklin E Barton. J Agric Food Chem, 1998, 46 (3): 854
- [3] Campbell M R, Brumm T. J Cereal Chem, 1997, 74 (3): 300
- [4] 陆婉贞等编著. 现代近红外光谱分析技术. 北京：中国石化出版社，2000：193~226
- [5] Morisseau K M, Rhodes C T. Drug Development and Industrial Pharmacy, 1995, 21 (9): 1071
- [6] Merckle P, Kovar K A J. Pharm and Biomed Anal, 1998, 17 (3): 365
- [7] W Plugge, C van der Vlies. J Pharmceut Biomed Anal, 1993, 11: 435
- [8] Olof Svensson, Mats Josefson, Frans W Langkilde. Applied spectroscopy, 1997, 51 (12): 1826
- [9] Ulmschneider Michel, Barth Gabriele Trenka Eszter. Drugs Made Ger, 2000, 43 (3): 71
- [10] 布鲁克公司近红外用户培训手册

11. 异常样品的分析

本章主要讨论在近红外分析中，异常样品（Outlier）形成的原因、重要性及找到异常样品后应当采取的办法。提出根据某样本的影响系数和参差大小作为标准来发现异常样品，并用实例（石蕊溶液定量分析的数据）对各种标准进行检验的模型。重点在于叙述双线性模型。

11.1 近红外光谱分析中异常样品的概述

在现实世界中错误和意外现象是不可避免的，近红外光谱（NIRS）分析也不例外。在扫描中的操作错误、实验室环境例如温度或电压的异常变化、仪器的噪声、波长的漂移等都可能产生异常的光谱，这个样品就称为异常样品或奇异样品，它的光谱也就是奇异光谱。另外，近红外的计量方法是根据已知样品建立拟合模型，而且通常为线形模型。如果某样品的光谱与定标样品差别较大，则该样品也是异常样品。前一类异常样品产生的原因是误操作或仪器异常，发现后可以简单地通过重新扫描得到纠正；而后一类异常样品是由于样品本身所产生的，不能简单地通过重新扫描得到纠正，对该样品的预测值是否可靠取决于其光谱变异与模型的拟合程度。因此，这里就出现两个关键的问题，第一是如何发现异常样品，第二是如何区分该异常样品属于哪一类，本章将围绕其进行讨论。

另外，需要强调的是：任何一个奇异光谱都是相对的，如相对于一个定标模型，或相对于一个光谱库等，也就是说，同一个样品对于定标模型 A 是异常样品，而对于定标模型 B 可能就不是异常样品。从另一个角度讲，如果一个定标模型或一个光谱库所包含样品的变异性越大，则产生异常样品特别是第二类异常样品的可能性就少。

通常 NIRS 分析中，异常数据有 4 大类：X 错误、Y 错误、定标系中的错误和检验系中的错误。当然，并不是说异常数据就是无用的错误，有些异常数据给出了极其有价值的信息。

11.1.1 异常样品的危险性

11.1.1.1 可预期的误差

在现实中，错误和未预料到现象是不可避免的。无论在对未知领域的研究中还是在日常的实验分析或在工业流程检测中都是如此。许多现象如操作错误、噪音尖峰信号、仪器偏差及异常样品类型等都会影响定量分析的结果。找出问题所在要比对这些问题视而不见好得多。因此对于异常的样品的检测是非常重要的。

就这点而言，传统的单变量的定标曲线是非常有危害的。多变量的定标为自动检测奇异点提供更有效的工具。这常有几种不同类型的基本误差能被发现和处理好：变量 x 中的误差， y 中的误差，定标中的误差及预测值的误差等。并不是所有不正常的现象都是无

用的错误，事实上有些不正常现象会提供非常有用的信息。

首先用自己的分光光度计——眼睛进行类比分析来发现异常样品，然后再举出一个用真正分光光度计的例子，并用各种计算方法把异常样品检测出来。

11.1.1.2 学会从咖啡中找出茶的一种类比分析

假设要学会怎样准备一杯浓度适宜的茶，当然可以现场查书看用几勺茶叶、用多少开水及泡多久时间，但观察每杯茶的颜色是最好的检查方法，因为按书本上做时，经常出现一些操作的错误。

如何才能学会鉴别一杯茶的颜色和浓度？必须通过比较每杯茶的颜色（ x ）和它们的浓度（ y ）来训练我们的眼睛。考虑一下茶杯中会出现的一些不正常现象，首先假设在实验过程中有一杯是白开水，其他各杯是不同浓度的茶水，这杯白开水自然被看成一个极端值。因为比较其他各杯茶水来说含有太少的溶质。相反，如果一杯茶水颜色特别深，并且苦的不能喝，那么它也可以被发现为一个极端值，因为它的颜色太深了。

像这样极端的样品会严重影响对于棕色茶和茶水浓度的判断。但只要这种极端值在其他方面都是好的，这种极端值的差距范围还是有用的。因为它使建立由颜色变化判断浓度的模型更方便一些。如果将那些非线性因素包括在模型中，这种极端值与正常值的太大差距就会使得无法观察到茶水正常浓度范围内的细微差别。

在这个训练眼睛过程中，一些不相关的或错误的数据会使得对一杯茶看起来究竟是什么样的产生错误的印象。如果一杯浓咖啡放在这些茶水中，在学习怎样通过颜色来判别茶水浓度时，最好把这杯咖啡当作不相关的因素。如果错误地将一杯淡红茶作为是一杯浓茶（绿茶）时，同样也会产生问题，在这种训练中，这种错误的极端值产生的不良影响有两个方面：

① 对不同浓度茶水的颜色产生错误印象，使得在以后预测未知茶水浓度时做出错误判断。预测模型 $Y=f(X)$ 会出现系统错误。

② 错误的定标数据使得设计一个非常错误的模型，在这个模型中，茶水的颜色变化范围可以很大，这样不利于预测杯中盛的是否是茶水。在这个极端的例中，由于不能学会识别出一杯真正的棕色茶，使在以后会把绿茶、橘子汁或咖啡当作棕色茶水。

还有一种情况，一些人在茶水中加入牛奶而另一些人不喜欢这样做。那么在这种情况下，怎样才能判断出这些茶水的浓度呢？如果在训练过程中没有见过加入牛奶的茶水，在以后的时间里，由于它的混杂就会把它当作一种非正常现象。

另一方面，如果进行了一些已知茶水浓度，并在这些茶水中加入牛奶和不加牛奶的识别训练，那么就不会把牛奶作为一种干扰因素：可以建立一个有棕色因素和白色因素的双因素模型。尽管受混杂现象的影响，还是可以根据棕色浓度的增加来判断茶的浓度的增加，而随着牛奶加入量的增大，茶中白色的增加可以补偿茶棕色的减弱。

还可以建立多个茶水浓度测量的模型。一个是加牛奶的茶水浓度模型，一个是不加牛奶的茶水浓度模型。

怎么来进行极端数据的检测？无论如何，白色的过浅或过深都会显示杯中都含有少量或大量的牛奶。棕色的深度依然是说明茶水浓度大小的基础，但应当考虑牛奶的白色对棕色的减弱作用，这样才不会把无色的绿茶和橘子汁当作红茶献给客人。

这里不讨论应该运用哪种定标模型。所举的例子仅仅说明了感觉器官进行多因素判

断。正因为人类有这种判断极端数的能力，正常生活才得以维持。

11.1.1.3 在实验室和生产加工中的极端值

通常意义上的极端数是指那些与数值的绝大部分相比，是不相关的、在总体上错误的或在某些方面异常的数据。观测目标，变量和单个的数据却可能是极端数。

一个奇异的样品，其 X 、 Y 向量与其他样品有很大的差异，其出现的原因可能是数据出现错误（噪音、仪器偏差、操作错误、印刷错误等），也可能是因为所分析样品的结构或物理状态的特殊性。这里所说的突出的或极端的变量 x_K 或 y_I 是指在观测过程中变量数据与模型中所使用的其他变量的数据有很大差别的变量。或者还因为高水平的随机噪音，或者因为它单独反映了事物的个别特性（特殊的化学、物理现象、或特殊的非线性类型），一个极端值数据 x_{ik} 或 y_{ij} 可能是噪音尖峰信号，也可能是印刷错误。

在传统的化学分析中，极端值的检测往往是单凭化学家们的主观直觉，而且这对于获得正确的分析结果又是一个不容忽视的重要步骤。但是存在许多类型的极端值，有些极端值不是轻而易举能发现的，就像演奏乐曲，一根弦发出的杂音是不能显示出整个乐曲的不和谐的。有些类型的关系模型是通过人的大脑或是通过计算机建立起来的。

这一点对于日常的分析特别重要：现代可运用的多变量计算工具产生了太多的数据用于计算，这样仅靠人的大脑是无法做到的；更糟的是，许多人仅仅因为是计算机分析出来的结果而绝对地相信，丝毫不加怀疑，为了保证正确的运用现代的分析工具，极端值检测理所当然的成为该仪器使用过程中必不可少的组成部分。

11.1.1.4 定标和预测中的极端值检测

在定标过程中，发现并删除和改正 x 和 y 数据中的极端值是很重要的，否则将降低定标模型的预测能力。在对未知样品进行预测的过程中备有检测出异常现象的方法也很重要，这样可以提高预测结果的准确度，有时未预测到的极端值的发现可能会提供新的科学信息。

运用多变量定标模型可以根据不同的标准来检测极端数。再来看一下定标模型的本质：选用数学模型来表示 x - y 之间， x - x 之间的关系（如果再加上一个成分时，还可以表示出 y - y 之间的关系），模型的选择取决于对这些数据关系的假设（预期的干扰因素、噪音水平、概率分布和线性），运用定标系中提供的数据，通过一些数学方法可以估计出模型的参数，运用建立的定标模型可以得到用 x 、 y 作为变量计算出的这类目标的数值以及正常的可接受的噪音水平。

与那些建立定标模型的数值进行比较，就可以发现某个样品、变量或数值是异常的，这种检测方法不仅可用于定标过程，也可用于求未知样品的预测中。

在本章的开头谈到通过感官的极端刺激，简单地发现异常样品。在定标过程中可通过对模型参数（例如载荷向量、回归系数向量）和预测残差的图像化来发现由于掩饰作用而表面看不出的奇异点。另外可以运用更先进的工具如库克和威茨柏格（Cook and Weisberg）在 1982 年提出的来找出复杂的极端值。

11.1.2 怎样来处理极端值

11.1.2.1 极端值的优劣

极端值非常重要并且在实际分析过程中是不可避免要出现的，必须学会处理极端数并

从中学会一些新的知识。因此极端数必须被严格处理：一旦发现极端数不利于建立模型，必须修改或删除它们。

如果有可能应当尽量了解每一个被检测出来的极端数产生的根源，有时做这些是比较容易的。例如，含量最高或最低的样本经常被当作是极端值，如果仪器显示为正常的线性，这些极端值是非常有价值的，应该保留在定标模型中。有过多或过少不同干扰因素的样品也可能轻易地被作为极端值检测出来。这些都是“好的”极端数。

劣的极端数通常是那些不能理解而且严重影响模型的极端数，对于这些极端数，特别是错误较大的极端数最有效的方法是检查一下是否出现打印错误、扫描错误等简单的错误，并纠正它们。但是要了解可疑的极端值往往要花费更多的时间，例如把样品再做一次化学分析和仪器分析。

然而，明智的做法就是首先把这些不良的极端值从数据中删除掉，重新建立定标模型以保证极端值不影响定标和预测。对于那些由于操作错误和设备出错所造成的极端值，这是一个明智的做法，当有大量的有用数据而极端数数量较少时，可以首先考虑这种方法。但除非是在探索性的实验初期的数据分析中，一般不宜过多地删除极端值。如果读者希望对关于如何删除极端数的数理统计方法有更多的了解，请参考 Huber 的《Robust statistics》。

11.1.2.2 极端值的分类和分割模型

极端值的检测、分类和分割模型之间有极其紧密的关系。多元定标就是根据定标系中的数据，用个简单数学关系来反映未知的 X 与 Y ， X 与 X 之间关系。用一个线性定标模型来反映不同类型的样品，总体上是可行的，但也会出现一些不能解释的样品或数据，也就出现了奇异点。

假设所取得的样品不是来自一个自然总体，例如用 NIR 光谱 (x) 来分析肉类和鱼类产品中的脂肪含量 (y)，由于两类样品的不同， X 与 Y 之间及 X 与 X 之间的关系当然有很大的差别。

如果我们只对肉类产品定标，那么在肉类样品中所混杂的一二种鱼类产品可能被当作是肉类测量模型中的极端值。为了得到最优的肉类产品模型和预测，必须把鱼类极端值删除掉，重新建立模型。同样的，在鱼类样品中的一二种肉类产品也会被认为是鱼类观测模型中的极端值，在重新建立模型时，也应当先把它们删除掉。

两个定标模型对极端值的检测是以样品的光谱数据 (x) 的分析为依据，来判断新的样品：一种未知的肉类产品可以划分为符合肉类模型，而对于鱼类模型来说它则是一个极端值，因此这种肉类产品的脂肪含量可以从肉类模型中估测出来。同样，一种未知的鱼类样品应当被作为鱼类产品，而不是肉类产品。

如果在定标中有同等数量的鱼类和肉类样品，可以对两种类型的产品放在一起进行定标。但是这样会产生一个比较复杂的定标模型，这个模型与单个肉类或鱼类模型相比较，其预测误差不是最低，但是有较高的极端值容纳能力。

在这种情况下，最好把定标分成若干具有相同性质的小类（例如将肉类产品和鱼类产品分开），然后对各个局部的小类分别建立一个专门的模型。这种分割可以通过对样品有充分的了解，并且结合着定标模型参数的图形分析（如每个因子的得分图）进行。但建议最好通过聚类分析来进行模型分割。

11.1.3 含有极端数的数据组

下面的例子用于说明发现极端值的不同的检测方法。假如我们想要通过波幅范围为 400~800nm 的光谱来分析一下 pH 为 10 的石蕊溶液浓度，用 17 个样品建立定标模型，用 6 个样品来检验定标模型。

首先研究一下定标系：17 个样品光密度与浓度的关系，其中纵坐标为 576nm 处光密度（在该处石蕊溶液吸光度最大），横坐标为石蕊溶液浓度。数据表明大多数数据的样品的溶解度在 0~50% 之间，大多数溶液的样品正好落在一条直线上，而观测目标 3 作为一个极端值落在直线的延长线上，观测值 1、2、7 则明显不在直线上。

17 种石蕊溶液的光密度光谱，最低的光谱代表最低的溶解度（ $y=0$ ）。把 17 个样品描绘在 50 个波长点的其中两个上，直接用眼睛就可以发现一些极端值。但怎样才能自动地把它们检测出来？还有其他的极端值吗？这些问题将在下一节的双线性方法检测极端值中进行阐述。表 11-1 给出了已知这 17 个样品所设定的背景特征。

表 11-1 实例中各样品（pH 为 10 的石蕊溶液）的背景说明

定标系	17 种样品,其中 11 个是正常的,含有不同量的溶质,不含杂质。并人为制造 6 个异常样品	
	样品号	奇异值类型
	1	物理性状干扰:光的散射和混浊度。原因是加入白色 ZnO 粉末引起
	2	化学性状干扰:分子状态变化,其原因是加入 pH=4 的酸液使其颜色变红
	3	极端含量,含有 100% 的溶质
	7	y 数据出现错误:真实的溶质溶解度有错误,印刷出现错误,将 $y=33\%$ 打印成 3%
	11	数据 x 出现错误,25 波段即位于溶质吸光度顶峰上出现测量噪音
	14	数据 x 出现错误,4 波段,不在吸光度峰值上出现测量噪音
检验系	6 个独立样品,样品号 18~23,不存在极端值	

预测系中的 6 个样品是正常的蓝色石蕊溶液，浓度处于正常范围（ $0 < y_i < 50\%$ ），这些观测目标是不会产生极端值的。

11.2 双线性模型中的极端值检测

11.2.1 定标中的极端值

对于只有一个 y 变量的双线性模型也可以写成：

$$X=1X'+TP+E; Y=1Y'+Tq+f$$

(11-1)

在这个模型中最基本的检测极端值的方法是建立在影响系数和残数的基础上。下面分别讨论一下这两种情况。并且研究一下它们各自所能检测出来的极端值的类型。

首先用上节所举例子中的 17 个定标样品所构建测定石蕊溶液浓度的双线性 PLS 模型。图 11-1 表示随着 PLS 因素数的增多，定标模型对溶液浓度分析能力的变化：交叉检验曲线表明在观测数据中存在着明显的极端值，因为这条曲线远离正常模型曲线，这就是

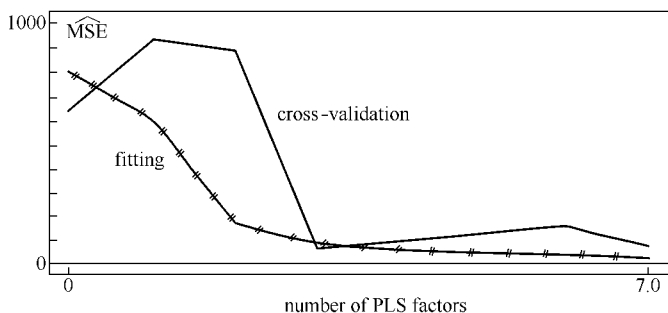


图 11-1 在石蕊溶液浓度例中，定标（点线）和检验（直线）的误差随模型复杂程度变化图

说明在建立最终观测模型时必须改正或删除一些极端值。

如果完全了解数据出现奇异的原因通常可以运用交叉检验曲线和具有最佳预测能力的拟合模型曲线相比较的方法检测出极端值。在这个实例中，即 PLS 因子=3。

这将有希望检测出特别严重的极端值，然后可以删除这些极端值并重建模型。PLS 因子分别为 1, 2, 3 时，载荷向量 $\hat{P} = (\hat{P}_a, a=1, 2, 3)$ ，它们表示 x 数据对构建模型的重要性。表明 3 个载荷向量图都是平滑的，绝对值差别不大，因此这 3 个向量对于 x 模型的建立是同等重要的。因此至少说明表面看来这 3 个因素不是随机噪音。

11.2.1.1 定标中具有极端影响系数的数据

样品的影响系数 (Leverage) 这个概念来自于库克和威斯伯格。它表示观察中各个独立变量（或 x 变量）相对于其他变量的影响，这里对上面所用的线性模型的最初概念进行修正：

$$h_i = 1/I + \hat{t}_i' (\hat{T}'\hat{T})^{-1} \hat{t}_i \quad (11-2)$$

这里 $\hat{t}_i = (\hat{T}_{ia}, a=1, 2, \dots, A)$ 是样品 I 的因子向量； $\hat{T}'\hat{T}$ 为整个定标系中因子得分矩阵， $1/I$ 表示截距的影响，它可以从公式中删掉，在这里保留它仅仅是与标准统计著作中的 h_i 点相一致。

正规的 h_i 点又称为马氏 (Mahalanohis) 距离，表示加权的协方差矩阵。当 $A=K$ 时，这里的影响系数 h_i 与通常的影响系数点又相一致，它表示出样品 i 与定标中心值的全程 Mahalanohis 距离。作为有最佳预测能力的回归方法，影响系数 h_i 应和表示截距的 Mahalanohis 距离一致，因为影响小并且不确定的 x 数据中的变量已经删除掉。建立这个定义最基本的原因是双线性模型中，只用 A 个因子做回归分析。在这种情况下，建立能评判 A 维变量空间标准是非常重要的。如何建立这个 A 维空间，将在下一部分中叙述。

从库克和威斯伯格的研究来看， h_i 的范围是在 $1/I$ 与 1 之间，如果一个样品的影响系数低（接近 $1/I$ ），则表明在 A 维空间，该样品接近定标系 x 测量值的中心值。因此这个样品对于定标模型的影响微乎其微。但接近 1 的影响系数表明该样品远离中心值，它对于建立 A 维测量模型有很大的影响。产生这样高的影响系数的点的原因可能是：样品有极端浓度的溶液或 A 维双线性模型中包含了较强的干扰因素。

因为它是由 x 和 y 对回归计算而获得，而且最小二乘法的估计对判断奇异点是较敏感的，所以有较高影响系数的样品会对 x 和 y 的载荷阵 $\hat{p}$ 和 $\hat{q}$ 产生很大的影响，此外它

还会对权重 $\hat{V}$ [通过 $\hat{p}$ 和 (或) $\hat{\omega}$] 产生主要影响, 从而对 X 如何导出产生影响。因此如果有高影响系数的样品, 它的 Y 或 X 数据出现错误或其他不希望的不正常现象, 那么在定标中使用该样品是不利的, 它会降低所获得的定标模型的预测能力, 另一方面, 如果奇异点反映的是有效的信息, 这样的极端数最好不要被删除掉。

定标系中 h_i 的平均值为 $(1+A)/I$, 提示异常样品的预警值通常为 $(1+A)/I$ 的 2 或 3 倍, 即当某样品的 h_i 大于这个预警值时, 就应该引起注意。(如想继续了解请读 vellemann 和 velsch 于 1981 年的论文)。

对于 PCR 和 PLSR 方法的定标模型, $\hat{T}'\hat{T}$ 为矩阵对角线矩阵, 给出一个正交得分, 因此影响因子 h_i 的计算可简化为:

$$h_i = 1/I + \sum_{a=1}^A (t_i \hat{a}^2 / \hat{a}' \hat{a}) \quad (11-3)$$

这里可以看到影响系数仅仅是 X 得分加权值之和, 每个因子的权重都是定标系中 T 平方和的倒数。用这种方式很容易就可以看出每个因子都进行了加权, 结果它们对影响系数大小的影响是可比的: 在 X 中因为某个小因子而产生的奇异值与某个较大因子而产生极端值是同样重要的。

尽管根据 h_i 分析来发现奇异点时, 利用各个因子得分总和的优势, 但它并不能告诉极端值的产生来自于哪个因子。如果每个单独因子表示一个特定现象 (这在控制的实验或通过轴旋转可以达到)。那么对这些单个因子进行极端值检验是非常有用的。极端值检验可以通过计算每个因子的偏影响系数 h_{ia} 来检验, 对于正态因子, 可简化为:

$$h_{ia} = \hat{t}_i \hat{a}^2 / \hat{t}_a' \hat{t}_a \quad (11-4)$$

平均的偏影响水平由定标系的 $1/I$ 决定, 因此在这里实用的极端值检测范围是 $h_{ia} > 3/A$ 。

研究一下测定石蕊溶液浓度的例子, 在这个例子中只用了一种光谱变量 (不同的溶质浓度) 和线性模型, 如果数据中不出现异常, 随着溶质浓度的增加, 在 K 维 X 空间上每个样品将落在一条直线上。如果该变量在 A 因子上进行模型化, 那么在 K 维 X 空间上成一条直线排列的点在 A 维 T 空间上也应排列成直线。

在这个例子中, 单独地看一下 $A=3$ 时的模型, 运用包含三个因子的最初定标模型中 17 个样品的得分, 在该模型空间里大多数的值实际落在一条直线上, 但是有些样品特别是 1、2、3 号样品明显与其他样品有所不同。在最初的因子为 3 的 PLSR 模型中, 仅为一种浑浊溶液的 1 号样品和仅为红色溶液的 2 号样品具有很高的影响系数, 这两个异常样品都是异常的系统干扰, 其 $h_i > 0.99$ 。而浓度异常高的 3 号样品也被警告作为异常样品, 但其影响系数较小, $h_i = 0.67$ 。

在 Martens 和 Naes (1987) 的著作中也可以看到运用 h_i 的例子。可以用不同的方式对模型进行调整来估计样品对整个定标的重要性, 其目的是突出对 Y 有重大影响作用的因子。例如用 $\hat{t}_i \hat{a} \hat{q}_{ja}$ 而不仅仅用 $\hat{t}_i \hat{a}$ 来进行 Y 权重的影响系数计算。

同样地, 可以研究一下不同 X 变量影响系数的计算。在 PCR 中计算式可以用正交矩阵 $\hat{p}$ 的形式表示:

$$h_K = \hat{p}_K' \hat{p}_K \quad (11-5)$$

在来源于正交载荷权重 ω 的 PLSR 中,

$$h_K = \hat{W}'_K \hat{W}_K \quad (11-6)$$

h_i 接近零时, 表示该 X 变量对定标模型的影响不大; 而其接近 1 时表示该 X 变量对模型的影响很大, 这个变量可以单独作为回归因子。 h_k 的平均值为 A/K , 当 $h_k > 2A/K$ 或 $3A/K$ 时, 则发出奇异值的警告。

对于以上的例子, h_k 的警告限为 $2 \times 3/25 = 0.24$, 而最大的 $h_k = 0.14k$ (对波段 $k = 23$ 而言)。因而在三因子的 PLSR 模型中没有发现特殊的 X 变量。因为异常高浓度的石蕊溶液、红颜溶液和浑浊溶液这三种现象没有一个进入该模型。

11.2.1.2 光谱数据 X 中未加模型化的残差

在多元定标中, 从分析仪器中 (如连续扫描型分光光度计) 得到的 X 变量的数目总是比建立 BLM 定标模型所需要的 X 变量数多得多。这表明会产生变量过多的模型, 因此也就可能产生 X 残差, X 残差说明了定标数据 X 与 $\hat{V}$ 矩阵确定的模型的不适合性、 X 中的随机噪音、非模型的干扰、或非线性的影响等。 X 变量的定标残差可以简单地表示为:

$$\hat{E} = X - 1\bar{X}' - \hat{T}\hat{P}' \quad (11-7)$$

对每个样品来说, X 残差表示为:

$$\hat{e}_i' = X' - \bar{X}' - \hat{t}_i' \hat{p}' \quad (11-8)$$

表中的每个数据的元素残差表示为:

$$\hat{e}_{ik} = x_{ik} - \bar{x}_k - \sum_{a=1}^A \hat{t}_{ia} \hat{p}_{ka} \quad (11-9)$$

X 残差对目标变量有多种意义, 首先, 它们可以检验出与其他数据不符的样本、单独的记录值。进一步研究 X 残差可以知道其原因是由于测量错误还是由于特殊因素造成的系统错误。

一般来说, 若在某个样品测量过程中存在较大的残差或残差形式异常, 说明该测量过程中有其他样品所不具有的干扰因素, 或说明该观测没有经过双线性模型修正。具有不正常形式的残差可以被解析为比其他变量较高的噪音水平, 单个的随机残差值 $\hat{e}_{ik}$ 大则表明由于特殊样品造成的 X 的异常值。

最简单有效的检验残差的方法是画图分析, 对此有许多不同的画图方法, 而且有许多有用的技巧。最直接的方法是在每个观测值 (或波长点) 旁注上各自的残数值。

所列举的 17 个样品经过三因子的 PLSR 模型化后的光谱参差对每个波段 ($K = 1, 2, \dots, 50$) 作图。大多数的样品光谱残差几乎在所有波段上都落在 $\pm 0.01\text{OD}$ 之间。但是在 25 波段上, 11 号样品表现出极大的残差尖峰; 在 4 波段上 14 号样品表现出极大的残差尖峰。这并不奇怪, 因为这是人为加上取得, 只是通过对光谱参差的分析, 将其显现出来。

也许更令人感到奇怪的是 11 号样品在 25 波段两旁出现了两个负波谷。因为在这些波段上并没有特别的错误。原因是由于 11 号样品的误差尖峰正好落在位于 PLS 因子 1 和因子 3 具有非零载荷向量 $\hat{p}_{kni}$ 的波长范围内, 因而 25 波段明显的正误差尖峰影响 PLS 因子的得分, PLS 因子 1 得分太高, 而因子 3 得分太低, 这就造成了另外的 X 残差。14 号样品的误差尖峰所处的位置内, 各因子没有重要的载荷向量, 因此只产生了较小光谱残差,

这说明确了对已经取得的残差值的解释不能只限于它们是由于最初的测量造成的这一说法，因为它们已经过了双线性模型的修正。

为了进一步检验这种 X 光谱的残差值，可以把它们已知光谱值代入相同的双线性模型获得的各自的光谱残数值进行比较。这一过程可以提供大量的信息，在平常操作过程中或其他观测过程中，如果出现残差值较大的样品时，有必要进行完整的检查。

(1) X 的总残差标准差

含有一系列样品 ($i=1, 2, \dots, I$) 的定标模型，在确定了因子数后，总残差标准差 $S(\hat{e})$ ，可以定义为

$$S(\hat{e})^2 = \sum_{i=1}^I \sum_{k=2}^K e_{ik}^2 / df \quad (11-10)$$

其中 df 表示残差 $\hat{e}_{ik}$ 的剩余自由度。

df 决定于定标的样品数、 X 变量的数量、所选用的定标模型以及定标的方法。因此要准确计算 df 是很复杂的。幸好 df 的准确度对于极端数检测并不是特别重要。Marten 和 Jensen (1983) 对 PLS 回归的 df 做了粗略估计，公式为：

$$df = IK - K - A(\max(I, K)) \quad (11-11)$$

请注意：由于 $S(\hat{e})$ 是对 E 项方差的近似估计，因此可以从公式中省略掉。但是，对于双线性模型 E 的定义和假设在总体上比较模糊，在此章中还保留了 $\hat{e}$ ，为的是与其他严格的误差项估计相区别。

(2) 定标系中每个样品的 X 残差标准差

单个样品的残差标准差定义为：

$$S(\hat{e}_i)^2 = \sum_{k=1}^K \hat{e}_{ik}^2 / (df/I) \quad (11-12)$$

df 是上面所说的自由度， I 是定标中样品的个数。

如果 $S(\hat{e}_i)^2$ 明显大于定标样品平均的残差方差 $S(\hat{e})^2$ ，这表明该样品是异常的。从理论上很难确定这个差距有多大就提出异常警告，但是在实践中一般用 F 检验来做近似估计：

$$[S(\hat{e}_i)/S(\hat{e})]^2 \quad (11-13)$$

如果分子在 df/I 自由度，分母在 df 自由度的 F 分布中， F 值大于某个百分比，则说明这个样品是有问题的。

更敏感的异常样品检验可以由下式得出：

$$[S(\hat{e}_i)/S(\hat{e}_{-i})]^2 \quad (11-14)$$

其中 $S(\hat{e}_{-i})^2$ 是除去 I 号样品的所有样品的平均残差方差。比较粗略但有效的方法是：当 $S(\hat{e}_i) > S(\hat{e})$ 的 2~3 倍时，就可以认为这个样品是异常的。在所举例子中，通过把样品 11 和 14 的 $S(\hat{e}_i)^2$ 与它们未经修正的误差尖峰相比较，就可以认为二者是奇异值。

(3) 单个 X 变量中的残差标准差

与前面相似，不同 X 变量的残差标准差被定义为：

$$S(\hat{e}_k)^2 = \sum_{i=1}^I \hat{e}_{ik}^2 / (df/K) \quad (11-15)$$

其中 df 依然为自由度。一个比 $S(\hat{e})^2$ 大的 $S(\hat{e}_k)^2$ 表明该 X 变量因为参差大，而有优先权重。可以通过近似的 F 检验，而通常较粗略的检验 $S(\hat{e}_k) > S(\hat{e})$ 的 2 或 3 倍，就可说明该 X 变量为异常。

$S(\hat{e}_k)^2$ 值不正常变的原因可能是：① X 的变异跨度没有被包括到 A 维模型中，② 它被 X 变量中的大的噪音所掩盖，③ 它与所用模型化不相关，④ X_K 有太高的随机噪音。

在双线性定标模型中的理想状态是：不同波段处 X 变量的残差水平应与该点噪声水平相一致。如果不同波段 X 变量间出现非常不同的残差方差 $S(\hat{e}_k)^2$ ，这表明在许多 X 变量中的噪音将影响双线性因素的估计。可以用每点处的残差作为除数，对该点吸光值进行校正，这样可以使所用的双线性模型更适合于观察数值，所以就改善了定标模型的预测能力。

11.2.1.3 目标变量（因变量） Y 中未模型化的残差

在定标或验证过程中知道数据 y_{ij} 的化学分析值，因而可以得到相应的残差 $\hat{f}_{ij}$ ，与自变量 X （光谱）残差分析类似，这些 Y 残差 $\hat{e}_{ik}$ 可以揭示某些类型的奇异点。例如， y_{ij} 的化学分析错误，或定标模型与某个因变量相不适应的错误。定义：

$$\hat{f}_{ij} = Y_{ij} - \bar{Y}_j - \sum_{a=1}^A \hat{t}_{ia} \hat{q}_{ja} \quad (11-16)$$

或

$$\hat{f}_{ij} = Y_{ij} - \bar{Y}_j - \hat{t}'_i \hat{q}_j$$

在这个例子中只有一个因变量（石蕊浓度 $J=1$ ）。因此用图示法估计 Y 残差是相对容易的。用因子为 3 的 PLS 得到的 17 个因变量值与化学分析测定值（“真值”）的残差。只有 7 号样品的 Y 残差为 -30 光密度，显示其为异常样品，其他的残差都落在 0~5 光密度单位之间。在前面的 X 变量分析中，样品 1、2、3、11 和 24 成功地被识别为奇异值，另外 7 号样品现在看来也是奇异值，实际上，这个样品在化学分析上就有 -30 光密度单位误差（参见表 11-1）。

怎样使基于 Y 残差上的奇异值自动被发现？对每个样本（ $i=1, 2, \dots, I$ ）的 Y 残差进行总结， $S(\hat{f}_j)(=RMSEE_j = \sqrt{MSEE_j} = S_f)$ 可以根据以下公式计算：

$$MSEE_j = S(\hat{f}_j)^2 = \sum_{i=1}^I \hat{f}_{ij}^2 / dfI \quad (11-17)$$

其中 df_2 是拟合 Y 对 $\hat{T}$ 的回归模型后的自由度，该值不仅随着所用的定标模型中因子 A 的变化而变化，而且随着所用的回归方法、 X 数据结构及 Y 数据结构的关系而变化。简化其运算可得到 $df_2 = I - 1 - A$ ，在本例中 A 恰好等于 K 。

如果 $A < K$ ，这样可能低估 df_2 的值，在 PLSR 中 Y 数据被使用了两遍（为了得到 $\hat{V}$ 和 $\hat{R}$ ），另一种表达如 $df_2 = I - 1 - 2K$ 可能是更真实的。当 K/I 较高且 X 变量包含更多的噪音时，需要一个更为保守的 df_2 估计值。但自由度的准确性对奇异值的发现并不是非常重要的。

如果 $\hat{f}_{ij}$ 的残差比 $RMSEE_j$ 更大，表明了一个奇异点的存在。在这个实验中我们需要校正 $\hat{f}_{ij}$ 的自由度：

$$\hat{d}_{ij} = \hat{f}_{ij} / \sqrt{df_2/I} \quad (11-18)$$

可以再次利用 F 检验, 例如 df_2/I 和 df_2 自由度中, 在 F 分布中用一个百分比来检验 $\hat{d}_{ij}/MSEE_j$ 。

我们可以利用除去 j 号样品之外的所有定标样品得到相似数据来替代 $MSEE_j$, 这样检查奇异点更敏感。

一个简单的措施是在下列情况下发出奇异点警告:

$$|\hat{d}_{ij}| > \text{常数} \times RMSEE_j \quad (\text{常数一般为 } 2 \text{ 或 } 3)$$

利用这种方法, 7 号样品被自动的发现是奇异点。在这个例子中, 产生奇异点的原因是 Y 数据中有打印错误。在其他例子中这个 $\hat{f}_{ij}$ 中的奇异点可能是由于 $X-Y$ 不同的关系。

经常, 由于 X 数据中的误差, 在 X 向量中被认为是奇异点的样品也会经 $\hat{f}_{ij}$ 参差分析被标出为奇异点。原因是 $\hat{t}_i = X'_i \hat{V}$, $\hat{y}_{ij} = \hat{t}'_i \hat{q}_j$, X 误差影响 $\hat{f}_{ij}$ 残差计算。

11.2.1.4 残差的杠杆修正

如果不考虑真正的错误水平, 一个具有较高的杠杆 (或称系数) h_i 的样品必定具有一个较低的残差 $\hat{e}'_i$ 和 $\hat{f}_i$ 。在含氧化锌的石蕊溶液的 1 号样品的 $\hat{e}'_i$ 和 $\hat{f}_i$ 中可以明显看出这个现象。因此如果只单独检验 $\hat{e}_i$ 或 $\hat{f}_{ij}$ 很难发现具有较高 $\hat{e}$ 或 f 的奇异值。修正的办法是将这个残差除以某个接近 0 的因子, 随着 h_i 的增加, 该样品的残差就增大了。这种修正可通过多种方法获得。

最常用的方法是用回归残留标准差的估计值除以因变量 Y 残留, 也就是得到所谓的学生化的残差, 这个标准差可以写成 $S_f \sqrt{(1-h_i)}$, 多元线性回归的学生化残差可以被定义为:

$$r_{ij} = \hat{f}_{ij} / S_f \sqrt{(1-h_i)} \quad (11-19)$$

其中 S_f 是通常的 RMSEE, 学生化残差可以简单通过对比正态分布来发现奇异点。这对于正态分布的误差 f 是非常有效的。

为了强调具有较高杠杆系数的样品, 可以对 $\sqrt{1-h_i}$ 进行平方, 这样可以得到杠杆系数的修正残差:

$$r_{lev} = \hat{f}_{ij} / S_f (1-h_i) \quad (11-20)$$

这个更严格的杠杆系数的修正可以用来模拟因变量的预测残差。在双线性回归中也可以用 r_{ij} 和 r_{lev} 进行残差的修正。

与因变量残留的修正相似, 杠杆系数修正可以用于 X 残差, 不但可以用于单个样品 $\hat{e}_i$, 而且也可以用于单个观察值 $\hat{e}_{ik}$, 这些杠杆系数修正的 X 残差现在就可以用来代替原来的残差发现新的奇异点。在本章所举的例子中, 2 号样品在整个光谱 $\hat{e}_i$ 和在特殊范围 ($19 < K < 30$) 内 $\hat{e}_{ik}$ 的杠杆修正 X 残差都显示, 该样品为异常样品。

11.2.1.5 影响力的测定和图示

在回归分析中的另一个重要概念是对分析结果影响力的评价。这种影响力的度量对于估计回归参数中奇异点的重要性是非常有作用的。有较大影响力的奇异点是非常危险的, 应给予足够的重视。

在普通的回归分析中, 通常用库克 D_i 值来度量 I 样品的影响力, 此外, 还有其他方

法（见 Belsley 1980），库克的 D_i 被定义为：

$$D_i = \frac{(\hat{b}_{(-i)} - \hat{b})' X' X (\hat{b}_{(-i)} - \hat{b})}{S_f^2 (k+1)} \tag{11-21}$$

其中 $\hat{b}$ 和 $\hat{b}_{(-i)}$ 分别表示包括所有样品或去除 i 号样本多元线性回归系数。参数 S_f 代表欠拟合的因变量 Y 的残差标准差。

也可以写成

$$D_i = \frac{1}{K+1} \frac{h_i}{1-h_i} r_i^2 \tag{11-22}$$

可以看出 D_i 是学生化残差 r_i 和杠杆系数 h_i 的简单函数。

一般来讲，仅靠杠杆系数或仅靠学生化残差来判断影响力都是不充分的，至少还需要关于那些数据的一般分布。例如，在前面的例子中，3 号样本因子为 3 的 PLSR 模型有较小的影响力，尽管它有较大的杠杆系数 ($h_i = 0.67$)，原因是它的残差 $\hat{f}_{ij}$ 太低，也由于 $X-X$ 之间的相关性。另外对于样本 7，尽管它有较高的残差 $\hat{f}_{ij}$ ，但是由于其杠杆系数 ($h_i < 0.1$) 太低，因此该样本对于模型的影响力也较低。也就是说，对模型产生较大影响力的样本必须有两个条件：一是非常不同于其他样本，二是被模型所描述。

作每个样本作杠杆系数 h_i 对因变量残差 $\hat{f}_{ij}$ （或学生化残差）的二示图，是显示该样本影响力的有力方法。它对双线性模型也适用，这种方法对于寻找影响因素也是非常有用的。比较各因子 $a = 0, 1, 2, \dots$ 间图，可以大略看一下定标模型怎样随着因子的增加而变化。

在本章所举的例子中，定标系中 17 个样本产生的影响力图 11-2 所示。样本 1 和 2 的杠杆系数 h_i 非常接近于理论上的最大值 1.0 和较低的残差方差 $S(\hat{e}_i)^2$ ，它们对该定标模型都有较大影响。对结果的进一步分析可以看出它们代表坏的奇异值。样本 3 与其他大部分观察目标属于同一类，它有中等大的杠杆系数和相对较低的残余方差。由于样本 11 和 14 的误差尖峰，所以它们都有相对较高的残余方差。但是由于它没有被包含在前三个因子的 PLSR 模型化中，所以它们的杠杆系数较低，也就是说它们不是那些非常危险的奇异样本。但是我们如果要充分地考虑到它们的不良作用，就可以在重新定标前把它们删除掉。

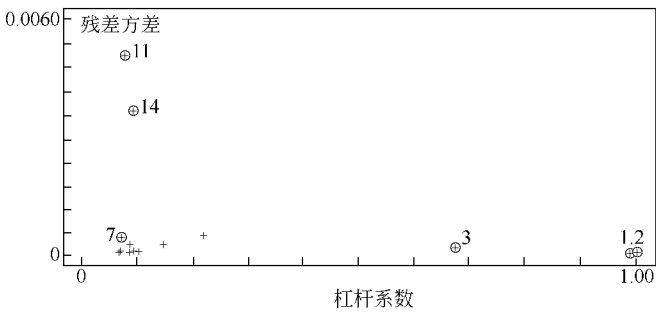


图 11-2 在 17 个石蕊溶液的定标例中，杠杆系数对浓度残差方差的二示图

X （光谱）残差方差对应于相应的杠杆系数的图 11-3 所示。即残余方差 $S(\hat{e}_k)^2$ 对应于因子为 3 的 PLSR 模型中杠杆系数 h_k 作图，可清楚显示各光谱波段在因子为 3 的定标

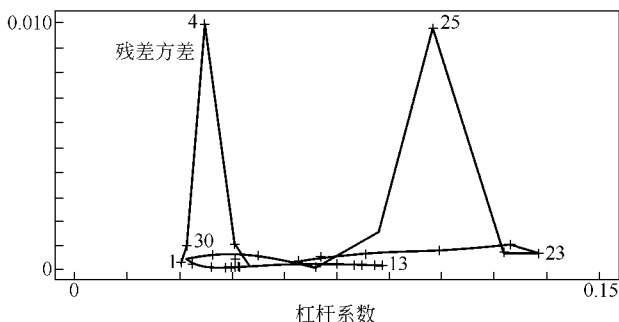


图 11-3 在 17 个石蕊溶液的定标例中，杠杆系数对光谱残差方差的二示图

模型中的影响力。各光谱波段的杠杆系数都较小，最大 h_k 值为 0.14，即波段 23（对应于光谱为在该波段中溶质的吸光度最大）。最显著情况是在波段 4 和 25 周围，出现高的残余方差。自然，这反映了由样本 11 和 15 带入的噪音信号。所以我们可以给 X 变量 4 和 25 较低的优先权重，再进行定标。

11.2.1.6 去除奇异点后重新定标

在本节例子中我们通过做影响力图自动判别了 6 个奇异点，并且通过分析知道其中 5 个（即样本 1, 2, 7, 11 和 14）是坏的奇异点。现在我们将这 5 个奇异点去掉，再进行重新定标，这剩余的 12 个样本将反映一个单一的现象：石蕊浓度的变化与光谱的关系。

从交叉检验所得到的内部预测误差结果反映，仅用因子数为 1 的 PLS 模型就能获得很好的内部预测能力，增加了第 2 个因子后，预测能力得到了一个微小的改善。

在第二次定标中我们得到了一个新的奇异点：样本 3，它具有较高的杠杆系数，即当 $A=1$ 时 $h_3=0.71$ ， $A=2$ 时 $h_3=0.82$ ， $A=3$ 时 $h_3=0.95$ 。因为样本 3 中石蕊的浓度比其他样本高出 2 倍以上，所以对它的警告被再一次提出来，但该样本 X 和 Y 数据是正确的。

定标模型所用的二个因子的载荷向量 $\hat{p}_1$ 和 $\hat{p}_2$ 及 12 个样品的平均光谱。这里非常清楚地显示，绝大多数的光谱变异包含在 $\hat{p}_1$ 中，它与平均样本的光谱一致。第二个载荷向量，虽然较小且明显包含有噪声，但它也明显表现了一个非随机的峰谷模式。定标模型中增加第二个因子后明显改善了预测能力，因此我们将它包含在最终的 PLSR 定标模型中，作为对那些我们尚不清楚干扰的修正。它在最大吸收处有一个峰谷，而在最大吸收峰的两侧形成两个小峰，因此该因子的载荷向量被看作是对仪器非线性反应的修正。

至此，定标模型已经确定，我们用这个二因子模型来推测未知溶液的石蕊浓度。

11.2.2 预测过程中的奇异点

对于双线性定标方法，在预测过程中的奇异点的分析是以 X 残差和预测杠杆系数为基础。因为多数情况下，我们不知道 y_{ij} 的实测值，因此 Y 参差并不存在，相应的以 $\hat{f}$ 为基础的分析方法也就不能使用。

11.2.2.1 预测中具有极端杠杆系数的样本

现在，我们用上面建立的二因子 PLS 定标模型分析新的样本。要预测的样品有 12 个，其中 5 个是原定标模型中坏的奇异样本，1 个是好的奇异点，另 6 个是普通的未知

样本。

因子 1 和 2 中这 12 个样本的得分二维图显示，低浓度样本分布在左边，高浓度样本分布在右边，所有 5 个普通未知样本都紧靠近因子 1 的得分向量轴。

那么在这 12 个样本中哪些是奇异点呢？首先看看样本 7，它是由 Y 残差异常而被作为奇异点，但从光谱数据分析上，它没有异常，因此在图中，样本 7 与其他样本没有多大区别。由 X 向量分析发现的奇异点分别为样本 1、2 和 3。从数学理论上，如何来辨别这些 X 奇异点呢？

杠杆系数 h_i 概括了模型中所有因子的极端性，与定标过程中杠杆系数的计算相似，预测样本的杠杆系数可以定义为：

$$h_i = 1/I + \hat{t}'(\hat{T}'\hat{T})^{-1}\hat{t}_i \quad (11-23)$$

式中 $\hat{t}_i = (\hat{t}_{ia}, a=1, 2, \dots, A)'$ 是预测样本 i 的回归因子的向量， $\hat{T}$ 为定标样本的回归因子得分矩阵。如同定标样本的杠杆系数，预测样本的杠杆系数定义也为马氏距离。而 $\hat{T}'\hat{T}$ 表示加权协方差矩阵。

由于 PCR 和 PLSR 方法的正交得分已经在定标系中给出 ($\hat{T}'\hat{T} = 1\hat{t}'_a\hat{t}_a$)，以上公式可以简化为：

$$h_i = 1/I + \sum_{a=1}^A \hat{t}_{ia}^2 / \hat{t}'_a\hat{t}_a \quad (11-24)$$

在奇异点的分析过程中，将预测样本 i 的杠杆系数与定标系 I 个样本的平均杠杆系数进行比较。例如 $h_i > \text{常数} \times (A+1)/I$ ，这里常数可以是 3。

但是要记住，定标样本的杠杆系数不可能大于 1.0。但对于预测样本则没有这种限制。有些未知样本具有特殊结构的光谱，可能会或多或少地产生一些很大的得分。

在上述例子中，样本 1 和 2 具有很高的杠杆系数 (h_i 分别为 62 和 34)，样本 11 (在溶质的最大吸收峰附近误差尖峰 (X_{25}) 的杠杆系数并不大 (h_i 为 1.4)。

如果有兴趣做进一步了解，请参考 Nas 和 Martens (1987) 及 Nas 和 Isaksson (1988) 的著作。

11.2.2.2 光谱数据 X 中未模型化残差

预测样本光谱残差可以由下式计算出来：

$$\hat{e}' = X' - \bar{X}' - \hat{t}'\hat{p}'_i \quad (11-25)$$

用因子为 2 的 PLSR 模型计算出的 12 个预测样本的光谱残差。正如预期的那样 (表 11-1)，异常的样本 1 和 2 显示出巨大的非模型化特征，而样本 11 和 14 分别在波段 $K=25$ 和 $K=4$ 周围产生误差的地方出现了较小的残差尖峰。而其他样本都很适合于定标模型，用定标模型计算出的光谱与实际光谱之间残差很小。

与 5.2.1.2 所讲内容相似，通过数学运算可自动找出预测样本的光谱奇异点。 X 中的平均残数方差 $S(\hat{e}_i)^2$ 定义为：

$$S(\hat{e}_i)^2 = \sum_{k=1}^K \hat{e}_{ik}^2 / (k-A) \quad (11-26)$$

每个元素的 $\hat{e}_{ik}^2$ 或平均残数方差 $S(\hat{e}_k)^2$ 可以与定标中平均噪声水平 $S(\hat{e}_k)^2$ 和 $S(\hat{e})^2$ 相比较。

在本例中，根据 $S(\hat{e}_i)^2$ 计算，样本 1、2、11 和 14 被认为是奇异点。除此之外，样本 11 在 25 波段的吸光度 ($X_{11,25}$) 和样本 14 在 4 波段的吸光度 ($X_{14,4}$) 也被认为是很异常的 X 数据。

将每个待测样品的光谱残数水平 $S(\hat{e}_i)^2$ 和杠杆系数 h_i 对应作图，直观显示。但这也不能确切地显示出样本 $i=1$ 和 $i=2$ 为奇异点，它们的坐标分别为 (62, 1.2) 和 (34, 0.2)。但该图反映了非常有趣的现象：尽管样本 11 和 1 都有相近的误差尖峰，但由误差所处的位置不同，如样本 11 的误差尖峰在波段 $K=25$ 附近，该波段在 PLS 因子限定的主要波幅范围内；而样本 14 的误差尖峰落在波段 $K=4$ 附近，这已经超出了 PLS 因子的主要变动范围，因此误差尖峰对样本 14 的影响要比对样本 11 的影响小。

因为我们已经知道了 12 个待测样品的实测值 (y_{ij})，很容易得出预测的误差。很明显对奇异样本 1 和 2 的预测误差很大，说明定标模型对它们的预测能力很小。而 6 个正常样本（标“+”号）得到了非常准确的预测。

样本 3 也得到非常好的预测，这可能是一个假象，因为该样本是一个定标样本，而且在所建立的定标模型中有较高的杠杆系数。但是通过对 6 个正常样本很好的预测这一点标准，说明即使预测样本的浓度超过定标样本的浓度范围，也可能得到很好的预测结果。

如期望的那样，样本 7 显示一个较高的预测误差，但是这是一个假象，因为预测的浓度实际上是很准确的，而实测值却有 30% 的误差。样本 11 和 14 也给出了很好的预测结果，尽管它被认为是奇异点，原因是它们的误差峰值相对较小并且只涉及到定标模型中的 2 个波长段的 X 变量。

12. 近红外光谱分析的误差源

从表面看，似乎 NIR/NIT 技术的误差源很少，因为在应用 NIR/NIT 技术进行分析时，只是一个简单的过程：样品的干燥、粉碎、装样、读数，而不需要称重及使用任何化学试剂。其实不然，有研究表明：在 NIR/NIT 分析技术中有多达 40 个误差源，它们在不同程度上影响着分析结果。但要说明的是，很多误差源，特别是对于来自仪器部分的误差源，能够得到有效的控制。

这些 NIR/NIT 分析技术的误差源可归结为三类：来源于仪器、来源于样品、来源于操作。笔者认为对于实际分析者，最主要的误差源来源于建立定标模型时样品的选择、采样和样品的制备、经典实验室分析值的误差等。表 12-1 是加拿大 Williams 总结的 NIR/NIT 分析技术误差源，有 * 号表示该因素也将影响分析结果的精确度。但要说明的是，表 12-1 所列并不能覆盖所有的 NIR/NIT 技术误差源，而且也不是按照误差源影响的大小顺序排列。另外，对于不同类型的 NIR/NIT 仪器，某一误差源对分析结果的影响力不同。

表 12-1 NIR/NIT 分析技术的误差源 (引自 Williams)

	来源于仪器
仪器的噪声 *	
散射光	
信号的非线性	
原始信号的数学处理 *	
波长的选择 *	
平滑的片段长度	
导数处理 *	
静电 *	
仪器的温度控制 *	
电源的稳定性 *	
仪器的几何光路 *	
光斑大小	
仪器元件的寿命	
	来源于样品
样品的物理组织	
粉碎样品的密度	
影响样品物理性质的一些外部环境,包括气候、灾害、植物的病虫害	
样品的温度	
环境温度	
样品类型(主要指样品的均质性)	
转换因素	

12.1 来源于仪器的误差源

12.1.1 仪器的噪声

在一般光谱分析中，噪声被定义为待测信号的波动或称干扰，这种波动信号将掩盖该

待测信号的测定。在 NIRS/NITS 分析中，信号与噪声比（简称信噪比，SNR）将影响着分析的效率。仪器分析中存在着噪声是不可避免的，为提高仪器的分析精度，应尽量提高仪器的信噪比，但是，应注意的是：在近红外光谱分析中，SNR 高并不能保证最终分析结果的准确性。

噪声可分为基本噪声和非基本噪声两种。基本噪声源于物质和光的粒子性，且不能被完全消除。非基本噪声也被称为附加噪声，源于仪器和元件的不完善，在理论上，它可以被完全消除。

在近红外光谱分析中的纯仪器噪声可来源于以下几个方面：仪器或环境的温度、仪器或环境的振动等。

12.1.1.1 仪器或环境的温度

仪器的温度是引起仪器噪声的主要原因之一。因为温度影响着波长的稳定性，如果仪器的温度发生变化，就使得仪器实际测量时的最佳波长不同于仪器所设定的波长。例如，随着仪器或环境温度的变化，某仪器设计中的最佳波长点，在实际测量中改变了原来的位置，或增或减了 $1\sim 2\text{nm}$ 。这样，对于一个定标模型，所选择波长点及相对应的回归系数是确定的。当波长发生变化后，首先，波长不是最佳波长点；其次，对应于实际波长点的回归系数也发生了变化。当然这种变化可通过对定标模型斜率的调整而恢复，但随着温度的恢复到原状态，波长点又回到原位置，就必须对已调整斜率的定标模型进行再调整。

仪器的温度发生变化的另一个影响对象是硫化铅检测器的输出信号量。检测器中的硫化铅作为一个电阻，随着接受到的光子其电阻量发生变化，这种变化是微小的，经过多级放大后进入记录仪。仪器或环境的温度的变化，也将使检测器中的硫化铅电阻量发生变化，这种变化同样被放大，作为信号进入记录仪。因此，现代的 NIR/NIT 分析仪，在硫化铅检测器中安装一个热电元件，以控制检测器的温度，并尽可能冷却检测器。

12.1.1.2 仪器或工作台的振动

当工作室、工作台及仪器可移动部分产生振动时，将使仪器产生噪声，这属于仪器的机械噪声。仪器可移动部分的磨损会引起仪器的振动和仪器元件间相对位置的变化。仪器的机械噪声产生的另一原因可能是，发动机位置产生微小变化。仪器磨损和老化的后果之一是影响狭缝大小，使得到达样品的入射光强不一致而产生噪声。

仪器的温度、振动等影响是间歇式的，而仪器的机械和电子元器件的磨损或老化引起的噪声是渐进的。这个磨损涉及到仪器的可移动部件如传动轮、光学元件、检测元件、放大元件、记录计算元件等诸多过程。对其的监测是通过经常性地分析稳定性样品如陶瓷标准样，或有特征性光谱带的聚苯乙烯。如果把这些标准或稳定样品近红外光谱带的位置和光密度记录下来，这样，当仪器的近红外光谱带的位置和光密度发生变化时就能很容易被发现。目前，大多数近红外光谱分析仪例如：DICKET-joinGAC III，Technicon InfraAlyzer，Percon Inframatic 8100 和 FOSS NIRSystem，都具备监测仪器波长位置和光密度变化的功能。通常的监测是在相同的波长位置，测定标准物的光密度，根据光密度的变化判断噪声大小及引起噪声的原因。表 12-2 显示一台近红外光谱仪在 3 个月内，标准物陶瓷光密度的变化，在这段时间内，该仪器很好地分析了小麦、大麦中的水分和粗蛋白质含量。

表 12-2 DICKET-join 近红外光谱仪在 3 个月内标准物陶瓷光密度的变化

	log0	log1	log2	log3	log4	log5
9 月 5 日	150.71	165.09	133.59	95.501	60.791	140.46
9 月 17 日	149.97	164.45	132.89	94.781	60.210	140.05
9 月 25 日	150.03	164.52	133.00	94.770	60.618	140.18
10 月 14 日	146.60	161.02	129.39	90.861	55.490	135.00
11 月 18 日	147.09	161.57	130.02	91.609	56.912	135.51
12 月 10 日	151.72	166.29	134.77	96.468	61.445	142.23

在现代近红外光谱分析仪中，各制造商努力通过稳定检测和放大元件、减少振动、冷却检测器及使用多个检测器等一系列措施来减少仪器的噪声。

在早期滤光片型 NIR 仪器中，每个样品扫描时，仪器仅记录所选择的波长点的信号，此外，每次扫描持续的时间较长。因此，电源瞬间的异常波动在转为检测器可测试的信号前可被抵消。然而，对于连续扫描的 NIR 仪器，它们扫描的范围是 1000～2500nm，扫描间隔为 2nm，这样每个样品可获得 700 个以上波长点的信号。而且每次扫描在 0.2～2s 内完成，所以，电源在瞬间的异常波动也将被检测器捕捉并记录下来，也就会影响 NIR 分析的精确度和准确性。通常，可通过同一样品多次扫描来克服电源在瞬间的异常波动。

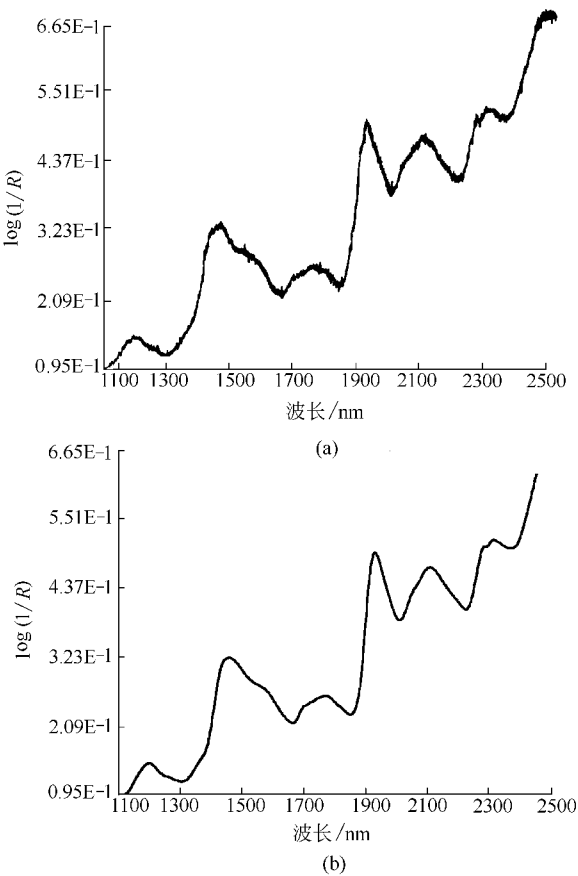


图 12-1 检测器的缺陷改正前 (a)、后 (b) 对小麦 NIR 光谱的影响

图 12-1 (a) 所示的是, 因检测器的缺陷而获得的含噪声较大的小麦的 NIR 光谱, 各点都含有细小尖锐的噪声信号, 当检测器的缺陷被改正后, 小麦的 NIR 光谱显得非常平滑, 没有细小尖锐的噪声信号 [图 12-1 (b)]。

12.1.2 散射光

在近红外光谱分析中, 散射光可能产生严重的误差。所谓散射光就是到达检测器的非仪器设定波长的光, 通常有 3 种类型的散射光, 即自然光的漏射、光源、单色光的重叠等。

自然光的漏射的原因是进样室的不密封, 使得自然光有可能到达检测器, 目前大多数台式近红外光谱分析仪都有密封的进样箱, 以尽可能地减少自然光的进入。

对于近红外光谱分析仪, 特别是工作范围为 $1100\sim 2500\text{nm}$ 的近红外光谱分析仪, 其理想光源的发光强度的峰值在 $1500\sim 2000\text{nm}$ 间。而目前常见近红外光谱分析仪的光源——钨卤灯发光强度的峰值在 900nm 左右。

由于技术上的缺陷, 由光栅或滤光片产生的单色光中还夹杂着少量或微量其他波长的光。对这些少量散射光的干扰, 大多数近红外光谱分析仪是通过使用第二个滤光片或光路中增加棱镜来克服。当然如果以近红外发光二极管 (NIRLED) 作为光源, 就很容易克服散射光的干扰。对于近红外的短波长区 ($800\sim 1100\text{nm}$), 这种技术的实现是没有问题的, 而且, 这种仪器已生产出来了。但是对于近红外的长波长区 ($1100\sim 2500\text{nm}$), 由于 NIRLED 的价格很贵, 目前还不能生产出具有商业价值的此类仪器。

12.1.3 非线性

众所周知, 在经典的透射式光谱分析中, 透明样品液的吸光度与吸收分子浓度线性相关, 遵循 Beer/Lambert 定律。近红外光谱分析与经典的透射式光谱分析不同, 其信号 ($\log [1/R]$) 与待测成分浓度并非呈严格的线性关系。在待测成分浓度较高时, NIR 反射光的曲线斜率增加的速度要低于浓度增加速度。对于吸收作用强的成分如水、油比吸收作用弱的成分如蛋白质, 这一点更明显。如样品中蛋白质含量从 8% 增加到 16%, 所引起的反射光密度仅增加 0.002 (OD), 水分含量以相同幅度增加, 却使反射光密度增加 1 (OD), 是蛋白质的 500 倍。

近红外光谱信号的非线性有三个方面含义: ①吸收成分不同引起光密度变化的非线性, 如水的吸收比蛋白质更强; ②不同波长点处信号的非线性, 如在扫描点的末端, 信号非线性程度更强; ③检测器与光源的相关性。

近红外光谱信号的非线性产生的原因主要有三个: ①仪器本身的缺陷; ②样品的影响; ③仪器与样品间的相互作用, 下面分别进行阐述。

(1) 仪器本身的缺陷 对于单色器型仪器, 并不是在所有波长点上, 硫化铅检测器及信号放大器所获光谱信号呈严格的线性变化, 这种现象的确存在, 但并不是错误的主要原因。由散射光产生的信号非线性要严重得多。例如, 对于波长在 1940nm 处信号, 由于 1940nm 单色光中夹杂着其他波长的光 (散射光), 这些散射光引起检测器产生较严重的信号非线性。

(2) 样品引起的信号非线性 样品所含的某一成分在不同浓度梯度时, 浓度的变化引

起光谱信号 OD 变化有时呈非线性。例如，样品中水分从 5%~10% 增加 5 个百分点所引起光谱信号 OD 的增加值，要大于由 30%~35% 或 75%~80% 同样增加 5 个百分点所引起光谱信号 OD 的增加。其主要原因是由于水分含量变化引起样品散射性能的变化。水分的折射指数 (RI) 为 1.3，而大多数农产品如小麦的 RI 为 1.45。样品中水分含量的变化引起了样品 RI 指数的变化，当样品含水量高时，该样品 RI 指数与水接近；当样品含水量低时，该样品 RI 指数与小麦接近。应该说明的是，目前已可通过软件来克服这种由样品引起的信号非线性。

(3) 仪器与样品间的相互作用 如果某样品对入射光完全吸收，理论上没有信号能跑掉，然而在表面仍有少量的反射光。由此可见，对于一些含水量变化大，吸收程度也大的样品将产生信号非线性。在 1940nm 处，光密度最高为 1.2，即仅有 6% 光被反射。而对于含水低的样品，OD 最低可为 0.4，有 40% 光被反射。通常表面反射（黑色表面反射）的错误为 0.5%，很显然 0.5% 占 6% 的比重要远大于 0.5% 占 40% 的比重，因此对于高含水量样品，表面反射的错误所产生的信号非线性较严重。另外，样品颗粒大小也能显著影响光谱信号。总的来说，样品越粗糙，OD 越高，黑色表面反射作用所产生的信号非线性越严重。当然，可通过测量炭黑的反射记录黑色表面反射，再通过软件来减少其产生的信号非线性。

12.1.4 波长的选择

在早期的 NIR/NIT 分析中，波长的选择是一个重要步骤。从前面几章的论述中知道，在 NIR 分析中，样品中某成分的含量是通过某些波长点处的光密度乘以该波长处的常数（定标中获得的回归系数）而获得。理论上，所选择的波长点和该点处的常数对相应的被分析成分应该是最佳的，但在实际分析中，这一点很难做到。从逻辑上讲，所选择的波长点越多，其中包含的非最佳波长越多。从这一点看，在定标模型中，选择的波长点越少越好。

用 NIR 分析同一系列（如小麦）的多个样品时，通常是选择主要的几个相邻波长点。但是因仅用单一波长的 NIR 模型很少，因为，对 NIR 光谱信息与样品成分间回归的迭代，至少需要两个以上的波长。仅用单一波长的 NIR 模型，Norris 博士的研究发现，对于一定的参与定标的样品，分析的准确性很高，但该模型对定标系内样品太敏感，对分析其他新样品则不敏感，分析的准确性很差。对单一波长的 NIR 模型，常用一阶或二阶导数光谱。

在一个样品变异性不大的定标系中，第一个波长点及其对应的回归系数解释了该定标系中样品的主要变异，对于该定标系，此模型是稳定的。但当用该模型（仅用一个波长点的模型）来预测变异性大的样品，如不同的产地、生长季节、水分含量、颗粒大小等时，由于各变异源的相互干扰，产生了很大的测定误差。这种情况下，可用多个波长点的模型来解决这个问题。主波长点以外的其他波长点称为模型的支持性波长点，在 Step-Wise 的程序中，对支持性波长点的选择同主波长点。这个包含众多波长点的模型所解释的样品变异要多于仅含主波长点的模型。在主波长点及较重要的支持性波长点（回归系数较大）处，其光密度随样品变化的变异小；而较次要的支持性波长点（回归系数较小）处，其光密度随样品变化的变异大。表 12-3 显示的是，20 个测定硬粒春小麦中蛋白质含量的定标

模型（每个模型是由 40 个样品回归所得到），在回归系数不同的波长点处，NIR 光密度的变异程度。

表 12-3 20 个样品不同的波长点处 NIR 光密度变异的比较（转引自 williams)^①

波长/nm	回归系数	光密度 CV/%	对测定蛋白质模型的贡献率/%	CV 比贡献率
2308	0. 0142	2204	3. 2	689
2230	－0. 1480	391	8. 1	48
2180 ^②	1. 3304	35	43. 8	0. 8
2100 ^②	－1. 0860	25	35. 9	0. 7
1940	0. 0308	105	1. 8	58
1680	－0. 2072	127	4. 6	28

注：① 该数据来自于 DICKY-john GACⅢ；

② 对该模型来说的主波长点。

回归系数最大的两点 2180nm 和 2100nm 处，其光密度的 CV 分别仅为 35％和 25％；而每个波长点对测定蛋白质含量（该模型）的贡献率（结合着回归系数和光密度大小）也清楚地显出，对该模型 2180nm 和 2100nm 最重要，其次是 1680nm 和 2230nm 处，最小为 1940nm 处。光密度变异与贡献率的比值表明，对模型的贡献越大，其光密度 CV 越小。回归系数次之的两点 1680nm 和 2230nm 处，其光密度的 CV 分别为 127％和 391％；而在回归系数最小的 2308nm 处，其光密度的 CV 达到 2204％。

12. 1. 5 NIR 光谱信号的数学处理

尽管许多仪器不对 NIR 光谱信号进行数学处理，也成功地分析了很多产品中的蛋白质、水分等成分的含量，但目前，对 NIR 光谱信号进行数学处理越来越普遍。Morrey（1968 年）首次在光谱信号的处理中引入数学模型，来解决光谱带重叠的问题。NIR 光谱信号的数学处理的目的是克服光谱带的重叠和基线漂移。目前，对 NIR 光谱信号主要的数学处理为：平滑、一或二阶导数、变数的增量、标准化、去偏差等。这些数学处理已在本书的其他章进行了详细的讨论，这里讨论的是由这些数学处理产生的分析误差。

12. 1. 5. 1 扫描的间距

扫描的间距（Segment Size，常以 *n* 表示），是由原美国太平洋公司（现称 NIRSis-tem 公司）引入。很明显扫描的间距是相邻两次扫描点间的片段长度。最近的研究表明，对于一或二阶导数光谱的平滑，当 *n*＝6 后，再增加扫描的间距，不能提高分析的精度。

12. 1. 5. 2 导数光谱的求导间距

在光谱的导数处理中，求导间距（Gap）是一个重要的参数。如果 Gap 太小，则算法不可能将系统的噪声降低到最低，这样处理后的光谱对来自样品和噪声的相互作用较敏感；如果 Gap 太大，则光谱对样品的敏感性也会降低，造成分析精度低。恰当的 Gap 值与仪器的分辨率和平滑程度有关。如果仪器的分辨率很高，则 Gap 越小效果越好，因为，其光谱对样品中成分变化很敏感；如果仪器的分辨率不高，则 Gap 较大时，分析的效果要好。在建立定标模型时，首先要确定最佳的扫描间距和 Gap。总的来说，对分析精度的影响，Gap 的作用要大于扫描间距。图 12-2 为 Gap 对

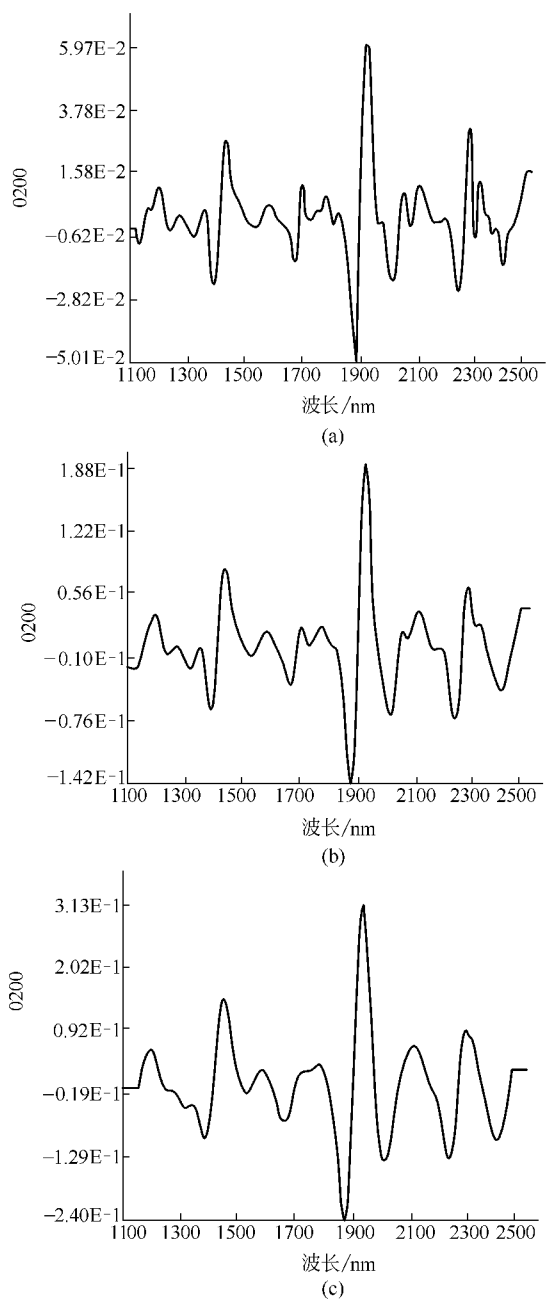


图 12-2 显示 Gap 对光谱的影响，(a)、(b)、(c)
的 Gap 分别为 20，50，90

光谱的影响。

表 12-4 揭示扫描间距和 Gap 对分析硬粒春小麦中蛋白质含量的影响。20 个样品测定结果的准确率说明，对一或二阶导数模型，随着扫描间距的增加，准确性提高，最佳的 Gap 为 20~30nm。

表 12-4 平滑片段 (Segment) 和求导间距 (Gap) 对分析小麦 CP 含量的准确度的影响
(仪器为 Analyzer 6250)

扫描间距

	2		4		6		8		10	
	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$
10	0.20	0.11	0.16	0.07	0.16	0.08	0.14	0.07	0.14	0.07
20	0.29	0.46	0.22	0.13	0.21	0.13	0.20	0.14	0.20	0.16

求导间距(Gap)

	2 ^b		5		10		15		25		35		45	
	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$	SET	$\bar{d}$
10	0.20	0.04	0.19	0.05	0.16	0.05	0.14	0.04	0.15	0.08	0.16	0.24	0.16	0.19
20	0.29	0.13	0.21	0.22	0.19	0.22	0.27	0.11	0.24	0.15	0.21	0.56	0.21	0.25

12.1.6 静电作用

静电作用是电荷的累积作用，其在相对湿度较低的地区较明显。电荷的积累决定于以下几个因素：相对湿度、纤维性物质如尼龙或毛。静电对 NIR 分析的主要的影响是：对样品槽中粉状样品颗粒的定向排列作用，这种定向排列作用影响 NIR 漫反射信号，结果影响了 NIR 分析的精确度和准确性。样品颗粒的形状和成分与受静电影响的程度有关，长型纤维质颗粒所产生的定向排列作用比圆形、方形淀粉质颗粒所产生的定向排列作用更强，因此，对于粉碎的稻草可很强地定向排列于样品槽表面。以石英作表面的样品槽比玻璃更易受静电的影响。另外，静电作用使灰尘的颗粒附于样品槽表面难以清扫，其产生的后果有二：①产生了样品间的相互污染；②灰尘附于样品槽表面，使灰尘的信号进入检测器，从而增加了仪器的噪声。因此，在样品装入样品槽前必须清除残留的灰尘和样品。由于毛刷易产生静电，因此，最好用光学仪器专用纸或丝巾作为工具。

12.1.7 仪器的温控

温度控制对于 NIR 仪器特别是早期产仪器是必须的，环境温度较大的变化可影响仪器的检测器及其他元件。对于以半导体元件为检测器的仪器，仪器本身都有 1~3 级温度控制装置，作为仪器的必须元件在仪器生产时就已安装。这样的仪器都要求一定的工作环境温度，对工作环境温度限制的宽松程度与仪器本身温度控制装置有关，仪器本身温控装置越多越好，则对工作环境温度要求越宽松；相反，仪器本身温控装置越少越弱，则对工作环境温度要求越严格。

仪器本身电机工作所产生的热量通过其配备的风扇来散热，如果由于某种原因，风扇停止工作致使仪器元件温度升高，其后果有二：①由于检测器工作的影响，仪器的噪声增加；②单色光波长准确性发生水平漂移。因此，NIR 分析结果的精确度和准确性皆下降。这种变化可通过对仪器的日常监控来发现。

12.1.8 电源

几乎所有台式 NIR/NIT 分析仪都以 60Hz 的 110~112V，或 50Hz 的 220~240V 为

电源，每天电压的变动幅度一般在 $\pm 5\%$ 。但是在一些地区，由于供电量不足，或某些大型设备启动阶段，产生灯泡变暗现象，这将严重地影响 NIR 分析的结果。

电压的忽高忽低引起的尖峰是潜在的噪声源。在实验室的小环境中，引起电压的忽高忽低的原因有多种：冰箱的启动、吸尘器、粉碎磨、烘箱等设备的开闭。

12.1.9 仪器的光路

NIR 仪器是一个很精密的分析仪器，从光源到样品整个光路的各部分的距离和位置是严格确定的。在进行 NIR 的定量分析时是以假设这个距离保持恒定为前提的，如果这个距离发生变化（哪怕只是微小的变化），也将引起由样品到达检测器和由内标到达检测器的信号强度发生变化。

在仪器的光路中有多个镜面，如果镜面的位置有微小的变化，将导致信号偏移。仪器的光路越复杂，位置变化产生的危害越大。

众所周知，从样品到检测器间的距离即光程发生变化，将产生分析误差。

12.1.10 样品槽的表面

几乎所有 NIR 仪器在样品槽的表面都是一个玻璃或石英的盖子，其目的是保证样品有一个匀质的表面。由于各个玻璃或石英的盖子间的厚度很难做到完全的一致，因此，盖子的厚度和折射率的不同将影响分析的准确性。在使用仪器前，应对所有样品槽进行检测，确定各样品槽的玻璃或石英的盖子间是等质的。此外，保持样品槽的清洁对减少分析误差是必须的，这一点对于分析含油量较高（超过 8%）的样品尤为重要。

目前，有少数仪器配置了一种开放式样品槽，这种样品槽没有玻璃或石英的盖子，也就消除了由于盖子的不一致而产生的分析误差。它也有缺点：在装样到上机扫描过程中，样品表面由于蒸发或吸湿作用，而产生的含水量的变化。

12.2 来源于样品的误差源

样品的颗粒大小、形状及密度是影响 NIR 分析的主要因素，而样品的这三个特征由样品的质地和样品干燥、粉碎等的前处理过程共同决定。表 12-5 总结了影响样品（主要是植物性样品）质地的 20 多个因素，下面就几个大的方面进行讨论。

表 12-5 影响农产品、食品物理性质的因素

化学组分,特别是纤维和油的含量
水分含量
植物种类,特别对于粗饲料
植物的变异性
植物的成熟阶段,特别对于粗饲料、农作物秸秆等样品
植物的木质化程度
植物的生长环境,主要是气候条件
土壤的肥力

续表

收获期内是否有异常气候的发生
植物特别是谷物的成熟期是否经历过霜冻
收获后样品的贮存情况： 贮存期的长短、温度、通气、容积类型、贮存前样品的水分含量
是否有霉菌的侵害、虫害的发生
加工处理：如热处理、冷冻处理
植物鲜样的干燥处理
影响样品温度的因素： 环境温度、粉碎中样品的温度变化（主要与磨的结构、粉碎细度等有关）

12.2.1 样品的化学成分

这里所指的化学成分是那些在近红外区有吸收峰的样品的组分。在农产品、食品中，近红外区最易产生交互作用的成分为：水分、蛋白质、油和纤维性碳水化合物。尽管在某些样品中有些成分含量较低，但通过下面的计算，会发现它们产生的 NIR 信号是足够大的。

近红外光可穿透 1mm 干燥粉碎后的样品。对于入射光来说，样品是由空间组成的而不是由颗粒组成的。在 1mm 厚的样品内约有 0.5~1.0g 样品，根据阿佛加德罗定律，每克分子量有 6.025×10^{23} ，1g 含水量为 12% 的粉碎后小麦含有 4×10^{21} 分子水和 7.8×10^{16} 分子蛋白质（假设蛋白质的分子量 100 万道尔顿），大多数这些分子都在近红外区有吸收。相对来说，分子量越小，单位容积内分子越多，则吸收越强烈。

近红外区最易产生交互作用的程度取决于化学成分的不平衡性和各成分的吸收强度的特征。在羽毛粉、鱼粉、肉骨粉和血粉等样品中，蛋白质是主要成分；而在大多数植物性样品中，淀粉、油、水分和纤维性碳水化合物是造成近红外区吸收峰交互作用的主要成分；在肉、鱼、水果和蔬菜等新鲜食品中，水是产生 NIR 光谱最主要的成分。

样品主要成分在近红外区吸收峰交互作用的表现形式主要有三个：①一个或几个化学成分对决定其他成分最佳波长的影响；②化学组分对样品物理性质和粉碎中颗粒结构的影响，例如，蛋白质、油、水分和纤维性碳水化合物都会影响样品的粉碎，在一定条件下，矿物质也影响样品的粉碎；③样品中水分对其他成分水合作用的影响。

（1）化学成分间的交互作用 在食品、农产品、饲料等样品中最常见的有交互作用的化学成分为：油和蛋白质、淀粉和蛋白质、纤维与蛋白质、油和纤维、淀粉与纤维，这些样品很少同时含有高淀粉和高油。这种互作的影响在 NIR 分析中表现为：高含量或高吸收的成分掩盖了低浓度或弱吸收的成分，结果，所选择的波长对低浓度或弱吸收的成分缺乏敏感性。对于固定波长式滤光片的仪器，在分析低浓度或弱吸收的成分时准确性低，因为对这些成分，波长是非最佳点。

单色器型仪器则可能选择较高（3~4 倍频）的倍频吸收点。例如，在油含量较低的样品中，测定纤维含量的最佳波长点之一是—CH₂—基团的二倍频 2275nm。而对于高油样品，就可能选择更高倍频的波长点 1700~1720nm，2275nm 波长点将被忽略。在实际

分析样品的蛋白质含量时，常常并没有选择纯蛋白质有主要吸收峰的 2150～2160 nm 点（一阶导数光谱），所选择的波长点往往是其他成分干扰较小的点。目前，对食品、农产品、饲料等样品水分、蛋白质、油、脂肪和纤维含量 NIR 分析的最佳波长点已有了很多的报道，表 12-6 总结了这些研究的结果。需要说明的是，在分析实际样品时，还要做适当的调整。

表 12-6 分析食品、农产品、饲料等样品常规成分时推荐的主波长点

化学成分	样品种类	波长点/nm	NIR 光谱的数学处理
蛋白质	小麦	2180	log(1/R)
	小麦加工产品	1680,2310	log(1/R)
	大麦	2180	log(1/R)
	小麦	2150	D10D
	粗饲料	2164	D20D
	粗饲料	2100	log(1/R)
	豆类	2310	log(1/R)
	油菜籽	2156	D10D
	大豆	2060	log(1/R)
	猪肉	2226,2470	D10D
水分	鱼产品	1684	log(1/R)
	小麦	1940	log(1/R)
	小麦、粗饲料	2020～2026	D10D
	小麦加工产品	2230	log(1/R)
	油菜籽	1390	D20D
	大豆	2010	D10D
	猪肉	1752,1468	D10D
油	油菜籽	2306	D10D
	大豆	2302,2334	D20D
	鱼产品	2310	D10D
	向日葵	2393	D20D
脂肪	猪肉	1750	D10D
中性洗涤纤维	粗饲料	2294,1674	D20D
酸性洗涤纤维	粗饲料	1666,1555	D20D
离体消化率	粗饲料	1555,1674	D20D
灰分	面粉	2345	log(1/R)

（2）样品的化学成分对其物理性状的影响 样品中水分、蛋白质、油、脂肪和纤维含量也将影响其物理性状。对于大多数 NIR 仪器，在分析前需要将样品粉碎，以获得较均匀的颗粒。当样品中淀粉含量高、蛋白质含量中等或偏低（8%～12%）、水分和油含量低，则该样品易于粉碎。在实际分析中，低蛋白质含量的小麦易于粉碎，而由于蛋白质的黏接作用，对于高蛋白质小麦的粉碎需要特殊的粉碎机。从小麦内在质地上看，对于非常硬质的品种，在高蛋白质含量时，由于质地变软而易于粉碎；但是对于软质或普通品种，在高蛋白质含量时，由于质地变得太软而难以粉碎。

样品中的油分像黏接剂作用一样，在粉碎时易形成糊，影响粉碎效果。当样品中含油量低于 5% 时，其对粉碎作用的影响很小；在含油量达到 7%~8% 时，其对粉碎的阻碍作用开始显现；当在含油量达到 20%~30% 时，粉碎后样品的结块作用非常明显，很少有硬质颗粒存在；在含油量达到 45% 或更高时，粉碎实际是一个糊化作用，这对于含纤维量低的样品更为明显。样品粉碎的物理性状直接影响其 NIR 光谱，也就影响分析的准确性和精确度。此外高含油量样品，由于大量油外溢黏于样品槽的表面，因此样品扫描后必须用酒精等有机溶剂清洗。对于这些含油量高的样品，粉碎后贮存时，即使保存在冰箱内也要结块。因此，建议对于含油量超过 10% 的样品，应现粉碎现做 NIR 扫描、分析，粉碎后样品最好不超过 1~2d。此外，同时应保留粉碎前的原样，便于复测备用。

纤维的含量也能显著地影响样品的粉碎和容重。在食品、农产品、饲料等中，纤维含量最高的是农作物秸秆和皮壳，如稻草、麦秆、豆壳等，其粗纤维含量一般在 35% 以上，甚至达 50%。相对而言，这些样品也易于粉碎，但粉碎后样品颗粒细长，明显不同于谷物的粉碎颗粒。粉碎后的纤维质样品容重较低，而且易于在样品表面形成多个小窝，由这样表面形成的 NIR 漫反射光谱不同于谷物样品。纤维质样品颗粒易于在表面形成定向排列，这主要是静电产生的结果，因此，其表面不如谷物样品均匀。对于含纤维量较高的谷物如大麦、燕麦，NIR 分析的准确度比含纤维量较低的谷物如小麦低，原因就在于此。

(3) 样品的水分状态 样品粉碎前后水分存在状态的变化影响着 NIR 分析的准确性。对于含水量高的样品如鲜肉和蔬菜，进行粉碎是不可能的，而最好以原样如切片、糊状进行分析。实际上，当样品水分超过 20% 后，粉碎就很困难。

样品中水分含量显著地影响样品粉碎后颗粒的大小、形状和其分布。应注意的是：尽管筛孔是一样的，但是样品在粉碎后并非有一致的颗粒大小，颗粒直径的变异一般从 1 μ m 到 500 μ m 或更高。均一性颗粒大小是不存在的，理想上，各种大小颗粒的分布应是随机的，但对富含纤维的样品难以做到。表 12-7 显示了 3 种不同含水量的小麦样品经旋风磨（筛孔为 1.0mm）粉碎后，平均的颗粒大小和其分布。表 12-7 的前半部分是 12 次测定的平均值，后半部分是 12 次测定中的极端值。

表 12-7 小麦含水量对粉碎后平均的颗粒大小和其分布的影响

小麦品种	含水量	500~1000 μ m	500 μ m	354 μ m	210 μ m	149 μ m	74 μ m	H ₂ O/%	平均颗粒 大小/ μ m
				平均值	(n=12)				
Durum	中	1.6	4.4	15.9	15.1	28.0	34.9	12.6	193
	高	1.4	6.4	15.1	16.9	23.5	36.7	16.2	197
HRS	中	2.0	4.6	11.6	11.6	22.0	48.8	12.2	178
	高	2.2	6.3	11.6	13.6	21.0	45.2	17.0	185
SWS	中	3.4	5.7	8.7	8.5	15.8	57.4	12.7	176
	高	4.8	5.0	9.4	8.4	15.9	56.4	15.4	189
				极端值					
Durum	低	1.6	5.0	17.4	19.6	29.4	27.0	7.2	202
	高	5.1	11.7	22.6	18.0	20.6	22.0	15.9	267
HRS	低	2.2	6.6	15.2	16.0	21.4	38.7	8.7	203
	高	4.4	11.5	18.7	14.9	21.3	29.2	15.4	253
SWS	低	3.8	5.8	8.3	8.5	15.1	58.4	10.0	181
	高	5.2	6.4	9.4	9.4	15.3	54.2	15.7	200

总的来看，随着含水量的增加，样品粉碎后平均颗粒增大；各种大小颗粒的所占的比例变化的趋势为：直径为 74~149 μm 的小颗粒比例减少，而直径为 210~1000 μm 的大颗粒比例增加。另外，12 次测定中的极端值显示，样品含水量的增加，也使粉碎后颗粒大小的变异增大。

样品含水量过高影响其粉碎，那含水量低又有什么影响呢？对大多数样品，含水量低，则粉碎更易。但是，对于软质谷物如软质小麦，含水量为 6%~7% 时，对粉碎作用比含水量为 9%~12% 样品有较强的阻碍。

含水量对样品 NIR 光谱的另一个影响是通过其他成分的水合作用而产生的，因为某成分的最佳波长点与该成分的水合程度有关。这一点对于含水量高的样品例如鲜肉、蔬菜尤为重要。样品中蛋白质、纤维等成分都是与水结合态形式存在，不同成分对水的亲和力不同，例如纤维素对水的亲和力要高于淀粉。在含水量较低时，纤维素的水合程度要高于淀粉。

纯净状态化合物与在植物性样品中该化合物的 NIR 光谱是不一样的。其中可能的原因：植物中的化学成分是以水合状态存在，由于水分子的存在，影响了该化合物分析的空间排列，从而影响其分子振动，也就影响其 NIR 光谱。例如蛋白质在不同水合程度下，NIR 光谱不同，在定标选择最佳的波长点也不一样。因此，由低含水量样品建立的定标模型就很难适应于高含水量样品。然而，植物中的主要成分如蛋白质、纤维、淀粉等都是非常复杂的物质，对其 NIR 光谱的解释还远远不够，水合程度可能只是其中一个部分。

12.2.2 样品的容重

样品的容重即单位体积样品的重量，样品的容重影响样品槽中样品的数量、紧密程度和表面颗粒的排列。影响样品容重的主要因素为其化学成分，总的来说，样品中含的低密度成分（例如，纤维素、戊聚糖等）越多，则该样品的容重越小。

样品的容重越小，则样品槽表面的均匀越差。表 12-8 显示了几种常见样品的容重（样品容重的测定方法为：25mL 圆柱性容器内靠地球引力装入旋风磨粉碎后样品的重量）。

因此，对于容重较小的样品如秸秆、干草、麦麸豆壳等，为了减少因样品不均匀而造成

表 12-8 几种常见样品的容重（旋风磨粉碎筛孔 1.0mm）

样品名称	容重/ $\text{g} \cdot \text{cm}^{-3}$	样品名称	容重/ $\text{g} \cdot \text{cm}^{-3}$
FABA 豆	0.639	大麦	0.390
小扁豆	0.542	小麦麸	0.349
小麦粉(HRS)	0.514	燕麦	0.345
玉米	0.506	脱脂油菜籽	0.338
小麦(HRS)	0.495	干苜蓿	0.240
大豆壳	0.489	大麦秸	0.142

成 NIR 光谱的差异，建立使用两次粉碎的措施。第一次粉碎用锤式或切割式磨，筛孔较大；经过第一次粉碎的样品，再由旋风磨处理一次，筛孔较小。尽管两次粉碎增加了劳动量，但可提高 NIR 分析的准确性。

12.2.3 样品的物理结构

影响样品物理质地的因素同样也影响样品的粉碎、颗粒、形状，因此，也就影响到样品的 NIR 光谱。样品的物理质地包括两个方面：化学成分的含量和其结构，其中物理结构的影响可通过对样品的粉碎处理而减少。现以农产品中容重最大、而且种子最坚硬的豆科种子为例：样品越坚硬越难粉碎，样品体积越大则意味着，引入粉碎槽中样品间含有的遗传变异性小。因此，要获得均匀性的样品，两次粉碎的措施是必须的。第一次粉碎用锤式或切割式磨（例如 Christy-Norris 8 英尺锤式磨），筛孔 8~10mm，它可产生较均匀的大颗粒，而后再经过旋风磨（筛孔为 1mm）处理，可获得均匀性的小颗粒样品。棕榈籽的粉碎也应同豆科种子一样用两步粉碎处理，而硬粒小麦（Durum）尽管质地硬，但由于体积小，可直接用旋风磨一次粉碎。

总的来说，尽管质地越硬的样品较难以粉碎，但是一旦它们经旋风磨或相似磨粉碎后，其颗粒越均匀；而质地软的样品粉碎较容易，但要获得较均匀颗粒却很难。

12.2.4 影响样品物理质地的因素

不同植物由于其茎叶比和木质化程度不同，因此，粉碎后样品的质地不同。植物物理质地的不同在样品 NIR 光谱上也能充分体现。因此，可根据 NIR 光谱很准确地区分小麦和大麦。

因为茎的木质化程度比叶高，样品中叶的比例越高越容易粉碎成均匀颗粒。植物的成熟程度与木质化程度高度相关。植物木质化程度增加，以支撑逐渐变重的头部（植物种子），因此，植物的种子越大，则其茎越硬，木质化程度越高。

植物的种植环境影响植物中化学含量（谷物种子）或茎叶比（饲料），从而影响样品的物理质地。总的来说，在低到中等环境温度和湿润的环境中，获得的饲料茎叶比要高于高温（30℃ 之上）和干燥的环境。

在植物收获期时的多雨环境，易引起谷物的脱色和发芽，这将使得谷物变软；对于饲草则易引起发霉。在植物成熟期发生霜冻，易引起谷物变硬。

贮存处理不当引起细菌、霉菌的滋生，从而改变样品的物理质地。当然，贮存对样品的影响与贮存期长短、环境温度、通风等参数有关。在低温、通气好的条件下贮存样品，对其物理质地影响很小。但在生产实际中很难达到这样的条件，通常控制样品的含水量是一项实用的措施。在南方地区，样品贮存前水分含量一般低于 12%，在北方地区，一般控制在 13% 以下。

12.2.5 样品和环境的温度

在粉碎时，样品（特别是谷物）温度升高有利于其提高粉碎效果。在连续粉碎 5~7

个样品后，粉碎机就达到其最高温度。因此，为了提高粉碎质量，在样品粉碎前应让粉碎机工作 1~2min，或粉碎 3~4 个无用的样品。

对于动物性样品如奶、奶制品、肉制品、血粉、鱼粉、肉骨粉等，受温度的影响很大，而且许多样品无法粉碎。因此，近红外分析仪器制造商为此提供了一些特殊的进样设备，使得液体、悬浊液、半流体等可直接获得 NIR 光谱。

样品温度对于 NIR 分析准确性影响的研究已有很多报道，以 1982 年 Williams 对小麦分析的研究比较透彻。在其研究中发现，随着样品温度从 -30°C 上升到 45°C ，测定的小麦中蛋白质含量呈线性下降，原因是定标模型中的主要波长对温度变化的敏感性。温度影响最大的是水分敏感波长，即 $1925\sim 1945\text{nm}$ 。因为温度决定了水的状态，水分子中 O-H 的振动和转动发生变化，因此，其近红外光谱的吸光度产生变化。如果用导数光谱进行分析，则温度的影响程度加大。

环境温度通过影响仪器、样品、子样品的温度和粉碎机温度上升速度等作用，来影响 NIR 分析。在我国北方地区的冬季，谷物贮存在低温条件下，在粉碎时，粉碎机的温度上升较慢；而在南方地区的夏季，环境温度达到 $30\sim 40^{\circ}\text{C}$ 。由于上述的原因，会给近红外分析结果带来误差。

12.2.6 样品的类型

目前的研究进展显示，近红外分析的准确性较差的样品的特征为，其 1~2 种化学成分在样品间分布很不均匀，例如油料种子，其含油量非常高，高纤维含量的饲草，鱼粉，蔗糖渣等。因为，在定标中，要收集到变异范围比较广的样品是很困难的，这一点在本书的第 2 部分有详细的阐述。

12.2.7 转换因子

在分析农产品、食品、饲料等样品质量时，有一些指标是通过转换系数间接获得，例如氮-蛋白质转换系数对于大部分谷物、种子和饲料为 6.25，对于奶制品为 6.38，对于小麦和小麦粉为 5.7。另外，粗纤维这个指标并不仅仅包含纤维素，它是前人制订的一种化学分析程序获得的混合物。而 NIR 光谱却直接根植于样品的分子，因此，用 NIR 分析方法测定一些需转换因子的指标时，可能会产生一定的分析误差。

因为粗蛋白质是农产品、食品、饲料等品质分析时最常用的指标，氮-蛋白质转换系数就是最重要的转换因子。凯氏定氮法测定的是总氮含量，而不同种类、不同样品的蛋白质分子中氮的含量并不都是一个常数。据研究，样品间蛋白质含氮量变化 $\pm 0.2\%$ ，则用 NIR 分析蛋白质含量为 13.5% 的小麦时，由此而来的分析误差为 $\pm 0.16\%$ ；而分析蛋白质含量为 40% 的豆饼时，由此而来的分析误差上升为 $\pm 0.5\%$ 。Tkachuk (1977 年) 曾报道，小麦间氮-蛋白质转换系数变化可达 0.13，鹰嘴豆为 0.09，这些转化系数上的变化可导致 NIR 分析小麦和鹰嘴豆蛋白质含量时，分别产生 0.2% 和 0.3% 的误差。

12.3 来源于操作过程的误差源

12.3.1 定标程序

12.3.1.1 定标系中样品的数量

定标系中样品的数量可大可小，一般在 30~200 之间。在早期 NIR 技术中，仪器制造商不愿意提倡增加定标系中样品的数量，因为随着样品数量的增加，计算中对计算机内存和速度的要求增加很多，而当时这样要求的 CPU 价格很贵。早期的 NIR 研究者证实，定标系中样品的数量太少，所建立的定标模型是定标系样品依赖性的，应用该模型对实际样品进行分析时误差大。为了使所建立的定标模型有更强的实际应用价值，定标系应尽可能多地包含不同质地的样品。但是，样品量增加过多时，运算速度降低，同时增加了化学分析的工作量，通常一个定标模型中样品的数量为 70~100 个。因此，定标系要达到尽可能包含不同质地样品的目的，又要控制其数量，就必须对样品进行选择，选择主要从样品背景和 NIR/NIT 光谱这两个方面。

12.3.1.2 样品的选择

定标系样品选择的原则为：在搞清将要进行分析样品所包含的质地变异的前提下，尽可能地收集不同化学含量、物理性质和 NIR 光谱特征的样品，形象化地表示为“先定边疆”，然后在此域内收集异质样品。

定标系中样品不应该是随机选择的样品，其数量由样品集合内所包含质地的变异程度所决定的。若选择到样品质地的变异较窄，则仅通过简单地增加此类样品，对提高该定标模型的精确度和准确性没有帮助。例如，在一个为分析面粉中粗蛋白质含量而进行的定标工作中，由于样品的颗粒大小比较均匀，若可选择样品蛋白质含量变化范围仅为 0.5%，则由 100 个样品建立的定标模型的分析能力不会优于 30 个样品。

对于比较简单的农产品例如软质春小麦，样品质地的变异主要来源于生长季节、地域、播种时间、灌溉及施肥措施、品种、成熟和收获期内病虫害和异常气候等的影响；对于较复杂的样品如饲草，还应考虑到植物木质化程度的作用；而对于更复杂的配、混合饲料，不仅要考虑到各组单独的作用，还要考虑到配合比例及各组间的交互作用。

对于动物源的样品如奶粉、奶酪、黄油、肉及毛发等，质地变异源的主要影响因素是样品的化学含量、加工温度和加工工艺。在化学成分中脂肪、蛋白质和水分的影响力最大。对于鱼、鱼粉、肉骨粉还有动物的成熟状态的影响。另外，由于动物源产品水分、脂肪和蛋白质含量较高，贮存易发生变化，因此贮存条件也是一个重要的影响样品质地的因素。

在一个定标系样品的选择中，每个变异源至少选择 3~5 个样品。上述的选择方法都是根据样品的背景进行的。然而，一个优秀的选择技术最终还应针对样品 NIR/NIT 光谱的分析，并以样品为中心建立一个动态选择的机制。

12.3.1.3 化学分析数据的准确性

从本书的第 3 章（NIR 分析的数学基础）中，可以知道定标是以化学分析数据为准

确值，来选择波长点、确定各所选择波长点处的回归系数，建立 NIR 分析的模型。因此，一旦在定标或检验中化学分析数据有误，则建立 NIR 分析的模型就不是最佳的，以此模型分析样品必定产生误差。在应用 NIR/NIT 技术进行实际分析中，对很多产生较大误差的事例分析发现，化学分析数据的错误是罪魁祸首。

12.3.2 样品的前处理

12.3.2.1 采样和分样程序

对所有的农产品和大多数食品等固体样品的分析，无论是用物理的还是化学的方法，采样都是一个非常重要并可能产生误差的过程，近红外分析技术也不例外。目前现有的采样方法很多，这里不一一叙述，但是无论使用哪种方法，都必须使获得的样品可代表样品的化学含量、物理质地和杂质含量等方面的特征。

表 12-9 显示了不同质量小麦所含麦粒的数量，显然质量、体积越大，则所含个体数量越多。一捆约 30 公斤的干草约含有 200~10000 不同的植株。显而易见，取样的质量或体积越大，则代表性越好。

表 12-9 不同质量小麦所含麦粒的数量

质 量	颗粒数量	质 量	颗粒数量
1kg	29000	$50 \times 10^3 \text{ kg}$	$1428570000(1.43 \times 10^9)$
$1 \times 10^3 \text{ kg}$	29000000	$10000 \times 10^3 \text{ kg}$	$285714230000(2.86 \times 10^{11})$

12.3.2.2 粉碎机

样品的粉碎通常有两种程序：一次粉碎或称直接粉碎；此外是在粉碎前增加预粉碎的两步粉碎，这主要对于一些不能直接进入粉碎室的样品例如饲草、颗粒饲料、大颗粒植物种子等。目前，人们对质量分析中粉碎及粉碎机的影响作用研究越来越多，甚至有此方面的专著。有意更多了解这方面知识的读者可查阅学习。

在粉碎某种样品前，应根据该样品的特征选择最适合的粉碎机。表 12-10 是加拿大科学家 Williams 根据研究，总结各类样品所适宜的粉碎机，然而这里指的各种粉碎机都是在加拿大市场出现的商品名，这一点读者在使用时应注意。

表 12-10 各类样品所适宜的粉碎机 (Williams 推荐)

工作量	样品种类	分析类型	方法	粉碎机
大	谷物、豆子	常规 ^①	标准 ^②	KT3303,旋风磨,Cyclotec
大	谷物、豆子	常规、硬度	NIR/NIT	旋风磨,Cyclotec
中等	谷物、豆子	常规	标准、NIR/NIT	KT3303, KT3100 旋 风 磨, Cyclotec, Retsch, Christy/Norris8-in
中等	谷物、豆子	水分	标准	Wiley
中等	油类种子	油、蛋白质、水分、纤维	标准、NIR/NIT	Krups, Culatti, Moulinex, Braun
中等	谷物	物理指标如硬度	标准、颗粒指数	KT3303,Cemotec

工作量	样品种类	分析类型	方法	粉碎机
中等	饲料	常规	标准、NIR/NIT	Christy/Norris, Wiley No. 4, 旋风磨
中等	饲草、硬质种子	常规	标准	Christy/Norris, Wiley No. 4

注：① 常规指蛋白质、水分、油、纤维、矿物质、碳水化合物等；

② 标准指常规的化学分析方法。

12.3.2.3 粉碎操作

样品粉碎的效果以粉碎后平均颗粒粒度和不同大小颗粒的比例表示，它不仅取决于粉碎机的种类，也取决于粉碎条件和粉碎时的操作。对于旋风磨应定期更换粉碎室的碳化硅环和筛子，因为在多次粉碎中，碳化硅环将被磨损，而筛子孔径也在逐渐增加，长期使用一个筛子磨出的样品，其平均颗粒粒度增加。KT3303 磨槽中刺状表面随着粉碎次数增加，也将磨损，应定期更换。通常，每粉碎 50000 个左右样品后就应更换旋风磨的粉碎室的碳化硅环和筛子，每粉碎 100000 个样品后应更换粉碎室的刺状表面。样品中夹杂着铁器或石子，易损坏粉碎机（尤其是旋风磨）。在所有类型的粉碎机中，锤式磨最耐用，一般在粉碎 100000~250000 个样品后，只需更换筛子。

样品的颗粒大小显著地影响着其近红外光谱，颗粒度越大（越疏松）则其吸光度（OD）越大，颗粒度越小（越致密）则其吸光度（OD）越小（见图 12-3）。

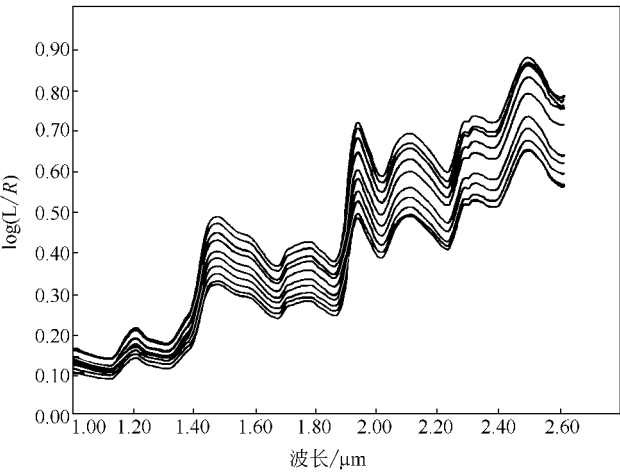


图 12-3 不同粉碎粒度小麦的近红外光谱，粗糙者吸光度大

12.3.2.4 粉碎后混合

粉碎后再混合尽管很简单，但对提高 NIR 分析的准确性是必要的。因为在粉碎机上的盛样容器中，样品是分层的，必须进行混合。方法很简单：在粉碎机的盛样容器上加上盖子，上下、左右振荡 15~20 次后即可。

12.3.3 样品的贮存

12.3.3.1 粉碎前贮存

样品在粉碎前贮存最重要的是防止昆虫、霉菌和细菌的侵害。样品贮存在牛皮纸袋中，水分含量会随环境条件发生显著变化。在干燥地区像澳大利亚、印度的次大陆、中东、北部和南部非洲、拉丁美洲和我国北方地区，样品贮存过程中水分的流失非常快。以小麦为例，在贮存了3~4周后，样品含水量达到8%时，才达到平衡湿度（即样品中水分与空气中水分的交换达到平衡）。而潮湿地区比如北部欧洲，美国和加拿大的东部，及北美的西北海岸和我国的南方地区，样品贮存在包装袋则可能增加水分含量。真菌和细菌的蔓延最可能发生在相对湿度高的区域，然而昆虫蔓延普遍的是在炎热的区域。

样品贮藏在密封的、厚0.1 mm的塑料袋中，将保护材料中水分的损失。但如果样品的水分含量较高，则该塑料袋不能保护其中样品免受真菌、霉菌的侵害。如果样品温度和湿度较低，则贮藏在金属、玻璃或硬质塑料容器中，可有效保护其不变质。如果温度和湿度高，则真菌和细菌的侵害有可能发生，并且如果存在昆虫卵，则昆虫的蔓延也将发生。

通常，粗饲料在贮存中不易发生变质，因为，粗饲料中的化学成分不易受到昆虫或真菌的侵害。通气也是良好贮藏的有效办法，贮藏中通气时，能保护许多种类型的材料长久贮存。样品贮存的理想条件是冷藏或冰冻，但是，在实际生产中，由于涉及贮存的物质量很大，冷藏或冷冻是做不到的。

在粉碎前，如果必须使样品调制成为不同含水量，则该调制时间至少为两天，并将其用金属或硬质塑料的容器或厚塑料袋贮存，以防止其水分的变化。如果调制的时间多于两天，则应该冷冻贮存，以防止细菌或霉菌的侵害或遭受虫害。

样品长期冷冻贮存情况下，则可能显著地增加样品的含水量，因此，在进行NIR分析前（定标或日常分析），应对该样品进行风干处理，即在空气中干燥。

12.3.3.2 样品粉碎后的贮存

样品粉碎后贮存的原理和方法与样品粉碎前的贮存大体是一致的。其主要的差别是，粉碎后样品的量较少，而且贮存期较短。因为样品粉碎后要不断进行混合，并用于分析。一般贮存期为48h。如果在某些情况下，要求样品粉碎后贮存期延长，则要采取必要的保护措施。

贮藏方法取决于样品的大小和化学组成（特别是含水量）、贮藏的期限、通风条件、贮存空间和可供贮藏的经费。显而易见，对于大量样品的贮存要比对于小量样品的贮存要更困难、费用也更高。一般情况下，对于粉碎后的样品，建议贮存期不要太长，特别是对于那些需要验证结果的样品，更是如此。

粉碎后需要贮藏一段时间的样品，应贮存在金属容器中，并用适合的盖子盖紧。此外，还应将该容器低温保存，以防止其遭受虫害和霉菌的侵害。盖子可以防止样品中水分与大气水分互换。对于样品量较多例如100~1000 g，最好用金属容器贮存或塑料桶贮存。建议使用圆形（圆柱状的）塑料的容器，而避免使用方形容器。因为圆形容器更易净化。而且，方形容器有角，易形成结块等弊病。最理想的贮存区是阴凉、干燥、有空气环流并配有可动的单元架子。在这种类型贮藏室，能贮存几吨样品，而且取样非常方便。其他类

型的贮存区域需要空调装置。

以 15~25 克样品为例，对于粉碎后待分析的样品，短期贮存可用小的圆柱状金属容器。这些金属容器容易清洁、贴标签和在浅盘上堆放，并且它们也易于加盖，以保护其中样品水分不与空气中水分发生互换作用，在称量前它们也便于样品的混合。

这些金属容器最好的替代物是可加盖的塑料小瓶或塑料袋子。小纸袋可用于将样品传递至实验室，但是它们主要限于传递，因为其多孔的质地，允许水向内或向外的运动。

塑料小瓶可存放在箱中，但是不能像金属容器一样贮存耐久。另外，塑料小瓶的静电作用更强，因此清洁更困难。对于分析量很大的操作，转变盛样容器将占用许多人工。

大气条件主要是指室温和相对湿度。在干燥气候中，样品中水分的遗失非常快，对于谷物类样品水分到 7% 时，才能与大气中水分达到平衡。所以，对于贮存在纸包或薄塑料袋中的样品，放置一段时间后，其水分含量降低。水分这些损失能引起分析结果产生重要的错误。即使贮存在小金属容器的样品，放置一段时间后，也会遗失水分。其快慢程度主要取决于粉碎后，样品中水分的起始含量。样品贮存罐上加上盖子并覆盖胶带，能在短期内保护样品水分的损失，但是在湿度和室温都较高时，霉菌的繁殖很迅速。

表 12-11 举例说明，经旋风磨粉碎的小麦贮存在加盖或不加盖的容器，样品水分的损失率。该研究是在加拿大的 Vancouver（英国哥伦比亚大学）进行的，在试验时，那里的相对湿度是非常高的。如果是在干燥区域，样品中水分的损失会更大。

表 12-11 粉碎后小麦贮存于用胶带密封或不密封金属罐中的水分损失

试验天数	起始的水分含量					
	14.3%水分损失/%		12.6%水分损失/%		10.4%水分损失/%	
	胶带密封	不密封	胶带密封	不密封	胶带密封	不密封
1	0	0.1	0	0.1	0	0
2	0	0.9	0	0.4	0	0.1
4	0	3.3	0	1.4	0	0.3
6	0	4.0	0	2.1	0	0.3
21	0.4	/	0.2	/	0.1	/
42	0.6	/	0.5	/	0	/

12.3.4 样品的进（装）样

12.3.4.1 进样

进样带来的 NIR 分析误差主要取决于分析前样品的混合、样品的相对密度、样品的颗粒大小和外形和进样器中样品的数量等因素。威廉（1975 年）曾研究了进样前混合样品对 NIR 分析的影响。经 Udy 磨粉碎的样品，再进行 15 次混合后进行分析，并不能提高分析的精确度。研究的对象主要为红色春小麦，软、硬小麦，大麦，燕麦和黑麦。分析误差随着样品中纤维含量的增加而增加，在所研究的样品中，燕麦产生的进样误差最大。对于富含纤维的样品，纤维含量越高，则样品加载错误越大。在某些情况下，样品的加载错误比粉碎产生的错误更高。此外，还存在样品/进样器间的相互加载作用。

装入进样器中样品数量不同，影响样品的疏密程度，也就影响了入射光的漫反射率，从而影响了 NIR 分析的结果。表 12-12 举例说明，由于进样器内样品没填充或过于充满时，对光谱信号 $\log(1/R)$ 的影响。对于谷物类样品，如果进样器过分充满样品，则光

谱信号 $\log (1/R)$ 增加；向日葵也是如此，原因可能是向日葵的油被挤出溢于表面。而对于秸秆类样品，如果进样器过分充满样品，则光谱信号 $\log (1/R)$ 反而减少，因为其密度很低。这种影响很容易克服：先将进样器完全充满，然后轻刮去多余部分，使得进样器完全充满，再加上盖子。这样，实际进入进样器样品的重量是不一致的，例如对于秸秆和其他富含纤维质样品大约为 2 g，而对于玉米几乎达 6 g，但是对于同一类样品，装入进样器内样品无论是体积还是重量都是一样的。

表 12-12 样品槽的充实程度对 NIR 光谱的影响

产品	log 0	log1	log2	log3	log4	log5	粗蛋白	水分	硬度
大麦									
不满	132.39	99.73	107.55	118.42	114.92	52.38	10.7	9.5	
正好充满	134.14	101.60	109.50	119.92	116.11	54.11	11.4	9.6	
过分充满	139.62	107.48	115.07	125.22	120.87	59.75	11.3	10.1	
硬质小麦									
不满	180.92	139.65	150.73	162.51	155.18	73.34	12.1	9.1	46.7
正好充满	183.67	142.04	153.17	164.98	157.40	76.90	12.1	9.1	45.2
过分充满	184.92	143.63	154.49	165.98	158.35	76.54	12.1	9.2	41.5
稻草									
不满	132.78	104.10	106.19	113.84	100.94	54.97			
正好充满	128.78	100.88	102.72	110.14	97.53	53.19			
过分充满	127.91	100.27	101.99	109.29	96.45	52.75			
向日葵									
不满	369.05	238.90	238.46	226.90	226.18	148.70			
正好充满	370.38	240.78	240.17	228.76	227.74	150.23			
过分充满	434.28	291.36	290.18	276.92	274.32	183.34			

太平洋公司（Pacific Scientific）仪器的实验也给出了相似的结果，其进样器内样品量大约是 DICKEY-John III 仪器的 2 倍。对于 Technicon InfraAlyzer 系列仪器，其加载错误也是可比较的，并且其进样器内样品量几乎与 DICKEY-John III 仪器一样。对于 Percon Inframatic 8000 系列仪器，由于其没有样品槽，样品直接填塞到仪器内的进样器中。填塞必须是均匀一致的，以保证精确。如果样品没有充分地填入，则可能产生相当大的误差（例如对于分析小麦蛋白质时，产生 1% 的分析误差）。

12.3.4.2 清洁

如果在扫描两个样品间没有很好的清洁，则可能产生样品的交叉污染。对于含油量低的比如禾谷类、粗饲料等样品，将前一个样品倒空后，把进样器内粉尘刷干净即可。但如果进样器的窗口是石英而不是光学玻璃，清洁粉尘则比较困难。因为石英的静电作用比较强，粉尘易黏着于其表面。对于一般光学玻璃，则静电作用相对较小。

对于含油量较高的谷物种子和一些配合的饲料（例如加油的肉鸡饲料），除了要清洁进样器内的粉尘外，还要用一定的醇溶液清洁表面的油渍。这听起来是件麻烦的工作，其实仅仅占用约 30s 的时间，而且对于克服样品的污染是不可缺少的。

13. 近红外光谱分析的光谱库与网络技术

前几章已经提到建立分析性能稳定、结果可靠的数学模型是实现近红外光谱分析的关键。建立优秀的近红外光谱分析数学模型需要以下几个基本条件：

- ① 具有优秀的近红外光谱仪，准确测定样品的近红外光谱；
- ② 具有丰富的样品资源，以取得样品复杂和重叠的背景信息；
- ③ 准确测定化学值的水平以及合理利用样品资源的技术，譬如挑选代表性样品等；
- ④ 具有优秀的定量、定性分析软件与有经验的建模人员。

建立模型必须同时具备上述几个条件，但一般近红外光谱分析用户，特别是刚开始的近红外光谱分析用户很难同时具备，这使得该项技术难于掌握，因此建立优秀的数学模型成为制约我国近红外光谱分析技术发展的主要因素。上述几个建模条件说明建立数学模型有两个要素：即建立数学模型的信息源（用于建模的、有代表性的标准样品集与光谱）与处理信息的技术（即算法）。近红外光谱分析用户建立数学模型时首先遇到的困难是建模信息源。

为了在我国推广应用近红外光谱分析技术，成立近红外技术的建模中心是非常必要的；建模中心可以帮助用户建立数学模型而不需要用户提供建模用的标准样品，即实现近红外用户的无标样分析服务。这样的建模中心必须拥有丰富的近红外信息源，可称为“有源建模中心”。“有源建模中心”的近红外信息源可以由近红外光谱库来实现。以农业生产领域近红外光谱分析谷物样品为例，我国各地作物的品种、栽培方法、气候与土壤条件千差万别，这些信息直接或间接反映在各地生产的谷物样品光谱中，收集我国各地有代表性的谷物样品，并测定其标准的近红外光谱，组成较大规模的谷物样品近红外标准光谱库，这样的标准光谱库好比是生物学研究中的基因库。只要光谱库的信息足够丰富，对于任何一个近红外用户要分析的谷物样品，“有源建模中心”总可以通过在光谱库中搜索，找到合适的代表性样品，由库中的这些样品光谱建立数学模型后，就可准确测定未知样品，从而避免了用户必须提供标准样品才能建立数学模型的困难。运用近红外标准光谱库，还能够通过模型转移的技术使不同类型、不同厂家的仪器共享光谱库的信息资源，使我国不同领域、不同档次的仪器都充分利用起来。光谱库建立后，还可以应用网络技术让用户通过访问光谱库的服务器实现近红外网上分析，建立网上实验室，极大方便国内用户运用近红外技术，也便于实现国际交流。通过近红外光谱库和网络技术可以实现近红外信息资源的共享，推广近红外技术的应用。

13.1 近红外光谱分析的光谱库技术

13.1.1 近红外光谱库技术的国内外发展现状概述

自 20 世纪 80 年代特别是 90 年代中期以来，近红外光谱库技术在国际上得到了非常

迅速的发展，一些大的近红外仪器公司（譬如美国的 FOSS 公司等）开始在全球范围内收集他们近红外光谱库中的标准样品，并把利用该光谱库所产生的定标数学模型固化在仪器上一起出售，既减轻了用户建模困难的负担，又取得了经济效益。

同时国际上也出现了许多专门开发各个不同领域近红外光谱库的公司、科研机构等单位，并通过网络技术来共享这些光谱库资源，为许多近红外用户提供各种服务。譬如美国的 FDM（Fiveash Data Management, Inc.）建立了食品和农产品的近红外光谱库；美国农业部的农业研究服务中心专门分别在 BRUKER 公司和 FOSS 公司的仪器上建立了大米的近红外光谱库；德国的 Buchi 公司建立了化工领域内许多工业在线产品譬如塑料、化妆品等的近红外光谱库；NeoLink 公司收集了大约 15000 个许多工业资源产品的近红外光谱库，并通过网络共享这些资源；Galactic 公司也建立了近红外光谱库，并通过网络免费提供该库中的资源；国际上在诸多领域从事近红外光谱库开发的公司和单位还有许多。

国内对近红外技术的应用热潮虽然已经有几次，第一次是在 20 世纪 70 年代，第二次是在 20 世纪 80 年代，第三次是从 20 世纪 90 年代中期开始到现在。可前几次的应用热潮中近红外技术在国内并没有真正发挥其应有的作用，这与国内没有及时地从事近红外标准样品光谱库开发的单位有直接的关系；从 20 世纪 90 年代以来国内也出现了许多从事近红外仪器研发的单位，但专门从事近红外标准样品光谱库开发的单位却几乎是没有，只有在国家的“十五”科技攻关中资助建立了中国部分农产品近红外光谱库。

13.1.2 开发近红外光谱库的必要性和重要性

与传统分析技术不同，近红外光谱分析是从复杂、重叠、变动的背景中提取弱信息，这种技术必须借助数学模型，优秀的数学模型必须从大量样品中选择代表性样品才能建立，因此作为各个行业中发展应用近红外技术的基础，必须建立具有一定规模标准样品的近红外基础光谱库，才能实现近红外光谱分析技术的充分应用。

在我国建立自己的、拥有独立知识产权的近红外标准基础光谱库更是非常必要，近红外仪器虽然可以进口，但样品资源却无法引进，只有建立自己样品资源的近红外光谱库才能真正实现充分利用近红外光谱分析技术为我国国民经济服务的目的。

同时利用近红外基础光谱库建立的数学模型，通过网络共享可实现近红外用户的无标样分析服务或用户可进行不定标分析，从而大大降低近红外用户使用近红外分析技术的要求，极大的方便近红外用户；并可将有模型捆绑在国产的近红外光谱仪上，带动国内近红外光谱分析仪器事业的发展。

13.1.3 中国部分农产品近红外光谱库简介

近红外光谱库技术实质是一种充分合理利用资源的软件技术，中国是国际上拥有农产品资源最丰富的国家，“部分农产品近红外基础光谱库”中丰富的农产品资源可为我国培育优良品种提供基础；此外，近红外光谱包含了样品中几乎全部有机成分的信息，通过建立光谱库过程中测定的以及建完库后用近红外模型分析的各种品质性状参数（即各种化学成分含量、物理特征等信息）可以用来建立部分农产品的品质性状参数库，为研究部门、政府部门、农产品加工、贸易厂商等提供农产品品质信息咨询服务，为继续开发利用奠定

基础。

该光谱库现有在全国范围内广泛收集的小麦、玉米、水稻、油菜籽、大豆等有代表性样品共 10000 个左右，并测定其中代表性最好的 2000 个左右样品的化学值，包括蛋白质、淀粉、脂肪等指标；在下一个 10 年内争取能基本覆盖全国所有农作物的代表性样品，使该光谱库的规模达到 10 万个。这样通过利用该光谱库可以构建我国近红外技术平台，以此沟通我国近红外技术的用户、近红外产业的厂家与近红外技术的研究单位，以及其他从事优质农业科技工作者的需求，并为有关人员提供技术咨询与技术服务，从而有力推动我国优质农业的发展。

为了充分合理的利用开发这些样品资源，必须具备为该光谱数据库服务的一系列软件系统，下面就简要介绍一下该光谱数据库的软件系统。

13.1.3.1 该光谱数据库中基本物理表（库）的组成

① 农产品资源信息库：表中的基本字段包括：统一编号（光谱文件名称）、贮藏号（可判定有无贮藏标样）、其他编号（国家品质资源统一编号等）、作物类型、状态、样品量、品种、年份、产地、有无标定化学值等。

② 农产品光谱资源库：表中的基本字段包括：光谱类型（傅里叶、光栅、单波长点）、仪器型号、统一编号、起点波长、终点波长、光谱点数、扫描日期、分辨率、光谱数据等。

③ 农产品近红外模型库：表中的基本字段包括：模型名称、模型类别、建模样品数目、模型成分名称、模型说明等。

④ 农产品光谱特征信息库：表中的基本字段包括：统一编号、特征信息类型、特征量的个数、特征信息数据等。

⑤ 农产品品质性状参数库（化学值信息库）（包括化学方法测定、近红外标定两大类）：表中的基本字段包括：统一编号、化学值标定类型、成分名称、成分含量等。

13.1.3.2 该光谱数据库中基本应用程序的组成

① 农产品资源信息以及品质性状参数（化学方法测定）数据库的建立、维护、使用、查询的实验室管理系统软件。能使操作使用人员能够方便、快速地向数据库中添加样品资源信息、品质性质参数信息记录，能够根据数据库中的特征字段实现对样品的分类贮藏、统一管理、快速查找等功能。

② 能够把近红外分析结果根据统一编号字段自动写入农产品品质性状参数库中的软件。

③ 为近红外光谱库中光谱资源充分利用的外围软件：这些软件是课题组近年来已完成开发的、可以独立运行的系统子软件模块，其中包括：a. 定量分析软件模块；b. 定性分析软件模块；c. 挑选样品（包括代表性、匹配性等）软件模块；d. 模型转移算法软件模块等。

④ 能够对不同类型模型进行管理、资源共享的软件。

⑤ 能够将农产品资源信息库以及品质性状参数信息库实现网络共享、查询的网页服务器软件。在该网页中，访问者能够根据作物类型、品种、年份、产地、成分名称等特征进行样品资源的信息查询，并且能够根据查询结果显示每种组分含量的分布曲线图（对于有国家品质资源统一编号的样品，显示出查找该样品农艺性状的方法）。

⑥ 为光谱数据资源充分利用相应软件提供光谱库数据接口平台的软件。

13.1.3.3 该光谱库与应用软件之间关系

光谱库与应用软件之间关系的示意图见图 13-1。

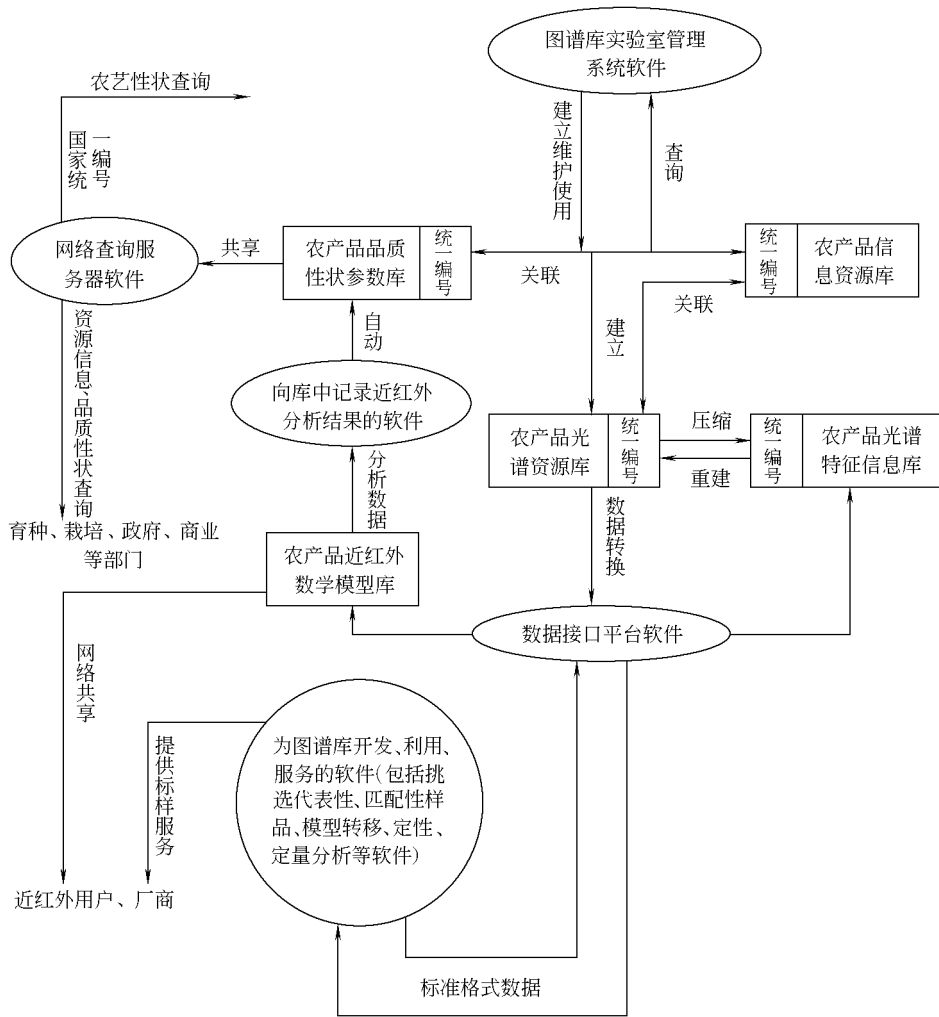


图 13-1 光谱库与应用软件间关系

13.1.3.4 与其功能相对应的该光谱库软件的主界面

光谱库主界面如图 13-2 所示。

该软件主要用以下 6 个主要菜单工作窗体实现该光谱库的主要功能：
样品光谱信息库管理菜单窗体：实现对样品光谱信息库管理工作窗体的调用；
光谱模型库管理菜单窗体：实现对光谱模型库管理工作窗体的调用；
品质性状库管理菜单窗体：实现对品质性状库管理工作窗体的调用；
光谱特征库管理菜单窗体：实现对光谱特征库管理工作窗体的调用；
数据查询与输出菜单窗体：实现对数据查询与输出工作窗体的调用；
信息处理软件菜单窗体：实现对信息处理软件工作窗体的调用。



图 13-2 中国部分农产品近红外光谱库主界面

13.2 近红外光谱分析的网络技术

13.2.1 概述

建立优秀数学模型不但需要好的仪器、优秀的分析软件、建模经验丰富的人员，还要有丰富的样品资源、合理应用样品资源技术和准确测定化学值水平等，因此，建立一个优秀的数学模型往往需要花费大量的人力、物力、财力。所以，在没有损失定标模型预测能力的前提下，应在近红外光谱分析仪器间实现模型的不断维护、转移与传送，充分达到资源的共享，它可使近红外光谱分析仪的用户避免建立、选择适当的定标模型等问题的困扰，减少对使用者的技术要求；通过一个定标服务中心和各个最终用户间网络的联系、数据的传递，使每个用户都可使用最佳的定标模型。由此可见，应用网络技术，可使近红外光谱分析的信息资源得到最大限度的共享与充分利用，有利于近红外分析技术在各个领域的推广应用。

近红外光谱分析的网络技术包括两部分，一部分是利用计算机的网络通讯技术，实现近红外模型、光谱、数据等信息资源的传输与共享，通过建模中心来实现对一大批近红外用户的服务、监控与管理；另一部分是不同近红外仪器之间的标准化问题，它是实现网络化的前提，本章对近红外仪器之间的标准化问题做简单介绍，实现不同仪器之间共享数学

模型即模型转移技术的更详细介绍请读者参见本书第 10 章“近红外光谱分析数学模型的维护”。

13.2.2 应用网络技术实现近红外信息资源传输与共享

现在的计算机网络技术得到了飞速的发展，可以充分利用局域网、电话网、无线等诸多通讯工具，为近红外信息资源的充分共享提供了有利条件；但通过网络技术实现信息资源的共享，并不是简单只借助于计算机网络的通讯技术，而应该是通过网络技术实现近红外信息资源的有效管理与共享，为近红外用户提供技术支持与服务。

完整的近红外分析网络系统，应该包括近红外分析网络系统客户端软件及近红外分析网络系统服务器软件。前者供一般近红外用户使用；后者供建模中心使用，用来对不同用户进行管理，实时的响应并处理用户的请求，并配合近红外光谱库，定性定量分析软件，数学模型等，进行模型的不断建立、维护、转移与传送。

13.2.2.1 基于 TCP/IP 协议的 WinSocket 网络编程简介

可以应用许多网络编程技术来实现近红外分析网络系统客户端以及服务器软件的编写，本书简单介绍一下灵活性比较好的基于 TCP/IP 协议的 WinSocket 网络编程，希望能对有兴趣的读者提供一些参考。

WinSocket 是一组用 C 语言写的 API，用于通过 Internet 进行数据传输。虽然现在有很多现成的工具如 WWW 浏览器、FTP 程序可以实现在 Internet 上传输数据和文件，但是，通过 WinSocket 编程可以获得更大的灵活性。编写 WinSocket 应用程序本来是很麻烦的，不过有些开发工具譬如 C++ Builder 等，并不需要直接与 WinSocket 中的 API 打交道，因为它提供了 TClientSocket 元件和 TServerSocket 元件，这两个元件封装了 Windows 的有关 API，使得对 WinSocket 的访问大大简化。

(1) Socket 概述

要通过 Internet 传输数据，至少需要一对 Socket（意为插座），其中一个 Socket 在客户端，另一个 Socket 在服务器端，一旦客户端和服务器端都接通了 Socket，客户端和服务器端就可以相互通信，就好像墙上的电话插孔一样。用 Socket 建立的连接是建立在 TCP/IP 协议基础上的，同时也支持其他相关的协议，如 XNS、DECnet 以及 IPX/SPX 等。

C++ Builder 分别用 TClientSocket 元件和 TServerSocket 元件来操纵客户端 Socket 与服务器端 Socket 的连接和通信。要说明的是，这两个元件用于管理服务器和客户的连接，本身并不是 Socket 对象，操纵 Socket 对象的是 TCustomWinSocket 及其派生类，如 TClientWinSocket、TServerWinSocket 和 TServerClientWinSocket 等。

根据连接启动的方式以及本地 Socket 要连接的目标，Socket 之间的连接可以分为三种类型：客户端连接、监听连接以及服务器端连接。所谓客户端连接，是指由客户端的 Socket 提出连接请求，要连接的目标是服务器端的 Socket。为此，客户端的 Socket 必须首先描述它要连接的服务器端 Socket（主要是指服务器端 Socket 的地址和端口号），然后再定位所要连接的服务器端 Socket，找到以后，就向服务器端 Socket 请求连接。当然，服务器端的 Socket 此时未必正好处于准备好状态，不过，服务器端的 Socket 会自动维护客户请求连接的队列，然后在它认为合适的时候向客户端 Socket 发出“允许连接”（Ac-

cept) 的信号, 这时客户端 Socket 与服务器端 Socket 的连接就建立了。所谓监听连接, 服务器端 Socket 并不定位具体的客户端 Socket, 而是处于等待连接的状态。当服务器端 Socket 监听到或者说接收到客户端 Socket 的连接请求, 它就响应客户端 Socket 的请求建立一个新的 Socket 句柄并与客户端连接, 而服务器端 Socket 继续处于监听状态, 还可以接收其他客户端 Socket 的连接请求。所谓服务器端连接, 是指当服务器端 Socket 接收到客户端 Socket 的连接请求后, 就把服务器端 Socket 的描述发给客户端, 一旦客户端确认了此描述, 连接就建立了。

(2) 建立服务器端 Socket

把一个 TServerSocket 元件加到 Form 或数据模块上, 应用程序就变成了一个 TCP/IP 服务器。当服务器处于监听状态时, 服务器端的 Socket 对象用 TServerWinSocket 来操纵, 当客户提出连接请求, 服务器响应了连接请求并建立了连接, 此时用 TServerClientWinSocket 来操纵服务器端 Socket 与客户端 Socket 的连接。

要使服务器能监听客户的连接请求, 就必须首先指定端口号 (设置 TServerSocket 元件的 Port 特性), 当然, 如果服务器提供的是某些标准的服务如 FTP、HTTP, 这些标准服务的端口号是事先约定的, 也就是说, 只要指定了服务类型 (设置 TServerSocket 元件的 Service 特性), 端口号就已经隐含确定了。如果同时设置了 Port 特性和 Service 特性, C++ Builder 将忽略 Port 特性。指定了端口号或由 Service 特性隐含指定了端口号后, 就可以调用 TServerSocket 元件的 Open 函数开始监听。如果在设计期就把 Active 特性设为 True, 服务器就自动进入监听状态, 也就是说, 把 Active 特性设为 True 和调用 Open 是等价的。一旦进入了监听状态, 就可以通过 TServerSocket 元件的 Socket 特性返回服务器端的 Socket 对象 (TServerWinSocket), 并由此对象获得有关连接的信息。

服务器进入监听状态后, 当客户提出连接请求, 服务器就自动接受请求建立连接并触发 OnClientConnect 事件。此时, 服务器端 Socket 与客户端 Socket 的连接由 TServerClientWinSocket 来操纵。TServerWinSocket 的 Connections 特性可以返回一个数组, 该数组由服务器当前活动的连接 (TServerClientWinSocket) 组成。如果要断开连接, 只要在服务器端调用 Close 函数或者把 Active 特性设为 False, 这将导致所有与客户的连接都断开, 并退出监听状态。如果在客户端调用 Close, 将只断开该客户与服务器的连接, 不影响其他客户的连接。TServerSocket 的直接上级是 TCustomServerSocket, 而 TCustomServerSocket 又是从 TCustomSocket 继承下来的, TCustomSocket 又是从 TComponent 继承下来的, 因此, TServerSocket 元件是典型的非可视的元件。要注意的是, TCustomSocket 是 TServerSocket 元件的虚拟基类, 所以不能直接创建 TCustomSocket 的实例。

(3) 建立客户端 Socket

把一个 TClientSocket 元件加到 Form 或数据模块上, 应用程序就变成一个 TCP/IP 客户。此时可用 TClientWinSocket 来操纵客户端的 Socket 对象。也可以把 TServerSocket 元件和 TClientSocket 元件加到同一个 Form 或数据模块上, 这样, 应用程序就兼具 TCP/IP 服务器和客户两种角色。

要建立与服务器的连接, 首先需指定要连接的服务器。指定服务器可以有两种方式, 一是设置 Host 特性指定服务器的主机名如 `http: //www. wSite. Com`, 这种方式

比较直观，但要花费域名解析的时间。如果不知道主机名，或者又想节省域名解析的时间，可以设置 Address 特性直接指定主机的 IP 地址如 123.197.1.2。这两种方式的效果是等价的，但如果同时设置了 Host 特性和 Address 特性，C++ Builder 将忽略 Address 特性。要建立与服务器的连接，还要指定客户要连接的服务器上的端口号，这里也有两种方式，一是直接设置 Port 特性指定端口号如 7000，也可以设置 Service 特性隐含确定端口号。如果同时设置了 Port 特性和 Service 特性，C++ Builder 将忽略 Port 特性。指定了服务器以及要连接的端口号后（这称为在客户端描述服务器），只需调用 Open 函数或者把 Active 特性设为 True，客户端 Socket 就向服务器端 Socket 提出连接请求。如果服务器此时处于监听状态，就会自动接受请求并建立连接。一旦建立了与服务器的连接，就可以用 TClientSocket 元件的 Socket 特性返回客户端的 Socket 对象（TClientWinSocket），并由此对象获得有关连接的信息。要断开连接很简单，只要调用 Close 函数，此时，服务器端将触发 OnClientDisconnect 事件，而客户端将触发 OnDisconnect 事件。TClientSocket 元件的直接上级是 TCustomSocket，而 TCustomSocket 又是从 TComponent 继承下来的，因此，TClientSocket 元件也是一个典型的非可视的元件。

（4）怎样在网络上传输数据

一旦服务器端 Socket 与客户端 Socket 之间建立了连接，就可以通过 Internet 传输数据和文件，这里先要介绍 WinSock 中“阻塞”与“非阻塞”的概念。

在 Windows 环境下，一般采用“非阻塞”方式。对于客户端 Socket 来说，如果把 ClientType 特性设为 ctNonBlocking，表示采用非阻塞方式进行连接。当位于另一端的服务器端 Socket 试图进行读或写时，客户端 Socket 就会得到通知（OnRead 事件或 OnWrite 事件）。对于服务器端 Socket 来说，如果把 ServerType 特性设为 stNonBlocking，表示采用非阻塞方式进行连接。当位于另一端的客户端 Socket 试图进行读或写时，服务器端 Socket 就会得到通知（OnClientRead 事件或 OnClientWrite 事件）。

在响应这些事件的句柄中都有一个 Socket 参数，就是服务器端或客户端的 Socket 对象（TClientWinSocket 或 TServerWinSocket 或 TServerClientWinSocket），这些 Socket 对象的共同基类是 TCustomWinSocket，TCustomWinSocket 提供了在 Internet 上传输数据的方法，例如，要读取数据，可以调用 ReceiveBuf 或 ReceiveText，要发送数据，可以调用 SendBuf、SendStream、SendText 或 SendStreamThenDrop，其中，SendStreamThenDrop 能够在发送完数据后自动断开 Socket 的连接。SendStream 和 SendStreamThenDrop 适合于发送流对象，例如文件，不过，并不需要显示删除流对象，因为当 Socket 之间的连接断开时，Socket 会自动释放所有的流对象。

与“非阻塞”方式不同的是，在“阻塞”方式下没有诸如 OnRead 或 OnWrite 等异步事件，Socket 必须主动去读或写数据，在读或写操作完成之前，应用程序处于等待状态，这样就不可避免地会影响整个应用程序的性能。对于客户端 Socket 来说，如果把 ClientType 特性设为 ctBlocking，表示采用阻塞方式进行连接。为了尽可能地减少“阻塞”方式的副作用，可以把所有涉及到读写的操作放在一个单独的线程中，这样，在 32 位的 Windows 环境下其他线程能够继续得到执行。当然，如果在读写函数中客户程序本身没有其他事情要做，也就不必创建一个线程了。对于服务器端 Socket 来说，如果把 Server-

Type 特性设为 `stThreadBlocking`, 表示采用阻塞方式进行连接, C++ Builder 为每一个阻塞方式的连接自动分配一个新的线程, 这样, 即使有某个客户正在进行读写操作, 其他客户也不必等待。服务器端的 `Socket` 用 `TServerClientThread` 对象来操纵每一个线程, 当线程开始执行时, 它就检查位于另一端的客户是否正在试图进行写, 如是, 就触发 `OnClientRead` 事件, 如果检查出客户不在试图写, 它就触发 `OnClientWrite` 事件让服务器去写。实际上, 这种用线程对象来管理连接的方式可以认为是“假阻塞”方式, 就编程而言, 与“非阻塞”方式是相似的。比较麻烦的是客户端, 它怎么知道对方(指服务器端)已准备好读或写呢? 这就要用到 `TWinSocketStream` 对象, 调用这个对象的 `WaitForData` 可以保证服务器和客户的同步。用 `TWinSocketStream` 对象进行读或写时, 如果在 20s 内操作还没有完成, `TWinSocketStream` 对象就认为超时了, 它就会自动断开连接, 这样就能避免应用程序锁死。正是由于 `TWinSocketStream` 对象具有超时保护机制, 也可以把 `TWinSocketStream` 对象用于服务器端进行读或写。要注意的是, `TWinSocketStream` 对象不能用于“非阻塞”方式。也许有的读者要问, 既然阻塞方式可能会影响应用程序的性能, 虽然用多线程技术可以加以弥补, 但又使编程变得复杂, 为什么还要用“阻塞”方式呢? 这是因为, 在阻塞方式下, 读或写是主动的, 可以在读或写操作完成后安全地关闭 `Socket` 的连接。而在“非阻塞”方式下, 读或写是随机的, `Socket` 必须经常保持在连接状态才能捕捉到异步事件。

13.2.2.2 近红外分析网络系统应具备的主要功能介绍

近红外分析网络系统软件包括两部分: 近红外网络远程分析服务器端(此服务器位于建模中心)和近红外网络远程分析客户端(位于各个近红外分析站点、实验室、生产车间、原料收购站和库房等)。

(1) 近红外分析网络系统服务器端应具备的主要功能

- ① 为不同的客户提供共享模型、在线模型、共享光谱文件。
- ② 通过客户端上载的光谱和模型, 对各客户端测试样品所用的仪器参数、分析方法、光谱扫描、模型应用和测试结果等进行监控, 确保数据的规范, 以保证测试结果的高可靠性。
- ③ 对不同客户的上载模型进行优化。
- ④ 将上载光谱进行优选并且添加进光谱库, 丰富光谱库以提高模型的预测精度。
- ⑤ 对客户的应用进行审批, 并对各客户端授予不同的等级, 按照不同级别提供网络服务。

(2) 近红外分析网络系统客户端应具备的主要功能

- ① 客户端根据对测试的要求向服务器提出服务申请, 经服务器端审批同意后可以按一定的级别运用服务器提供的共享模型进行网上即时的远程分析, 或者从服务器上下载最新版本的模型进行本地分析。
 - ② 定期的用标样检测服务器端的数学模型, 保证近红外分析结果的准确性和可靠性。即提供优秀的数学模型。
 - ③ 将光谱和数学模型上载到服务器, 有专业人员对其进行分析和优化。
 - ④ 将光谱和数学模型下载到客户端, 对客户已建立的模型进行检验和优化。
- (3) 实现近红外分析网络系统客户端主要功能的实例

① 近红外网络远程分析客户端主界面（图 13-3）。由中国农业大学设计开发的近红外网络远程分析客户端主要具备以下功能：



图 13-3 近红外分析网络系统客户端主界面

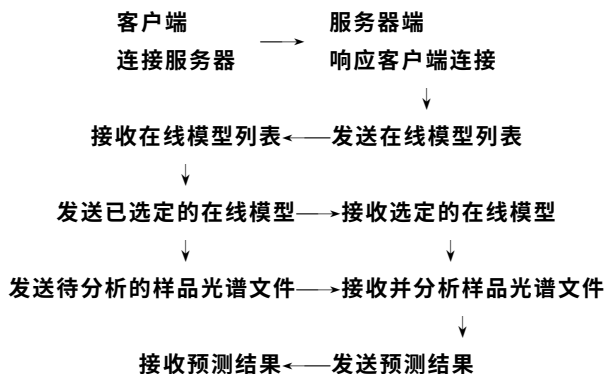
申请用户名、远程分析、本地分析、下载模型、下载光谱、上载模型、上载光谱、检验模型和转移模型等主要功能，另外，还有对分析结果的处理部分，包括：保存分析结果、打开分析结果、打印分析结果和清空分析结果。点击“离线”按钮退出远程分析。

② 申请用户名。客户要想拥有近红外网络远程客户端的相关使用权，就必须首先申请到一个合法的用户名，在这个用户名中包括“用户名”、“相关信息”和“模型和光谱详细描述”这三部分。整个申请过程如下：

点击“申请用户名”按钮，跳出如图所示的对话框，填入“主机名”、“用户名”、“相关信息”、“模型和光谱详细描述”后，点击“确定”；如果主机连接成功，则跳出“申请成功，请等待注册！”的提示框，否则提示“连接失败”或者“该用户名非法，请重新申请”。

③ 远程分析。客户通过近红外扫描仪器得到的样品的扫描光谱以文件的形式保存，要知道样品的成分含量，客户要做的只是将待分析的光谱文件发送到服务器，选择好数学模型（由服务器端发送给客户），余下的只是等待服务器将预测的成分含量结果传送过来了，这正是各近红外用户想得到的。这就实现了即时分析的功能。

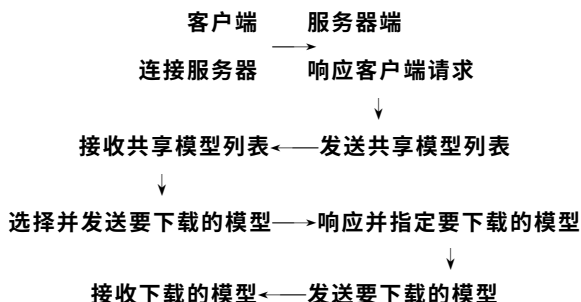
整个分析流程如下：



④ 本地分析。如果客户拥有自己的数学模型，就可以在客户端进行近红外本地分析。数学模型可以由客户自己建立，也可以从服务器下载。本地分析的操作十分简便，客户只要选择好待分析样品和相应的数学模型，就可以开始分析了。

⑤ 下载模型和下载光谱。有源建模中心（服务器端）提供了共享的模型和光谱供不同级别的客户使用（有“下载模型”和“下载光谱”使用权限的客户才可以使用）。模型的类型有两种：建模中心建立的数学模型和客户上载的已经由服务器优化了的模型。用户下载后可以直接使用。供下载的光谱都是各种样品中具有代表性的光谱，用户下载光谱可以用来检验客户机上的数学模型，是否符合有源建模中心的规范。通过检验可以提高分析结果的准确性。

下载流程如下：（以下载模型为例）



⑥ 上载模型和上载光谱。上载模型和光谱有两大用处：第一，通过上载的模型和光谱，有源建模中心可以对各客户测试样品所用的仪器参数、分析方法、光谱扫描、模型应用和测试结果等进行监控，了解各客户端仪器和软件使用的特性，为有效的修正和优化客户端的数学模型收集资料，以保证客户端分析结果的高可靠性。第二，如果客户有比较典型的具有代表性的光谱文件，可以将光谱文件上载给服务器加入光谱库；有比较优秀的可共享模型，可以上载给服务器实现共享，让拥有“共享光谱”权限的客户使用（这样就给其他用户节约了建立相同模型的资源），客户有模型需要优化，可以将模型上载给服务器，利用服务器端丰富的光谱库资源对它进行优化，在较短时间内就完成了对模型的优化，为客户节约了大量的人力、物力和时间。

以上载光谱为例，点击“上载光谱”按钮，出现如图所示的界面：点击“添加光谱文件”按钮，选择想要上载的光谱文件，则光谱文件路径显示在“光谱文件路径列表”框中，选择好要上载的光谱文件后，选择“主机名”和用户名，然后点击“上载”按钮。如

果连接成功，则弹出“连接成功！上载光谱吗？”提示框，按“确定”按钮：上载光谱开始，通过主窗口中的计数器可以掌握上载的情况。上载完成后，跳出“光谱文件上载完成！”提示框。

⑦ 检验模型。建立模型是一项非常费时、费力的工作，一个好的模型可能要经过好几个月甚至好几年的时间去修正，在这一过程中必须不断地检验模型。另外，一个建立好了的数学模型也不可能具有长期的稳定性、可靠性和动态适应性，对相同种类的不同样品在进行光谱分析时，由于光谱测试仪器、分析方法、人为因素等各方面的不同，分析结果可能有一定的偏差，解决这一难题的方法如下：在分析前，用标准的样品光谱对数学模型进行检测，查看该模型是否适合该类型样品的近红外分析，如果分析结果不理想，可通过“模型转移”方法（3.1.6 中详细解释）对该模型进行处理，以达到合适的分析结果。对服务器共享的数学模型，用户也可以进行检验（在线检验），将检验结果传送给建模中心，建模中心再对该共享模型进行相应的处理，以提高模型的适应性和可移植性。只有这样客户才会放心地使用该模型。

点击“选择在线模型”，向服务器发送用户名和模型相关信息（客）→接收用户名和相关信息，发送共享模型列表（服）→接收共享模型列表，选择并发送待检验模型（客）接收检验模型并且发送模型检验成分名（服）→接收检验成分名，选择并发送标准样品光谱文件（客）→接收标样光谱文件并且预测成分含量，发送预测的成分含量（服）接收预测值，与真实值比较出检验报告（客）。

⑧ 转移模型。因为光谱扫描仪器性能的不同，或者扫描条件、扫描时间的变化，以及样品状态的差别，使得客户端采集得到的光谱可能与建模所用光谱之间存在着差异，这样在经过模型检验后，若发现已有模型不符合预测要求（如 RSEP 过大等），这就必须进行模型转移，使转移后的模型能有较强的适应性，这样各个客户之间才能真正意义上实现模型的共享。

模型转移界面如图所示，流程如下：

打开待转移模型→添加标准样品集→选择模型转移后的路径名→转移模型开始→转移模型结束，查看转移模型分析报告

（4）实现近红外分析网络系统服务器端主要功能的实例

近红外网络远程分析服务器端主要功能是：管理用户、限制用户权限、监控客户端；为用户提供共享模型、在线模型、共享光谱文件；对不同客户的上载模型进行优化；将上载光谱进行优选并且添加进光谱库。网络系统服务器主界面如图 13-4 所示。

① 管理用户、限制用户权限、监控客户端。对客户端申请的用户名，如果申请成功的话，将列在服务器“待建用户”列表框内。如果某用户享有利用服务器端资源的权利，服务器端管理员就会通过“查看用户资料”了解此用户想要享有的资源内容。通过客户端上载的模型和光谱，服务器端能够监控客户测试样品所用的仪器参数、分析方法、光谱扫描、模型应用和测试结果等情况。服务器端管理员通过用户的级别可以限制用户的权限，并且有注销用户的权力。

② 提供共享模型、在线模型和共享光谱文件。服务器端管理员通过“查看用户资料”为此用户分配好相应的共享模型、在线模型和共享光谱文件。然后点击“添加用户”，将“待建用户”添加到“已建用户”列表中，那么该用户名就成了合法用户名，该用户可以



图 13-4 近红外分析网络系统服务器主界面

享有他具备的一切权限。

服务器端提供的各种资源有一定的针对性，不同的客户享有的资源都会是不同的：对小麦模型的用户，服务器不会给他提供水稻的数学模型。这样就方便了客户端的操作，也方便了服务器端管理员对用户的管理。

③ 对不同客户的上载模型进行优化和建立光谱库。对不同客户上载的模型，通过服务器端强大的光谱库进行优化，这是一个比较复杂的过程，尤其是光谱库的建立，首先要收集到大量的样品光谱文件，然后从中选取优秀的光谱文件用于建立光谱库。这些样品光谱文件不可能由建模机构单独去收集，这就必须通过各个近红外用户在实际过程中将优秀的样品或者样品光谱文件提供给建模机构。将样品光谱文件提供给建模机构可以通过近红外网络客户端的“上载光谱”这一功能实现，这样就大大加快了光谱库的建立。建立好光谱库后，对模型的优化就变得很方便了。

13.2.3 实现近红外光谱分析的标准化

实现近红外光谱分析标准化有两条途径：①调整定标模型的截距和系统偏差，使得不同仪器对同一样品的分析，获得相同的结果。这是一种早期简单的标准化处理，其缺点是对仪器间同一样品获得不同光谱信号无能为力，这种技术要达到分析间标准化程度低。②通过调整仪器的波长准确性、光密度和带宽，使得不同仪器对于同一样品，能获得相同的光谱信号。这是一个先进的标准化技术，仪器间经过这种处理后，无需再对定标模型的截距和系统偏差进行调整。

有关实现标准化即实现不同仪器之间共享数学模型即模型转移技术的更详细介绍请参

13.3 小结

应用近红外光谱库和网络技术可以帮助用户解决数学模型的建立和维护的困难，大大降低对分析人员的要求，将近红外光谱分析技术充分推广应用起来，使原本使用难度较大的近红外技术逐步发展为使用方便的“傻瓜”式分析。

参 考 文 献

- [1] 李军会等. 中国农业近红外分析技术软件及网络系统的研制与开发. 现代仪器, 2000 (5)
- [2] 金钦汉. 从'98 匹茨堡会议看分析仪器的发展趋势. 现代科学仪器, 1999 (3): 10~13
- [3] 金钦汉. 从'99 匹茨堡会议看分析仪器的跨世纪发展动向. 现代科学仪器, 1999 (3): 8~10
- [4] 范世福. 新形势下我国分析仪器事业的发展. 现代科学仪器, 1999 (3): 10~13
- [5] Cheung J Y, Hull S J. Detection of abnormal eleitrocardiograms using a neural network approach. In Proc Ann Int Conf. IEEE Eng Med Biol Soc, 1989: 2015~2016
- [6] J Zupan, J Gasteiger. Neural networks: a new method for solving chemical problems or just a passing phase? *Anal Chim Acta*, 1991. (248): 1~30
- [7] Dr Chrles R Hurburgh. Performance of a network of whole grain NIRS Units for Coarse Grains Proc Of the 9th International NIRS Conf P16 AGRO
- [8] Peter Tillmann, Christian Paul. Networking of NIRS instruments for rapeseed analysis using different procedures Proc Of the 9th International NIRS Conf P38 AGRO
- [9] James H Duckworth. SPECTROSCOPY AND THE INTERNET: A NEW MODEL FOR REASEARCH Galactic industries Corp, No 349, PITTCO' 99 Book of Abstracts
- [10] Henry Buijs, Daniel Gravel, Thomas Buijs et al. THE NEW NETWORKIR, THE NEXT GENERATION FT-NIR BASED FIBER OPTIC COUPLED MULTIPOINT PROCESS ANALZER; RAISING THE STANDARDS OF RELIALIABILITY AND EASE OF USE. Bomen Inc450 St-Jean-Baptiste Ave, No 1333, PITTCO' 99 Book of Abstracts

应 用 篇

近红外光谱分析可以用于与含氢基团有关的各种分析，而且测定样品不需要前处理，因此应用范围极为广泛，可以用于众多的产业和一些学科与技术中。在本篇第 14~20 章中介绍了近红外光谱分析在农业等九个产业中的应用，第 21~24 章介绍了该技术在两个学科（生物学和矿物学）和两个技术（遥感技术、过程控制技术）中的应用。

14. 近红外光谱技术在作物品质分析中的应用

14.1 概述

随着光学、计算机数据处理技术、化学计量学理论和方法的不断发展，以及新型 NIR 仪器的不断出现和软件版本的不断翻新，近红外光谱技术的稳定性、实用性和准确性不断提高；其分析快速、简便、非破坏性以及可同时测定多成分的优点不断为人们所认识；不仅可用来测定样品的水分、粗蛋白、脂肪、淀粉等常量成分，还被用来测定氨基酸、脂肪酸，以及对生产加工过程进行在线质量控制；分析对象也从粉样样品扩展到分析完整籽粒样品。

另一方面，农业和农产品加工业（如食品、饮料、饲料、烟草等）的发展，人们越来越关注的不是农产品产量的高低，而是质量的好坏。品质育种在发达国家早已十分重视，目前我国也已开始重视。在育种工作中每年对大量组合的分析，在农产品加工中对大批原料的分析及成品、半成品的即时抽检，单靠化学方法是难以完成的，近红外光谱分析是满足这一需要的有力工具。

农作物籽粒中的主要组分是碳水化合物（纤维素、淀粉、糖）、脂类、蛋白质和水分等有机化合物，这些组分中含有 CH、OH、NH、CO 基团。上述官能团振动的谐波及其组合带落在光谱的近红外区（波长 700~2500nm），所以作物样品的近红外光谱包含了各有机组分的大量信息。通过对这些信息分析、处理、计算就可以得出各组分的含量，这就是近红外光谱定量分析的理论基础。同一作物不同品种会显示出不同的光谱特性，通过谱图的比较就可以辨认不同的作物品种，使得近红外光谱分析不仅可以做到定量，又可以做到定性。根据这一原理，在种子市场上可以用于打假，在遗传育种上可以识别不同品种间的亲缘关系。这就进一步拓宽了近红外光谱的应用范围。

现代近红外光谱分析是 20 世纪中叶首先从农业分析开始发展起来的一项高新技术。

早在 20 世纪 60 年代, 美国农业部贝茨维尔农业仪器研究室 K. H. Norris 博士及其同事开展了固体样品近红外漫反射光谱的研究, 并且首先开始了对谷物材料中某些组分(如水分)的定量分析^[1]。继而扩展到了谷物中的其他组分, 如蛋白质、油分等^[2,3]。这一分析方法的出现, 由于其独特的优越性——方便、快速、不损耗样品、多组分同步测定、经济、无污染等, 引起了化学工作者的重视, 也受到了广大农业分析测试人员的欢迎。30 多年来, 随着仪器的不断更新换代, 近红外光谱分析在作物品质方面的应用研究不断深入, 分析精度越来越高, 应用范围越来越宽。十几年前, 在美国就已经把近红外光谱法测定小麦籽粒中的蛋白质含量作为国家认定的标准方法——“Near-Infrared Method for Protein Determination AACC Approved Method 39—10”^[4]。在分析内容上每年都有所增加, 1989 年在日本东京召开的第二届国际近红外光谱会议上收集到的论文中已涉及到了各类农作物和多种组分。涉及到的作物有谷类作物、油料作物、经济作物、豆类和块根类等, 测试的组分有水分、蛋白质、脂肪、灰分、硬度、面筋、氨基酸、戊聚糖 (Pentosan)、可溶性氮、 β -葡聚糖 (Glucan)、淀粉、果糖 (Fructose)、蔗糖、纤维素、木质素、可食纤维、 β -硫代葡聚糖苷、叶绿素、丹宁和类胡萝卜素等^[5]。近红外光谱分析在作物品种资源、遗传育种和农产品加工中显示出越来越大的应用潜力。

14.2 近红外光谱分析在我国作物品质分析中的应用概况

与国际近红外分析发展的历史一样, 我国近红外光谱分析也是首先从农业分析, 特别从作物品质分析开始的。20 世纪 80 年代初, 国内出版的《仪器分析原理以及在农业中应用》一书就以较大的篇幅介绍了近红外分析在农业方面的应用^[6]。

20 世纪 80 年代以来, 我国陆续从国外引进一批近红外光谱分析仪, 这些仪器主要分布在农业分析及检测部门, 用于农作物的品质分析。测试的主要内容是谷物和油料作物中的蛋白质、脂肪和水分。1985 年中国农科院农业测试中心吴秀琴在 51A 型近红外仪上成功地测定了小麦中赖氨酸含量^[7]。漫反射技术在近红外分析中应用广泛, 原北京农业大学从 20 世纪 80 年代开始研究了近红外漫反射光谱分析的基础研究^[8], 包括物理基础、数学基础以及建立定标方程的技术等, 同时还进行了一批应用的研究^[9]。其他一些部门开展了对谷物中的淀粉、高粱中的单宁^[10]、小麦的硬度^[11]等组分的测试实验, 也都不同程度地获得了成功, 渐渐拓宽了仪器的使用面。

我国是个农业大国, 有着丰富的作物品种资源。据统计到目前为止已经收集到的各类作物种质资源已超过 30 万份。为了摸清这些作物种质资源的品质特性, “七五”、“八五”和“九五”期间及正在进行的“十五”计划期间, 由中国农科院作物品种资源研究所主持了国家重点课题“粮食作物种质资源主要品质鉴定”, 曾组织全国 26 个农业科研单位和大专院校联合攻关。近红外光谱分析在该课题中发挥了重要的作用, 提供了大量的有用数据, 并为国家节约了开支。协作单位引进的近红外仪有美国 NEOTEC 公司 (后来被

FOSS/TECATOR 公司兼并) 的 GQA31 型、51A 型、6250 型, 德国 BRAN+LUEBBE 公司的 IA450 型, 瑞典 PERTEN 公司的 Percon 8100 型。这些仪器与化学分析方法相结合测定了小麦、大麦、大豆、水稻、玉米、谷子、高粱和食用豆中的蛋白质、脂肪、淀粉、赖氨酸、水分、单宁、沉淀值和硬度等项目, 获得了二十几万个数据, 基本上摸清了这几种作物的主要品质性状。大量数据作为宝贵资料输入了作物品种资源国家数据库。另外, 每种作物都筛选出了一批品质优异的种质资源, 从 20184 份小麦样品中筛选了蛋白质含量高于 18% 的 1637 份, 赖氨酸含量高于 0.50% 的 1988 份; 从 6036 份大麦样品中筛选出蛋白质含量高于 19% 的 241 份; 从 22250 份大豆样品中筛选出蛋白质含量高于 48% 的 1317 份, 脂肪含量高于 23% 的 25 份; 从 18083 份水稻样品中筛选出蛋白质含量高于 16% 的 6 份, 赖氨酸高于 0.55% 的 20 份; 从 5072 份玉米样品中筛选出蛋白质含量高于 15% 的 47 份, 脂肪含量高于 7% 的 22 份; 从 17337 份谷子样品中筛选出蛋白质含量高于 16%、脂肪含量高于 5%、赖氨酸含量高于 0.3% 的 49 份; 从 3009 份高粱样品中筛选出蛋白质含量高于 15% 的 35 份^[11]。近几年还从食用豆(包括绿豆、菜豆、豌豆、豇豆、蚕豆、木豆、小豆等)中筛选出一批高蛋白、高淀粉的种质材料。这些优异的种质资源已经和正在或将被育种工作者利用, 这必将会尽快改善我国粮食作物的品质状况。

作物品质近红外光谱分析可以用漫反射光谱法, 也可以用透射光谱法。前者应用的谱区主要在近红外光谱中样品的吸收相对较高、透入样品深度较浅的长波近红外区 ($1 \sim 2.5 \mu\text{m}$, $4000 \sim 10000 \text{cm}^{-1}$), 仪器的价格相对较高; 后者应用的谱区主要在近红外光谱中样品的吸收相对较低、透入样品深度较深的短波近红外区 ($0.75 \sim 1.1 \mu\text{m}$, $909 \sim 13333 \text{cm}^{-1}$), 该谱区用的检测器与光导纤维等光学器件价格较低, 导致整个仪器的价格也相对较低。

中国农科院作物品种资源研究所于 1981 年、1987 年分别引进 GQA31EL 型(倾斜滤光片式, 具有 3 个滤光片)和 IA450 型(固定滤光式, 具有 19 个滤光片)近红外光谱仪。十几年来在仪器上建立了小麦^[10]、大麦、大豆^[11](包括栽培大豆和野生大豆)、高粱、玉米、谷子^[15]、水稻和食用豆等作物中, 包括蛋白质、脂肪、水分、淀粉、赖氨酸和单宁等多种组分的近百个定标方程(即数学模型), 在仪器上利用这些方程每年都分析大量样品, 为育种家提供大批优质的材料。多年来用离散波长(滤光片)型近红外光谱仪对各类作物不同组分定标(Calibration)和预测的(Prediction)结果列于表 14-1。由表 14-1 看出, 对于蛋白质、脂肪和水分的测定, 通常只用 3~5 个滤光片就可以得到比较满意的效果。相关系数 R 都在 0.97 以上, 标准差相对较低, 已经接近或达到了化学方法的分析精度。淀粉的结果次之, 赖氨酸、单宁的结果稍差。之所以如此, 我们认为除近红外光谱本身的原因之外, 还可能与下列因素有关: ①组分含量较低, 作物籽粒中的赖氨酸含量通常在 1% 以下, 高粱籽粒中单宁含量在 2% 以下, 在干扰因素较多的情况下难以准确测量; ②测定定标样品的化学方法欠精确, 如测淀粉的旋光法, 测赖氨酸的 DBL 法和测单宁的比色法, 这些问题有待进一步研究解决。随着近红外仪器硬件和软件系统的不断完善, 随着化学方法测定精度的提高, 对一些低含量有机化合物的测定精度将会得到改善。

表 14-1

离散型近红外光谱仪分析各类作物不同组分定标和部分预测结果

组分	样品 种类	定标结果					预测结果		
		K	N_c	含量范围/%	R_c	SEC	N_p	R_p	SEP
蛋 白 质	小麦	3	64	12.14~19.34	0.99	0.25	19	0.98	0.27
	大麦	4	90	7.30~20.91	0.99	0.31	20	0.98	0.33
	水稻	5	82	8.51~18.03	0.99	0.22			
	栽培 大豆	4	51	36.99~48.86	0.99	0.34	20	0.99	0.34
	野生 大豆	4	43	40.50~46.72	0.98	0.37			
	高粱	5	66	9.63~16.62	0.99	0.31	38	0.97	0.50
	玉米	4	40	8.85~16.39	0.98	0.238	22	0.98	0.33
	谷子	6	100	10.14~21.07	0.995	0.245	40	0.994	0.245
	蚕豆	4	66	24.02~31.20	0.95	0.49			
	绿豆	3	47	22.22~26.94	0.97	0.32			
	木豆	3	45	17.67~22.26	0.97	0.25			
	豌豆	3	40	22.21~26.26	0.93	0.32			
	豇豆	3	46	22.29~30.34	0.99	0.23			
	菜豆	3	48	20.76~28.52	0.99	0.33			
淀 粉	大麦	5	66	46.93~65.55	0.99	0.69	40	0.95	1.43
	高粱	5	56	65.57~78.16	0.98	0.51			
	玉米	6	58	50.26~74.50	0.98	0.77	22	0.94	1.25
	绿豆	4	47	46.99~53.73	0.94	0.60			
	木豆	4	44	40.64~46.99	0.92	0.68			
	豇豆	5	47	29.14~46.34	0.95	1.27			
脂 肪	栽培 大豆	3	67	16.63~22.91	0.99	0.28	20	0.99	0.20
	野生 大豆	4	43	8.59~17.45	0.996	0.213			
	谷子	6	100	2.82~5.33	0.974	0.124	40	0.975	0.197
	玉米	3	55	2.90~14.38	0.99	0.21	22	0.97	0.19
水	小麦	4	39		0.998	0.202	16	0.986	0.390
	大麦	3	42		0.997	0.086	15	0.999	0.048
	大豆	3	273	5.90~9.50	0.945	0.13			
分 赖	玉米	3	77	15.28~30.28	0.99	0.32			
	大麦	5	90	0.36~0.54	0.97	0.01			
氨 酸	小麦	4	30		0.95	0.015			
	高粱	5	61	0.29~0.39	0.87	0.021			
	谷子	6	50	0.17~0.31	0.98	0.06			
	玉米	5	58	0.31~0.58	0.94	0.02			
	高粱	6	56	0.014~1.406	0.93	0.17			
单宁	高粱	6	56	0.014~1.406	0.93	0.17			
沉淀 值	小麦	5	44	12.0~67.0	0.97	2.48			
硬度	小麦	5	55	10.4~16.9	0.97	0.32			

其中： K 为定标波长数； N_c 为定标样品数； R_c 为定标相关系数； SEC 为定标标准差； N_p 为预测样品数； R_p 为预测相关系数； SEP 为预测标准差。

下面以连续波长透射或反射近红外法为主，分别介绍几类主要作物的近红外光谱分析。

14.3 小麦、大麦的近红外光谱分析

小麦是作物品质近红外分析中应用最为广泛的几类作物之一。小麦近红外分析被用于小麦作物栽培、育种以及粮食加工、食品行业等许多领域。应用近红外分析技术可分析小麦的蛋白质、脂肪、淀粉、氨基酸、水分、面筋、沉降值、硬度、灰分和色泽等多种指标。

用近红外技术分析小麦品质可是籽粒，也可是粉样。分析的方法可采用漫反射法也可用透射法。由于漫反射法特别是应用积分球可在很大程度上消除装样条件（如样品的高度、厚度、表面状态）对测定的影响，分析效果较好。

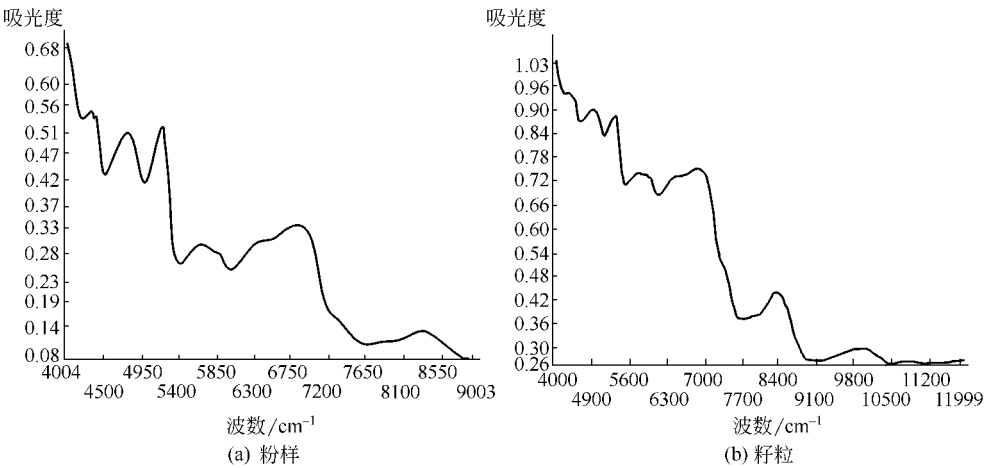


图 14-1 小麦粉样（a）及籽粒（b）的近红外光谱

图 14-1 中的（a）和（b）分别为小麦粉样和籽粒在布鲁克 22/N 傅里叶变换近红外光谱仪扫描的近红外漫反射光谱，图中 4100~4500cm⁻¹ 为 CH 的合频区，4500~4700cm⁻¹ 为 NH 的合频区，4900~5200cm⁻¹ 为 OH 及 H₂O 的合频区。图中 5400~6000cm⁻¹ 主要为 CH 的一级倍频（二倍频即基频的 2 倍），6300~7200cm⁻¹ 主要为 CH 的倍频的合频区，7600cm⁻¹ 以后为二级以上倍频（三倍频以上即基频的 3 倍或更高倍数）。由图可见，谷物近红外光谱的合频和二倍频的吸收较强，信息强度也较高，小麦近红外漫反射分析主要用 4000~8000cm⁻¹ 谱区。表 14-2 和图 14-2 是在傅里叶变换近红外光谱仪上用 PLS 算法分析小麦粉样及籽粒的蛋白质含量的结果。由相关数据可见，粉样样品近红外分析的精度比籽粒高。籽粒样品用近红外分析时需要用旋转的直径较大的样品池。

表 14-2 漫反射光谱分析小麦粉样及籽粒校正报告摘要

成分	状态	样品数	成分含量范围	波长范围/cm ⁻¹	相关系数	SEC	R _{SEC}
粗蛋白	粉样	66	12.14%~19.34%	4000~8000	R=0.992	0.24%	1.53
粗蛋白	籽粒	27	12.82%~19.31%	4500~9800	R=0.935	0.65%	4.19

小麦样品的近红外光谱分析也可以在短波近红外区用透射光谱进行分析，图 14-3 分别是小麦粉样及籽粒的短波近红外光谱图。

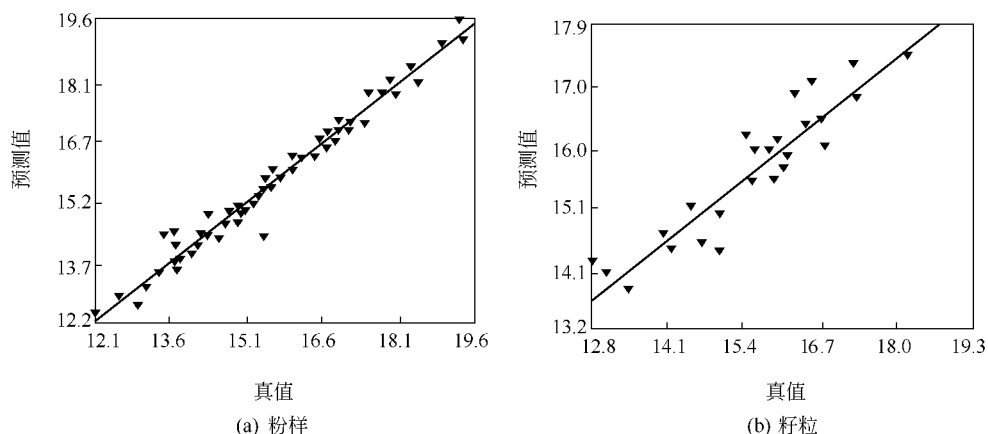


图 14-2 长波近红外漫反射光谱分析小麦粉样 (a) 和籽粒 (b) 样品粗蛋白散点图

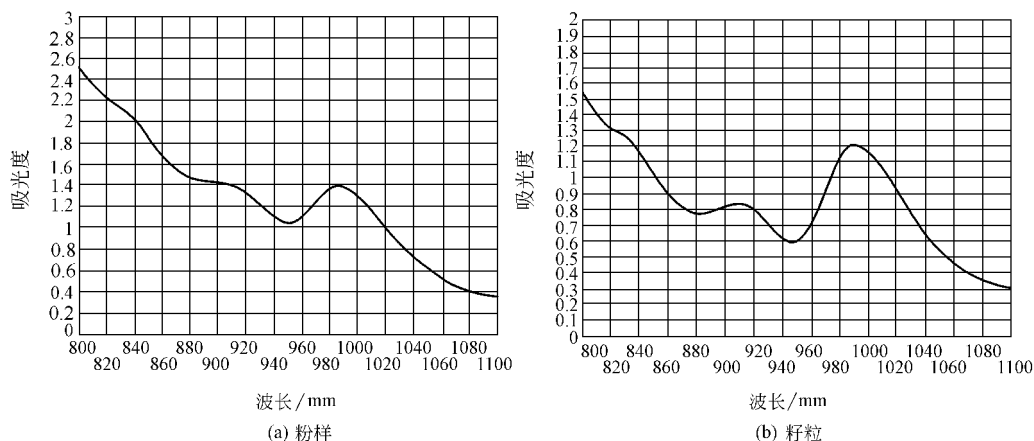


图 14-3 小麦粉样 (a) 和籽粒 (b) 的短波近红外光谱图

由图 14-3 中可见籽粒和粉样的短波近红外透射光谱的吸光度图大体相同。透射光谱的吸光度强度与样品的厚度关系极大，样品的厚度增加一倍，光谱吸光度能量降低大约一个数量级，因此样品的厚度需严格控制。表 14-3 和图 14-4 是短波近红外透射光谱用 PLS 算法分析小麦粉样和籽粒的结果。由数据可见，分析结果大体和近红外漫反射光谱分析的结果相当。

表 14-3 短波近红外透射光谱分析小麦籽粒和粉样校正报告摘要

成分	状态	样品数	成分含量范围	波长范围/cm ⁻¹	相关系数	SEC	R _{SEC}
粗蛋白	粉样	64	12.14%~19.34%	850~1050	R=0.970	0.46%	2.98
粗蛋白	籽粒	27	12.82%~19.31%	850~1050	R=0.9512	0.54%	3.51

除了上述小麦近红外光谱分析的基本研究以外，还有很多其他的研究。Chmann 等 (2002 年) 利用一个共享的欧洲非线性人工神经网络 (ANN) 模型，测定 1998 年份收获的丹麦谷物样品的水分和蛋白质含量，以评价该模型性能。将结果分别与丹麦和欧洲 PLS 模型的测定结果进行了比较，发现 ANN 模型的分析结果准确性高、稳定性好、偏

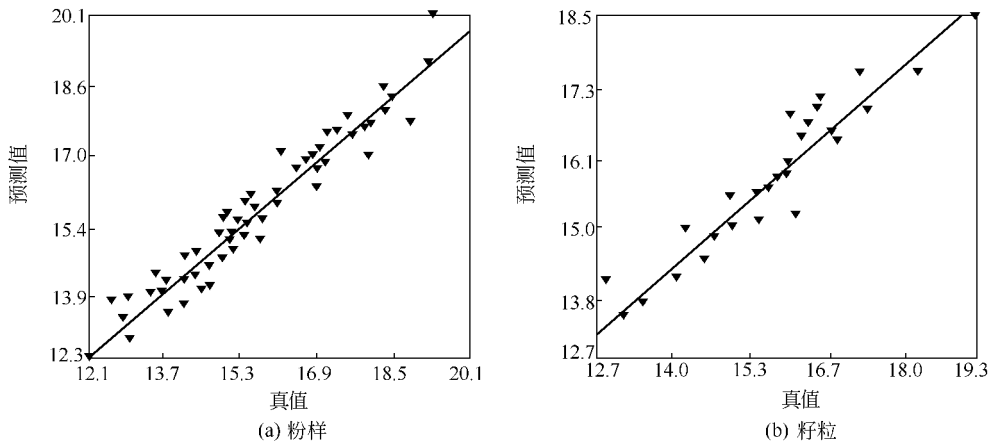


图 14-4 短波近红外分析小麦粉样 (a) 和籽粒 (b) 粗蛋白散点图

差小。

由于存在价格差异，对硬粒红小麦进行冬、春性分类是必要的。但在实际育种中，难以根据籽粒的形态学，对小麦进行冬或春性的直观分类。以 1987~1989 年的 2403 份硬粒红小麦的粉样样品，Chen 等（1995 年）应用近红外反射光谱 BP 神经网络技术研究鉴定硬粒红小麦的冬性或春性的方法。结果表明，利用全波长区（1100~2500nm）所建神经网络模型，其校正、验证集分别具有 97% 和 96.8% 的准确性。当采用窄波长区（2142~2472nm）的光谱数据建立模型时，模型的准确性降低到 97% 和 95.9%。还分别利用全波长和窄波长的校正模型，对 1990 年份样品进行预测，预测准确性分别为 95.1% 和 95.6%。因此，BP 神经网络优于以前结合马氏距离分类法的主成分回归预测模型。作者指出，二阶导数处理与神经网络技术相结合，是一种非常有效的 NIR 鉴定方法。

Delwiche（1993 年 a）利用透射光谱（740~1140nm）、单粒法测定完整小麦粒的硬度。结果表明，5 个波长点的 MLR 和 8 个主成分的 PLS 回归模型，都具有最佳预测效果。还利用所建模型测定两个小麦品种硬度的粒间变异，发现较硬粒小麦和软粒小麦的硬度在粒间存在更大的变异。

Delwiche 等（1993 年 b）利用近红外反射光谱对硬粒红小麦进行分类。结果表明：
①单独用一个参数如蛋白质含量或硬度所建模型的性能最差，错误鉴定率大于 25%；
②联合采用蛋白质含量和硬度 2 个参数建立的模型，其鉴定准确性稍微提高；
③当所分析的未知样品来自不同年份或不同于建模样品时，模型鉴定准确性则降低。

Delwiche 等（1994 年）采用近红外反射光谱技术测定面团形成特性、面包制作等面粉品质。结果表明，由于蛋白质与水分吸收正相关，近红外模型可预测水分吸收。但对和面时间、耐柔性、面包体积、烘烤评分和面包粒外观等 5 个性状的预测准确性差。

Delwiche 等（1995 年 a）应用近红外反射光谱对硬粒红小麦的整粒混合样品进行分类。比较了多元线性回归、主成分分析（含马氏距离）、偏最小二乘和人工神经网络 4 种化学计量学方法的建模效果。结果表明，所有模型的分类准确性都大于 88%。其中，以 ANN 模型的性能最好，具有 95%~98% 的准确度。

Delwiche 等 (1995 年 b) 建立了近红外透射光谱 (NITS) 单粒法测定整粒小麦蛋白质含量的 PLS 模型。外部验证结果表明, $R^2=0.85\sim0.93$, $SEP=0.4\%\sim0.9\%$ 。作者指出, 单粒样品扫描时, 不同的籽粒放置方法影响分析结果的准确性。可通过对单粒样品的多次扫描或从不同方向测定样品光谱、取平均值的方法, 来提高模型的性能。

Delwiche 等 (1996 年) 以可见光和近红外光谱、单粒法对小麦粒色进行鉴定。采用的化学计量学方法为偏最小二乘和多元线性回归法。结果表明, 所建模型的准确性取决于建模时所用的谱区 (可见光+短波 NIR)、样品数目和样品类型。例如, 当模型包括两个不同粒色类型的小麦样品时 (如白粒与红粒小麦), 模型的鉴定准确度达最高 (99%); 但对于具有相同粒色的小麦样品, 如红粒 (如硬红冬 vs 软红冬小麦) 或白粒小麦 (硬粒白小麦 vs 软粒白小麦), 模型的分类准确性降低为 78%~91%。

Delwiche 等 (1998 年) 研究近红外反射光谱测定硬红冬小麦粉样品的蛋白质组成和面团形成特性。验证结果表明, 谷蛋白、醇溶蛋白含量、SDS 沉淀值、和面仪峰值的化学值与 NIR 预测值间的相关系数 (R) 分别为 0.84~0.94。由于含量低, 白蛋白、球蛋白含量不能由近红外反射光谱准确地预测。

Delwiche 和 Williams (2000 年) 利用配备光纤探头的单粒 NIR 反射光谱仪, 测定小麦籽粒混合样品的蛋白质含量。结果表明, 除 SRW (软粒红冬小麦) 外, 由其他 4 种类型小麦样品所建立的模型, 其验证决定系数 (R^2) 均大于 0.90。在分析准确度上, 单粒 NIR 反射光谱仪与传统混合籽粒 NIR 光谱仪相当。

Delwiche 等 (2002 年) 以包括糯粒型、野生型的 200 个小麦品系样品, 采用主成分回归方法, 用近红外光谱对糯、部分糯和野生小麦型进行分类。结果表明, 近红外方法区分糯小麦、非糯品系的准确率接近 100%。但难以在 3 个非糯类型中做进一步分类。

由于红、白粒小麦在磨粉和烘烤品质上存在差异, Dowell (1998 年) 采用可见光/近红外反射光谱对小麦粒色进行单粒分类。结果表明, 与可见光 (450~700 nm) 或 NIR 光谱 (700~1688nm) 比较, 全波长光谱 (450~1688nm) 具有较高的正确分类率 (>99%)。采用可见光/近红外光谱法对小麦粒色进行分类, 可消除感官分类法的主观性。

在小麦成熟、收获季节, 潮湿、凉爽的环境利于霉菌生长, 导致籽粒变灰暗、无活力或泛白甚至产生毒素, 影响食品安全和品质。Dowell 等 (1999 年) 应用近红外光谱分析法, 单粒检测霉菌疮伤、Deoxynivalenol 和 Ergosterol 的含量。作为一种快速检测方法, NIR 技术在监测食物安全和质量方面具有应用前景。

硬粒小麦的透明性影响面粉和蒸煮品质。目前, 视觉法区分籽粒的透明性常带有主观性, 分类结果因不同的检验者存在很大差异。Dowell (2000 年) 以 62 个小麦样品, 用近红外反射光谱法区分透明性或非透明性小麦籽粒。结果表明, 对于用视觉法可明确分类的透明或非透明性籽粒, NIRS 的分类结果与感官法一致。但对于用视觉法难以明确分类的籽粒, NIRS 的区分准确性则较差。

Fontaine 等 (2002 年) 以 213 个小麦粉样品, 采用 MPLS 分析技术, 建立了测定蛋白质和氨基酸含量的 NIRS 校正模型, 并用 98 个样品作外部验证。结果表明, 蛋白质含量的预测决定系数 (R^2) 达 0.98, 预测标准差 (SEP) 为 0.290; 9 种氨基酸的预测决定系数 (R^2) 介于 0.71 (色氨酸) ~0.96 (亮氨酸), 预测标准差 (SEP) 0.009 (蛋氨酸) ~0.033 (精氨酸)。

Fontaine 等 (2002 年) 还用 185 个大麦粉样品, 研究 NIRS 测定蛋白质和氨基酸含量的方法。结果表明, 蛋白质含量的交叉验证决定系数 (R^2) 0.97, 交叉验证标准差 (SECV) 0.240; 9 种氨基酸: (R^2) 介于 0.87 (色氨酸) ~0.96 (缬氨酸), 交叉验证标准差 (SECV) 为 0.006 (色氨酸) ~0.021 (精氨酸)。

McCaig 等 (1992 年) 用近红外光谱测定硬粒小麦的色素含量。采用逐步回归法, 以 1990 年的 53 个小麦样品建立校正模型, 用 1991 年的 28 个样品进行验证。对样品光谱数据进行分析, 发现在 500nm 波长处, 色素具有最大吸收。利用单一波长 (498nm) 和对光谱数据做一阶导数处理, 建立了校正模型, 其验证决定系数 (R^2) 为 0.94, 预测标准差 (SEP) 为 0.77。

在小麦加工成面粉的过程中, 淀粉粒因受压而损伤。Morgan 和 Williams (1995 年) 利用 83 份样品, 根据多元线性回归方法, 建立了淀粉损失的 NIRS 与化学法测定值间的校正模型。用 132 份样品作外部验证, 其预测决定系数 (R_p^2) 达 0.90, 预测标准差 (SEP) 为 0.615。表明, NIRS 是监测淀粉粒损伤的一种快速方法。

Pasikatan 等 (2002 年) 以 35 个具有不同粒度的小麦粉样品, 研究近红外光谱法测定硬粒红冬小麦在磨粉过程中的粒度变化。化学计量学方法采用偏最小二乘法。结果表明, 对于粒度分别 >1041 、 >375 、 >240 和 $>136\mu\text{m}$ 的样品, 以单位面积标准化和一阶导数两种光谱处理方法所建模型的预测效果最好, 其决定系数 (R^2) 为 0.80 ($>136\mu\text{m}$) ~0.93 ($>240\mu\text{m}$), 预测标准差 (SEP) 为 0.43~3.80。因此, NIRS 可作为一种在线粒度感应器 (Sensory), 自动调节滚轴磨的滚轴间距。

在人工干燥或贮藏过程中, 小麦籽粒易发生热损伤 (Heat Damage) 引起蛋白质变性, 降低加工品质。Wang 等 (2001 年) 用近红外光谱、单粒法鉴别热损伤和非损伤籽粒。结果表明: ①光散射和颜色变化是热损伤籽粒光谱特征的主要因素。②以 750~1700nm 波长区、采用偏最小二乘法建立的 NIR 校正模型, 具有最高的分类准确性 (100%)。由可见光波长区 (400~750nm) 建立的模型, 其鉴定准确性最差。③所建立的两波长模型, 其分类准确性可以满足工业和粮食检验的要求。

籽粒的透明性影响蛋白质含量、籽粒硬度、磨粉特性和烘烤特性。因此, 黑色硬质透明 (DHV) 籽粒的多少是硬红春小麦分类的一个重要指标。Wang 等 (2002 年) 采用偏最小二乘回归分析法, 对硬粒红春小麦透明度的可见光/近红外反射光谱 (Vis/NIRS) 测定方法进行了研究。结果表明, 以 Vis/NIR 或 NIR 区单独建立模型, 其分类准确性高于可见光模型。

将 42 个小麦品种 (包括硬红冬、硬红春、软红冬和软白冬小麦), 分别贮藏 in 20%, 40%, 60% 和 80% 的相对湿度下, Windham 等 (1993 年) 研究不同湿度对近红外反射光谱 (NIRS) 测定小麦硬度的影响。结果表明, 在 80% 相对湿度下贮藏的样品, 硬度评分 (Hardness Score, HS) 最高; 20% 湿度下贮藏的样品, 其 HS 低于对照。在每一类型小麦品种内, 硬度评分随湿度的增加而增加。

Wang 等 (1999 年) 应用近红外反射光谱法对小麦粒色进行单粒鉴定。采用偏最小二乘、多元线性回归两种校正技术, 建立 3 个不同波长区 (500~750nm、500~1700nm 和 750~1900 nm) 的分类模型。对光谱数据分别进行一阶导数、二阶导数和光谱不处理。结果表明, 对于 PLS 模型, 500~1700nm 区分类模型, 其分类准确性为 98.5%。其中,

采用一阶导数/光谱数据不处理方法建立的模型，其准确性高于二阶导数模型。对于多元线性回归模型，在可见光/近红外波长区，以光谱数据不处理法所建模型分类准确性可达 98.1%。

郑咏梅等（2002 年）以 160 个小麦粉样品，通过选择不同波长间隔的光谱数据参与回归，剔除无用波长的信息，建立了 4 个波长点（1756nm、1692 nm、1604 nm、1550 nm）的多元线性回归模型，并用 32 和 24 个未知样品对所建模型进行预测，预测标准差为 0.331~0.384。

吉海彦和严衍禄（1998 年）利用国产近红外光谱分析仪实验样机，以 68 个大麦粉样品，建立测定蛋白质、淀粉含量的偏最小二乘校正模型。蛋白质、淀粉含量的预测相关系数分别为 0.961 和 0.949。

高文淑等^[16]（1990 年 a）利用傅里叶变换近红外光谱仪，以 40 个小麦粉样品，通过对样品光谱进行一阶导数处理，建立了 3 个波长点（4256cm⁻¹、5312 cm⁻¹和 4608cm⁻¹）的校正模型。蛋白质含量的预测值与化学值间的相关系数 R 为 0.972。

高文淑等（1990 年 b）以 30 个小麦粉样品，通过一阶导数处理和选择 5 个波长点，建立了测定 SDS 沉淀值的模型。对 30 个样品的预测相关系数（ R_p ）为 0.812，标准差为 0.301。

李大群（1990 年）提出一种对样品光谱进行一阶差分变换的数据处理新方法，对解决在不连续型近红外光谱分析时，存在的样品中其他成分和背景干扰问题具有较好效果。并通过 52 个小麦粉样品，建立测定蛋白质含量的校正模型，其定标与预测相关系数分别为 $R_e=0.957$ 和 $R_p=0.972$ 。定标与预测标准差分别为 $SEC=0.570$ 和 $SEP=0.496$ 。

张录达等（2000 年）以 66 个小麦粉样品，采用主成分—所有可能回归法，对蛋白质含量进行分析。根据一阶导数光谱数据，建立测定蛋白质含量的 NIR 回归模型，预测相关系数与 PLS 相当。

彭玉魁等（1997 年）用近红外光分析技术对 124 个小麦品种的营养成分进行了比较测定。结果表明，水分、粗蛋白、粗纤维、赖氨酸含量的 NIRS 测定值与化学值间的相关系数（ R ）分别为 0.971、0.986、0.955 和 0.985。

景茂等（1991 年）用傅里叶变换近红外漫反射光谱法测定了单粒小麦种子的蛋白质含量，其分析结果与经典的化学值间不存在系统偏差，相关系数（ R ）达 0.92 以上，达到了农业分析的要求。

李庆春等（1992 年）采用逐步回归分析方法，建立测定小麦、大麦粉样品的粗蛋白、赖氨酸、总淀粉含量和小麦的水分及硬度值的 NIR 定标方程（19 波长点）。NIR 法与参照方法所测各组定标和预测样品数值间回归相关系数均大于 0.968，变异系数均小于 3.2%。经 t 测验，NIRDRSA 与参照法之间无显著差异（ $p<0.05$ ）。

小麦是世界上许多国家的主要粮食作物，研究快速品质测定方法与技术，对于粮食检验、贸易、食品加工和制粉业都具有重要意义。在欧美等国，用近红外法测定小麦蛋白质、水分和硬度已是法定方法。自 20 世纪 90 年代以来，近红外光谱分析的应用主要是对籽粒硬度、透明性、粒色和冬春性等感官品质、磨粉品质（在线监控磨粉质量）和烘烤品质进行测定。分析对象已从粉样样品向分析整粒样品扩展。多元校正技术方面，主要是 ANN 等非线性校正技术将逐渐得到应用。随着互联网的应用和普及，在全球或某个国家

内，建立一个测定小麦品质的近红外光谱分析网络，实现资源、仪器共享，将是今后的发展趋势。

14.4 玉米的近红外光谱分析

玉米的近红外光谱分析研究资料丰富，图 14-5 是玉米粉样的近红外漫反射光谱。

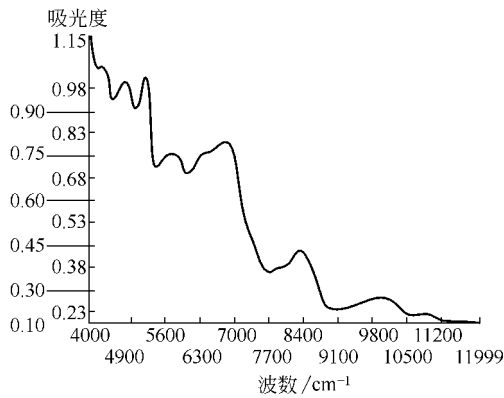


图 14-5 玉米粉样的近红外漫反射光谱

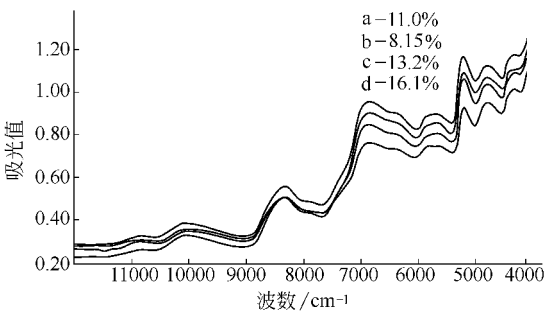


图 14-6 4 个玉米完整籽粒样品的近红外光谱图谱

图 14-6 为 4 个玉米完整籽粒样品的 NIRS 图谱，与粉样光谱大体相同。从图 14-6 中可见，虽然 4 个样品的蛋白质含量不同，但它们的谱图十分相似；在 4200cm^{-1} 、 4600cm^{-1} 、 5200cm^{-1} 、 5500cm^{-1} 、 6900cm^{-1} 、 8300cm^{-1} 附近共出现 6 个较强的吸收峰。在近红外漫反射光谱中，各种基团存在特定吸收峰，而且不同基团产生的光谱在吸收峰位、强度上不同，随着样品组成的变化，其光谱特征也将发生变化。因此，根据测定样品的近红外光谱与其组成或性质数据，通过合理的样品光谱处理及校正方法，可建立样品光谱与组成或性质间的校正模型。

图 14-7 与表 14-4 是用傅里叶变换近红外光谱分析玉米粉样蛋白质含量结果，由分析数据可见，分析结果的精确度大体和小麦结果相当。

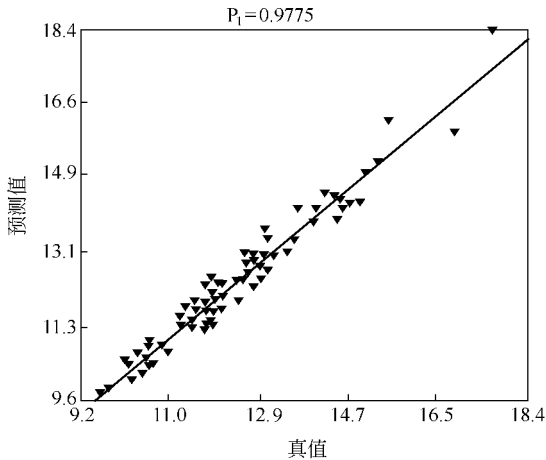


图 14-7 近红外光谱分析玉米粉样散点图

成分	状态	样品数	成分含量范围	波长范围	相关系数	SEC	RSEC
粗蛋白	粉样	80	9.163%~17.62%	4000~8000	$R=0.978$	0.36%	2.92

玉米的近红外光谱分析除了以蛋白质为对象外，其他的资料也很多。

近红外光谱分析是从复杂、重叠、变动的背景中提取弱信息，必须依靠有代表性样品建立的数学模型才能进行分析，代表性样品需大量收集各类样品才能选择确定。

杂交优势利用是作物育种的一种重要手段，在品质育种过程中需要对成千上万的杂交后代样品进行品质分析，近红外分析可以实现快速与无损检测，是对杂交后代品质分析的最佳选择。数学模型的建模样品必须有一定的代表性，由于遗传变异，建立预测杂交后代的数学模型有一定困难。

近红外光谱特征反映了样品的结构和组成的特征，而作物的品质特性决定于籽粒结构和组成成分。由于作物的品质性状是可遗传的，因此籽粒的某些近红外光谱特征也可能具有遗传特征，即玉米杂交后代的近红外光谱特征应具有亲本自交系近红外光谱的主要特征。中国农业大学与国家玉米改良中心合作在主成分空间研究了玉米亲本自交系和杂交后代的光谱特征，证明玉米亲本自交系的光谱大体上能覆盖其杂交后代。

图 14-8 是自交系、杂交种玉米样品光谱的第 1、2 主成分空间分布。

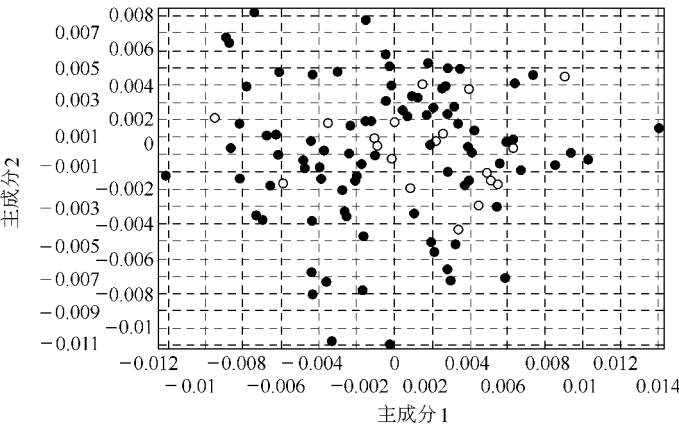


图 14-8 自交系、杂交种玉米样品光谱的第 1、2 主成分空间分布图

图 14-8 表明，亲本自交系样品光谱的主成分特征分布空间（实心点表示）可以完全覆盖杂交后代样品（空心点表示）。说明：自交系的结构和组成特征大体可覆盖其杂交后代。因此，通过选择一定的自交系作为建模的代表性样品，建立近红外光谱定量分析模型，来预测大多数由这些自交系产生的玉米杂交种乃至全国范围的样品在理论上是可行的。

用 89 个自交系玉米建立数学模型，对 40 个杂交后代材料进行了实际预测，结果表明实际值和预测值最大相对误差分别为 2.7%（粗蛋白）、2.46%（粗淀粉）和 7%（油分）。如散点图 14-9 所示，预测值与化学值的相关系数（ R ）达 0.93（粗淀粉）、0.98（粗蛋白）和 0.97（油分）。

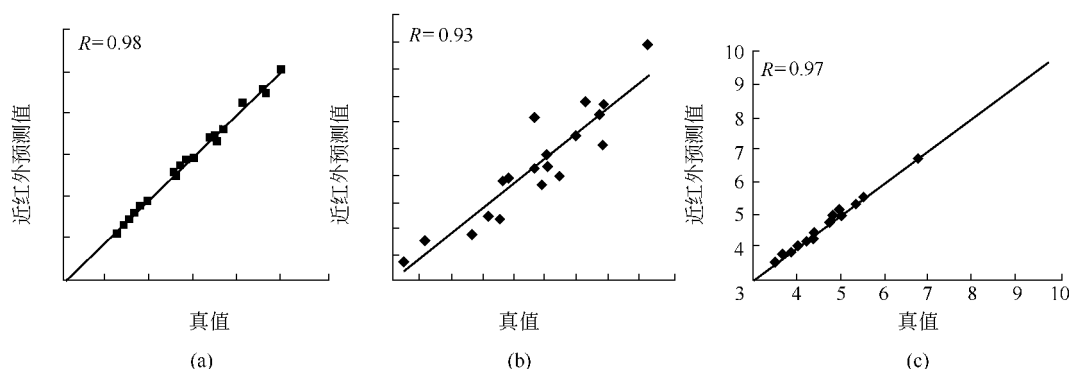


图 14-9 用 89 个自交系玉米建立数学模型预测 40 个玉米杂交后代的蛋白质、淀粉、油分
(a) 蛋白质 (b) 淀粉 (c) 油分

国内外研究玉米品质近红外光谱分析的资料相当丰富。Albanell 等 (1995 年) 应用近红外反射光谱测定玉米半外来群体秸秆的粗蛋白、中性洗涤纤维、酸性洗涤纤维、离体干物质消化率等品质性状。样品包括 1989~1991 年的 3 个季节、3 个地点的 84 份秸秆样品, 采用多元线性回归法。验证结果表明, $R^2=0.88$ (粗蛋白), 中性洗涤纤维 (0.85), 酸性洗涤纤维 (0.90) 和离体干物质消化率 (0.88), $SEP=0.24$ (粗蛋白) —1.88 (离体干物质消化率)。作者指出, 利用 NIRS 预测半外来玉米群体的秸秆品质是可行的。

Campbell 等 (1999 年) 以近红外透射光谱预测完整玉米籽粒中直链淀粉含量 (GAC) 和总淀粉中的直链淀粉含量 (SAC)。样品包括单基因、双基因突变体的一系列胚乳基因型, 淀粉中直链淀粉含量为 8.5%~76%。采用 PLS 和人工神经网络分析技术。对于 SAC、GAC, NIRS 预测值与化学值间的相关系数 (R) 分别为 0.96 和 0.96, 预测标准差 (SEP) 都为 2.7%, 二者精度相同。缩小样品集中直链淀粉含量范围, 一般不能提高模型的性能。用包括所有样品建立的 ANN 模型, 表现最佳性能 ($R=0.96$), PLS 模型次之 ($R=0.94$)。剔除 SAC 含量大于 40% 的样品, 建立 PLS 模型, 具有最低的预测标准差 (SEP)。结果表明, NIRS 可作为一种快速分析方法, 用于在特用玉米的运输、贮藏过程中鉴定混杂籽粒或在育种实际中对大量样品进行快速筛选。

Cozzolino 等 (2000 年) 以在 3 个年份和不同地点种植的 400 个植株样品为材料, 对样品光谱采用一阶导数和散射校正处理, 采用改进的偏最小二乘 (MPLS) 技术, 建立了测定干物质、粗蛋白、酸性洗涤纤维、中性洗涤纤维和离体有机质消化率的校正模型。其交叉验证标准差 (SEC) 介于 7.7 (粗蛋白) ~34.3 (中性洗涤纤维)。用 42 个独立样品对模型进行了验证。结果表明, NIR 与化学值间的决定系数 (R^2) 为 0.61 (干物质) ~0.98 (酸性洗涤纤维), 预测标准差 (SEP) 为 10.4 (粗蛋白) ~42.5 (干物质)。

Dijkhuizen 等 (1998 年) 以 IHP×ILP 杂交后代的 200 个 F_2S_1 家系, 研究近红外反射光谱测定玉米 100g 湿磨分析。结果表明, NIR 淀粉含量与湿磨淀粉值间高度相关 ($R=0.80$), NIR 可用来鉴定与湿磨淀粉组分有关的染色体区域。

Orman 和 Schumann (1991 年) 对 NIRS 测定玉米粉样、整粒样品的蛋白质、淀粉和油分含量效果进行了比较。结果表明, ①由粉样样品的反射光谱建立的模型, 具有最佳的预测结果。其 3 种成分的预测决定系数 (R^2) 都大于 0.70, 预测标准差与参考方法相当。

以整粒样品的漫透射或漫反射光谱所建立的校正方程, 则具有较高的预测标准差 (SEP) 和较低的预测决定系数 (R^2)。②对于 3 种光谱数据, 多元线性回归、偏最小二乘回归法的建模效果相当。

Wehling 等 (1996 年) 利用近红外反射光谱测定整粒黄马齿玉米的干磨品质。采用弦式磨粉脱粒装置 (TADD) 和面粉评价因子 (MEF) 评价干磨品质。利用多元线性回归 (MLR) 与偏最小二乘 (PLS) 分析法, 分别建立 TADD 和 MEF 与 NIR 测定值间的校正模型, 并做外部验证。对于 TADD、MEF 校正模型, 其预测相关系数 (R) 分别为 0.87 和 0.90, 预测标准差 (SEP) 为 1.55%。对 NIR 光谱数据进行导数处理, 可消除样品间因散射引起的基线漂移, 降低预测标准差, 提高预测效果。

Campbel 等 (1999 年) 用近红外透射光谱 (NITS) 整粒测定玉米总淀粉和淀粉中直链淀粉含量。以 136 个包括不同单、双突变体基因型的自交系与杂交种籽粒样品, 采用偏最小二乘回归法建立校正模型并进行外部验证。预测结果表明, GAC 的 NIR 与化学值间决定系数 (R) 为 0.94, 预测标准差 (SEP) 为 3.5%。

Wehling (1993 年) 用近红外光谱测定黄马齿玉米的湿磨淀粉产量。结果表明, 以整粒玉米的二阶导数反射光谱所建多元线性回归方程, 具有最佳的性能, 其相关系数为 0.8~0.9。

Fontaine 等 (2002 年) 以 258 份粉样样品, 采用改进的偏最小二乘法, 建立 NIRS 测定玉米必需氨基酸和蛋白质的校正模型, 并用 78 份样品对模型做外部验证。其中, 蛋白质含量的预测决定系数 (R^2) 0.95, 预测标准差 (SEP) 0.165。9 种氨基酸的预测决定系数 (R^2) 介于 0.55~0.92, 预测标准差 (SEP) 为 0.03~0.045。

Kim 和 Williams (1990 年) 利用 NIRS 对饲用玉米的淀粉、能量进行测定, 并比较了 2 种数学处理方法、3 种近红外谷物分析仪的定标和预测效果。

赵环环等 (1999 年) 采用偏最小二乘法, 建立了 NIRS 测定玉米粉样样品蛋白质含量的校正模型。对 17 个样品的预测结果表明, NIR 与化学值间的相关系数 (R) 为 0.982, 预测标准差 (SEP) 为 0.587%。

高文淑等 (1990 年 a) 以 40 个玉米粉样样品, 采用逐步回归法, 建立 3 个波长点的测定蛋白质含量的预测方程 ($R = 0.964$), 29 个未知样品的预测决定系数 (R^2) 为 0.941。

高文淑等 (1990 年 b) 利用 NIRS 对玉米籽粒中赖氨酸、苏氨酸、蛋氨酸等 9 种氨基酸进行测定, 相关系数多大于 0.85。

段民孝等 (2003 年) 利用 FOSS 公司透射型近红外光谱仪 (自带定量分析模型) 对 425 份自交系、125 份杂交种样品的粗蛋白和粗淀粉进行整粒非破坏性测定。

张晔晖 (1995 年) 以 84 粒样品的含油量 (由核磁共振仪测定) 为参考值, 用光纤探头采集单粒玉米胚部的光谱数据, 采用偏最小二乘, 建立了单粒样品光谱与含油量的数学模型, 并对模型进行交叉验证与外部验证。预测结果表明, NIR 值与参考值间的相关系数 (R) >0.95 , 预测标准差 (SEP) 为 1.22%, 平均相对误差为 12%。

魏良明 (2003 年) 对近红外反射光谱 (NIRS) 测定玉米完整籽粒混合样品的化学成分的方法进行了研究。实验材料包括自“八五”、“九五”以来, 我国玉米生产上广泛利用的 89 份普通玉米自交系、13 份杂交种和部分高油玉米杂交种。采用国家标准方法测定上

述材料的 3 种化学成分。采用偏最小二乘回归分析法, 通过比较不同光谱预处理方法和谱区的校正、外部验证结果, 剔除个别异常样品, 建立了 NIRS 测定粗蛋白、粗淀粉和油分含量的校正模型, 并对模型进行交叉验证与外部验证。模型的外部验证决定系数 (R^2) 介于 0.87 (粗淀粉) $\sim$ 0.96 (粗蛋白), 预测均方差 (RMSEP) 为 0.31 (粗蛋白) $\sim$ 0.88 (粗淀粉)。此外, 还用 40 个玉米杂交后代材料对 3 个化学成分的校正模型的实际预测效果进行检验; 结果表明, NIRS 预测值与化学值间的相关系数 (R) 可达 0.98 (粗蛋白)、0.93 (粗淀粉) 和 0.97 (油分), 最大相对误差为 2.46% (粗淀粉) $\sim$ 7% (油分)。由此说明, NIRS 法准确性与化学方法相当。作者认为, 可利用上述模型对育种材料进行快速的鉴定、筛选。

以往研究结果表明, 可利用近红外光谱技术, 对玉米籽粒的淀粉加工品质 (总淀粉和直链淀粉含量)、氨基酸与脂肪酸含量以及秸秆品质等进行准确的分析。考虑到玉米品质育种发展的实际需要, 研究整粒 (整粒混合样品或单粒)、非破坏性的测定方法将是今后近红外光谱分析在玉米上应用的重点。

14.5 稻米的近红外光谱分析

Barton 等 (2000 年) 比较近红外光谱 (NIR) 测定、拉曼光谱和质子核磁共振测定大米的直链淀粉、蛋白质、透明度、研磨程度、碱扩散值等品质的效果, 认为, 3 种光谱技术在蛋白质含量测定方面都取得满意的效果; 但对直链淀粉含量、碱扩散值等的测定效果较差。

Li 等 (1997 年) 以在不同地点和季节种植的 140 个水稻样品, 采用偏最小二乘回归法, 建立 NIRS 测定整粒稻谷、米粉脂肪酸度的校正模型; 外部验证决定系数 (R^2_{val}) 分别达 0.87 和 0.94; 预测标准差 (SEP) 则为 0.83 和 0.73。

由于 NIRS 与常规的 RVA 参数不相关, Meadows 和 Barton (2002 年) 探索可由 NIRS 测定的新黏度参数, 以替代快速黏度仪 (RVA) 参数。初步研究结果表明, NIR 光谱 (1100 $\sim$ 2500nm) 与 RVA 曲线上 212 $\sim$ 228sec 处的黏度成线性变化。据此建立的偏最小二乘回归模型, 其相关系数 (R) 介于 0.961 (212sec) $\sim$ 0.903 (228sec)。

Chen 等 (1997 年) 以可见光/近红外光谱测定精米表层脂含量。对可见光和近红外谱区的光谱数据进行标准正态变量转换和导数法预处理, 以逐步回归、主成分、偏最小二乘和改进的偏最小二乘等 4 种多元回归技术, 建立了校正模型。其中, 以改进的偏最小二乘方法所建模型的预测准确性最高 ($SEP=0.04$)。

直链淀粉含量是影响大米蒸煮品质和加工特性的重要因素。Delwiche (1995 年) 以美国 5 个水稻主产州的 50 个商品水稻品种或品系为材料, 研究近红外反射光谱测定精米的表观直链淀粉含量 (AAC) 的方法。结果表明, 依据全区光谱数据建立的 PLS 模型, 具有最佳的性能; 外部验证决定系数 (R^2) 为 0.95, 标准差为 1.0% AAC。因此, 近红外法的分析误差与比色法相当。

Meullenet 等 (2002 年) 采用偏最小二乘回归分析法, 通过对光谱数据进行一阶或二阶导数处理, 建立了 NIRS 测定 7 个米饭质地的校正模型。米饭质地由经训练的 9 人评审小组进行描述性评价。结果表明, 所建模型对 4 个性状 (米饭黏唇性、硬度、饭团聚合性

和黏齿性)的预测效果较满意 ($R_{\text{val}}^2=0.79\sim0.88$), 其余 3 个性状的预测效果较差。因此, 可利用 NIRS 预测米饭质地和对高代育种材料进行非破坏性分析。

Villareal 等 (1994 年) 以籽粒大小、粒型、垩白度和透明性差异大的 250 个样品为材料, 研究近红外透射光谱测定大米中直链淀粉含量 (AC) 的方法。直链淀粉含量采用碘比色法进行测定。数学方法采用主成分分析和偏最小二乘回归法。结果表明, 近红外透射光谱模型可用于对育种材料直链淀粉含量进行快速分析。

Barton 等 (1998 年) 以 3 种样品、2 种仪器和 2 种光谱数据 (反射、透射), 评价大米品质和建立测定大米化学成分与物理性状的化学计量学模型。结果表明, 对于直链淀粉、蛋白质和水分含量, 在 $1100\sim2500\text{nm}$ 谱区建模效果较好; 但物理性状 (如白度、透明度、研磨程度) 仅在 $750\sim1050\text{nm}$ 得到较好的预测。作者指出, 分析复杂的品质性状, 需要采用一个或多个预处理方法或谱区范围。

Champagne 等 (2001 年) 以包括短、中和长粒型品种的 87 个样品, 根据偏最小二乘回归法, 建立近红外反射光谱与米饭质地性状 (硬度、粉质性、饭团黏性、光滑性和黏滞性) 间的校正模型。模型的校正决定系数 $R_c^2=0.71\sim0.96$ 。其中, 对饭团黏聚性的预测效果最好 ($R_c^2=0.96$, $R_{\text{val}}^2=0.90$)。

Windham 等 (1997 年) 对 NIRS 分析技术及其与理化测定方法相结合, 测定米饭质地性状的可行性进行了研究。结果表明, 可利用 NIRS, 对感官质地性状进行成功的预测。NIRS 结合理化测定方法不能提高米饭质地性状的预测效果。

Wu 等 (2002 年) 用 NIRS 测定米粉中不同氨基酸和氮含量。结果表明, 对于 14 种氨基酸、氮和总氨基酸含量, 以 3g 样品所建的校正模型, 具有较高的决定系数和较低的标准差。

舒庆尧等 (1999 年 a) 比较 NIRS 测定 0.5g 糙米粉和 3g 精米粉样品的表观直链淀粉含量 (AAC)、糊化温度 (碱消值) 和蛋白质含量的效果。以标准正态变量转换方法矫正光谱在样品间的差异, 采用改进的偏最小二乘 (MPLS) 回归法。结果表明: ①用精米粉、糙米粉样品测定蛋白质含量都有很好效果, 模型的验证决定系数分别为 0.99 和 0.91, 标准误差为 0.20 和 0.21; ②以小样品量糙米粉较精米粉的建立 AAC 回归方程, 模型效果差; ③以小样品量建立的测定蛋白质和 AAC 的校正模型, 可满足实际育种工作的需要。

舒庆尧等 (1999 年 b) 以精米粉为样品, 研究不同光谱处理技术 (降势变换、标准正态变量转换、多元散射矫正、加权散射矫正)、不同数学方法 (一阶导数、二阶导数、三阶导数) 和回归统计方法 (MPLS、PLS、PCR) 对近红外反射光谱测定直链淀粉 (AAC) 的影响。结果表明: ①光谱预处理对校正结果的影响较小; ②回归统计方法对回归方程的影响最大, 以 MPLS 的建模效果最好; ③通过对光谱做散射矫正和一阶导数处理, 采用 MPLS 技术建立测定 AAC 的回归方程, 其验证决定系数 $R=0.94$, $SEP=0.84$ 。

刘建学等 (2001 年 a) 以全国水稻主产区的 97 个粉样为样品, 采用主成分回归分析法, 建立近红外反射光谱测定蛋白质含量的校正模型。模型的验证决定系数 (R^2) 达 0.94, 标准差为 (SEP) 0.43, 相对误差为 2.1%。还比较了不同粒度样品对建模的影响。发现样品粒度越小, 模型的预测效果越好。

刘建学等（2001 年 b）先对定标集光谱数据提取主成分，再用 BP 神经网络方法，建立测定大米直链淀粉含量的 NIRS 预测模型。结果表明：①样品粒度对神经网络模型的预测效果无显著影响；②精米直链淀粉含量的 NIR 预测值与化学值间决定系数 $R^2=0.90$ ，预测标准差 $<1\%$ ，平均相对误差为 2% 。

稻米是亚洲一些国家的主要粮食作物。从以上研究结果看，今后，近红外光谱技术在稻米品质测定方面的应用，将主要集中在蒸煮或食用品质（米饭质地、淀粉的黏度、直链淀粉含量）、感官品质（透明性、研磨程度等）以及蛋白质和氨基酸含量等营养成分方面。

14.6 油料作物的近红外光谱分析

油料作物中大豆的近红外光谱分析研究较多。图 14-10 是大豆粉样的短波近红外透射光谱。

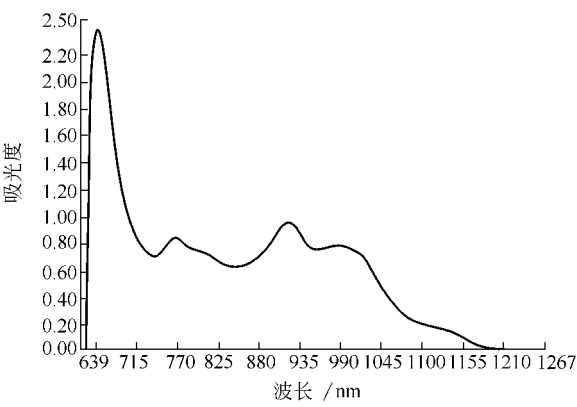


图 14-10 大豆粉样的短波近红外透射光谱

大豆近红外光谱分析的主要对象是蛋白质、脂肪和水分。表 14-5 是短波近红外分析上述三种成分的结果，图 14-11 是其散点图。

表 14-5 短波近红外透射光谱分析大豆粉样校正报告摘要							
成分	主成分数	样品数	成分含量范围	波长范围/nm	相关系数	SEC	RSEC
粗蛋白	8	131	35.6%~48.18%	850~1050	$R=0.973$	0.52%	1.28
粗脂肪	8	101	16.23%~23.48%	850~1050	$R=0.969$	0.37%	1.88
水分	8	108	6.4%~9.57%	850~1100	$R=0.985$	0.21%	2.71

除了近红外光谱分析大豆的上述的三组成分以外，近红外光谱分析其他油料作物中的国内外文献也相当丰富。

Velasco 等（1999 年 a）对近红外反射光谱（NIRS）测定单粒完整油菜籽的粒重、含油量及含油量的脂肪酸组成的方法进行了研究。采用改进的偏最小二乘法，外部验证结果表明，对于油酸、软脂酸含量，其 NIRS 测定值与化学值间的相关系数分别为 0.92 和 0.94；亚油酸、硬脂酸含量的 NIRS 测定值与化学值间的相关系数分别为 0.75 和 0.73。

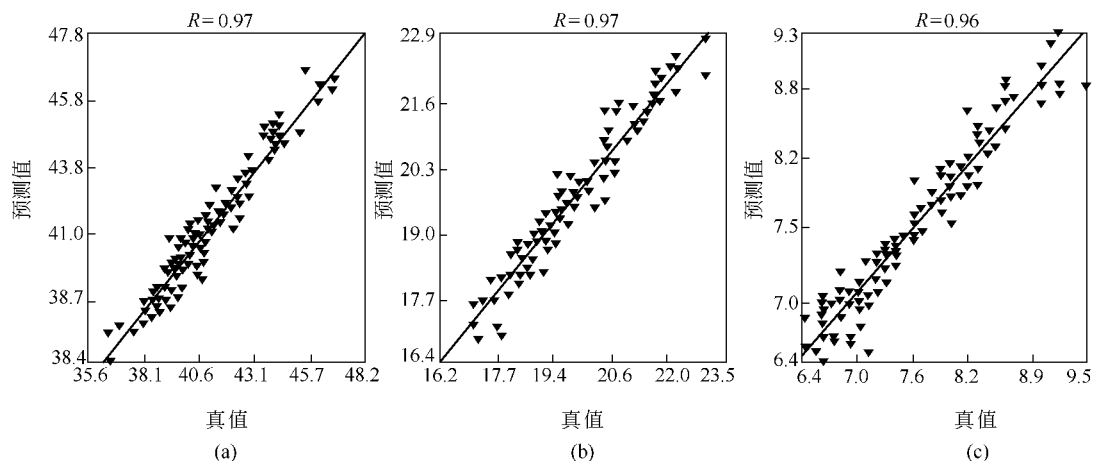


图 14-11 短波近红外透射光谱分析大豆的蛋白质 (a)、脂肪 (b)、水分 (c) 散点图

对于粒重和含油量，外部验证相关系数都为 0.92。根据对 2 个 F_2 分离群体的油酸、软脂酸含量的分析结果，作者指出可利用近红外光谱技术，对大批育种材料进行快速的筛选、鉴定。

Velasco 等 (1998 年 a) 以油菜的完整油菜籽样品，对 Sinapic Acid Esters (SAE) 含量进行 NIRS 非破坏性测定。采用偏最小二乘法，对样品光谱数据进行二阶导数、散射校正法预处理，建立了校正方程。其交叉验证决定系数 (R^2) 为 0.91，标准差为 1.1。后又添加 187 个低 SAE 含量样品重新建立模型。其外部验证决定系数 (R^2) 为 0.92；还利用 57 个独立样品，对模型进行再验证；预测决定系数 $SEP=1.1$ 。上述结果表明，所建校正方程是稳健、可靠的，可用于预测油菜籽的 SAE 含量。

Velasco 等 (1998 年 b) 以 3 种不同样品量的样品和 NIRS 测定整粒油菜籽含油量的脂肪酸组成。化学计量学方法为改进的偏最小二乘法。用气液色谱法 (GLC) 测定脂肪酸组成。测定油酸、亚油酸、亚麻酸含量的校正方程，其交叉验证决定系数 (R^2) 介于 0.95~0.98 (3g 样品)、0.95~0.97 (300mg 样品) 和 0.84~0.96 (60mg 样品)。由于棕榈酸、硬脂酸含量的变幅范围较窄，NIRS 测定这 2 种脂肪酸的准确性较低，其交叉验证决定系数 (R^2) 通常低于 0.8。作者认为，NIR 作为一种高效分析技术，可用来预测含油量的脂肪酸组成。

Velasco 等 (1998 年 c) 以籽粒密度为 262~676kg/m³ 的一组样品，对埃塞俄比亚芥菜籽粒容重的 NIRS 测定方法进行研究。结果表明，采用光谱数据不处理和改进的偏最小二乘回归法，建立的校正模型具有最佳性能。其外部验证决定系数 (R^2) 为 0.92。随后，增加 56 个样品对上述模型进行扩展，对所建模型进行外部验证，预测决定系数 (R^2) 为 0.92。还利用 19 个波长的普通滤光片式仪器上，建立了籽粒容重的多元线性回归模型。

Velasco 等 (1998 年 d) 利用 NIR 法，对 1487 个油菜种质资源样品和 1361 份育种材料进行非破坏性的评价。还对具有最低 Sinapic Acid Esters (SAE) 含量的 112 份种质资源和 75 份育种材料进行化学法测定，以验证 NIRS 分析法的可靠性。

Velasco 等 (1999 年 b) 以 NIRS 单粒法非破坏性的分析向日葵瘦果中油酸、亚油酸含量。用二阶导数、多元变量变换和去势 (Detrending) 散射矫正等方法, 对 465 个样品光谱数据 (1100~2500nm) 进行处理, 建立基于 MPLS 的校正模型。外部验证结果表明, 两种脂肪酸含量的 NIRS 测定值与 GLC 测定值之间的决定系数 (R^2) 为 0.88。以 GLC 法作对照, 还测定了 2 个 F_2 分离群体的油酸含量。

Velasco 等 (1996 年) 以 106 份样品、采用 MPLS 和降势散射矫正处理方法, 建立分析埃塞俄比亚芥菜的软脂酸含量校正模型。外部验证决定系数 (R^2) 为 0.96。用所建模型对 1334 份芥菜资源样品进行测定。结果表明, 在 90% 的样品中, 其 NIRS 测定值与 GLC 法测定值相近或相同。

Velasco 等 (2002 年) 利用 NIRS 非破坏性地测定单粒油菜籽的蛋白质含量。以 451 粒种子建立校正模型, 用 117 粒种子进行外部验证。其预测决定系数 (R_p^2) 为 0.94, 预测标准差 (SEP) 为 0.77%。作者还利用 NIRS 方程, 分析了 3 个品种的蛋白质含量在株内变异情况。

Pazdernik 等 (1997 年) 比较了用 NIRS 测定粉样和整粒大豆样品中氨基酸、脂肪酸组成的效果。化学计量学方法为 MPLS。采用二阶导数、散射矫正、标准正态变量变换和降势散射矫正对光谱数据进行处理。粉样样品、完整籽粒的预测决定系数 (R_p^2) 分别为 0.38~0.85 和 0.06~0.83; 偏差分别为 -1.7%~3.2% 和 -3.1%~4.2%。表明 NIRS 测定粉样样品的准确度较高。

Misra 等 (2000 年) 用常规的索氏抽提法测定了 64 个花生仁的含油量 (变幅为 40.3%~53.6%), 采用近红外透射光谱 (NITS) 对含油量做多元线性回归分析。预测值与化学值之间的预测标准差、决定系数分别为 5.82 和 0.865。对于含油量 48%~51% 的样品, 索氏法和 NIT 法测定值之间的误差小于 2%; 对于含油量低于 48% 或高于 51% 的样品, 两种方法之间的误差则大于 2%。结果表明, 作为一种非破坏性的分析方法, 近红外透射光谱 (NITS) 可用于测定花生含油量。

Letzetter 和 Wilson (1995 年) 以偏最小二乘 (PLS) 分析技术和傅里叶变换近红外光声光谱 (PAS), 单粒测定豌豆的淀粉、蛋白质、脂类。结果表明, NIR 法的预测值与化学法相近。

吴建国等 (2002 年) 以近红外反射光谱 (NIRS) 整粒测定 258 份油菜籽的 3 种不同样品量 (8g、3g 和 0.6~0.8g) 的含油量。结果表明, 光谱预处理方法和回归统计方法对建立校正方程的影响较大。其中, 以三阶导数、标准正态变量变换、标准正态变量变换+趋势变换、标准多元散射矫正和降势多元散射矫正法处理的效果较好。3 种回归方法中, 以改进的偏最小二乘 (MPLS) 回归法最好。研究 3 种不同样品量对校正、预测效果的影响, 其中以 8g 样品的分析效果最好, 3g 样品次之, 0.6~0.8g 样品最差。根据两种小样品量建立的回归方程, 可直接用于品质育种中筛选育种材料。

王文真, 付翠真 (1994 年) 利用固定滤光片式近红外反射光谱仪测定了大豆籽粒中的蛋白质、脂肪和部分氨基酸含量。对于蛋白质、脂肪含量, 只需用 3~4 个滤光片, 即可获得定标、预测相关系数均高于 0.95 的分析精度。测定天门冬氨酸、谷氨酸、赖氨酸和精氨酸含量, 则需要 5~7 个滤光片, 其定标、预测相关系数在 0.86 以上。

张录达等 (2002 年) 用 40 个大豆样品的近红外光谱数据, 建立测定大豆蛋白质含量

的特征根回归模型。对 32 个样品的预测结果与 PLS 模型的效果相当。

Fontaine 等 (2001 年) 用近红外光谱法对大豆粉、油菜粉的粗蛋白和必需氨基酸含量进行分析。对于大豆粉样品, 粗蛋白: 交叉验证决定系数 (R_{cv}) 为 0.99, 标准差 (SECV) 为 0.545; 各种氨基酸: 交叉验证决定系数 (R_{cv}) 介于 0.81 (胱氨酸) ~0.98 (亮氨酸); 标准差 (SECV) 为 0.015 (色氨酸) ~0.092 (精氨酸)。对于油菜粉样品, 粗蛋白的交叉验证决定系数 (R_{cv}) 为 0.93; 各种氨基酸的交叉验证决定系数 (R_{cv}) 介于 0.65 (苏氨酸) ~0.90 (亮氨酸)。

综上所述, 分析大豆、油菜等油料作物中氨基酸、脂肪酸组成已成为近红外光谱分析的主要方面。考虑到油料作物育种的实际需要, 研究整粒、非破坏性的测定方法将是今后近红外光谱分析的主要方向。

参 考 文 献

- [1] Albanell E, Plaixats J and Ferret A. Evaluation of near-infrared reflectance spectroscopy for predicting stover quality trait in semi-exotic populations of maize. *Sci Food Agric*, 1995 (69): 269~273
- [2] Barton F E, Himmelsbach D S, McClung A M, et al. Rice quality by spectroscopy analysis: precision of three spectral regions. *Cereal Chem*, 2000, 77 (5): 669~672
- [3] Barton F E, Windham E T, Champagne E T, et al. Optimal geometrics for the development of rice quality spectroscopic chemometric models. *Cereal Chem*, 1998, 75 (3): 315~319
- [4] Büchmann N B, Josefsson H and Cowe I A. Performance of European artificial neural network (ANN) calibrations for moisture and protein in cereals using the Danish near-infrared transmission (NIT) network. *Cereal Chem*, 2001, 78 (5): 572~577
- [5] Campbell M R, Mannis S R, Port H A, et al. Prediction of starch amylose content versus total grain amylose content in corn by near-infrared transmittance spectroscopy. *Cereal Chem*, 1999, 76 (4): 552~557
- [6] Campbell M, Brumm T J, and Glover DV. Whole grain amylose analysis in maize using near-infrared transmittance spectroscopy. *Cereal Chemistry*, 1997, 74 (3): 300~303
- [7] Champagne E T, Bett-Garber K L, Grimm C C, et al. near-infrared reflectance analysis for prediction of cooked rice texture. *Cereal Chem*, 2001, 78 (3): 358~362
- [8] Chen H, Marks B P, and Siebenmorgen T S. Quantifying surface lipid content of milled rice via visible / near-infrared spectroscopy. *Cereal Chem*, 1997, 74 (6): 826~831
- [9] Chen Y R, Delwiche S R and Hruschka W R. Classification of hard red wheat by feedforward back-propagation neural networks. *Cereal Chem*, 1995, 72 (3): 317~319
- [10] Cozzolino D C, Fassio and Gimenez A. The use of near-infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants. *J Sci Food Agric*, 2000, (81): 142~146
- [11] Czuchajowska Z and Pomeranz Y. changes in maturing wheat as determined by near-infrared reflectance spectroscopy. *Cereal Chem*, 1989, 66 (50): 432~435
- [12] Delwiche S R and David R M. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chem*, 1996, 73 (3): 399~405
- [13] Delwiche S R and Norris K H. Classification of hard wheat by near-infrared diffuse spectroscopy. *Cereal Chem*, 1993, 70 (1): 29~35
- [14] Delwiche S R and Weaver G. Bread quality of wheat flour by near-infrared spectrophotometry: feasibility of modeling. *J of Food Science*, 1994, 59 (2): 410~415

- [15] Delwiche S R and Williams R H. Protein content of bulk whet from near-infrared reflectance. *Cereal Chem*, 2000, 77 (1): 86~89
- [16] 段民孝, 郭景伦, 王元东等. 利用近红外透射分析仪快速测定玉米籽粒品质的初步研究. *华北农学报*, 2003, 18 (1): 37~40
- [17] 高文淑, 景茂, 严衍禄. 傅里叶变换近红外漫反射光谱法测定谷子、玉米中多种氨基酸含量. *北京农业大学学报*, 1990, 增刊: 88~93
- [18] 高文淑, 张录达, 王万军. 应用傅里叶变换近红外反射光谱法测定几种谷物籽粒中蛋白质含量. *北京农业大学学报*. 1990, 16 (增刊): 72~79
- [19] 高文淑, 张录达, 王万军. 应用傅里叶变换近红外漫反射光谱法测定谷子中的粗脂肪含量. *北京农业大学学报*, 1990, 增刊: 94~97
- [20] 吉海彦, 严衍禄. 在国产近红外光谱仪实验样机上用偏最小二乘法定量分析大麦成分. *分析化学*, 1998 (5): 607~611
- [21] 景茂, 严衍禄, 刘广田. 傅里叶变换近红外漫反射光谱法测定小麦单籽粒中蛋白质含量. *光谱学与光谱分析*, 1991, (3): 20~22
- [22] 李大群. 近红外漫反射光谱分析中新的数据处理方法. *分析仪器*, 1990 (3): 39~42
- [23] 李大群等. 近红外反射光谱法测定大豆和小麦中的蛋白质. *分析仪器*, 1989 (4): 45~50
- [24] 李庆春, 王文真, 张玉良. 近红外漫反射光谱分析法 (NIRDRSA) 在作物品质育种中的应用. *作物学报*, 1992, 18 (3): 235~240
- [25] 李庆春等. 近红外漫反射光谱分析法 (NIRDRSA) 同时测定小米中的粗蛋白质、粗脂肪及赖氨酸含量. *卫生研究*, 1991 (1): 34~38
- [26] 刘建学, 吴守一, 方如明. 基于近红外光谱的神经网络预测大米直链淀粉含量. *农业机械学报*, 2001 (2): 55~57
- [27] 彭玉魁, 李菊英, 祁振秀. 近红外光谱分析技术在小麦营养成分鉴定上的应用. *麦类作物*, 1997, 17 (2): 33~35
- [28] 舒庆尧, 吴殿星, 夏英武等. 稻米表观直链淀粉含量近红外光谱测定技术校正设置的优化. *浙江大学学报 (农业与生命科学版)*, 1999, 25 (4): 343~346
- [29] 舒庆尧, 吴殿星, 夏英武等. 用近红外反射光谱技术测定精米粉样品表观直链淀粉含量的研究. *中国水稻科学*, 1999, 13 (3): 189~192
- [30] 舒庆尧, 吴殿星, 夏英武等. 用近红外反射光谱测定小样本糙米粉的品质性状. *中国农业科学*, 1999, 32 (4): 92~97
- [31] 王文真, 付翠真. 利用近红外反射光谱快速测定大豆籽粒蛋白质、脂肪和部分氨基酸含量. *作物品种资源*, 1994, (1): 31~32

15. 近红外技术在面粉加工业中的应用

近年来国内外近红外技术除了在农业部门得到应用之外，在粮油加工行业也有了长足的进展，其中在面粉加工行业应用最为广泛。

在面粉加工行业中，近红外技术有三个方面的应用：①原料小麦质量的检测；②制粉过程的质量控制，包括离线检测和在线检测两种；③最终加工产品面粉的质量检测和监控。近红外技术在面粉加工行业最重要的应用是对小麦原料和面粉成品的检测。目前面粉加工业测定项目有水分、蛋白、灰分、白度、硬度、面筋、颗粒度、吸水率、损伤淀粉、沉淀值、快速混合实验（RMT）等。其中前6项是制粉行业中质量控制的关键指标。

15.1 原料小麦质量检测

目前我国市场上的各种专用面粉，对原料小麦品质有不同的要求。面包专用粉要求高蛋白含量和强筋力蛋白质的硬质小麦，而饼干专用粉要求低蛋白质含量和弱筋力蛋白质的软质小麦。在面粉厂生产过程中，为了保证面粉产品的质量，还要对各种质量的小麦进行科学、合理的搭配，再进行研磨加工，由此可见小麦质量在面粉生产过程中的重要性。

Law（1977年）、Williams（1985年）^[1~3]等用近红外技术测定了小麦蛋白质、水分等含量。刘广田（1990年）^[4]用 Perten 8620 近红外仪测定小麦蛋白、SDS 沉淀值、面粉蛋白、Zeleny 沉淀值、湿面筋、干面筋含量，近红外测定值与常规测定值的相关系数分别为 0.981、0.870、0.993、0.940、0.939 和 0.953，标准差分别为 0.38、0.512、0.218、0.308、1.724 和 0.393，结果表明可以满足小麦籽粒品质改良及种质资源特别是小麦杂交种早代品质测定的需求。彭玉魁等（1997年）^[5,6]用 Neotec 51A 近红外测定仪对 124 个小麦品种的成分进行了比较测定，结果表明：近红外仪小麦水分、粗蛋白、粗纤维、赖氨酸测定值与常规化学分析法测定值之间的相关系数分别为 0.9714、0.9862、0.9548 和 0.9847，估测误差分别为 0.41%、0.32%、0.84% 和 0.22%。说明两种方法的测定值之间具有显著的相关关系。近红外法具有快速、简便的优点。研究证明采用近红外技术测定我国小麦成分含量已达到实用程度。陈锋、何中虎（2003年）^[7]利用近红外透射光谱技术测定 426 个小麦品种的水分、蛋白、硬度和面粉的干、湿面筋、灰分、SDS 沉淀值、Zeleny 沉淀值，结果表明近红外技术用于测试小麦品质是可行的，能够用于育种的早代选择。

经过近年来，国内外近红外技术的发展和实际应用的积累，近红外技术对小麦测定项目不断增多，目前，被面粉加工业重视的测定项目有小麦水分、蛋白、硬度、颗粒度、面筋、沉淀值、快速混合实验（RMT）、水吸收率（Water Absorption）等。

其中水分和蛋白质含量的测定最为成熟和准确，大量资料报道近红外法测定结果与常规基准法之间误差均为 0.2%。近红外蛋白质测定方法已列入 ICC（国际谷物协会）标准 No. 159。美国在小麦定购合同中已规定采用近红外法测定小麦的蛋白质含量。世界一些小麦出口国美国、澳大利亚、加拿大也把小麦蛋白质含量作为合同指标和定价标准，例如美国出口的硬红春麦规定蛋白质最低含量不少于 13.5%，从蛋白质含量 13.5%起，每增

加 0.10% 蛋白质含量，其离港价可提高 0.5%，加价范围到蛋白质含量 14% 为止。这样近红外测定结果已作为小麦贸易标准和商品价格的依据。

使用近红外分析仪用来接收小麦，不仅可迅速得知此批小麦的质量，保证按质存放、分类贮存、合理利用，而且可连续监测整批小麦接收过程的质量变化，及时发现质量上的波动和差异，为生产提供确切的质量数据。小麦接收后，贮存期间由于受外界湿度、温度的影响，其品质会发生劣变。通过近红外仪定期监测小麦品质，可简单方便、直观迅速地了解小麦的质量变化情况。

小麦硬度反映了小麦胚乳内部组织结构的紧密程度和蛋白的含量，与小麦的用途密切相关。小麦分为硬麦和软麦两类，根据种植季节、表皮颜色分为春麦和冬麦、红麦和白麦，所以有硬红冬、硬红春、软白春等类别之分。硬度的常用测定方法是目测法，用肉眼观察小麦胚乳的透明度（玻璃质）。硬麦胚乳透明、蛋白含量高；软麦胚乳不透明、蛋白含量低。由于是目视，判断易存在人为误差。小麦硬度另一种测定方法是 PSI 法（Particle Size Index），它利用软、硬麦磨粉特性的不同，硬麦在研磨过后得到粗的粉粒多，软麦得到粗粉粒少，小麦磨粉后用一定孔径的筛来筛理，硬麦筛上物比例高、软麦筛上物比例低。采用近红外技术可以测定小麦的硬度，根据波通仪器公司提供的资料表明：近红外硬度值与常规 PSI 基准法硬度值的相关系数为 0.933，测定范围为 35~65，标准误差为 3，近红外硬度测定值越大表示小麦硬度越高、近红外硬度测定值越小表示小麦硬度越低。该测量精度已满足了加工行业评定小麦硬度的要求，一些面粉厂已应用于小麦原料的分类、选择和评定，取得明显的效果。

周艳华、何中虎（2002 年）^[8]实验得出：近红外仪（NIR）硬度值与单粒谷物特性测定仪（SKCS）的单粒小麦压碎硬度之间存在较高的相关性， $R=0.85\sim0.94$ 。任红波（2003 年）^[9]使用 8620 型近红外仪测定小麦蛋白及硬度，实验验证：近红外仪测定小麦蛋白与硬度具有很好的精密度和准确度，可以替代常规实验室方法，且具有测定速度快、效率高的优点，能够很好地为农业和粮食部门服务。

近红外技术近年来已发展到一些小麦“内在”质量和功能特性的测量，沉淀值（Sedimentation）是测定小麦蛋白质“内在”质量的测定项目，快速混合实验（Rapid Mix Test）是测定小麦烘焙体积的测定项目。

沉淀值是用乳酸-异丙醇溶液与小麦粉混合振荡成悬浮液，蛋白质在乳酸作用下发生溶胀现象，使悬浮液稠度增加，面粉下沉速度变慢，静置 5min 后观察沉降高度来表示沉淀值。蛋白质含量高、筋力强的小麦沉降值高；相反的，蛋白质含量低、筋力弱的小麦沉降值低。中国农业大学刘广田教授（1990 年）^[4]测定了小麦样品，近红外测定值与常规 Zeleny 沉淀值测定值的相关系数为 0.940，标准差为 3.0，测量范围为 30~73。波通仪器公司报道：测定 300 份小麦沉淀值与 Zenely 沉淀值相关系数为 0.857，标准差为 6，测量范围为 11~66。

快速混合实验（RMT）是德国评价小麦烘焙质量的方法，采用面团快速混合方法，经烘烤后测量其面包的体积，根据波通仪器公司报道，150 份小麦样品快速混合实验的近红外值与常规实验室值的相关系数为 0.948，标准差为 22，体积测量范围为 570~789mL，利用近红外技术可以直接快速地由小麦的预测出面粉烘烤成面包的体积，以此方法来评价小麦的烘焙质量。

测定小麦有两类仪器，一类要把小麦粉碎成粉状样品才能检测，如瑞典波通 Perten 86 系列近红外分析仪、帝强 Dickey-john Instalab 600 系列近红外分析仪和美国太平洋公

司 6250 近红外仪；另一类可以直接检测整粒小麦样品，如瑞典波通 Perten 9100 型整粒谷物近红外分析仪、Perten DA 7200 固定光栅连续光谱近红外分析仪、福斯 Foos Tecator 1241 近红外整粒分析仪和美国 Zelxt 公司 ZX-800 近红外仪等。

15.1.1 粉样近红外检测仪

早在 1995 年，刘凤华、刘志华^[10]使用美国太平洋公司 6250 近红外仪测定进口小麦的蛋白、水分，获得良好的结果。

Perten 86 系列近红外仪器是目前我国面粉行业广泛应用的一类近红外仪器。



图 15-1 Perten 86 系列近红外分析仪

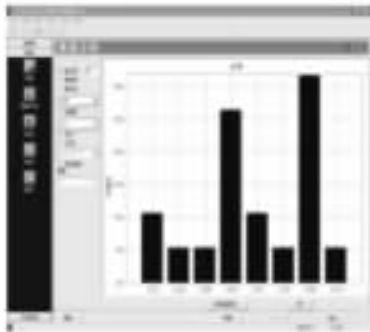


图 15-2 PICS 中文近红外分析软件

Perten 86 系列近红外分析仪采用滤光片技术、快速样品槽设计和覆膜防尘键盘（图 15-1），可常年不关机和在粉尘条件下工作。配套中文计算机软件（图 15-2），内存容量大。测定项目、测量范围和测量精度见表 15-1。

表 15-1 小麦测定项目、测量范围与测量精度

项目	测量范围	相关系数	标准差	基准方法
水分/%	12.8~21.6	0.983	0.2	烘箱法
蛋白/%	11.2~15.4	0.989	0.18	凯氏法
硬度/%	35~65	0.933	3	德国方法
沉淀值/cm	17~66	0.857	6	Zeleny
快速混合实验/cm ³	570~789	0.948	22	德国方法
水吸收率/%	49~60	0.919	0.85	Brarender

Perten 86 系列近红外仪配有 3100 和 120 实验粉碎磨，用于小麦样品的粉碎，3100 和 120 磨为高速旋风磨，转速为 16800r/min，粉碎细度为 80~100 目，30~50s 粉碎 300g 样品。

15.1.2 整粒小麦近红外检测仪

Perten 9100 型近红外谷物分析仪是典型的为整粒谷物样品分析设计的滤光片型近红外光谱分析仪（图 15-3），为了克服小麦颗粒间的差异，可以自动把 120g 样品分为 120 次通过光学测量窗口，采用高速测量技术，1 分钟后就显示测定结果。由于不必进行样品粉碎，加快了测量速度，降低了劳动强度，特别适合于小麦的收购部门。

Perten 9100 型整粒谷物近红外分析仪符合美国谷物化学协会 AACC 标准 39—25。
Perten 9100 具有联网功能，可以实现远程联网控制、检测分析和故障诊断（图 15-4）。



图 15-3 Perten 9100 型整粒谷物近红外分析仪

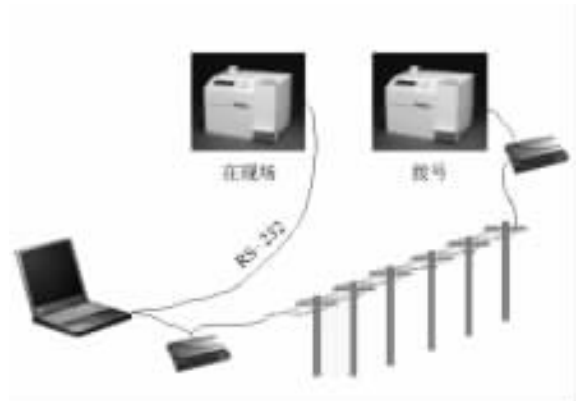


图 15-4 近红外远程网络示意图

Perten 9100 型和 Perten 86 系列近红外仪是获得中国国家质量监督检验检疫总局型式批准证书的多成分近红外仪器。

西北农林大学胡新中等（2002 年）^[11]利用 Perten 9100 近红外仪对陕西省 624 户农民 2000 年夏季上缴粮库的公购小麦进行现场品质分析，检测指标有水分含量、籽粒硬度、蛋白质含量、泽伦尼沉淀值、烘烤体积。结果表明，Perten 9100 完全能够满足小麦检验的需要（表 15-2）。

表 15-2 小麦商品粮品质调查结果

项目	含水量/%	硬度	蛋白质/%	沉淀值/mL	烘烤体积/cm ³	小麦等级
抽检样品平均值	10.94	56.05	13.98	32.64	534.10	4.13
混合样品平均值	11.1	59.0	13.9	47	557	4
极差	8.5~14.3	41~67	11.1~17.5	10~73	433~644	3~5
变异系数/%	3.26	4.62	12.16	7.19	12.79	1.80

福斯 Foos Tecator 1241 近红外整粒分析仪采用光栅分光技术，可测定整粒小麦样品，具有远程联网控制功能。朱虹、陈斌等（2001 年）^[12]使用阵列发光二极管为光源的 ZX-888 近红外仪测定整粒小麦的水分含量，取得良好的结果。

15.1.3 连续光谱固定光栅近红外小麦分析

Perten DA 7200 是连续光谱固定光栅连续光谱近红外检测仪器（图 15-5、图 15-6），采用固定光栅和二极管阵列检测器新技术，可同时检测和并行处理分光后样品连续光谱信息，具有波长精度更高、检测速度更快（1~3s）的优点。

仪器采用非接触式检测方法，样品放在旋转或固定样品盘上，仪器光源照射到样品上，其反射光由固定光栅分光，被二极管阵列检测器接收，信号经放大电路和软件处理，在显示屏上显示结果。仪器可以检测固体颗粒、固体粉末液体和膏状体，而不必更换装样装置。样品盘有效检测面积大，样品不需要前处理，非常适合对不均匀样品的检测。固定光栅连续光谱近红外检测小麦结果见表 15-3。

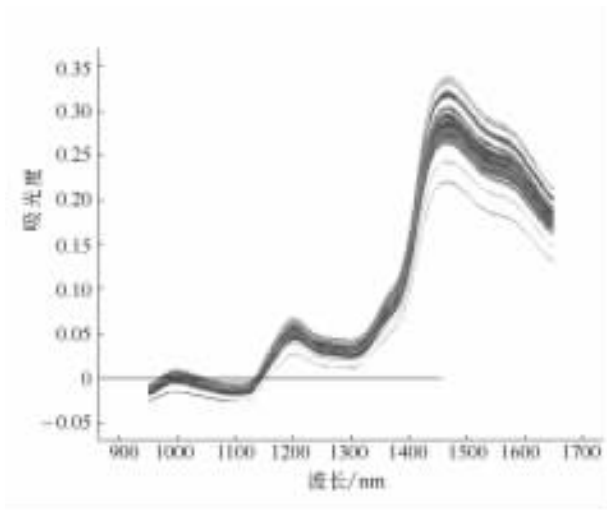


图 15-5 小麦连续近红外光谱（900~1700nm）

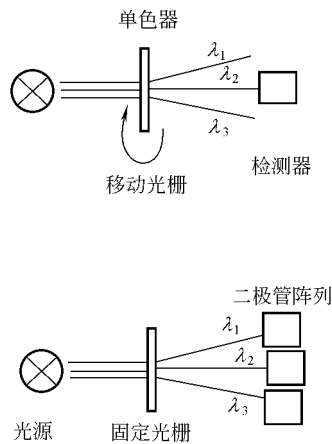


图 15-6 移动光栅与固定光栅区别

该仪器在美国农业部 USDA、加拿大、欧洲、澳大利亚等地得到应用。

表 15-3 固定光栅连续光谱近红外检测小麦结果

项目	范围	样品量	R^2	SECV	数据处理	因素	波长范围/nm
蛋白	9.2%~22.2%	786	0.98	0.35	MSC	10	1100~1400
水分	9.5%~21.6%	874	0.98	0.34	MSC, 2SG5	10	1100~1500
淀粉	61.5%~72.7%	683	0.98	0.88	MSC, 2SG5	14	1100~1500

15.2 制粉过程中的质量控制

在制粉工艺中，由于原料、设备情况及生产与销售的要求，进行工艺调整和变更工艺方案是很经常的，因此，快速及时地了解工艺变更后，其产品质量是否能达到预期的要求是十分重要的。现代化制粉工业要求在最短的时间内，用最快的方法反映工艺改变的结果，是质量管理和质量控制的关键，也是制粉工业现代化的要求。

制粉过程质量控制有离线控制（在生产线上人工取样，用近红外仪检测，进行质量控制）和在线控制（近红外仪进行检测部件安装在生产线上，进行质量控制）两类方式。

15.2.1 近红外离线检测

Perten 86 系列灰分型近红外分析仪是美国谷物化学家协会批准测定面粉灰分的仪器（AACC No. 08—21）。李巍（1998 年）^[13]报道：北京面粉八厂利用 Perten 8620 灰分型近红外仪每天随时多次取样，对产品进行质量监控，主要检查项目有面粉水分、灰分、白度、面筋等项目，如发现不合格或超标面粉，就及时调整设备或粉路，使产品质量合格。颜璐（1999 年）^[14]、刘继明、王丽等（2000 年）^[15]分别介绍：秦皇岛面粉厂、沈阳东大面粉公司使用 Perten 8620 灰分型近红外仪测定面粉水分、蛋白、灰分、颗粒度，监控质

量的经验认为：近红外仪具有测定速度快、效率高、大大节约费用和时间等优点，在生产中起重要指导作用。面粉灰分是评价质量等级和加工精度的最重要指标，它与面粉加工设备优劣、工艺参数选择、面粉产率有十分密切的关系。传统的化学方法测定灰分需要 4h 左右，分析结果就会滞后生产过程 4h。以一个中型面粉厂一天处理 300t 小麦为例，4h 生产 50t 面粉。一旦发现质量不合格的产品，再行调整，已生产 50t 不合格产品。不合格产品处理起来也比较麻烦，如果回机或降等，都会带来不必要的损失。

灰分是样品经高温炉燃烧灰化后的残留物质，即无机物。由于无机物在近红外光谱区域无明显的吸收谱带，所以无法用近红外技术进行测定。瑞典波通仪器公司采用了专利“相关”技术开发了面粉灰分测定的近红外谱段，采用两个特定波长的滤光片用于面粉灰分测定，NIR 测定结果与高温炉测定值的误差值在 0.02% 之内。相关系数达 0.95。Perten 86 系列灰分型近红外分析仪是美国谷物化学家协会惟一推荐测定面粉灰分的仪器（AACC No. 08—21）。

李巍（1998 年）^[13]介绍北京面粉八厂使用 Perten 8620 近红外仪测定面粉灰分、湿面筋、白度指标，开发面粉新品种或改变生产工艺过程的经验，用常规方法要 1~2 天调整时间，才能完成新工艺或新产品的调整过程，投入正常生产。使用近红外仪就可以在 1.5h 内完成整个过程。1997 年 9 月该厂要从生产特种面粉工艺路线改为生产精粉、专用粉两种面粉联产工艺。准备生产的两种粉的质量要求如表 15-4。

表 15-4近红外仪调整工艺过程（1）

		水分/%	灰分/%	湿面筋/%	白度
质量 要求	特精粉	<14	<0.5	>29	>80
	专用粉	<14	<0.65	>30	>79

在工艺调节过程中，随时取样用 Perten 8620 灰分型近红外仪测定，发现某一指标不合要求及时对生产工艺参数进行调整，过程如表 15-5。

表 15-5近红外仪调整工艺过程（2）

14:00		水分/%	灰分/%	湿面筋/%	白度
开车	特精粉	13.3	0.42(小)	32	81.9
	专用粉	13.3	0.60(小)	34	79.2

开车一段时间后用近红外仪测定面粉成分。发现特精粉和专用粉灰分太低，影响产品出率，进行第一次工艺调整（表 15-6）。

表 15-6近红外仪调整工艺过程（3）

14:30		水分/%	灰分/%	湿面筋/%	白度
第一次 调整	特精粉	13.0	0.44(小)	32	81.8
	专用粉	12.0	0.62(小)	35	79.1

经过第一次调整后，特精粉灰分和专用粉灰分略有提高，但仍偏低。进行每二次调整（表 15-7）。

表 15-7

近红外仪调整工艺过程 (4)

14:50		水分/%	灰分/%	湿面筋/%	白度
第二次调整	特精粉	12.8	0.45(小)	33	80.6
	专用粉	12.8	0.67(大)	35	79.0

经过第二次调整后，特精粉灰分还低，但专用粉灰分太高，进行第三次调整（表 15-8）。

表 15-8

近红外仪调整工艺过程 (5)

15:20		水分/%	灰分/%	湿面筋/%	白度
第三次调整	特精粉	13.1	0.47(小)	32.5	81.6
	专用粉	12.9	0.64(小)	34.5	79.1

经过三次调整，两种面粉灰分、湿面筋、白度已接近预期要求。调整过程共用 1.5h，工艺调整完毕后，转入正常生产，两种产品均达到质量要求。常规法测定也证明两种面粉符合质量标准。

另外，采用不同面粉搭配（配粉）工艺来生产专用面粉，可以根据近红外仪测定各种被搭配面粉的测定结果，计算出合理搭配比例，生产出符合质量要求的专用粉，保证产品质量的稳定。

15.2.2 布勒-波通在线近红外检测系统

1993 年瑞典波通仪器公司研制成功了 MYRA-PC 近红外在线连续测定系统（Online Measuring and Control Unit）^[16]，该系统由在线检测部分（图 15-7）和控制、显示两部分组成。由瑞士布勒公司负责安装到生产线上，现已在全世界布勒面粉生产线上广泛使用。我国黑龙江大庆面粉厂已安装了 MYRA-PC 近红外在线连续测定系统。

系统采用 6~12 个滤光片反射检测技术。检测波长为 540~2500nm，每分钟检测样品最多可达 22.7 次。光学系统工作原理见图 15-8。



图 15-7 在线近红外测定系统

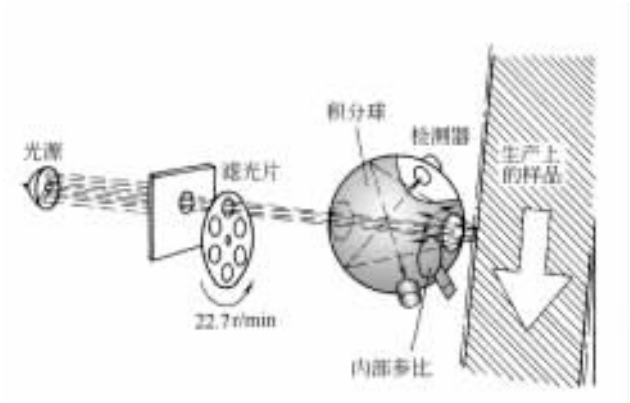


图 15-8 在线近红外测定光学系统工作原理

15.2.2.1 检测及控制系统组成

检测部分安装在面粉厂的不同生产线上，控制、显示部分安装在中央控制室内。光学

部件测定管路中流动的面粉，把测量信号传递到中控室，最大间距 1200m。

15.2.2.2 检测项目和精度

瑞士布勒公司保证在线检测系统测量精度为水分 0.20%，蛋白质 0.20%，灰分 0.02%。

- 面粉：
 - 蛋白、水分和灰分（6 个滤光片）
 - 蛋白、水分、灰分、白度和吸水率（12 个滤光片）
- 颗粒粉：
 - 蛋白、水分、灰分及颗粒度（12 个滤光片）
- 小麦：
 - 蛋白、水分和灰分（12 个滤光片）

15.2.2.3 自动控制和显示

测量信号反馈到中控室计算机后，计算机会自动运算，并在显示屏上显示各指标的平均值、最大值和最小值（图 15-9），用图示曲线表示随时间变化曲线，自动生产趋势分析图（图 15-10），当超过设定值会发出报警信号，及时通知技术人员，当检测系统与生产控制系统相连时，会自动调整工艺参数，确保产品质量的连续稳定。

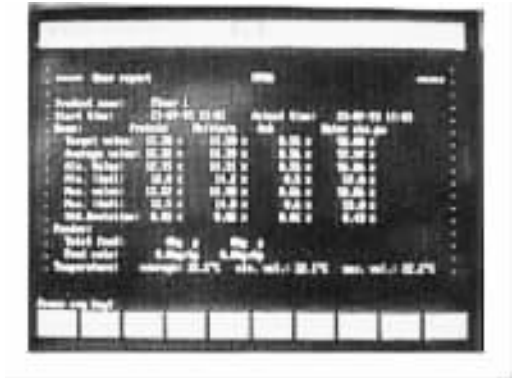


图 15-9 质量报告

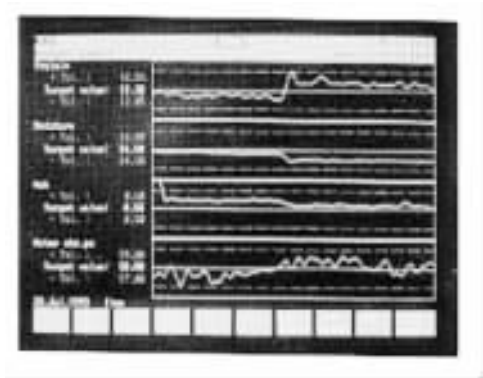


图 15-10 生产趋势分析

张广军、Suranjan Panigrahi (2001 年)^[17]报道：采用近红外透射光谱分析技术，建立了谷物加工过程蛋白质含量在线监测系统。用宽光束照射、透射光谱二阶微分处理和人工神经网络的标定方法保证监测系统有效工作。同时建立小麦蛋白质含量现场在线监测的最佳 BP (Back Propagation) 网络结构，其 $SEC=0.12\%$ ， $SEP=0.15\%$ 。该监测系统不仅可以直接应用于谷物成分含量的在线监测，也同样适合于其他光谱分析应用。

15.3 面粉和副产品的检测

近红外分析仪适合于面粉质量测定，也适用麦麸和面团的检测。

15.3.1 面粉近红外检测

据 Perten 公司介绍^[18]：Perten 86 灰分型近红外测定仪可测定面粉多种组分和项目，

除了检测面粉水分、蛋白质、灰分含量外，还能对面粉加工特性和面粉“内在功能”指标如水吸收率、粗细度、损伤淀粉进行检测。

粗细度（颗粒度）表达小麦胚乳在磨粉机中粉碎的程度，不同类型和不同等级的面粉有不同粗细度要求，如果面粉研磨得太细，会使小麦胚乳的淀粉粒结构造成损伤，影响面粉的理化特性，如面粉水的吸收率、吸水速度、对淀粉酶的抵抗能力等。粗细度的常规测定方法是用不同孔径的筛网进行筛理，计算留存或通过一定孔径筛下物占全部被筛理样品的百分比。

水吸收率表示面粉加水混合，使面团达到一定阻力或稠度时的加水量，它与面粉原始水分、蛋白质含量、损坏淀粉的程度都有密切的关系。一般面粉吸水率 56%～65%，蛋白质含量越高、淀粉损伤程度越大的面粉吸水率越大。

小麦由外层麸皮和内部胚乳组成，麸皮是褐色的，通过小麦磨粉加工和筛理使麸皮从面粉中分离出去，麸皮分离的越完全，面粉的颜色越白，所以精制粉的白度明显高于普通粉，白度不但反映面粉加工精度，也反映面粉中麸皮分离的程度，对于精制粉来说，胚乳中的胡萝卜素，黄色素的含量也会影响面粉的颜色和白度。

由表 15-9 可以看出近红外检测面粉水吸收率、粗细度、损伤淀粉指标在测量精度上不如水分、蛋白和灰分那样精确，但对面粉厂的生产和质量控制有很好的指导意义。

表 15-9面粉测定项目

项目	范围 Range	相关系数 R	标准差 SEE	基准法
水分 / %	12.6～15.7	0.99	0.12	烘箱法
蛋白 / %	10.5～12.3	0.96	0.18	凯氏
灰分 / %	0.3～0.75	0.95	0.02	高温炉法
白度单位	3.2～4.8	0.878	0.48	Kant-jones
水吸收率 / %	56～65	0.954	0.79	Brabenoter
粗细度单位	35～65	0.9	4	德国
损伤淀粉 / %	19～39	0.861	2.73	法兰德法

15.3.2 麦麸近红外检测

近红外技术除了对面粉产品测定外，也可以用于面粉厂副产品麦麸成分的测定。1992 年 Perky 报道了近红外技术成功测定 35 份麦麸的水分、蛋白、脂肪、纤维成分的测定范围、相关系数、标准差性等指标，见表 15-10。

表 15-10麦麸测定项目

项目	范围 Range	相关系数 R	标准差 SEE
水分 / %	7.6～11.4	0.99	0.11
蛋白 / %	15.7～21.6	0.97	0.34
脂肪 / %	1.5～4.7	0.94	0.29
纤维 / %	2.5～3.5	0.91	0.14

15.3.3 面团近红外检测

面包生产过程中，首先要把面粉加水混合成面团，经发酵成型后再入炉烘烤成面包。比利时研究者采用近红外技术测定面包中间产品面团的各**种成分**，测定项目包括面团的蛋白质、盐、糖、脂肪和干物质含量，表 15-11 列出了测定范围、相关系数、标准差等指标。

表 15-11		面团测定项目		
项目	范围 Range/%	相关系数 R	标准差 SEE	基准法
蛋白质	8.5~16.7	0.984	0.04	凯氏
盐	0.96~1.80	0.980	0.05	Potentiometric
糖	1.20~7.40	0.979	0.44	Bertrand
脂肪	0.68~32.26	0.998	0.68	Gerber
干物质	59.3~74.8	0.996	0.33	计算法

15.4 小结

近红外分析技术在面粉加工行业已得到广泛的应用，在原料小麦、加工过程、成品面粉质量控制方面取得不少成功的经验，充分发挥了近红外技术简便、迅速的优点。在原料小麦收购中可保证质量，贯彻优质优价政策；在面粉加工过程中，保证面粉质量的稳定，减少次品的产生；在产品销售方面，把好面粉质量关，为创名牌，提高经济效益提供了可靠的技术保证。

随着对近红外技术的发展和进步，对近红外分析技术的逐步深入认识和了解，近红外光谱分析技术在制粉领域将发挥越来越大的作用，必将为我国面粉加工行业技术进步和现代化做出重要贡献。

参 考 文 献

[1] Law D P. Determination moisture content in wheat by NIR diffuse reflectance spectrophotometer. Cereal chem, 1977, 54 (4) : 874~881

[2] Law D P. Near infrared diffuse reflectance spectra of wheat and wheat components. Cereal chem, 1977, 54 (2) : 256~265

[3] Williams P C, Norris K H, Sobering D C. Determination of protein and moisture in wheat and barley by NIR transmission. J Agric Food chem, 1985 (33): 239~244

[4] 刘广田, 尤明山. 应用近红外分析技术测定小麦籽粒的性状. 近红外测定仪技术培训班资料, 1998

[5] 彭玉魁, 李菊英. NIRS 法同时测定小麦种子水分、粗蛋白、赖氨酸和粗淀粉含量研究. 西北农业学报, 1996, 5 (3): 31~34

[6] 彭玉魁, 李菊英, 祁振秀. 近红外光谱分析技术在小麦营养成分鉴定上的应用. 麦类作物, 1997, 17 (2): 33~34

[7] 陈锋, 何中虎. 利用近红外透射光谱技术测定小麦品质性状的研究麦类作物学报, 2003, 23

(3): 1~4

- [8] 周艳华, 何中虎. 中国小麦硬度分布及遗传分析. 中国农业科学, 2002, 35 (10): 1177~1185
- [9] 任红波. 近红外仪快速测定小麦蛋白及硬度. 黑龙江农业科学, 2003, (3): 17~18
- [10] 刘凤华, 刘志华. 近红外光谱分析技术在进口小麦的应用. 现代商检科技, 1995, 5 (1): 53~54
- [11] 胡新中, 郭波莉, 魏益民等. 近红外技术 (NIR) 在小麦商品粮收购中的应用研究. 西北农林科技大学学报 (自然科学版) 增刊, 2002 (30)
- [12] 朱虹, 陈斌, 付西光. 用 ZX-888 型近红外分析仪快速检测小麦水分. 食品科技, 2001 (6)
- [13] 李巍. 8620 型近红外测定仪在粉厂中的应用. 近红外测定仪技术培训班资料, 1998
- [14] 颜璐. 浅谈近红外分析仪在制粉行业中的应用. 西部粮油科技, 1999, 24 (5): 51
- [15] 刘继明, 王丽等. Perten 8600 近红外仪在面粉厂的重要应用. 粮油食品科技. 2000, 8 (5): 28~29
- [16] Perten Instruments NIR Online measuring and control unit MYRA-PC (technical leaflet), 1998
- [17] 张广军, Suranjan Panigrahi. 近红外透射式谷物蛋白质含量在线监测系统. 光电工程, 2001, 28 (2): 18~22
- [18] Perten Instruments: Application sheet of Perten 86 series Model Near-Inframatic NIR Analyzer

16. 近红外光谱技术在饲料分析上的应用

饲料工业中饲料品质的控制包括饲料原料、饲料成品料及加工工艺的控制。用作饲料的原料范围很广，主要包括植物性饲料、动物性饲料、发酵产业的副产品和各种饲料添加剂。植物性饲料中，主要的一类为谷物及谷物的加工副产品，目前在饲料业中常用的有玉米、小麦、大麦、燕麦、高粱、稻谷、麦麸、米糠等。各种豆类及其加工的副产品，特别是油脂加工业的副产品，如豆粕、棉粕、菜粕、花生粕、胡麻粕等也是最常用的植物性蛋白性饲料。动物性的饲料主要有鱼粉、肉粉、肉骨粉、血粉、羽毛粉、蚕蛹粉等。饲料工业标准中规定的对原料品控指标包括水分、粗蛋白质、粗纤维、粗灰分，不同的等级规定了不同的营养指标值。饲料全价料是饲料工业的终产品，在配制全价料前需首先测定出各种原料各营养指标的含量，再进行饲料配方，然后按照一定的生产工艺生产出终产品。终产品出厂前要进行各种指标的检验。饲料全价料出厂时的检验指标包括感官指标、成品粒度、水分、粗蛋白质；判定成品饲料合格指标包括感官指标、水分、混合均匀度、粗蛋白质、粗灰分、粗纤维、钙、磷；而能量（猪为消化能、鸡为代谢能、牛为净能）、粗脂肪、食盐、成品粒度为参考指标。因此饲料成品的品质控制是一个比较复杂的过程。在饲料工业中，饲料质量常表现在以下几个方面：

① 饲料中营养素的含量，用化学方法进行测定。饲料中的营养素有几十种之多，把这几十种营养素分成六大类：水分、蛋白质、碳水化合物、脂类、矿物质、维生素。对饲料中营养素的测定分为概略养分分析和纯养分分析。概略养分分析（Feed Proximate Analysis）是1864年德国的Hanneberg提出的常规饲料分析方案，把饲料中的营养成分分成六大类：水分、粗蛋白质、粗脂肪、粗纤维、粗灰分和无氮浸出物，通过该方案可测出上述六种指标。该方案一直沿用至今，简单、适用。一般的小型 and 中型饲料厂仅进行概略养分的测定，测定的指标包括水分、粗蛋白质、粗灰分、钙、磷、食盐。但随着饲料工业的发展，仅进行概略养分分析不能满足需要，需要进行纯养分的分析，如测定蛋白质中各种氨基酸的含量，如赖氨酸、蛋氨酸、苏氨酸等18种氨基酸的分析，粗脂肪中必需脂肪酸（如 ω -3脂肪酸等）的分析，纤维中中性洗涤纤维、酸性洗涤纤维及木质素的分析，锌、铁、锰等微量元素的分析，13种维生素的分析等，但这些分析往往需要价格较高的设备，并且试剂昂贵。纯养分分析一般在科研机构、大专院校、饲料质检部门及大型的饲料厂内进行，一般的饲料生产企业常送样到这些机构进行测定，但测定的费用昂贵，并且需要训练有素的操作人员，否则测定的误差比较大。

② 饲料的消化率和生物学价值。饲料中含有的营养成分不能全部被动物消化，不同饲料的消化率不同，因此常进行消化率的测定来评定饲料的品质。被消化吸收后的养分也不能全部被动物利用，可利用的养分占吸收养分的百分比称为饲料的生物学价值。因此，决定饲料品质的第二方面即是饲料的消化率和生物学价值。饲料中含有的营养素在动物体内的消化率和消化吸收后的营养素被动物的利用率最终决定了动物的生产性能，因此饲料工业中常进行消化率和生物学价值的测定，消化率常利用动物的消化实验进行测定；测定

的周期长，操作复杂，费用昂贵。

③ 工艺质量。饲料原料有些需要进行粉碎，粉碎后的原料需在搅拌机中进行搅拌，有些饲料需要制粒和膨化，因此饲料的粉碎粒度、搅拌均匀度、制粒的温度等工艺参数会影响饲料的品质。饲料工业标准规定了全价料的粉碎粒度、混合均匀度和成品粒度的测定方法。

④ 卫生质量。新版《饲料卫生标准》已于 2001 年 7 月 20 日经国家监督检验检疫总局批准发布，自 2001 年 10 月 1 日起实施。规定了砷、铅、氟、铬、汞、镉、霉菌、黄曲霉毒素 B₁ 等在各种饲料中的允许含量，规定沙门氏杆菌在饲料中不得检出。饲料中微生物含量用微生物法进行测定。

⑤ 抗营养因子和有毒因子的含量。饲料原料还含有各种抗营养因子和有毒因子，这些抗营养因子的存在会影响饲料的品质，严重时会造成动物的中毒。常见的抗营养因子有异硫氰酸酯，存在于菜籽饼、粕中，其允许量（每千克产品中）（以丙烯基异硫氰酸酯计） $\leq 4000 \text{ mg}$ （GB/T 13087）。棉籽饼、粕中游离棉酚的允许量（每千克产品中） $\leq 1200 \text{ mg}$ （GB/T 13086）；亚硝酸盐（以 NaNO_2 计）的允许量（每千克产品中），鱼粉 $\leq 60 \text{ mg}$ （GB/T 13085）；氰化物（以 HCN 计）的允许量（每千克产品中），木薯干 $\leq 100 \text{ mg}$ （GB/T 13084）。

为评定饲料的品质需进行上述各种指标的评定，上述这些评定饲料品质的指标都是应用传统的离线测定方法，用化学方法进行测定，常常需要破坏样品，耗费时间，需要装配良好的实验室和经过培训的操作人员，有的需要具有危险性的化学试剂。因此，这些测定方法是非常昂贵的。为控制一个产品的加工过程和质量，就需要一种快速、费用低和非破坏性的离线或在线的测量方法。因此，需要一种简单的方法来测定几种不同而有关联的质量指标。近红外光谱分析（NIR）这种方法适合于离线和在线（at-and on-line）测定，它能够完成以上所提到的各种有机指标的测定，NIR 不适合进行钙、磷、铁、锌等无机化合物的测定。近红外光谱技术正迎合了饲料行业需要快速进行饲料原料检测、产品的在线测定及产品出厂前进行品质控制的要求，在 21 世纪必将成为一个先导技术。

16.1 近红外分析在饲料分析中的理论特性及其优势

16.1.1 近红外分析在饲料分析上的优势

近红外在饲料分析上的优势是测定快速，1min 内测定出结果；一次可以同时测定几个甚至十几个指标，如水分、粗蛋白质、粗脂肪、粗纤维、氨基酸、维生素、能量、有毒有害因子等；不需要对样品进行破坏性的处理；不需化学试剂；预测消化率不需要进行生物学实验；另外其操作简便，不需训练有素的操作人员。

近红外测定的投入主要在仪器本身。近红外的维护和操作费用非常低，一台仪器一个人即可操作，一天可测定几百个样品，操作的过程很简单，一般可分为四个步骤如图 16-1~图 16-4 所示。第一步装样；第二步选择要进行分析的样品方程，按分析键，仪器开始分析，第三步仪器给出分析结果，通过打印机打印出来；第四步，清理样品槽内的样品，测样结束。



图 16-1 装填样品



图 16-2 选择要分析的成分，开始检测



图 16-3 1min 内读取检测结果



图 16-4 清理装样装置

仪器一般配有打印机，结果会自动打印出来。近红外也可与计算机连接，结果显示在计算机屏幕上。对于操作者来说，近红外像一个“黑箱”，放进饲料样品，仪器给出测定结果。那么近红外是根据什么样的原理进行测定的？下面阐述一下近红外进行饲料成分测定的原理。

16.1.2 近红外进行饲料成分测定的原理

近红外仪器使用 $700\sim 2500\text{nm}$ 的红外光，近红外对饲料样品的测定主要是根据饲料样品中的有机成分对近红外光的吸收。首先做一个简单的类比：想想花园中的绿草，假定它是绿色的，为什么它看起来是绿色的？因为当光照射到草上时，它吸收了红色的光谱区段，把黄色光和蓝色光反射回去，而反射回去的黄色光和蓝色光就形成了所看到的绿色。如果草是稀疏时，看到的是一点点的绿色，如果是生长旺盛而浓密的草地，看到的是深绿色。草越多，绿色越深。不仅得到草越多绿色越深的概念，而且同样可以根据物体反射光的颜色把草与其他物体区分开来。想象一下大街上的沥青，人的眼睛可根据颜色把沥青和草区分开，因为沥青和草反射了光的不同区段。

NIR 光谱的原理与此相似，正像所见到的不同的物质在不同的光谱区段有不同的颜色一样，不同的物质在近红外光谱区段具有不同的近红外光谱。因此，近红外光谱仪不仅

能把水分与其他成分区分开，而且能够根据吸收光谱的强度判定水分的含量。就像密集的草地比稀疏的草地看起来更绿一样，水分越多，吸收的特征光谱越多。1450nm 和 1940nm 的光特别会被水吸收，2 个水分含量不同的小麦样品对 1940nm 光的吸收情况是不同的。这说明，1940nm 的光可以用于检测小麦样品的水分含量。以此类推，蛋白质主吸收峰波长为 2180nm，脂肪为 2130nm 和 1759nm，淀粉为 2100nm。纤维为 2310nm（见图 16-5、图 16-6）。实际上有机物质化学组成是复杂的，所以吸收光谱也是十分复杂的。可以用包括这些波长的光来同时检测饲料水分、蛋白质和脂肪等的含量，各成分的响应波长见表 16-1。

纯水和蛋白含量为 12.4%水分含量不同的两种小麦近红外光谱

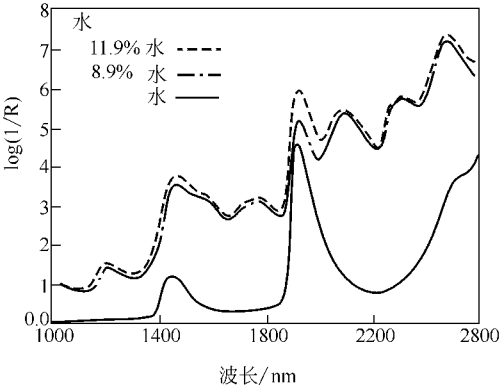


图 16-5 纯水和两种不同水分含量小麦的吸收光谱图

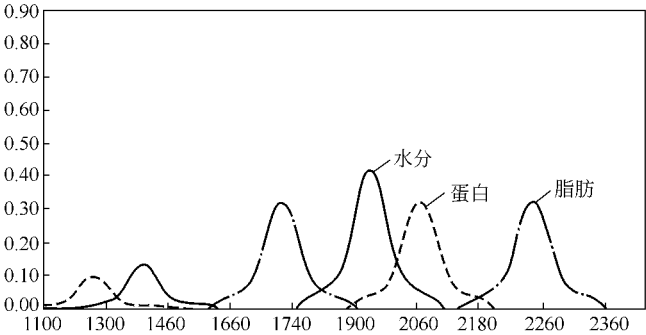


图 16-6 水分、蛋白质、脂肪的吸收光谱图

表 16-1 近红外波长与各成分的响应特性

波长 /nm	响应指标	波长 /nm	响应指标
2345	纤维素,脂肪	2050	蛋白质
2336	纤维素,纤维	1982	尿素,水
2310	脂肪,纤维	1940	水
2270	酒精,脂肪	1818	乳糖,纤维素
2230	参考	1778	纤维,淀粉
2208	尿素,乳糖	1759	脂肪
2190	蛋白质	1734	蛋白质
2180	蛋白质	1722	纤维素,淀粉
2139	淀粉,蛋白质	1680	参考
2100	淀粉	1445	水

有好几个滤光片对蛋白质含量进行响应。波长 2180 nm 和 2190 nm 经常用于像小麦或大麦这样低蛋白质含量的谷物。当测量蛋白质含量高的油菜籽时，可能要用波长 1734 nm。对于波长 2139 nm，则主要用于对淀粉的响应，但也用于蛋白质吸收。脂肪情况与

此类似。

1680nm 的波长为参考波长，主要的饲料组成成分在这个波长上不吸收 NIR 光。这个波长可作为像颗粒大小这样影响光谱特性的因素的补偿。

对于用近红外测定大麦水分，可以使用 2310nm 的波长作为参考波长。从表 16-1 中可以看出，这个波长实际上是用来测量脂肪的。因为，大麦、小麦和面粉的脂肪含量相对来说差别很小。所以，对于此类产品这个波长经常用于水分参考滤光片。用 2310nm 作为参考滤光片比用 1680nm 优越，因为它距离水分的特征吸收波长 1940nm 更远一些，并且同 1940nm 处的光吸收量相似。

一般来说，用近红外测量小麦蛋白质，选择的波长为 3 个，一个蛋白质的响应波长，一个淀粉的响应波长和一个参考波长。淀粉干扰对蛋白质的测量，需要测量淀粉来进行补偿。参考波长用来补偿其他因素的影响。对于小麦蛋白质，一般用 2180nm（蛋白质）波长，同时还用 2100nm（淀粉）和 1680nm（参考）波长。

如果直接测量干基蛋白质，应包括 1940nm 波长来对水分进行补偿。所以，对于测量小麦蛋白质，应当在实际水分含量下测量蛋白质时，用 3 个波长，当测量蛋白质干基时，用 4 个波长。

如果每一种物质的特征吸收光谱是惟一的，NIR 可能是一种最好的分析方法。不幸的是，事实并不是如此简单。每一种物质具有很多的 NIR 吸收峰，每一种物质对光谱的吸收与其他许多物质重叠在一起。所以，当试图测定一个包括很多成分的复合物，如豆粕中的蛋白质的吸收峰时，在选择的任何光谱点都包含了其他物质，如水分、纤维、油分在此光谱点的吸收。更糟的是，测定的吸收光谱受样品温度的影响。如此复杂的事情怎么能完全用作一种分析方法呢？答案在于定标的过程。通常几百个样品用于开发一个具有实际应用价值的良好的定标方程。定标样品应覆盖以后在分析过程中能遇到的所有的变异范围。一旦收集了所有的定标样品，将应用数理统计方法建立光谱值与实验室测定值之间的相关关系。为列出 NIR 的用途及 NIR 在饲料分析中的应用，需要了解 NIR 方法在另一个重要领域的应用。此法最初的应用基于样品表面对光的反射，检测器检测反射出来的光谱。此种方法被称作近红外反射光谱或 NIR。虽然进行 NIR 分析所做的样品准备比大多数传统的分析方法简单快速，但它还是需要把样品粉碎以获得均匀一致的样品粒度，因而为光谱反射提供光滑而均匀的反射表面。最近，许多近红外的应用是测定近红外透射光谱，它测定入射光穿过样品后的光密度。这种方法被定义为近红外透射光谱，或称 NIT。NIT 有这样的特点，它需要很少或不需要样品的制备，因它不需要样品粉碎，但却有很好的重复性。但这种制样的简便导致测定结果敏感性下降，因此许多特殊的分析只能用更敏感的 NIR 技术进行测定。

近红外光谱仪的工作原理就是把各种饲料表示为对各特定频率（或波长）近红外光的吸收。饲料对近红外光的吸收用 NIRS 仪发射光与样品反射光的差值表示。

为了预测饲料成分必须进行定标。定标就是建立 NIRS 光谱值与营养成分值（由湿化学法得到）间的相关关系。用一个假想模式很容易理解这一概念。假设三个豆粕样品除蛋白含量不同外，其余成分含量完全相同。它们的蛋白含量分别是 44%，46%，48%。在这个例子中，样品的光谱值仅在 1900~2000nm 的区域内不同，这一差异只能与样品蛋白含量的差异有关，因而不同的样品被分离出来。

将每个样品在波长 1950nm 下的吸光度与相应蛋白含量作相关图，结果显示，样品在 1950nm 处的吸光度与其蛋白含量呈线性相关。可以设想，这个相关可用于预测第四个豆粕样品的蛋白含量。

实际的光谱要复杂得多。蛋白不只在 1 个光谱区段产生吸收，当样品中的蛋白含量增加时，其他成分（如碳水化合物）的含量就会下降。相应地，样品对其他光谱区段的吸收也将改变。因此，这三个样品的光谱会在几个区段内表现不同，因此很难确定在哪个光谱区段中样品吸光度与蛋白含量相关。但不管怎样，上述这一理想模式是对确定样品光谱值与化学成分值间相关关系的概括，这一相关关系被用于开发定标方程来预测未参与定标的样品。

近红外光谱法定量测试有机物，主要应用近红外反射光谱法。光的反射有两种，一种是表面反射，另一种是漫反射。所谓表面反射，就是光线照射在光滑物体表面上被有规则地反射出来的现象，它没有携带入射光与样品相互作用的信息，与样品相互作用无关。所谓漫反射，就是光线照射到粗糙物体表面被无规则反射现象。当光线照射到由一定厚度颗粒物质组成的样品层时，一部分被吸收，一部分被反射出来，反射出来的光线反映了样品的吸收特性，更多地携带有样品的化学信息。在样品具有相当厚度的情况下，用红外光照射有机物，样品透射光可忽略不计，入射光主要被样品吸收或是被样品反射。

$$\text{反射率} = \frac{\text{反射光强}}{\text{入射光强}}$$

$$\text{光密度} = \log \frac{1}{R}$$

因此，样品在近红外区的不同的波长处便会产生相应的光密度值，光密度值的大小与样品某成分的含量有关。用波长及其对应的光密度值便可绘出样品的吸收光谱图。吸收峰位置与样品中组分结构有关。

近红外漫反射分析的基础在于不同样品的不同组分在近红外区有特征吸收。被测组分含量与特征吸收波长处漫反射率倒数的对数成线性关系，即：光密度 $A = \log 1/R \propto C$ 。

被测饲料样品的光谱特征是多种组分的吸收光谱的综合表现。对其中一个组分便可建立一个回归方程：

$$Y = C_0 + C_1 X_1 + C_2 X_2 + \cdots + C_n X_n$$

Y——有机物某成分的百分含量；

$C_0 \sim C_n$ 为回归系数；

$X_1 \sim X_n$ 为各有机成分的反射光密度值。

Y 值可通过常规法求得。 $X_1 \cdots X_n$ 值可用近红外光谱仪获得。

利用一套定标样品（50 个左右），利用常规法求得的化学分析值及其在近红外光谱区的吸收光密度值，通过多元回归计算，求出上述公式的回归系数，即可建立定标方程，即某饲料某成分的定标软件，对未知饲料样品进行检测。

16.1.3 近红外光谱分析过程

近红外光谱分析过程可分为三大步骤：定标、预测和应用。

16.1.3.1 定标

运用一套定标用样品，根据其化学测定值和特征波长处的光密度值，运用多元回归计

算求出定标方程的系数，即求出定标方程的过程，它包括以下几步：

- ① 标样筛选：从大量样品中筛选出一套用于定标用的样品。筛选出的样品必须具有代表性，包括物理学、化学、植物学特性和加工贮藏方法。选样的方法可根据样品的含量分布、结构特性或样品光谱特性随机选择。
- ② 常规分析方法测定成分含量：用常规化学法测定样品某成分含量，获得标样的某成分含量值。
- ③ 获取光谱数据：用近红外光谱分析仪获取定标样品的光谱数据，对其进行数学转换（一阶或二阶导数形式）。
- ④ 制作定标方程：利用样品成分含量及样品的光谱数据，运用多元回归计算产生某成分的定标方程。
- ⑤ 定标方程准确性的检验：定标方程优劣的评判可选用定标标准差 SEC 、变异系数 (CV) 、定标相关系数 R_c 和 F 检验的 F 值等对定标结果作初步评价，如 SEC ， CV 比较低而 R_c 、 F 值高说明其准确性较好，否则不好。这些参数的具体含义如下：

$$SEC = \sqrt{\sum (Y_{NCi} - Y_{RCi})^2 / (n_c - n_w - 1)}$$
$$R_c = \sqrt{1 - SEC^2 (n_c - n_w - 1) / SD_{RC}^2 (n_c - 1)}$$
$$F = R_c^2 (n_c - n_w - 1) / (1 - R_c^2 n_w)$$

其中， n 和 n_w 分别为定标样品数和波长数；
 Y_{NCi} 为第 i 个样品某成分近红外预测值；
 Y_{RCi} 为第 i 个样品某成分的化学分析值；
 SD_{RC} 为定标样品某成分化学分析值的分布标准差；
变异系数为 SEC 占平均数的百分比。
定标方程的最终衡量涉及到预测过程。

(1) 实际定标方程的举例

如果想用近红外测定大麦的水分含量，首先要做预测大麦水分的定标方程。
将 20 个大麦样品用化学法测定它的水分值，然后把已知水分值的大麦粉碎样品放在近红外仪器中，同时，记录其对 1940 nm 光吸收率，就可以得到类似表 16-2 的结果。

表 16-2 大麦水分化学值以及 1940nm 近红外光的吸收率

水分* 烘箱法	1940nm 吸收率	水分* 烘箱法	1940nm 吸收率
7. 55	0. 36342	10. 4	0. 46437
8. 45	0. 38214	10. 85	0. 40843
8. 70	0. 39585	11. 15	0. 40790
9. 05	0. 38507	11. 35	0. 42551
9. 40	0. 39972	11. 7	0. 46944
9. 50	0. 43819	12. 2	0. 49005
9. 55	0. 44372	12. 2	0. 49994
9. 80	0. 40061	12. 3	0. 43765
9. 9	0. 40973	12. 9	0. 45592
10. 1	0. 44301	13. 0	0. 48210

* 水分值为双试样平均值；资料来源于瑞典波通仪器公司。

通常来说，随样品中水分增加，1940nm 处的光吸收程度也增加。然而，这种关系不是完全的正比关系。接下来，可以用表 16-2 的数据得到类似图 16-7 的关系图。图 16-7 可以很形象地说明样品光吸收率和水分含量的关系。

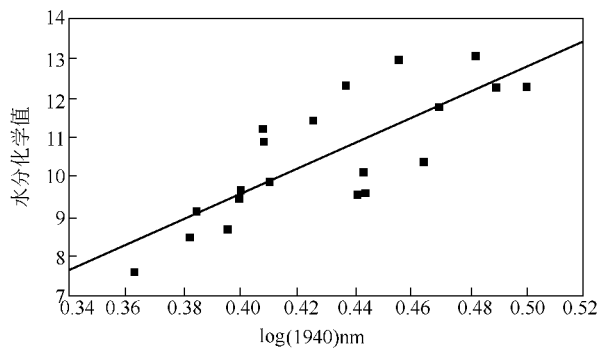


图 16-7 20 个粉碎大麦样品结果

然而，这一结果却不是非常可靠的。如图 16-7 所示，各点离斜线的距离相当大，这样处理并不让人非常满意。除了只用 1940 nm 水分响应波长外，还会用第 2 个波长的光，将水分含量与这 2 个波长吸收率的差异作成 X-Y 图，见图 16-8。由图 16-8 可见，极大地提高了检测结果的精度。

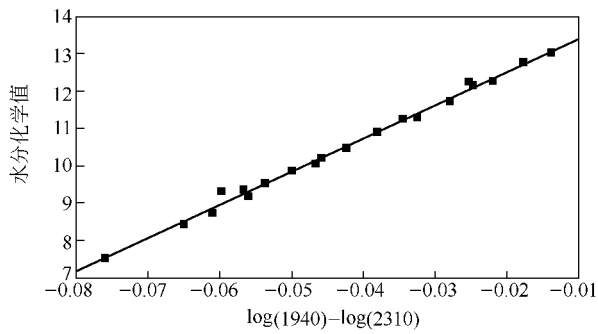


图 16-8 20 个粉碎大麦样品水分含量与二个波长的光吸收率关系

样品对 1940 nm 光的吸收率不仅受水分含量影响，而且受样品粒径大小的影响。虽然定标及检测的样品都是用同样的方法研磨的，但是颗粒大小还是有差异，这些差异将影响测定结果。

如果粒径导致样品特征波长处的吸收率偏高，那么就要用第二个光波对粒径的影响进行校正，因为第二个光波也会受粒径的影响，通过分析它们彼此间的差异，就能“消除”粒径对光吸收率的影响。如果用 1940nm 的光波测定水分，第二个光波就应该从不是水分的特征光波中进行选择，即使用一个测水分的特定波长，另一个波长作为参考。

由图 16-8 可见，这种差异和水分含量之间有着很好的一致性。测定结果与斜线偏差较小。可以用一种叫做最小平方或回归分析法的数理统计方法来进行分析。采用回归分析程序，可以真正地计算出与测定结果最接近的斜线方程。

采用此程序，图 16-8 中斜线的方程为：

$$\text{水分} = 14.3 + 88.8(\log 1940 - \log 2310)$$

例如，一未知成分的样品的 1940nm 的光吸收率为 0.49005，2310nm 为 0.51216，那么，水分值可预测为：

$$\text{水分} = 14.3 + 88.8(0.49005 - 0.51216) = 12.3$$

这样，就给近红外仪器建立了一条校准曲线。实际上，采用同样类型的回归程序，将得到各个波长的不同的常数。结果如下：

$$\text{水分} = 13.6 + 87.7(\log 1940) - 86.3(\log 2310)$$

这与上面的公式很类似。2 个波长的滤光片常数均为 88.8，而此公式中分别为 87.7 和 86.3。14.3 也略有变化，变为 13.6。对 2 个波长采用各自的常数可以得到更好的结果。

回归公式中的 13.6，称之为截距或补偿常数，写为 C_0 ，常数 87.7 和 86.3 称作滤光片常数。这些常数将被近红外仪器储存如下：

$$C_0 = 13.6$$

$$\text{水分滤光片常数} = 87.7$$

$$\text{参考滤光片常数} = -86.3$$

现在，就可以用近红外仪器来测定未知大麦样品的水分含量了。近红外仪器将根据测量的光的吸收率和储存的校准常数，自动计算出样品的水分含量。

(2) 回归分析数据和定标常数好坏的判定

定标的过程是建立实验室化学值和近红外光谱值之间的比例关系，NIR 软件提供较好的回归分析等数理统计功能，利用定标软件可制定定标方程。在制定定标方程的过程中，判断所收集样品记录和建立校准常数的好坏，一般应观察以下几个指标。

① 同一样品的 log 值的差值

同一样品几次扫描 log 值的差值反应样品的均匀性。log 值为小数点后 5 位，后 3 位用百位数表示的话，大多数谷物产品，如研磨过的小麦、大麦和小麦粉，每 2 次光谱差值在 0~500 范围内，一般在 200 左右，对于磨过的饲料样品差值情况也如此。

奶粉应在 200 范围内，乳清粉差值较大。酒类产品，差值一般较小，在 0~100 范围内。而不透明液体差值则相对较大。鲜奶制品差值一般都在 100 以内，偶尔出现 200 以内，也是可以允许的。一般复杂的饲料样品要求在 400 以内，在 200 左右最佳。

② Mah 距离

它是对某个特定记录的 log 值与全部 log 值平均值差值的量度。Mahalanobis 距离越大，这个特定记录或者说测量 log 值的差值也就越大。Mahalanobis 距离与基准实验室数据无关，只依赖于近红外仪的光学测量。一般国外将 Mahalanobis 距离大于 7 的样品，认为与整体有明显差异，应视为非同类产品。在我国，一般可以将此距离定得大一些。

③ SEE

SEE 值（误差）尽可能的小，它不可能小于基准分析法的误差。

对于研磨谷物（麦类作物、油料作物及其副产品）蛋白质的测量，SEE 值一般低于 0.25~0.30。对于研磨谷物水分的测量，SEE 值一般应低于 0.20。灰分为 0.02，豆粕、棉粕、菜粕一般小于 0.5，而对于动物性饲料样品一般小于 2.0。

④ 相关性 Corr

相关性是指近红外仪测量与基准分析法之间线性关系的强度。相关系数为 1 表明两者完全吻合，相关系数为 0 表明没有关系，相关系数越接近 1 表明相关性越好。

一般来说，小的 SEE 值比大的相关性更为重要。

⑤ F -值

F -值越高，校准的统计有效性越强。 F -值应一般高于 100。

⑥ C 值

C 值一般应比较小。大的 C 值，如大于几千或上万，则表明检测的重复性不好。

⑦ T -值

T -值为某滤光片的校准常数除以其标准偏差。如果误差与校准常数本身一样大，那么此滤光片的 T -值为 1。在这种情况下，此滤光片的校准常数没有意义，可以去掉这个滤光片重新定标。某个滤光片的 T -值表明该滤光片在定标方程中的重要性， T -值越大重要性越大。一般如果一个滤光片的 T -值小于 2 时，则不使用这个滤光片，在定标方程中把这个滤光片删除。

⑧ sum 值

sum 指除 C_0 以外，所用滤光片校准常数的和，用于判明校准曲线的稳定性。

⑨ $Index$ 值

$Index$ 值越高越好，数值为几表明数据被分成几类。如果 $Index$ 值为 2 时，表明所得到的校准曲线只能把样品区分为两类。但当 $Index$ 值为 2 时，对育种行业有用。

⑩ 数据删除量

参与定标的样品数据中总是有些异常的数据，这时需要删除某些异常数据，但必须有合理的理由。删除前最好重新检查被删除的样品，确认决定是否正确。

一般建议删除的数据量不要超过总数据量的 10%。对于异常的数据应进行重新的化学分析或扫描光谱，确认其确实是异常的数据再删除。若删除 10% 的数据后，对结果仍感到不满意，必须要寻找其他原因。但是，在实际应用中，一般删除的数据会多于 10%，这种情况也是正常的。

⑪ 滤光片组合数量

对于滤光片型的仪器，每个指标的定标都要选择几个滤光片的组合。除水分以外的大部分其他组分应选用多于 3 个滤光片的组合，所获得的结果会更好一些。所选滤光片应包含此成分的特征波长。如测检某一成分，必需至少包含一个对应该检测项目的响应滤光片在内，且 C 值为正值。为了决定最终选择哪一种滤光片组合，可用一组另选的测试样品来检查最好的 5 种滤光片组合，最后决定一组最佳的滤光片组合。

16.1.3.2 预测

预测是指考察定标方程用于测定未参与定标样品的准确性。预测所选样品应能代表被测样品的大致含量范围。用预测标准差 SEP 、变异系数、预测相关系数 R_p 和偏差对定标方程做最终评价。若预测样品的特性与定标样品相近，则预测效果好，否则效果差。

预测标准差和相关系数与定标值两者相差不应太大，才能说明定标是可行的。

$$SEP = \sqrt{\sum (Y_{NPi} - Y_{RPi})^2 / (n_p - 1)}$$

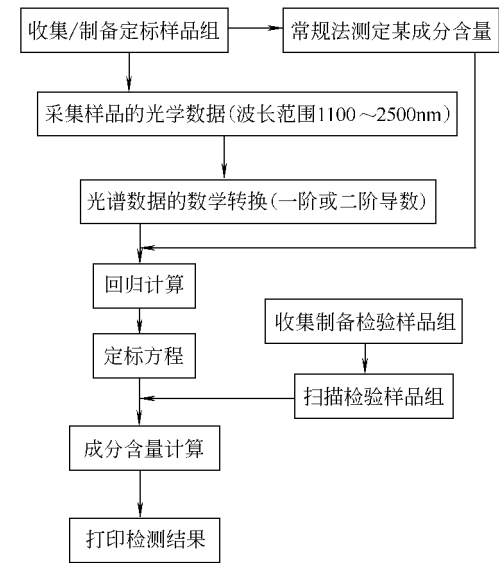


图 16-9 近红外光谱分析过程

首先就要按上面的方法开发定标方程。如果仪器中已有了定标方程，就需要对已有的定标方程进行校正。因为即使是同一公司同一型号的仪器也不是完全一样的，灯、滤光片和其他组件之间的差异必须逐个调整补偿。而且，参考实验分析方法也会同原始的做定标方程时的化学分析方法间有差异，可以按照自己的参考分析方法进行校准。样品也可能和原始的样品组之间存在一定的差异。单独对机器进行调整将有效提高近红外的检测结果。调整已有的定标方程远比建立校准曲线容易得多。

国际谷物科技协会推荐标准 No. 159 和 No. 202 认为调整已经存在的定标方程至少需要 30 个样品，一般不低于 20 个。样品制备方法应与建立曲线的样品制备方法一致。样品应有代表性，间隔尽可能均匀，能够涵盖所希望的检测范围。为找到合适的样品，可用已存在的校准进行测定或采用样品选择程序进行筛选。如进行大麦水分的测定，筛选完样品后就用化学法测定它的水分含量，然后用已有的定标方程测定它的水分含量，比较两者的差异，见表 16-3。

表 16-3 调整已经存在的定标方程

样品号	近红外结果	实验结果	实验与近红外差异	样品号	近红外结果	实验结果	实验与近红外差异
1	12.0	10.6	+1.4	11	14.7	14.1	+0.6
2	12.2	11.5	+0.7	12	15.0	14.0	+1.0
3	12.4	11.3	+1.1	13	15.2	14.2	+1.0
4	12.8	12.0	+0.8	14	15.5	14.3	+1.2
5	13.0	11.8	+1.2	15	15.7	14.9	+0.8
6	13.3	12.3	+1.0	16	15.9	14.8	+1.1
7	13.4	12.1	+1.3	17	16.2	15.5	+0.7
8	13.8	13.2	+0.6	18	16.5	15.1	+1.4
9	14.0	13.1	+0.9	...	16.6	15.7	+0.9
10	14.3	13.4	+0.9	30	17.0	16.2	+0.8

$$R_p = \sqrt{1 - SEP^2 / SD_{RP}^2}$$

其中： n_p 为被测样品数；

Y_{NPi} 为第 i 个预测样品某成分含量的近红外分析值；

Y_{RPi} 为第 i 个预测样品某成分含量的化学分析值；

SD_{RP} 为预测样品某成分化学分析值的分布标准差。

16.1.3.3 应用

如果预测的效果较好，就可把定标方程用于生产，进行实际监测。分析过程如图 16-9 所示。

16.1.4 已有定标方程的校正

如果购买的近红外仪器没有任何的定标方程，若想测定某一原料的某一成分的含量，首

从表 16-3 的结果可以算出，水分近红外值比化学法测值平均高 0.97，定标方程中的 C_0 值应下降 0.97。如果原来的 C_0 值是 13.6，则应调整为 12.63。

如果近红外值平均偏低 0.97，则 C_0 应相应增加，应变为 14.57。像这种只调整 C_0 ，叫作基准校正。如果高含量的样品被测得含量更高，低含量的样品被测得含量更低，那么，仅仅进行基准校正是不够的。这时，还应进行斜度调整——修正斜率系数。

16.2 影响饲料近红外测定准确性的因素

近红外分析技术尽管具有很多优越性，而且近红外在谷物、面粉的检测上取得了可信的结果，但谷物、面粉是相对单一的成分，变异不大，而饲料往往是由几种甚至几十种原料组成，原料和成品往往还经过不同的加工处理，其光谱的复杂性提高，因此其结果的准确性受到一定程度的影响。

近红外技术的原理是依据某一化学成分对近红外区光谱的吸收特性而进行的定量测定，所以应用 NIR 光谱进行检测的技术关键就是在两者之间建立一种定量的函数关系。确定回归模型的过程其实就是定标过程，定标的好坏直接关系到分析结果的准确性。因此，定标软件是近红外分析技术的核心。计算得到的定标方程必须通过实际测量来调整它的准确性和精确性。精确性是指重复测定时测值间的相近程度。准确性的度量通常用定标方程的预测标准误差（SEP）来表示。SEP 表示测定值与“真值”间的相近程度。

目前在饲料品质分析中常用的近红外光谱仪，为近红外反射光谱仪（简称 NIR）和近红外透射光谱仪（简称 NIT）。由于仪器结构不同，这两类分析法在被测样品的制备、分析灵敏度及适用范围方面存在差别。近红外光照射到被测样品后，从样品表面反射出来的光被检测器吸收，此为近红外反射光谱分析法（NIR）。NIR 分析法要求样品反射表面光滑一致，通常需对被测样品进行烘干、粉碎，并通过 1mm 筛网。另一类为近红外光穿过样品后，再被接受器检测到，即为近红外透射光谱分析法（NIT）。NIT 分析法无需制备样品，重复性较高。但该法的灵敏度较低，只适用于含量占干物质质量 1% 以上的成分分析，适用于在近红外区具强吸收的成分的测定。NIR 所用近红外辐射光谱带较 NIT 宽，因此 NIR 较 NIT 有较高的灵敏性，可用于测定样品中含量低的成分，被测成分含量占干物质质量 0.1% 以上的样品均可测定。由于受 NIT 分析灵敏度的影响，近 30 年 NIR 分析法的应用更为广泛。

近红外的分析准确性受很多因素的影响，主要表现在以下几个方面：

16.2.1 样品本身产生的误差

样品本身包括取样误差、样品不稳定和不均匀产生的误差，以及实验室分析产生的误差，或者测定对近红外没有响应的成分所产生的误差。

16.2.1.1 取样误差

如果用于定标的样品取样时不能代表原样品平均值，那么即使实验室值分析得再精确，也不能准确评定饲料的品质。在饲料分析过程中，取样技术不佳和次级样品数量不足是最大的误差来源。干草样品和青贮样品的取样要获得有代表性的样品就更加困难。

样品的粒度对 NIRS 光谱有影响, 供 NIRS 法测定用的样品须配有 1mm 或 0.8mm 的高速粉碎机, 粉碎机最好采用旋风磨, 因旋风磨提高了样品粒度的均匀性。采用 1mm 还是 0.8mm 的粉碎粒度可参考近红外厂家的建议, 原则是制作定标方程时和测定样品时采用的粒度是一致的。如果厂家推荐某一粉碎粒度, 它的大部分定标方程都是用此粉碎粒度开发的, 建议采用此粒度, 这有利于利用现有的定标方程及进行定标方程的转移。

16.2.1.2 样品贮藏误差

如果要重新定标或把样品提供给其他仪器定标, 则常需将样品贮藏较长一段时间。本质上这些样品的化学性质没有变化, 但由于样品含水量的变化或因厌氧或有氧微生物的活动, 样品仍有可能变质, 样品光谱也会产生相应变化。所以, 为阻止水分变化容器必须密封, 为限制微生物的活动, 样品的干物质应大于 88%。根据笔者研究的经验, 保存 1 年以上的样品再进行定标误差比较大。最好的办法是用新取来的样品, 用旋风磨粉碎后, 一边进行化学值的测定, 一边进行光谱扫描, 两者同时进行。

16.2.1.3 装样误差

近红外测定需把样品放在样品盒内或装在固定在仪器上的样品室内。盖子应和样品紧密地固定在盒子上, 盖子和盒子不应有空隙, 否则会造成杂散光进入, 应重新装样。如果盖子与盒子接触不能密封, 这时应更换盖子。样品盒玻璃片上的指纹和外来物都会造成测定误差, 常用骆驼发刷或其他吸尘器清洁样品盒。进一步清洁可用软薄纸或棉绒布擦。

准确取样至关重要, 先打开样品袋, 用样品勺小心地搅拌样品, 充分混合后, 从中随机取出样品, 装满样品盒或样品室, 盖上盖子, 压紧样品。每次装样压紧的程度应一致。如果出现反常现象应去掉盖子重新装样。

16.2.1.4 实验室分析误差

近红外测定的准确性取决于用于定标样品的测定的准确性和精确性, 因此实验室的分析准确性很重要。

16.2.2 仪器噪音产生的误差

仪器噪音主要因环境分长期噪音和短期噪音。长期噪音主要因环境温度变化和元件功能衰退而产生。温度变化 $10\sim 20^{\circ}\text{C}$ 就可引起读数的变化。所以, 近红外的测定一般要求在相当稳定的环境条件下工作, 一般建议温度在 $(25\pm 5)^{\circ}\text{C}$, 湿度在 $(60\pm 2)\%$ 下工作。仪器的预热很重要, 预热的时间与仪器的型号有关, 一般不能短于 15min, 最长需 1 个 h, 一般在 30min 左右。元件功能的衰退一般由参比样品石英窗、棱镜及其他光学元件上有灰尘, 或者由于运动部件机械损伤及电子元件功能衰退造成的, 这些不仅会产生误差可能还会导致仪器的全面故障。

短期噪音的一个主要的来源是波长不稳定, 另外一个十分常见的来源是检出器和输入放大器。仪器外面的噪音也可能是一个问题, 可能来源于刷子发动机、加热器开关以及能产生电脉冲的类似电流装置。其他的噪音来源还有杂散光、非线性光度和表面反射。这些噪音会产生非线性误差。

16.2.3 饲料近红外定标方程的准确性

要测定某一样品的某一成分前需要有这一成分的定标方程。一个成功定标方程的开发

需要的一个主要因素是选择合适的定标样品。NIRS 技术实际是一种经验性的技术，所以，NIRS 预测结果的准确性不可能超过仪器定标样品测定准确性。NIRS 是通过测定多个波长处的反射值来校正不同成分间的干扰，确定用于定标的基本样品及数量是非常重要的。Shenk 等（1976 年）认为：使用尽可能多的样品以便把所有影响饲料理化成分的因素包括在内。定标方程确定所需样品数目决定于方程的用途，即用于预测有限群体或无限群体。Shenk 等（1976 年）认为，在实际应用中，定标样品最低数量为 50 个，对大群体则选择总群体样品的 10%~20%。如果定标样品组能够代表预测样品群体，并选用了合适的数学处理方法，则预测值比较准确；如果定标方程用于分析较大且有多钟多样的变异群体，就需选择高的样品百分率，才可得到准确的预测结果。一旦建立了代表性较广的方程，那么样品随时可以增加到定标样品组中，重新定标以增加定标的适用范围。

除样品数量和选择样品方法外，用于定标的样品参比数据必须准确无误并能代表待测群体。预测动物反应的指标误差一般比预测化学成分的误差大，这主要是收集的动物反应数据有一定误差。简单地说，定标是一种数学技术来指令计算机依据 NIRS 光谱特性和化学成分含量之间的相关性正确地预测饲料品质的指标。因此，不能期望 NIRS 法能准确预测样品特性未包括在定标组内的任何样品或样品群体。

在实际使用过程中要获得准确的 NIRS 测定值，需要定期对 NIRS 法的测定结果进行校正。最明显的校正方法是应用参比方法分析样品，再用 NIRS 法测定，对两者之间的差异进行校正。另外，也可采取一些措施来确定化学分析结果的可靠性，并可反过来校正 NIRS 测值，一般可采用以下几个方法。第一，采用相同的分析方法，用经验分析方法有困难时，可根据不同分析者或实验室分析结果来校正。第二，采用公认标准的分析方法，可对实验室内和实验室之间的误差进行校正。第三，如果样品组内所有的样品都出现了问题，那么可用一组已知适合于定标的较好样品组来校正结果。

最后，使用一组参比样品。这些样品已通过常规分析法重复分析多次，因此其化学成分测定值很少或没有变化。此外，如果样品量大，实验室化学分析数据可能更准。参比样品作为较好的样品组用以校正 NIRS 法测值，这组样品需经常分析并保持相对稳定的测值，以便对仪器和定标的条件也做连续检验。

只有建立了一个基于大量，且具有代表性的样品和准确的参比分析值，并经适当的数学处理方法得到的定标方程，并且在使用过程中不断补充具有新理化特性的样品重新定标，用一组稳定参比样品不断进行校正，同时消除可能影响测定结果的各种因素，NIRS 法才能获得良好的预测结果。

16.3 近红外定标方程在饲料检测上的应用

现有的 NIRS 定标方程最有意义的应用是在购进饲料时通过对原料营养价值的监控帮助厂家进行决策，它可检出异常样品做进一步分析，并需对供应商和产品价格重新估价。应用 NIRS 一段时间后，当连续预测了几批饲料的营养价值，再结合实验室分析值，NIRS 数据可用于调整配方。

这一过程还可评价 NIRS 定标方程的价值。当试验室分析值肯定了 NIRS 的预测结果时，使用者就增加了对 NIRS 预测效果的信心。使用一段时间后，使用者会增加对 NIRS

技术的信赖，这样在试验室分析的样品会越来越少。

当一个饲料厂为某一变异大的原料准备许多料仓以使原料按质量分类贮存时，通过 NIRS 的预测就可判定某批原料属于哪个质量等级。用 NIRS 根据数据集中的样品对饲料进行分类。如肉骨粉根据其可消化赖氨酸含量分为三类，分别为 1%~2%，2%~3%，3%~4%。在这项试验中，41 个样品中有 30 个样品被正确归类，6 个样品“压线”（这种分类错误是因为它落在参数极限值之间）。另外有 5 个样品被错误分类。

如上所述，开发预测可消化氨基酸的 NIRS 定标方程的初步研究工作很成功。定标方程预测蛋白含量高的动物性饲料预测误差虽不完美但足以进行实际应用。

另外，也开发了预测玉米、豆粕、棉粕、菜粕氨基酸和可利用氨基酸的定标方程。定标的检验结果表明，用 NIRS 定标方程预测上述样品也可取得同样好的预测性能。然而，为了得到比较准确的预测值，需要增加样品数量使定标样本包含至少 60% 的变异率，使 NIRS 预测的范围更广和准确性更高。

NIRS 的另一个很重要的用途是进行原料掺假的鉴定。一张近红外光谱图如用一个点表示，按照化学计量学的方法，同一种类的产品，如不同厂家生产的人工牛黄（汤药、丸剂、复方本草药均可）可以在空间形成一球状或类似球状的区域（平面时为球形），在这一球状区域外的点为假产品，而在这一球状区域内的点代表的是真产品。由于疯牛病可以通过食物链传染给人类，因此对各类食品和饲料的检疫变得十分重要。Murray [2001 年] 等用 22 种标准鱼食和有意掺杂了牛肉和骨粉饲料（浓度 3%，6%，9%）的 45 种鱼食作标准建立了区分标准鱼食和掺杂了牛肉和骨粉的鱼食模型，该模型可以成功地识别鱼食中是否含有牛肉和骨粉掺杂物。

如上所述，近红外测定是经验型的，当测定时的条件和对象与建立校准回归方程时的条件和对象有所变化时，就会一定程度地影响测定结果。同一农产品由于品种、气候条件、生产年份、耕作条件不同，会在一定程度上影响该产品的化学组分或结构。将会引起测定过程红外光光密度的变化，影响测定结果。当这种变化不足以影响其测量精度时，原来的校准方程仍可适用，如果影响超过测量精度的要求，就要对原有的校准进行修正，改变 C_0 和 S 值，甚至要重新建立新的校准方程。

一般在出现下列情况时，就要对原定标方程常数进行校验和修正。

①改变样品粉碎方法，粉碎的粒度发生了变异；②检测的样品年份不同；③样品来自不同的国家和地区；④生产工艺条件明显改变。

使用者应不定期地对近红外仪的校准曲线方程进行校验，以确保测定结果的可靠和准确性。

16.4 进行饲料近红外分析的仪器类型

16.4.1 滤光片型仪器

滤光片型近红外测定仪由光学系统和电学系统两大部分组成，其光学系统由卤钨灯、透镜、滤光片、积分球、样品槽、红外探测器和参照基准板组成。卤钨灯发出红外光，由聚焦透镜聚焦成为平行光，经过滤光片形成一定波长的单色近红外光，照射到样品表面，

一部分红外光被样品吸收，另一部分反射进入积分球。吸收和反射的比例与红外光波长和样品中某种物质的含量有关。反射光进入内有镀金层的积分球后，反射达到红外探测器的表面，把红外光转换成电信号，由电路处理，进行软件运算。参照基准板是镀金的金属板，可以反射红外光，起到基准值的作用。测定样品前仪器先测定参照基准板的反射量，然后基准板移出，再测定被测样品的反射量。近红仪工作时，一个滤光片完成测量后，另一个滤光片进入光路中，这样继续下去，测出所有的波长的光密度值，经软件和电路运算处理，立即在显示屏上显示出各种成分的含量。

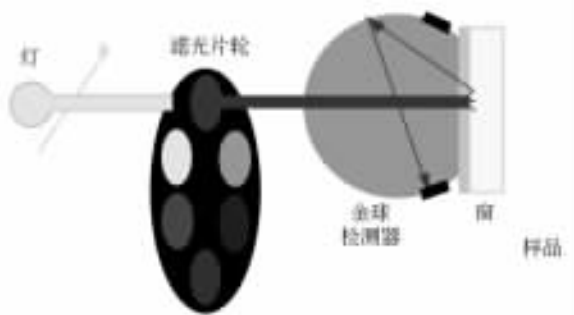


图 16-10 近红外测定仪光学系统原理图

16.4.2 扫描型仪器

以瑞典波通公司的 DA 7200 饲料分析仪（固定光栅连续光谱近红外分析仪）为例介绍一下扫描型仪器的特点。

DA 7200 是近红外连续光谱（950~1650nm）二极管阵列光谱分析仪，适用于快速无破坏分析液体、膏状、粉状和整粒样品的物理和化学特性，如脂肪、水分、蛋白质等。DA 7200 采用固定全息波长扫描和二极管阵列能量检测技术。光学系统完全固定。二极管阵列检测系统同时检测和并行处理所有波长信息，其检测速度和精度一般检测难以匹敌。DA 7200 系统设计可以使许多样品都能在初始状态下进行检测。样品在检测盘中完成检测。可配置不同的检测盘。分析检测面积大，对不均匀样品检测十分有利。分析不受环境影响，无接触检测。常见样品分析时间少于 5s。样品盘放置在检测平台，从 DA 7000 光学系统发出宽带白光照射样品盘。样品的漫反射光通过 DA 7000 光学窗口，进入感应系统。进入感应系统的光由固定衍射光栅进行分光。合成光谱落在波长间隔一样的光电二极管感应元件上。阵列中的每个二极管都有一个单独的同类信号通道，可以完成对光谱的动态“捕捉”。DA 7200 使用瑞典波通公司研究开发的 Simplicity™ 软件进行日常分析。校准曲线可以采用以下两个化学分析软件包——Unscrambler 或 Grams32 PLS+IQ 来建立。可选配标准校准曲线。校准曲线可在不同 DA 7000 类型仪器上传递。

16.5 近红外在饲料工业中的应用

近红外光谱分析技术最早应用于谷物的水分、蛋白质测定。在饲料分析方面，近红外光谱技术不仅能测定饲料中的常规成分，如水分、粗蛋白质、粗纤维、粗脂肪、粗灰分，而且能测定饲料中的微量成分，如氨基酸、维生素、有毒有害物质。现在近红外光谱技术已经向饲料营养价值评定方面发展，如测定可消化的氨基酸、消化能、代谢能、预测动物的反应（动物的采食量、消化率）等。从近红外的原理看饲料中需要测定的有机成分都可以用近红外进行测定。下面对 NIRS 在饲料工业中的研究和应用做一概述。

16.5.1 饲料水分、粗蛋白的测定

水分是饲料品质的一个很重要的指标，水分含量高意味着干物质少，价值就低。粗蛋白质是饲料原料和成品料的一个非常重要的指标，很多原料常以粗蛋白质和水分作为其评定等级的标准，因此在饲料行业中，不论是饲料原料还是成品料，粗蛋白质和水分的测定都非常重要。近红外在定量测定上的最早应用就是进行水分和粗蛋白质的测定。Norris (1976 年) 最早应用近红外反射光谱测定谷物的粗蛋白质、水分和脂肪的含量，随后 Law (1977 年)、Williams (1983 年)、Hartwig (1990 年) 等应用 NIR 法对谷实类、油料作物籽实的粗蛋白质、水分含量进行了测定并获得满意的结果。1978 年加拿大和美国的谷物协会已将 NIR 技术作为公定法被认可。中国也相继开展了近红外的研究，研究者们先后用近红外测定了谷实类、饼粕类、糠麸类、动物性蛋白质、饲料全价料、浓缩料的水分、粗蛋白质，获得了良好的结果，见表 16-4、表 16-5。从表 16-4 可看出，NIR 法估测饲料水分含量与化学法实测值间具有高度相关性，预测相关系数在 0.94~0.99；预测标准差为 0.10%~0.34%。变异系数为 0.9%~4.9%。从以上的结果可以认为，NIR 技术可用于饲料（饲料原料和成品料）水分含量的快速测定，其分析准确性基本上达到烘干减量法近似的水平。

表 16-4 NIR 技术测定饲料中水分含量的统计结果

饲 料	含量范围/%	N	R	SEP/SEC	CV/%
玉 米 ^①	9.8~18.4	68/41	0.990/0.989	0.20/0.16	1.7/1.4
玉 米 ^②	9.4~17.3	22/50	0.991/0.996	0.14/0.17	1.2/1.5
小 麦 ^①	10.2~14.6	20/58	0.954/0.978	0.33/0.25	2.6/2.0
高 粱 ^①	9.7~13.5	20/61	0.938/0.952	0.34/0.30	3.0/2.6
米 糠 ^①	3.5~7.6	20/40	0.982/0.990	0.25/0.18	4.9
麦 麸 ^②	8.9~13.6	47/76	0.994/0.998	0.10/0.09	0.9/0.8
豆 粕 ^①	8.7~13.7	71	0.991	0.17	1.5
豆 粕 ^②	7.3~12.7	49/89	0.995/0.998	0.09/0.09	0.9/0.9
棉籽粕 ^②	5.5~10.5	39/65	0.990/0.994	0.14/0.15	1.5/1.8
菜籽粕 ^②	6.4~11.4	40/65	0.972/0.987	0.15/0.18	1.7/2.0
鱼 粉 ^②	4.9~15.8	12/40	0.969/0.994	0.31/0.32	3.9/3.5

注：N 为样本数；R 为相关系数；SEP 为预测标准差；SEC 为定标标准差；CV 为变异系数。

在“/”左右两侧的 N 值及 R 值分别为检验与定标对应值。

①资料来自任鹏等；②资料来自丁丽敏。

从表 16-5 可看出，NIR 法可准确地测定谷实类饲料的粗蛋白质含量；对于饼粕类饲料、动物性蛋白质饲料的粗蛋白质含量的测定效果稍差，可达到粗估水平。从表 16-5 可以看出，丁丽敏研究结果表明，豆粕、棉粕、菜粕、麦麸、鱼粉、复合料的粗蛋白质 NIRS 测值与化学法测值的相关系数（R）都在 0.98 以上，而玉米为 0.881，最低；从标准差看，鱼粉最高，为 0.949，其他样品都小于 0.60；所有样品的变异系数（CV）都在 4% 以下；经 F 检验相关关系都达到了极显著水平（ $P<0.01$ ）。说明除玉米外其他样品的定标效果非常好，达到了实用程度。玉米的相关系数为 0.881，小于 0.90，这可能与玉米的粗蛋白含量低、实验室测定误差大有关，为提高预测精度应增多定标的样品数量，并提

高实验室的测定精度。但从变异系数（1.62）和 *F* 检验的结果（*P*＜0.01）看，也达到了实用程度。因蛋白质在 NIRS 光谱区内的吸收峰主要由 N—H 键的振动引起，对于不同来源的蛋白质以及相同来源但经不同处理的蛋白质，将引起 NIRS 吸收光谱的变化。大豆在取油过程中温度、压力、浸提方法、作用时间乃至水分的含量等都有可能直接或间接影响大豆粕的蛋白质分子结构，进而影响其在近红外光谱区的吸收。因此，饼粕的定标工作较谷实、油料籽实类自身的定标难度更大。为了提高 NIRS 法预测饼粕类饲料营养成分估测精度，应进一步根据不同加工条件将粕粕分类，分别定标。

表 16-5 NIR 技术测定饲料中粗蛋白质含量的统计结果

饲 料	Range/%	N	R	SEP/SEC	CV
玉 米 ^①	7.6～12.1	68/41	0.943/0.973	0.26/0.23	2.9/2.6
玉 米 ^②	7.7～8.7	19/39	0.854/0.881	0.10/0.13	1.2/1.6
小 麦 ^①	11.1～16.2	23/50	0.990/0.993	0.24/0.20	1.9/1.5
高 粱 ^①	15.0～20.1	18/24	0.986/0.997	0.34/0.18	1.9/1.4
米 糠 ^①	13.6～15.3	20/40	0.909/0.950	0.20/0.20	1.4/1.6
麦 麸 ^②	9.6～21.4	25/60	0.975/0.998	0.21/0.17	1.3/1.1
麦 麸 ^①	7.3～13.3	20/41	0.979/0.993	0.38/0.23	3.8/2.3
米糠饼 ^①	14.8～15.7	20/40	0.836/0.961	0.19/0.21	1.3
豆 粕 ^①	40.9～47.1	20/40	0.925/0.942	0.63/0.73	1.4/1.6
豆 粕 ^①	37.6～49.4	41/57	0.983/0.991	0.67/0.38	1.6/0.9
棉籽粕 ^②	33.1～40.2	52/61	0.963/0.986	0.47/0.32	1.3/0.9
鱼 粉 ^②	22.8～69.5	12/37	0.966/0.991	3.36/1.95	5.7/3.6
菜籽粕 ^②	33.13～39.84	52/61	0.963/0.986	0.470/0.323	1.26/1.62
浓缩料 ^②	13.31～50.89	11/40	0.983/0.998	0.398/0.582	3.81/2.98

注：Range 为含量范围；N 为样本数；R 为相关系数；SEP 为预测标准误差；SEC 为定标标准误差；CV 为变异系数。

在“/”左右两侧的 N 值及 R 值分别为检验与定标对应值。

①资料来自任鹏等；②资料来自丁丽敏等。

表 16-6 NIR 技术测定配合料粗蛋白质和水分含量的统计结果（鲜重/%）

	N	R	SEC	RSD	Range	Mean	CV	F
水分	37	0.986	0.348	0.333	4.74～14.51	10.57	3.29	393 * *
	12	0.982	0.198	0.206	9.00～12.47	10.15	1.95	217 * *
粗蛋白	40	0.998	0.582	0.551	13.31～50.89	19.51	2.98	2284 * *
	11	0.983	0.398	0.573	13.19～19.49	10.39	3.81	174 * *

注：N 为样本数；R 为相关系数；Range 为含量范围；SEC 为定标标准误差；CV 为变异系数；RSD 为残余标准误差；Mean 为平均值；F 为 *F* 检验的 *F* 值，*F* 值后一个 * 为显著水平（*P*＜0.05），两个 * * 为极显著水平（*P*＜0.01） 资料来源于丁丽敏。

从表 16-6 可以看出，近红外用于预测配合饲料的水分和粗蛋白质含量达到了可利用的程度，说明近红外用于终产品的品质控制是可行的。

16.5.2 粗纤维、粗灰分含量的测定

16.5.2.1 粗纤维的测定

从表 16-7 可以看出，麦麸和棉粕纤维定标和检验的相关系数都达到 0.90 以上，变异

系数小于 6%，达到可利用的程度。豆粕的纤维的定标结果不如麦麸和棉粕好，只能达到粗估水平。粗纤维本身不是一个固定的饲料实体，粗纤维的测定方法也是一个公认条件下的测定结果，本身的变异较大，所以近红外测定粗纤维的结果不如水分和粗蛋白精确。

表 16-7		饲料粗纤维近红外的定标方程及检验参数						
	<i>N</i>	<i>R</i>	<i>SEC</i>	<i>RSD</i>	Range	Mean	<i>CV</i>	<i>F</i>
豆粕	40	0.887	0.199	0.185	4.82~6.41	5.46	3.64	25 * *
	23	0.727	0.137	0.140	4.82~5.58	5.22	2.62	19 * *
麦麸	65	0.971	0.388	0.373	3.02~8.90	6.75	5.75	194 * *
	33	0.940	0.317	0.318	4.09~8.53	7.05	4.50	205 * *
棉粕	54	0.986	0.582	0.559	11.33~20.83	16.23	3.59	416 * *
	46	0.949	0.631	0.634	11.33~20.83	14.79	4.27	368 * *

注：表中各项含义同表 16-6 资料来源于丁丽敏（1997 年）。

16.5.2.2 粗灰分的测定

丁丽敏（1997 年）用 46 个豆粕样品、56 个玉米样品、63 个麦麸样品进行粗灰分的定标研究，定标结果如表 16-8 所示。

表 16-8		粗灰分的定标结果						
样品名称	<i>N</i>	<i>R</i>	<i>SEE</i>	<i>RSD</i>	Range	Mean	<i>CV</i>	<i>F</i> 值
豆粕	46	0.956	0.146	0.141	5.26~7.37	5.93	2.46	148 * *
棉粕	56	0.917	0.174	0.170	5.53~7.32	6.11	2.85	68 * *
玉米	56	0.890	0.073	0.070	1.01~1.72	1.23	5.93	49 * *
麦麸	63	0.987	0.205	0.199	2.93~7.77	4.96	4.13	567 * *

注：*N* 为定标的样品数；*R* 为相关系数；*SEE* 为标准误差；*RSD* 为残余标准误差；Range 为含量范围；Mean 为平均值；*CV* 为变异系数；*F* 为 *F* 检验的 *F* 值。

F 值后一个 * 为显著水平 ($P<0.05$)，两个 * * 为极显著水平 ($P<0.01$)。

从表 16-8 的结果可以看出，豆粕、玉米、麦麸的粗灰分 NIRS 测值与化学法测值的相关系数 (*R*) 分别为 0.956、0.890、0.987；残余标准差 (*RSD*) 分别为 0.141、0.070、0.199；变异系数 (*CV*) 分别为 2.46、5.93、4.13；*F* 检验的 *F* 值都达到了极显著水平 ($P<0.01$)。除玉米的相关系数为 0.890 以外，其他饲料的相关系数都达到 0.95 以上，说明定标效果非常好，达到了实用程度。玉米的相关系数为 0.89，小于 0.90，变异系数 (*CV*) 为 5.93，大于其他两种饲料，这可能与玉米中粗灰分含量低，实验室测定误差大有关，为提高预测精度应增多定标的样品数量，并提高实验室的测定精度。从变异系数和 *F* 检验的结果 ($P<0.01$) 看，玉米的粗灰分 NIRS 估测达到了粗估程度，但作为快速测定的方法是可行的。

用 30 个豆粕、32 个玉米、42 个麦麸对粗灰分的定标方程进行了检验，结果如表 16-9 所示。

表 16-9

粗灰分的定标结果检验

样品名称	<i>N</i>	<i>R</i>	<i>SEP</i>	<i>RMSD</i>	Range	Mean	<i>CV</i>	<i>F</i> 值
豆粕	30	0.903	0.13	0.133	5.28~6.44	5.75	2.26	115 * *
棉粕	36	0.850	0.099	0.106	5.54~6.54	5.95	1.66	81 * *
玉米	32	0.889	0.071	0.071	1.01~1.72	1.22	5.82	102 * *
麦麸	42	0.982	0.158	0.175	2.98~7.19	4.39	3.60	981 * *

注：*N* 为检验样品数；*R* 为 NIRS 估测值与化学值的相关系数；*SEP* 为估测标准误差；*RMSD* 为估测残余标准误差；Range 为含量范围；Mean 为平均值；*CV* 为变异系数；*F* 为 *F* 检验的 *F* 值。*F* 值后一个 * 为显著水平 ($P<0.05$)，两个 * * 为极显著水平 ($P<0.01$)。

资料来源于丁丽敏（1997 年）。

从表 16-9 的结果可以看出，豆粕、玉米、麦麸粗灰分 NIRS 测值与化学法测值的复相关系数 (*R*) 分别为 0.903、0.889、0.982；与定标的复相关系数相近 (0.956、0.890、0.987)；估测残余标准差 (*RMSD*) 分别为 0.133、0.071、0.175，与定标的残余标准差 (0.141、0.070、0.199) 相近；变异系数分别为 2.26、5.82、3.60，与定标的变异系数 (*CV*) 相近 (2.46、5.93、4.13)；*F* 检验都达到极显著水平。定标方程检验结果表明，豆粕、麦麸粗灰分定标效果都非常好，达到了实用程度，可用于实际监测。玉米粗灰分的定标变异系数大于 5%，估测精度不如豆粕、麦麸好，这与玉米本身的粗灰分的含量低有关，但 *F* 检验也达到了极显著水平，作为快速监测手段仍有其实用价值。按照近红外光谱产生的原理，无机成分在近红外光谱区并没有特征的吸收光谱，对其定标只能根据其与其他成分的相关来进行。所以，粗灰分的定标效果不如水分、粗蛋白质好。

16.5.3 饲料氨基酸含量测定

饲料工业需快速而准确地获得与饲料营养价值有关的数据。一方面可根据其营养价值对一宗饲料原料的合理价格进行谈判，另一方面可准确地将各种饲料原料配合在全价日粮中，在满足动物营养需要的情况下获得最低成本配方。所以，最好能快速地获取这些数据，使对饲料质量的常规监测得以实现。

目前，许多饲料厂家使用近红外光谱技术 (NIRS) 预测饲料中的粗蛋白质、水分、粗脂肪和粗灰分含量。然而，这些参数在应用时有很大的局限性，因其只反映了饲料真正营养价值的一小部分。所以，更好地评价饲料营养价值的参数是氨基酸、氨基酸利用率和净能。

Williams (1984 年) 应用近红外光谱技术测定了小麦及大麦的氨基酸含量，结果表明近红外光谱法预测小麦标准误差均小于 2.07%，变异系数除蛋氨酸外均小于 8%；大麦 14 种氨基酸中除蛋氨酸外，化学法测值与近红外测值间的相关系数为 0.93~0.99，变异系数均小于 6.5%。吴秀琴 (1985 年) 用近红外光谱法测定小麦中的赖氨酸含量，检验结果表明，化学法测值与近红外法测值间相关系数达 0.95，两法测值间相对误差最大不超过 5%。冯平 (1986 年) 进行了小麦麸的赖氨酸、精氨酸、苏氨酸、缬氨酸、亮氨酸、组氨酸的近红外的定标及检验。定标组的化学法测值与近红外法测值间的相关系数在 0.94~0.97，变异系数为 4.8%~7.0%。检验组化学法测值与近红外法测值间相关系数在 0.92 以上，两法间的相对偏差在 5% 以下，达到定量水平。陆国权等 (1993 年) 选用 40

个品种的甘薯进行 9 种氨基酸的近红外定标及检验。定标结果表明：9 种氨基酸的化学法测值与近红外法测值间的相关系数为 0.92~0.98，其中蛋氨酸为 0.84；9 种氨基酸的变异系数为 3.4%~11.4%，蛋氨酸的变异系数为 13.4%。检验结果表明：上述二值间的相关系数为 0.55~0.91；变异系数为 6.2%~22.6%。其中蛋氨酸的预测效果较差。杜荣等（1993 年）用 40 个裸大麦样品进行了 11 种氨基酸的定标及检验。定标结果表明：除色氨酸、胱氨酸外，其余 9 种氨基酸的近红外法测值与化学法测值间的相关系数均大于 0.95，变异系数均小于 6%。检验结果表明：上述两法测值的相关系数均在 0.88~0.99，预测标准误差在 0.01%~0.03%，变异系数为 2.25%~6.50%。丁丽敏（1997 年）应用近红外分析技术进行玉米氨基酸含量的定标及检验。除胱氨酸预测结果不能达到实用水平外，其他氨基酸均获得满意的预测结果。玉米赖氨酸的预测相关系数（*R*）为 0.745，预测标准差（*SEP*）为 0.02，变异系数（*CV*）为 6.7%；蛋氨酸的预测 *R* 为 0.909，*SEP* 为 0.01，*CV* 为 4.0%；蛋+胱氨酸 *R* 为 0.871，*SEP* 为 0.01，*CV* 为 3.2%；胱氨酸的预测 *R* 为 0.494，*SEP* 为 0.01，*CV* 为 3.5%。由此可见，近红外光谱分析技术可成功地用于谷实类饲料氨基酸含量的快速测定。除胱氨酸外，其他氨基酸的近红外法估测精度达到了化学法测值的要求。

表 16-10 3 种饼粕的主要必需氨基酸的近红外检验结果

	样本数	相关系数	预测标准误差	含量范围/%	变异系数/%
		<u>苏 氨 酸</u>			
豆粕	20	0.882	0.04	1.65~2.03	2.07
棉粕	31	0.920	0.04	0.68~1.42	3.46
菜粕		<u>未 测</u>			
		<u>赖 氨 酸</u>			
豆粕	16	0.893	0.02	2.62~2.78	0.77
棉粕	30	0.944	0.03	1.39~1.85	1.72
菜粕	30	0.903	0.114	1.52~2.16	6.30
		<u>蛋 氨 酸</u>			
豆粕	14	0.834	0.01	0.58~0.64	1.64
棉粕	28	0.849	0.03	0.33~0.64	4.58
菜粕	36	0.864	0.01	0.61~0.69	2.12
		<u>胱 氨 酸</u>			
豆粕	14	0.734	0.02	0.62~0.70	2.58
棉粕	27	0.631	0.03	0.32~0.59	5.71
菜粕	40	0.606	0.02	0.69~0.85	2.84

续表

	样本数	相关系数	预测标准误差	含量范围/%	变异系数/%
蛋+胱氨酸					
豆粕	15	0.912	0.03	1.07~1.33	2.64
棉粕	24	0.902	0.04	1.06~1.23	3.62
菜粕	34	0.832	0.02	1.36~1.52	1.35
色氨酸					
豆粕	14	0.801	0.02	0.47~0.60	3.27
棉粕	30	0.937	0.03	0.33~0.62	5.92
菜粕	22	0.760	0.01	0.41~0.46	2.09

色氨酸的 R 为 0.76), 变异系数小于 5% (棉粕色氨酸 CV 为 5.92%)。蛋+胱氨酸的近红外检验结果优于胱氨酸及蛋氨酸单独检验结果, 已达到了实用水平。对三种饼粕饲料而言, 豆粕的检验结果最好; 估测棉粕、菜粕的氨基酸含量效果稍差。丁丽敏等 (1997 年) 应用高压液相色谱法测定 50 个鱼粉的氨基酸含量, 并用 NIR 分析仪收集样品的光谱数据, 进行了鱼粉氨基酸含量的定标, 用未参加定标的样品对定标软件的精确度进行了检验。定标结果表明, 赖氨酸、蛋氨酸、胱氨酸、总氨基酸的相关系数分别为 0.965、0.876、0.939 和 0.940; 变异系数分别为 11%、22%、22%和 8%。检验结果表明, 赖氨酸、蛋氨酸、胱氨酸、总氨基酸的相关系数分别为 0.939、0.664、0.962 和 0.975; 变异系数分别为 11.6%、21.7%、22.1%和 7.5%。其中赖氨酸和总氨基酸的定标效果好于蛋氨酸、胱氨酸的定标, 达到了可利用的程度。蛋氨酸、胱氨酸的定标效果稍差, 可通过控制影响其精度的因素进一步提高其预测精度。

应用 NIR 分析技术预测饲料中的氨基酸含量较传统化学法更快、且能节省昂贵的试剂费用。在畜禽日粮配制中较常用的查饲料成分表获取原料的氨基酸含量的方法更准确, 因此具有很强的实用性。

对于蛋白质含量高的饲料, 在进行近红外定标时, 除采用专一定标样品 (只选同一种待测饲料) 进行定标外, 国外亦进行通用定标 (包含所有饲料的定标) 及产品群定标 (某一类饲料的定标)。表 16-11 列出了 Jackson (1996 年) 采用上述三种方法进行的肉骨粉的定标及检验结果。

表 16-11 肉骨粉三种氨基酸的通用定标、产品群定标和专一产品定标方程
的预测标准误 (RMSEP) 及预测相关系数 r_{rmsep}^2

	赖氨酸	蛋氨酸	苏氨酸
样 品			
平均值	2.85	0.85	1.87
标准差	0.74	0.26	0.42
	通用定标方程		
RMSEP	0.23	0.14	0.12
r_{rmsep}^2	0.90	0.69	0.92

续表

	赖氨酸	蛋氨酸	苏氨酸
	<u>动物性产品群定标方程</u>		
<i>RMSEP</i>	0.26	0.14	未测
r_{rmsep}^2	0.87	0.69	未测
	<u>肉骨粉定标方程</u>		
<i>RMSEP</i>	0.32	0.14	0.13
r_{rmsep}^2	0.81	0.72	0.90

从三种定标方程的检验指标——预测相关系数（ r_{rmsep}^2 ）及预测标准误差（*RMSEP*）来看，通用定标方程的预测效果优于或相当于后两种定标方程。其他动物性饲料如鱼粉及家禽副产品也表现类似结果；但对低蛋白质含量的饲料如谷物和油籽粕，专一定标方程的预测结果优于产品群及通用定标的预测效果（数据未列出）。Jackson（1996 年）在这三种定标方程的检验时，所用方法不同于通常采用未参与定标的样品组成检验样品群，再进行检验。而是将数据集中的一个样品删除，然后用数据集中的其他样品作定标来预测被删除的样品。对数据集中的所有样品都重复此过程，最后求出预测标准误差 *RMSEP* 及预测相关系数 r_{rmsep}^2 。但从实用角度考虑，应使用未参与定标的样品组成样品群进行预测，这样才能检验定标方程的估测准确性。在这种情况下，专一样品定标的检验效果则优于产品群及通用定标方程的检验结果。

16.5.4 进行饲料可消化氨基酸含量的测定

16.5.4.1 近红外预测可消化氨基酸含量

了解饲料原料的氨基酸含量，特别是可消化氨基酸含量，对合理利用饲料资源、降低饲养成本、满足家禽营养需要，具有重大意义。但由于饲料氨基酸消化率测定方法繁琐、费时、耗资，一般的饲料生产者难以做到在饲料配制前实测原料的可消化氨基酸含量，而常用经验数据或书本值作参考。而近红外光谱技术测定快速、相对准确，为测试原料的可消化氨基酸含量提供了可能。

NIRS 预测动物的消化率的准确性有赖于下面两个假定：试验内和实验间所有随机变异来源都保持在最低限度值，定标和检验样品群体使用相同的测定方法。在测定消化率数据中存在的任何偏差都将产生预测消化率的偏差。

魏瑞兰（1987 年）用近红外法进行了花生饼鸡可利用氨基酸的定标，结果表明生物学法测值与近红外法测值间呈强相关，相关系数为 0.85~0.98；估测变异系数为 3.2%~9.9%。除赖氨酸外，其他氨基酸在近红外法值与生物学法测值间十分相近。丁丽敏等（1997 年）进行了豆粕、棉粕、菜粕的可消化氨基酸的近红外的定标和检验（表 16-12、表 16-13、表 16-14）。除豆粕的真可消化胱氨酸、棉粕及菜粕的真可消化蛋氨酸和胱氨酸外，其他氨基酸的定标方程均具有良好的预测性能。然而用总含硫氨基酸（蛋+胱氨酸）进行定标及检验，效果好于单个含硫氨基酸。豆粕的蛋+胱氨酸的定标及检验的相关系数均大于 0.90，标准误差 0.05% 以下，变异系数小于 4%，可用于实际检测。

表 16-12		豆粕鸡真可消化氨基酸的近红外定标及检验结果			
氨基酸	样本数	相关系数	标准误差	范 围	变异系数
		<u>定 标 结 果</u>		1.21~2.08 1.71~2.99 0.47~0.63 0.41~0.72 0.88~1.36 0.21~0.61	4.97 4.02 4.04 6.14 4.00 7.89
苏氨酸	40	0.884	0.09		
赖氨酸	45	0.933	0.10		
蛋氨酸	40	0.845	0.02		
胱氨酸	41	0.883	0.04		
蛋+胱氨酸	50	0.906	0.05		
色氨酸	39	0.877	0.04		
		<u>检 验 结 果</u>		1.61~1.97 2.35~2.93 0.54~0.64 0.54~0.66 0.88~1.29 0.43~0.58	3.24 3.10 2.59 4.75 3.84 4.51
苏氨酸	16	0.849	0.06		
赖氨酸	15	0.903	0.08		
蛋氨酸	11	0.874	0.02		
胱氨酸	11	0.679	0.03		
蛋+胱氨酸	20	0.950	0.04		
色氨酸	12	0.886	0.02		

表 16-13		棉粕鸡真可消化氨基酸的近红外定标及检验结果			
氨基酸	样本数	相关系数	标准误差	范 围	变异系数
		<u>定 标 结 果</u>		0.55~1.10 0.76~1.13 0.42~0.50 0.23~0.39 0.51~0.89 0.23~0.49	6.33 3.58 4.78 13.21 7.54 10.32
苏氨酸	65	0.924	0.05		
赖氨酸	65	0.919	0.03		
蛋氨酸	62	0.812	0.02		
胱氨酸	67	0.648	0.04		
蛋+胱氨酸	58	0.852	0.06		
色氨酸	64	0.921	0.03		
		<u>检 验 结 果</u>		0.55~0.96 0.76~1.13 0.42~0.50 0.23~0.41 0.51~0.86 0.24~0.47	4.12 2.99 3.48 15.00 3.87 10.29
苏氨酸	33	0.826	0.03		
赖氨酸	26	0.874	0.03		
蛋氨酸	29	0.668	0.02		
胱氨酸	29	0.400	0.04		
蛋+胱氨酸	30	0.854	0.03		
色氨酸	24	0.912	0.04		

表 16-14		菜粕鸡真可消化氨基酸的近红外的定标及检验结果			
氨基酸	样本数	相关系数	标准误差	范 围	变异系数
		<u>定 标 结 果</u>		0.87~1.78 0.53~0.61 0.42~0.69 0.97~1.24 0.33~0.39	14.28 2.11 6.96 3.10 2.00
赖氨酸	68	0.847	0.20		
蛋氨酸	71	0.808	0.01		
胱氨酸	80	0.876	0.04		
蛋+胱氨酸	63	0.909	0.04		
色氨酸	58	0.892	0.01		
		<u>检 验 结 果</u>		0.87~1.79 0.53~0.59 0.42~0.66 0.97~1.24 0.33~0.39	11.64 2.32 6.67 3.36 3.33
赖氨酸	38	0.922	0.15		
蛋氨酸	37	0.680	0.01		
胱氨酸	32	0.821	0.04		
蛋+胱氨酸	32	0.817	0.04		
色氨酸	26	0.629	0.01		

另一种方法是先测定一批饲料的蛋白含量，然后根据回归方程预测氨基酸含量。这些回归方程的准确度变化很大。对玉米和小麦来说，统计结果表明该相关无效（ $r^2_{rmsep}=0$ ），对肉骨粉样品相关预测的变异率较大（ $r^2_{rmsep}=0.59$ ）。其他产品的预测效果介于两者之间。这说明即使是最好的回归方程也较 NIRS 的预测效果差。

因此，作为一个快速而有效的方法，用 NIRS 测定饲料中总氨基酸和可消化氨基酸的效果优于其他方法（表 16-16）。

在评价定标方程的效果时，考虑定标样品数据的准确性非常重要。因为定标方程的研制是根据光谱值和试验室值（可能含有误差）的相关产生的。

首先是测定饲料中可消化氨基酸时由分析饲料氨基酸产生的误差，第二步是氨基酸消化率体内试验部分产生的误差。用高压液相色谱分析氨基酸，由这一技术引起的误差一般占±5%。通过体内实验测定消化率，由动物引起的误差和真可消化氨基酸分析误差在总误差中增加了另外 2% 的误差。可消化氨基酸的分析总误差约为 7%，而对于肉骨粉，可消化赖氨酸误差为 0.2%。

表 16-16 肉骨粉和鱼粉通用定标方程的定标和预测性能参数
(带有“P”的氨基酸表示禽可消化氨基酸)

	平均值	标准偏差	RMSEP	r^2
肉骨粉				
赖氨酸	2.85	0.74	0.25	0.90
赖氨酸 ^P	2.21	0.62	0.30	0.77
蛋氨酸	0.85	0.26	0.14	0.72
蛋氨酸 ^P	0.70	0.21	0.13	0.63
苏氨酸	1.87	0.42	0.12	0.92
苏氨酸 ^P	1.42	0.35	0.14	0.83
鱼粉				
赖氨酸	4.88	0.66	0.22	0.89
赖氨酸 ^P	4.47	0.75	0.26	0.87
蛋氨酸	1.76	0.29	0.17	0.61
蛋氨酸 ^P	1.65	0.29	0.17	0.65
苏氨酸	2.74	0.36	0.16	0.80
苏氨酸 ^P	2.49	0.44	0.18	0.83

资料来源于 Theo Van Kempen 等（1996 年）

表 16-17 NIRS 预测动物性饲料可消化氨基酸和其他快速方法预测可消化氨基酸精度的比较

书面值			N 回归	NIRS	体内法测定值
			可消化赖氨酸		
肉骨粉	RMSEP	0.62	0.37	0.30	0.20
	r^2_{rmsep}	0.00	0.59	0.77	
鱼粉	RMSEP	0.75	0.60	0.26	0.34
	r^2_{rmsep}	0.00	0.41	0.87	
			可消化蛋氨酸		
肉骨粉	RMSEP	0.21	0.14	0.13	0.06
	r^2_{rmsep}	0.00	0.57	0.63	

		书面值	N 回归	NIRS	体内法测定值
鱼粉	RMSEP	0.29	0.16	0.17	0.12
	r_{rmsep}^2	0.00	0.74	0.65	
可消化苏氨酸					
肉骨粉	RMSEP	0.35	0.16	0.14	0.13
	r_{rmsep}^2	0.00	0.77	0.84	
鱼粉	RMSEP	0.44	0.30	0.18	0.19
	r_{rmsep}^2	0.00	0.52	0.83	

注：体内测定值仅作为对照，因为该方法不是快速方法。

资料来源于 Theo Van Kempen 等（1996 年）

由（表 16-17）数据可知，NIRS 产生的预测误差，接近体内试验方法误差或是其 2 倍。这一误差在 NIRS 技术的允许范围之内。可以认为这一初步通用定标方程是可行的。

丁丽敏（1997 年）的研究表明，豆粕 NIRS 预测氨基酸含量方程与以粗蛋白为自变量预测方程相比，相关系数远高于后者。NIRS 的 R 值一般在 0.90 以上，最小的为 0.823，而以粗蛋白为自变量预测氨基酸的方程的 R 值在 0.40~0.70 之间，远小于 NIRS 的 R 值。并且，NIRS 的预测标准误差是以蛋白为自变量预测氨基酸标准误差的 1/2~1/6。所以，NIRS 的预测性能好于以粗蛋白为自变量的预测方程的预测性能，且它具有快速、准确、经济和操作简单的特点，在预测氨基酸含量方面具有其他方法无可比拟的优势。

16.5.5 近红外进行饲料有效能值的测定

国内外不少学者尝试用近红外光谱技术测定饲料有效能值。杜荣（1993 年）用近红外技术进行了裸大麦鸡表现代谢能（ AME ）和猪消化能（ DE ）的定标，相关系数分别为 0.863、0.897；残余标准误差（ RSD ）为 60kcal/kg；变异系数为 2.0%和 1.9%，达到粗估水平。陆国权（1993 年）用近红外光谱技术进行了甘薯干鸡代谢能和猪消化能的估测，相关系数分别为 0.912 和 0.694，残余标准误差为 60kcal/kg 和 100kcal/kg；变异系数为 2.1%和 3.0%。Leeson（1997 年）用近红外法估测了玉米、豆饼、小麦、大麦以及配合饲料的鸡氮校正表现代谢能（ $AMEn$ ）。近红外分析结果与实测值间相差 10~50kcal/kg，此结果完全可作为饲料配制的数据。而配合饲料计算的 $AMEn$ 与近红外法估测的 $AMEn$ 比较，两者相差最大为 200kcal/kg（表 16-18）。由此可见，近红外光谱分析技术在饲料原料的有效能值的估测上有着广阔的应用前景，其估测准确性较应用常规成分估测模型估测出的能值更准确，可应用于日粮的配制。

表 16-18 饲料原料及配合饲料的近红外法估测值与实测值（计算值）的比较 单位：kcal/kg

原 料	实测值	NIR 值	原 料	实测值	NIR 值
玉米 ^{#1}	3.38	3.37	肉仔鸡中期	3.30	3.28
玉米 ^{#2}	3.45	3.43	肉仔鸡后期	3.40	3.37
豆饼	2.34	2.32	蛋鸡	2.70	2.90
小麦	3.27	3.23	育成鸡	2.65	2.86
大麦	2.72	2.67	种鸡	2.87	2.92
全价料	计算值	NIR 值	种鸡	2.70	2.72
肉仔鸡前期	3.20	3.32			

16.5.6 近红外进行饲料有毒有害成分的测定

应用近红外光谱技术可测定饲料中某些抗营养因子及有毒有害成分。Birth 等（1982 年）应用该技术测定了棉籽中的棉酚含量，检验结果表明：近红外法测值与化学法测值间相关系数为 0.99，预测标准误差为 0.07。Biston 等（1988 年）应用近红外法分析了菜籽中总葡萄糖苷的含量，以高压液相色谱法测值作为化学法测值，经检验近红外法值与化学法测值间的相关系数为 0.99，预测标准误差为 $2.15\mu\text{mol/g}$ 。Parrish 等（1990 年）应用近红外法及色谱法测定了棉籽中的植酸磷含量，检验结果表明，两法测值间的相关系数为 0.88，预测标准误差 0.12%。吴秀琴等（1993 年）、赵宏等（1993 年）用近红外分析法和铁沉淀法就小麦麸、玉米、米糠、菜籽饼、高粱、米糠饼的植酸磷含量进行了比较测定，近红外法测定小麦麸和高粱的植酸磷结果较为理想，两法相关系数分别为 0.93 和 0.92；预测标准误差为 0.05 和 0.02。但测定玉米、菜籽饼的植酸磷时，两法间存在较大误差。米糠、米糠饼的植酸磷测定准确性居中。

16.5.7 近红外进行饲料矿物元素含量的测定

饲料中矿物质，如 Ca、Mg、P 等在近红外区无吸收。因此应用该技术测定矿物元素含量似乎不可能。但 Shenk（1979 年）报道了近红外光谱法测定干草中钙、磷、镁含量的结果。Clark（1989 年）应用近红外光谱法测定了牧草中钙、磷、钾、镁等含量，测定结果变异范围在 10%~35%。Valdes（1985 年）应用近红外光谱法测定了干草、半干青贮、青贮饲料的钙、磷、钾、镁的含量。预测结果表明：近红外法与化学法测值间的相关系数为 0.02~0.75，变异系数 10.5%~34.5%。应用近红外光谱技术测定饲料矿物元素含量未见理想报道。由于近红外法测定矿物质含量是基于这些矿物质含量与饲料其他可测成分间具有一定相关关系，因此近红外法在饲料矿物元素含量测定的准确性方面欠佳，还无法应用于实际。

16.5.8 近红外进行粗饲料品质的测定

Norris（1976 年）首先提出用近红外光谱技术测定粗饲料的品质。此后国外许多学者进行过此类研究。Valdes（1985 年）应用 NIR 分析法测定了干草、半干青贮、青贮饲料的粗蛋白质、酸性洗涤纤维，预测结果表明：NIR 法与化学法测值间的相关系数粗蛋白质为 0.72~0.96，酸性洗涤纤维为 0.21~0.92；预测标准误差粗蛋白质为 0.43~1.17，酸性洗涤纤维为 1.44~2.53；变异系数粗蛋白质为 5.6%~10.4%，酸性洗涤纤维为 5.3%~7.9%。另外，近红外光谱分析法还可用于牧草的营养价值及采食量等的测定。Given（1991 年）应用近红外光谱分析法及三种体外法测定了谷实秸秆及氨化秸秆的有机物消化率，并将这些数据与体内法所测有机物消化率值进行回归，其中近红外分析法估测值与体内法测值相关最好，相关系数为 0.65，预测标准误差为 0.04。

近红外光谱分析技术在粗饲料品质分析上应用的另一分支是青贮的品质分析。青贮发酵终产物如乳酸、挥发性脂肪酸、氨态氮是青贮发酵后品质检测的重要指标，这些指标与青贮的营养价值有密切的联系，如有机物采食量。通过发酵终产物来预测有机物采食量仍有一定困难，主要是终产物的化学法测定费时且费用高。Reeves（1989 年）应用来自 3 个地区的首蓿、玉米及两种青贮混合定标、检验。用近红外光谱分析法估测湿样的某些指

标, 如干物质、粗蛋白质、酸性洗涤纤维、中性洗涤纤维、pH、氨态氮、体外干物质消化率、乙酸、丙酸、丁酸、异丁酸、异戊酸、乳酸含量。所用 146 个样品 2/3 用于定标, 1/3 用于检验。结果表明湿苜蓿青贮主要成分的近红外光谱分析值与化学法测值间的相关系数为 0.78~0.99, 湿玉米青贮的相关系数为 0.75~0.96。但对大多数有机酸而言, 上述两法间的相关系数很低。这主要是由于近红外光谱分析技术分析湿青贮的化学成分存在着许多干扰因素, 如湿样难以取到有代表性的样品, 高水分对近红外光谱有干扰作用, 干燥处理则会使挥发酸测定受到影响。Synman (1992 年) 通过化学处理使短链有机酸转变为相应的盐, 再干燥粉碎, 预测结果表明: 近红外分析法与化学法测值间相关系数分别为乳酸 0.85, 总挥发性脂肪酸 0.86, 乙酸 0.92, 丁酸 0.65, 氨态氮 0.855; 预测标准误差分别为 21.1、18.7、6.5、5.8 和 1.5g/kgDM。Dyer (1997 年) 应用化学法及近红外法对比测定了玉米青贮的粗蛋白质、中性洗涤纤维、酸性洗涤纤维含量。7 个实验室用化学法, 4 个实验室用近红外法。共分析 8 个样品, 每样品重复测定 3 次。计算每一实验室重复测值间的标准误差, 结果表明, 化学法重复样品测值间的标准误差是近红外法的 2 倍; 近红外法实验室间测值的变异与传统分析法的变异相近或更小。近红外法具有较高的精确性和重复性。

国内应用近红外分析法估测粗饲料的化学成分还较少报道。陆国权 (1993 年) 测定过甘薯茎叶粉的化学成分; 许彩萍等 (1993 年) 测定了苜蓿草粉的化学成分。分别用 60 个苜蓿样品定标、25 个样品检验, 进行了苜蓿草粉的粗蛋白质、粗纤维、粗灰分的测定。预测结果表明: 近红外法与化学法测值间的相关系数分别为 0.99、0.99、0.87; 预测标准误差分别为 0.43%、0.57%、0.53%; 变异系数分别为 2.3%、2.2%和 5.7%。作者认为近红外技术可快速、简便地估测来自不同地区、不同品种、不同成熟期的苜蓿草粉的粗蛋白质含量, 对粗纤维及灰分的估测效果稍差。赵环环等 (2001 年) 选取不同品种、不同成熟期、不同地块的黑麦草样品 65 份, 用 NIR 分析技术测定粗蛋白质含量。选取 30 个样品进行定标, NIR 法预测值与凯氏定氮法得到的测定值间的复相关系数 $R^2 = 0.99$, 定标准误差均方 $RMSEC = 0.34\%$ 。用 21 个样品作检验样品集, 凯氏定氮法得到的测定值与 NIR 预测值间的复相关系数为 $R^2 = 0.98$, 预测标准误差均方 $RMSEP = 0.42\%$ 。表明 NIR 技术完全可以达到快速测定不同品种、不同成熟期、不同地块黑麦草的粗蛋白质含量, 在实际应用中有广阔的前景。

近红外光谱分析技术测定牧草的化学成分的报道很多。从该法预测标准误差的范围来看, 粗蛋白质约 0.32%~1.15%; 酸性洗涤纤维约 1.00%~2.46%; 中性洗涤纤维约 1.24%~3.50%; 木质素约 0.30%~1.13%; 体外干物质 (有机物) 消化率 1.47%~4.10%。该法分析的准确性主要受所用仪器、采用的定标方法、粗饲料品种等因素影响。此外牧草的成熟期、刈割方法、加工贮存方法等因素也有影响。

16.6 近红外光谱技术在饲料工业中应用的发展趋势和展望

NIRS 技术有广阔的应用前景。测定成分由测定水分、粗蛋白质、脂肪、糖、淀粉、纤维素等成分发展到测定烟草中的尼古丁、咖啡碱及饲料中的某些有害成分, 如棉酚、总葡萄苷、植酸磷以及氨基酸、可利用氨基酸和有效能的测定。测试准确性来看, 在测定饲

料原料水分、大多数谷物及牧草的蛋白质含量、动物性副产品、豆饼及大多数饲料原料、脂肪含量低于 5% 饲料的脂肪及油分方面，均可达到较好的预测效果。对于动物性副产品，鱼粉或含较高动物性副产品及鱼粉的配合饲料的蛋白质的测定，准确性稍差。对饲料的粗灰分、粗饲料中的矿物元素的测定，虽然 NIRS 法与化学法存在一定相关，但仍不能达到实用程度。

除此之外，NIRS 还用于产品质量及原料的在线分析。

近红光谱分析技术虽是具有快速、简便、相对准确等优点，但该法估测准确性受多种因素的影响。其中以样品的粒度及均匀度影响最大，粒度变异直接影响近红外光谱的变异。定标样品选择对近红外分析估测准确性影响也很大。标样选择时不仅要考虑样品成分含量和梯度，同时应考虑标样的物理、化学及植物学特性，使定标软件具有广泛的应用范围。对于开放性定标方程应不断增加标样，使定标样品群尽可能包容各方面的变异特性，不断完善定标软件。特别是作为近红外光谱分析的基础方法——化学法的准确性也是影响其估测准确性的重要因素。

尽管近红外光谱分析技术在估测准确性方面存在一些问题，由于其快速、简便、相对准确等特点，应用范围越来越广泛，已被应用于农业、饲料及食品等行业。因此，近红外光谱分析技术在饲料质量监测方面有广阔的应用前景，该法的出现为现场快速监测饲料质量提供了可能。除此之外，近红外光谱分析技术还可用于产品质量及原料的生产线抽样 (on-line) 分析，通过监控饼粕类饲料加工过程壳的数量，保证饼粕的各项指标要求。美国明尼苏达州利用可移动近红外光谱分析实验车（大篷车）到田间或交易市场为用户分析牧草样品，提供适时收割的日期，提供饲料配方，在提高饲料、牧草的可消化性及其营养价值方面收到良好的效果。

1992 年 9 月，美国国家牧草协作组已同美国农业部及 Nebraska 大学合作，建立直播牧草的近红外光谱分析定标网络，该网络是中心实验室应用近红外光谱仪，选择数百个样品在近红外光谱区挑选最适宜波长制定定标方程，这些定标方程可通过电话线传送到边远农场设置的近红外光谱仪，农民就可利用该定标方程分析牧草的品质，同时农民测试结果又反过来通过网络输送到中心实验室，并可使定标软件进行修正。因此客户可定期地从中心实验室获得新版本的定标软件，用于田间牧草品质分析。

我国近年来进口了一大批各种型号的近红外光谱仪，“七五”期间还研制成功了用于农产品分析的近红外光谱分析。如前所述，软件是影响测值准确性的关键，因此有必要建立一套适用于我国国情的各种饲料原料及产品的定标软件，充分发挥已有的硬件的作用。

对于已购置近外仪器的饲料加工厂，应立足本场，在进行产品、原料质量监测的同时，不断收集样品，逐步建立适宜于本场的常用原料及配合饲料的定标软件，使该仪器充分发挥其在质量监测中的快速、简便、准确的优势。

参 考 文 献

- [1] 丁丽敏. 近红外光谱技术预测饲料中氨基酸和真可消化氨基酸含量的研究. [博士论文]. 北京: 中国农业大学, 1997
- [2] 杜荣等. 近红外光谱分析技术在饲料营养价值评定上的应用——饲料用裸大麦. 见: 中国农业科

- 学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 178~181
- [3] 杜荣等. 近红外光谱分析技术在氨基酸含量测定上的应用——饲料用裸大麦部分. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 196~203
- [4] 冯平. 近红外光谱分析技术在麦麸的饲料营养价值评定上的应用. 中国畜牧杂志, 1986 (2): 10~14
- [5] 陆国权等. 近红外光谱分析技术在粗饲料质量监测上的应用——饲料用甘薯茎叶粉部分. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 120~124
- [6] 陆国权等. 近红外光谱分析技术在饲料氨基酸含量测定上的应用——饲料用甘薯干部分. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 216~226
- [7] 陆国权等. 近红外光谱分析技术在饲料营养价值评定上的应用——饲料用甘薯干. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 182-186
- [8] 任鹏, 陈雪秀. 近红外光谱分析技术在饲料成分及营养价值评定上的研究进展. 动物营养研究进展, 1994: 222~238
- [9] 魏瑞兰, 张子仪. 近红外光谱分析技术在花生饼(粕)可利用氨基酸评定中的应用. 中国畜牧杂志, 1987 (6): 3~7
- [10] 吴秀琴. 应用 51A 型 NIRS 测定小麦种子中的赖氨酸含量. 农业测试分析, 1985, 2 (1): 15
- [11] 许彩萍, 任鹏, 陈雪秀等. 近红外光谱分析技术在粗饲料质量监测上的应用. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 112~119
- [12] 赵宏等. 近红外光谱分析技术在饲料植酸磷监测技术上的应用——饲料用小麦麸、米糠、玉米和菜籽饼. 见: 中国农业科学院畜牧研究所编. 近红外光谱分析技术. 北京: 中国农业科技出版社, 1993. 236~241
- [13] 赵环环. 近红外光谱分析技术在黑麦草粉粗蛋白测定中的应用. 动物营养学报, 2001, 13 (4): 40~43
- [14] Birth G S, H H Ramet Jr. Near infrared reflectance for analysis of cotton seed for gossypol. Cereal Chem, 1982, 59 (6): 516~519
- [15] Biston R, et al. Fast analysis of rapeseed glucosinolates by near infrared reflectance Spectroscopy. JAOCS, 1988. 65 (10): 1599~1600
- [16] Clark D H, et al. Analysis of trace elements in forage by near infrared reflectance spectroscopy. AgronJ, 1989 (81): 91~95
- [17] Daniel J D, Ping Feng. NIR destined to be major analytical influence. Feedstuffs, 1997 (10), 16~17
- [18] Given D I, C W Baker, A R Moss et al. A comparison of near infrared reflectance spectroscopy with three in vitro techniques to predict the digestibility in vivo of untreated and ammonia-treated cereal straws. Animal Feed Science and Technology, 1991 (35): 83~94
- [19] Hartwig P A, P H Charles Jr. Near Infrared Reflectance Measurement of moisture, protein and oil content of ground crambe seed. JAOCS, 1990 (67): P7
- [20] Jackson D A, et al. Determination of total and digestible amino acids by near infrared reflectance spectroscopy. Proc Aust Poult Sci Sym, 1996 (8): 46~52
- [21] Law D P. Determination moisture content in wheat by NIR diffuse reflectance spectrophotometry. Cereal chem, 1977, 54 (4): 874~881
- [22] Leeson S. Potential for real-time ingredient quality control procedures. J Appl Poultry Res, 1997 (6): 501~506
- [23] Murray I, Aucott L S, Pike I H. Use of Discriminant analysis on visible and near infrared reflectance spectra to detect adulteration of fish meal with meat and bone meal [J]. J. of near infrared spectroscopy, 2001, 9 (4): 297~311

17. 近红外光谱技术在食品分析中的应用

近红外在食品工业上的应用非常广泛。利用近红外光谱技术可以进行食品成分的定量分析、水分子中氢结合状态的解析、淀粉的损伤检测、加工适应性的测定和水果内部品质的测定，如表 17-1 所示。所测食品形态可以是固态、液态、粉状、糊状。因此，样品池也是多种多样。粉状物料的测量多用长波近红外反射法，液体样品多用短波近红外透射法，而固体既可以用反射又可以用透射法测定。图 17-1 是典型的

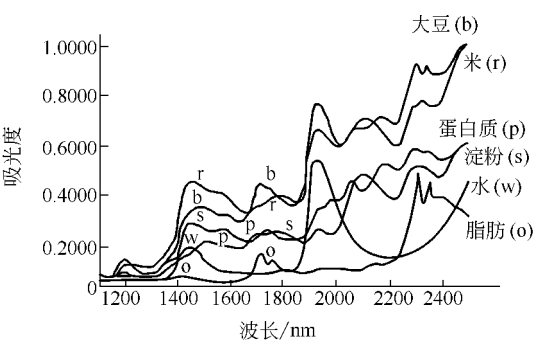


图 17-1 大米大豆的主要成分近红外光谱

食品近红外光谱图，由图 17-1 可以看出，大豆光谱是水、蛋白质、脂肪等物质的综合反应的结果。

表 17-1

近红外技术在食品工业上的应用

谷物、豆类、种子类	
米	淀粉(直链淀粉)、蛋白质、水分、灰分、氨基酸、风味值
小麦(小麦分)	淀粉、蛋白质、水分、灰分、软硬度、损伤淀粉、锈病、SDS 沉降性、淀粉酶、麸皮混入率、品种判别、氨基酸、制作面包适应性
大麦	蛋白质、水分、提取物、氨基酸、氮、外观最终发酵度
荞麦	蛋白质、水分、灰分
棉籽	棉籽酚、水分、葡萄糖、果糖
葵花籽	油脂、水分、纤维
大豆	蛋白质、水分、脂肪、7S、11S
油菜籽	油脂、叶绿素
豌豆	蛋白质、淀粉
辣椒	辣椒碱
啤酒花	水分、α 酸、精油
茶	蛋白质、水分、全氮、咖啡因、茶氨酸、全游离氨基酸
红茶	质量
乳制品	
牛奶	水分、脂肪、蛋白质、乳糖、TMS
干燥乳	水分、脂肪、蛋白质、乳糖、灰分
奶酪	水分、脂肪、蛋白质、乳糖、灰分、固形物、酪氨酸、pH
乳清	水分、脂肪、蛋白质、乳糖
冰淇淋	脂肪
奶粉	水分、脂肪、蛋白质
水畜产品	
鱼肉	鱼肉中水的存在状态
畜肉(肉制品)	水分、脂肪、蛋白质、盐分、热量、大豆粉、pH、调理适应性

续表	
饮料 果酒 啤酒 果汁 豆乳 茶	酒精、浸出物、糖、滴定酸度 酒精 葡萄糖、果糖、蔗糖 蛋白质、水分 咖啡因、茶氨酸、总氨基酸
一般加工食品 麦芽 面包 可可 巧克力 酱油 干燥蔬菜	纤维、糊化度 蛋白质、水分、脂肪 蛋白质、脂肪、淀粉 蔗糖 盐、氮、酒精、乳酸、转谷氨酸、葡萄糖 蛋白质、脂质、灰分、ADF
果蔬 葱头 生菜 甘蔗 桃、梨 苹果 西红柿 蜜橘	糖度、水分、干物质 成熟度 粗纤维、水分、糖度(Brix)、蔗糖、还原糖 糖度(Brix) 糖度(Brix)、酸度、硬度、水心病 糖度(Brix)、滴定酸度 糖度(Brix)、酸度

17.1 成分的定量分析

17.1.1 水分

水在 1940nm、1450nm、1190nm、970nm、760nm 处有较强的吸收。其中，1450nm、970nm 及 760nm 是 O—H 的倍频吸收带，1940nm 及 1190nm 是 O—H 的合频吸收带。

在水的近红外吸收峰中，1190nm 波长处的吸收率小，一般用于大于 50％的湿度测量；2950nm 波长处的吸收率过大，测湿的范围窄，一般用于小范围高精度的测量，而 1450nm、1940nm 波长的吸收率较为适当，可用于工业领域的测湿，实践表明，1940nm 波长比 1450nm 波长的灵敏度更高，线性更好，因此，得到了更广泛的应用。

水分含量测量所用定标主波长因样品水分含量范围、所含其他成分、光谱的测定方法的不同而各异。像谷物等低水分的样品选择 1940nm 或 1450nm，像水果类的高水分样品为 970nm 或 760nm。预测标准误差（SEP）随样品的不同而不同，大致为 0.1％～0.5％。

在实际测量中，被测物体的表面一般是呈现不规则的，具体表现在表面的反射率要发生变化。这个变化在效果上也表现为光反射强度的变化，因此形成了测量过程中的干扰。为消除这种干扰，林晓鹰采用所谓的三色法测量，即采用除水分吸收峰的波长测量水分含量外，还使用吸收峰两侧不吸收水的波长 λ_1 （1700nm）、 λ_2 （2100nm）作为参比波长。

水分基本上在 λ_1 、 λ_2 处无吸收，把 1940nm 波长的反射光强与它们的反射光强相比就可以得到物品的透过率了。并由此求出该物品的水分含量。

用近红外光谱法测定水分的缺点是对于微量水分（0.04%以下）的测量有困难。

17.1.2 蛋白质

蛋白质的特征吸收波长为 2180nm 和 2050nm。这些为胺化物Ⅰ的倍频和胺化物Ⅲ的合频振动，以及 N—H 的倍频振动、胺化物Ⅲ的合频振动的吸收波长。小麦蛋白质含量 C_p 可由下式求得：

$$C_p = 12.68 + 493.7 \times \log\left(\frac{1}{R_{2180}}\right) - 323.1 \times \log\left(\frac{1}{R_{2100}}\right) - 243.4 \times \log\left(\frac{1}{R_{1680}}\right)$$

其中， $\log(1/R_{2180})$ 、 $\log(1/R_{2100})$ 和 $\log(1/R_{1680})$ 为 2180nm、2100nm、1680nm 处的吸光度。2180nm 如前所述是蛋白质的吸收波长，2100nm 是淀粉的吸收波长，1680nm 与成分吸收无关，但当样品为粉料时，1680nm 反映粒度信息。所以，蛋白质含量定标方程是基于蛋白质、淀粉、样品的粒度建立起来的。预测标准误差（SEP）为 0.2% 左右。在多种食品中，2180nm 用于蛋白质测量主波长的理由是它的吸收不受盐类、糖类、氨基酸等成分及 pH 的影响。

郑咏梅利用逐步回归法对小麦的蛋白质进行了分析。在波长 1000~2500nm 范围内的最佳波长和数学模型如表 17-2 所示。利用这些模型对小麦粉蛋白质含量分别进行预测，在Ⅰ波段的预测标准误差 $SEP = 0.7422$ ，Ⅱ波段 $SEP = 0.2809$ ，Ⅲ波段 $SEP = 0.2487$ 。分析结果表明，在 1860~2500nm 范围内的预测效果最好。

表 17-2 不同波段选取的最佳回归波长及建立的数学模型

波段	Ⅰ 波段(1000~1400nm)	Ⅱ 波段(1400~1860nm)	Ⅲ 波段(1860~2500nm)
回归波长/nm	1237,1180,1159,1137,1082,1073	1756,1692,1604,1525	2197,2175,2111
回归模型	$C = -2270a_1 + 3018a_2 - 2096a_3 + 2523a_4 + 2356a_5 - 3569a_6 - 1.4497$	$C = -419.9a_1 + 1293a_2 - 1415a_3 + 553.9a_4 + 12.6852$	$C = -711.9a_1 + 1051a_2 - 333.3a_3 + 7.1478$
回归系数 R^2	0.959	0.9758	0.9675
F 值	97.39	273.3	278.2

刘建学等人对不同类型（糙米、精米）、不同粒度（0.15mm、0.20mm、0.28mm）的大米蛋白质进行了近红外光谱快速测定的研究。其结果表明，在粒度不均匀时，以二阶导数谱建立预测模型比用归一化谱建模效果更好，如表 17-3 所示。在用二阶导数谱建立大米蛋白质含量预测模型时，可以用 0.20mm 粒度样品代替 0.15mm 粒度样品，以简化磨粉程序。

表 17-3 二阶导数谱各预测模型检验结果

粒度/mm	样品	预测模型相关方程	R^2	F
0.15	精米	$y = 0.7818x + 2.0476$	0.9236	2.17
0.15	糙米	$y = 0.9323x + 0.6401$	0.9145	2.90
0.20	糙米	$y = 0.9048x + 0.9402$	0.9164	0.85
0.28	糙米	$y = 0.8478x + 1.2415$	0.8909	1.55

17.1.3 脂肪

脂肪在近红外区域的吸收主要是长链脂肪酸，吸收波长为 2345nm、2310nm、1765nm、1734nm 及 1200nm。2345nm 及 2310nm 是 CH_2 的伸缩振动和弯曲振动的合频振动，1765nm 及 1735nm 是 CH_2 的第 1 倍频，1200nm 是 CH_2 的第 2 倍频吸收。顺式不饱和脂肪酸的情况，除了前述吸收波长外，2180nm 和 2140nm 也被认为是吸收波长。

近红外法在脂肪含量测量上的应用开始于早期。Ben-Gera 测量了各种肉制品（厚 2mm）的透射光谱（1500~1800nm）。脂肪的吸收光谱 1725nm 处的吸收强度用 1725nm 和光谱低谷 1650nm 处的吸光度差 ΔOD （1725-1650）表示。求得与索氏法测量的化学分析值的相关关系，其相关系数为 0.97，标准偏差 2.1%。此时的误差主要是因为不能准确地调整 2mm 厚的样品以及由透射信号弱所产生的弱信噪比所致。如果这些方面得到改善，精度可以提高到 $\pm 0.5\%$ 。

测定油料油分含量和评价植物油品质是近红外技术的一大应用。顾伟珠（1995 年）对完整油菜籽含油量进行测定，得到以下多元回归方程：

$$C\% = 91.241045 - 125.7195X_1 + 286.9495X_2 - 109.4939X_3 + 41.1942X_4 - 139.3279X_5$$

其中： $C\%$ 为油菜籽含油量， X_1 、 X_2 、 X_3 、 X_4 、 X_5 分别为油菜籽光谱在 1680nm、1720nm、1940nm、2100nm 和 2180nm 处的吸光度。方程相关系数 $R=0.882$ ，标准偏差 $S=0.464$ ，利用该模型对验证集样品进行预测，得预测集样品相关系数 $R=0.882$ ，标准偏差 $S=0.489$ ，说明模型有较好的相关性。该方法预测一个样品的含油量只需几分钟，而作为定标的标准值所采用的索氏提取则需 8~10 小时，利用 NIR 技术大大提高了实验效率。

徐晓峰利用 GQA-31 型近红外光谱仪对整粒大豆进行扫描，经相关性测定，近红外光谱仪测得的结果与索氏法测得结果相比，相对偏差在 0~3.5% 之间，其中绝大部分在 1% 左右。为提高仪器测定结果的准确性，选用大豆油分含量在 18.7%~23.38% 范围作为仪器的有效区间。对有效区间样品进行两结果间相关性测定，相关系数 $R=0.98$ 。16 个验证集样品评定结果：相对偏差为 1%~2%，相关系数 $R=0.98$ ，因此相关性很好，可用于准确预测大豆油分含量。

汪旭升等研究了棉籽油粉中油含量的测定。选用棉花的不同品种（包括各种特色棉）材料 248 份，进行常规成分的定标和校验。定标结果表明，采用改进的偏最小二乘法（Modified PLS）和（3，5，5，1）的数学转换的方法得到的定标效果最好，其定标决定系数（ $RSQ=0.9335$ ）和可解释的变异（ $1-VR=0.9131$ ）最高，定标标准误差（ $SEC=0.7905$ ）和交互校验标准误差（ $SECV=0.9024$ ）最小。预测结果表明，其决定系数（ $RSQ=0.9780$ ）也明显高于其他方法，同时具有最低的预测误差（ $SEP=0.508$ ）和校准标准误差（ $SEC=0.4910$ ）。由此可见，用近红外光谱法测定棉籽粉的油分具有与化学法相近的准确性和精确性，可以在实践中应用。

亚麻子含油量也可以通过近红外反射光谱测定（Bhatty，1991 年）。经过计算机软件选择合适波长处理数据，完整亚麻子含油量的测定与传统方法得到的结果有相同的精度，而且利用 NIR 检测亚麻子含油量不受生长地区和种子含水量的影响。

油脂品质主要由碘值、过氧化值、游离脂肪酸含量等指标评价，通过 NIR 技术测定上述指标可以加快检测速度，更好的对油脂品质进行监控。

碘值是判断油脂不饱和度的指标，常规方法是先加成反应再用标准硫代硫酸钠溶液滴定。Hui Li 等用 NIR 对保持在 75℃ 的样品进行扫描，通过偏最小二乘法（PLS）对光谱数据和样品集碘值进行训练，得到校正模型。PLS 模型可以将四种类型的食用油清楚地地区分开，并且预测的碘值与气相色谱测量的值几乎一致。该方法能使油脂加工者更好控制其产品。

过氧化值（PV）是判断油脂氧化程度的指标。Yidiz 使用 NIR 透射光谱快速测定大豆油（SBO）氧化水平，定量分析了 PV、共轭二烯值（CD）和茴香胺值（AV）。通过对光谱数据进行 PLS 回归和向前逐步多元回归等处理，建立模型。当使用 2mm 路径时，光谱范围从 1100~2200nm 最适合测量氧化水平。使用光谱一阶导的 PLS 回归则给出了最好的 PV 结果。验证集中 PV ($R=0.99$)，CD ($R=0.95$)，然而 AV 测量结果并不好。棕榈油的过氧化值也可以通过 1350~1480nm 测定 (Moh, 1999 年)。

游离脂肪酸指标反映的是油脂的新鲜程度，快速检测油中脂肪酸含量意义很大。Che Man 等利用近红外反射光谱法测定了天然棕榈油的游离脂肪酸（FFA）。在 60℃ 下用 15% 的脂肪酶水解天然棕榈油，配制成不同 FFA 浓度的棕榈油，进行光谱扫描，根据多元线性回归分析基于 C=O 的振动范围（1850~2050nm），得到最好的 3 个波长 1882nm、2010nm 和 2040nm， $R^2=0.994$ 。只需要知道这 3 个波长的吸光度就可以在 5 分钟内测出 FFA 值。

利用近红外技术还可以进行品质分级、检测油脂制品的真假、测定不饱和度、识别顺式和反式等。

17.1.4 淀粉

淀粉由 O—H 和 C—H 构成，在近红外区域可以观察到多处吸收作为分析波长，通常，定量分析用 2270nm、2239nm、2190nm、2100nm、1778nm 及 1722nm 的波长。

刘建学等通过不同粒度（0.15mm、0.20mm、0.28mm）不同类型（精米和糙米）样品的近红外光谱分析，进行了 PCR 定标与校正建模。考察所建模型的预测效果，精米要好于糙米。这可能是由于糙米的颖糠部分与内部对光的吸收不同所致。精米模型的预测值与化学分析值的相关系数为 0.95，SEP、MER 和 BIAS 值分别为 0.56、3.1% 和 1.1。

舒庆尧等人研究了用近红外反射光谱技术测定稻米表观直链淀粉含量（AAC）的效果及影响校正的一些因素。对三种不同回归统计分析方法校正效果的差异比较表明，用修正的最小二乘回归法和主成分回归法（PCR）进行 AAC 校正时，校正标准误差（SEC）、交叉检验标准误差（SECV）和检验工作标准误差（SEPC）均呈上升趋势，分别为 0.83、1.75 和 1.09（MPLS）；1.73、1.98 和 1.74（PLS）以及 2.29、2.56 和 1.72（PCR）；相反，校正决定系数 R^2 和检验决定系数 RSQ 呈下降趋势，分别为 0.983 和 0.91（MLPS）；0.927 和 0.84（PLS）以及 0.870 和 0.84（PCR）。由此可见，用 MPLS 技术建立 AAC 回归方程效果最佳。该试验与先前相似报道比较，通过采用了 MPLS 和光谱分组法，所得回归方程预测 AAC 的精确度进一步提高，SEP 由 1.47 减小为 1.09，而 RSQ 由 0.865 上升为 0.91，从而将进一步推动 NIRS 技术在稻米 AAC 测定中的应用。

17.1.5 水果糖度 (Brix)

河野等人采用透射方式测量温州蜜橘的光谱，因光谱受果实大小影响，对利用二次微分进行了大小修正的光谱定标后的结果是，914nm、769nm、745nm、786nm 处的波长测量精度最高，此时的相关系数高达 0.99，校正标准误差（SEC）为 0.28°Brix，检验工作标准误差（SEP）为 0.32°Brix。所用波长中的 914nm 为蔗糖吸收带。

为了开发桃的无损检测装置，利用光纤进行了近红外分光法可行性研讨。由高中低糖度含量不同的果实二次微分光谱可知，在 906nm 的左右区域有明显差异。因为 906nm 与蔗糖的吸收带相近，因此可以判断为是蔗糖的吸收。含有二次微分光谱的 906nm，878nm，870nm 及 889nm 定标精度最高，其相关系数为 0.97，定标标准误差 SEC=0.48°Brix，预测标准误差 SEP=0.50°Brix，偏差为 0.01°Brix。由此证明了近红外分光法可以高精度地测量完整桃的糖度。

恩田等人对李子的糖度在 1100~2500nm 范围内进行了近红外光谱分析。其分析结果表明，采用了 8 个波长的定标方程精度最高。定标标准误差 SEC=0.41°Brix，预测标准误差 SEP=0.61°Brix，相关系数 R=0.92，具有实用价值。

$$\text{糖度 (°Brix)} = 3.03 - 1325.16A_{1316\text{nm}} - 1055.22A_{144\text{nm}} + 903.14A_{1604\text{nm}} + 1522.18A_{1204\text{nm}} \\ - 2769.06A_{1224\text{nm}} + 1988.43A_{1272\text{nm}} + 1576.08A_{1172\text{nm}} - 839.20A_{1608\text{nm}}$$

其中，A 为相应波长时的吸光度 lg (1/R) 值。

在上述方程中采用的 8 个波长有 4 个集中在 1200~1300nm 区域，通过与糖微粉末的 2 次微分光谱比较得知，1204nm 及 1278nm 处存在吸收峰，因与定标所用波长基本一致，则认为这 2 个波长即为糖的吸收波长。

17.1.6 水果酸度

水果的酸度值是水果风味的重要指标，测量柑橘的酸度实际上是测量游离酸的含量，游离酸在 850nm，900nm 处存在吸收峰，但其浓度为 0.4%~1.7%，与糖含量相比较低，难以准确测量。

17.1.7 硬度

恩田等人对李子的硬度进行了近红外分析。以 7 个波长为自变量的定标方程精度最高，R=0.83，SEC=0.11kg，SEP=0.21kg。

$$\text{硬度 (kg)} = 2.54 + 214.24A_{1864\text{nm}} + 283.18A_{1596\text{nm}} - 245.36A_{1860\text{nm}} - 514.56A_{1616\text{nm}} \\ - 137.24A_{1544\text{nm}} + 283.23A_{1580\text{nm}} + 115.43A_{1656\text{nm}}$$

因为果实硬度这个物理指标不能直接吸收光线，近红外法的可行性曾受到过质疑，但上述结果表明，即使是不具备近红外吸收的信息也可与光谱间接地相关。其原因是随着成熟度的增加，与硬度密切相关的果胶发生了变化，或者说，由果胶组成变化引起的物性变化充分地反映在光谱上。

小麦的硬度对小麦粉的加工粒度有影响，而小麦粉的颗粒粒度又是最终产品的品质的重要参数。根据粉体物料粒度的不同，近红外光谱利用上下移位的性质测量小麦的硬度。

当软质及硬质小麦的硬度平均得分（HS）分别为 25 和 75 个单位时，由反射法获得

的 *HS* 用下式表示 (K. H. Norris):

$$HS=a+b_1L_{1680}+b_2L_{2230}$$

其中: a 、 b_1 、 b_2 为常数; L_{1680} 、 L_{2230} 为各自下标波长的吸光度。1680nm 波长作为定量分析时的对照波长, 经常被采用, 是一个不受蛋白质、水等成分影响的波长。即, 1680nm 波长处主要反映粒度的不同。在相同条件下进行粉碎, 由粉碎物料的粒度的差异即可测出小麦的硬度。

17.1.8 果实成熟度对近红外光谱的影响

恩田等人对不同成熟度的李子反射光谱变化进行了测量, 由图 17-2 可以看出, 随着贮藏时间的推移, 光谱基线漂移。在 1410nm 左右处的吸收峰值明显增大。此时果实的糖度虽略有变化, 酸度也缓慢地呈直线降低, 但果实的硬度出现急剧下降趋势。另外, 果实硬度在贮藏 7~8 天时变化显著, 与此同时, 近红外光谱基线上移加大。由此可以推断, 由李子的后熟引起的近红外光谱的变化是果实硬度变化造成的。

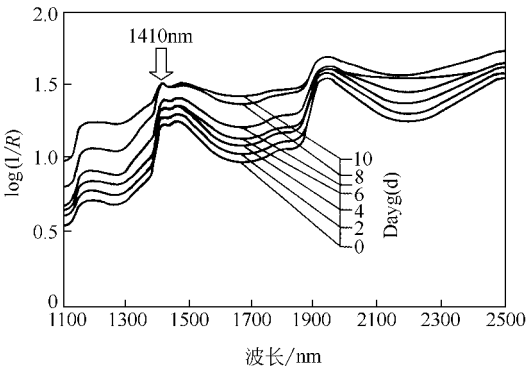


图 17-2 李子在贮藏过程中的近红外光谱

17.1.9 盐、其他

因盐在近红外区域没有吸收, 不能用近红外法测量。但是, 存在于食品中的盐可以测量。在酱油中的盐分析中, 由现行方法测量的成分值与近红外法测量的成分值的相关系数达到 0.999。此时所用主波长是 1445nm, 属于水的吸收波长。在火腿盐分分析中, 与 1806nm 处的吸光度 (二次微分) 相关系数高达 0.96~0.997。能够在近红外区域分析没有吸收的盐分的理由是随着盐浓度的变化, 水的吸收光谱的形状及峰值的位置发生了变动。

与盐分同样的还有小麦粉中灰分的测量。灰分由小麦粉麸皮的含量推算而得。

17.2 食品品质分析

17.2.1 大米食味分析

大米食味是指人们对食用米饭的视觉、咀嚼、嗅觉、味觉的综合感觉。由于是主观感觉, 所以食味因人而异, 没有绝对的尺度, 主要是从米的外观、香味、味道、黏度、硬度及综合指标来评价。大米食味与米的来源 (品种、产地土壤、气候、栽培水肥管理)、收割、贮藏、烘干、加工和煮饭方法、过程都有很大的关系。某个环节没有处理好, 均会影响大米的食味。

人们对大米食味的评价由来已久, 开始时评价食味采用感官检查实验法。方法是以标准品为基础, 采用一般人对大米评定分级标准, 每人做出评价并将每个人对食味评定编入

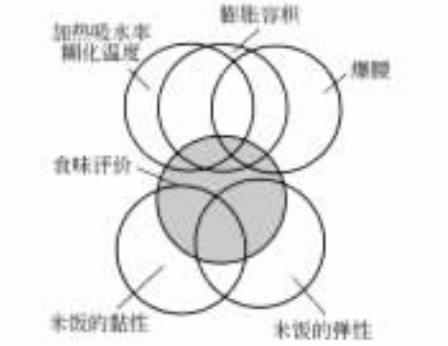


图 17-3 大米食味评价与理化要素的关系

每个人的最后评定，这种方法缺少统一性，并且参加人不同的年龄、性别、喜好、进行感官鉴定的不同时间、环境变化都会影响评定结果，因此，这种评价方法的结果不很准确。

影响大米食味的因素有米饭进入口后的黏度及硬度等物理特性，以及呈现香气及酯味的化学成分。长期的研究表明，如图 17-3 所示的理化特性与大米食味密切相关。

通过对多个理化特性与食味的关系进行的回归分析，得出以精米的蛋白质（干物质）、精米粉糊的黏度特性中最高黏度、最低黏度、爆腰率、精米蒸饭特性中的蒸饭液碘试剂显色度等 5 项为自变量的回归表达式：

$$\hat{y} = -0.1272x_1 - 0.0929x_2 + 0.0902x_3 + 0.0946x_4 - 6.5955x_5 + 2.6425$$

其中： $\hat{y}$ 为食味估计值； x_1 为蛋白质（干物质）； x_2 为精米粉糊的黏度特性中最高黏度； x_3 为精米粉糊的黏度特性中最低黏度， x_4 为爆腰率， x_5 为米饭液碘试剂显色度。

另外，米饭食味与米粒中的矿物质成分有关。即与米粒中的镁（Mg）成负相关，与钾（K）成正相关。Mg 的含量与 K 的含量的比值越大，大米食味越好。

大米食味计是日本的独特应用，现已普及 4000 多台，用于大米的分选、品质管理。利用近红外分光法的大米食味计有 6 种，如表 17-4 所示。不论哪一种都是先求出与食味相关的多个成分值，再通过事先导出的食味关系式算出食味值。

表 17-4 各生产企业近红外食味计测量指标

公司仪器型号	测 量 项 目	物料形态
佐竹(TB15A)	水分、蛋白质、脂肪酸度、直链淀粉、食味值	粉碎精米
静岡(RA-6101)	水分、蛋白质、黏度值(碘试剂显色度的反函数)、食味值	粉碎精米
山本(YAP)	水分、蛋白质、黏度值、食味值	粉碎精米
Nireko(HON4500)	Mg、K、N、直链淀粉	粉碎或未粉碎精米、糙米
Nireko(shon6500)	水分、黏度、蛋白质、整粒米食味值	
久保田(K-AS100)	直链淀粉、蛋白质、水分	未粉碎米
ketto 科学研究所	水分、蛋白质、直链淀粉、脂肪酸	未粉碎精米、糙米

最早由佐竹制作所开发出的食味计，是根据蛋白质、水分、脂肪酸、直链淀粉成分的平衡决定食味的。硬件部分采用了干涉滤光片方式的近红外装置（BRAN+LUEBBE）。借助设定的高精度的前处理装置，把样品通过碾白、恒温粉碎、恒温贮存等工序，在前处理中把样品一致化，将粉碎的精米放入专用的样品池中，在装置内数秒即可自动测出成分值。同时，结合在计算机中存储的有关大米食味的基础资料与与日本谷物协会的官能检查实验调查结果相关的食味资料，独自开发出食味关系式表示出食味的点数。佐竹食味计以 100 分为满分，通过测定样品各个成分并计算得出综合值，最后以百分制分数表示出来，即最后的食味值。因为粉料样品的粒度、温度影响近红外法的测量精度，为此专门开发出了保证粒度大小一致的粉碎机，调节温度的定温装置。新近开发的食味计操作更加方便，将整粒大米从样品口放入，瞬间即可获得数据。

影响食味测定的因素有：

① 白度。白度即大米的加工精度，通过白度计测定出来。白度是影响大米食味测定的最重要的因素。白度偏高，即加工精度偏高，将使测定的食味值偏高。而白度偏低时，将使测定的食味值偏低。正常的白度要求是待测样品的糙米白度加上 20 ± 0.5 。

② 样品的准备。为了保证测定的各个样品的准确，必须要求样品处理过程一致，尤其是温度和时间要准确。既要保证大米在恒温 40°C 下被粉碎，又要保证粉碎后在恒温箱中 25°C 下贮存，如果粉碎后立即测定，将会出现结露现象，使测定不准确。

③ 样品的均匀性。在测定前，要用小匙搅拌样品，使之均匀。在样品加入测定圆盘时，也要搅拌均匀。同时添加要适量，防止样品盘中样品过紧和过松，这些都会影响测定。

静冈制机开发的大米食味计是根据蛋白质、碘试剂显色度的平衡所决定的。硬件部分采用了干涉滤光片方式的近红外装置 (Dickey John)。将粉碎的精米放入专用的样品池中，在装置内数秒即可自动显示食味值。

Nireko 开发的大米食味计的食味基本由黏度决定，而黏度与 $\text{Mg}/(\text{K} \cdot \text{N})$ 值具有高的相关性。硬件部分为光栅式近红外装置 (NIRSystems)。将糙米或精米的粉碎样品或整粒样品放入专用的样品池，开机测量，自动测出 Mg 、 K 、 N 的含量，由 $\text{Mg}/(\text{K} \cdot \text{N})$ 值算出食味值。

Nireko 有粉碎型和整粒型两种机型。粉碎型是以大米的呈味成分均匀地分布在米粒中为前提，对粉碎后的大米米粉进行近红外光照射，通过反射光谱计算出食味值。整粒型则依据大米呈味成分存在于米粒表层至

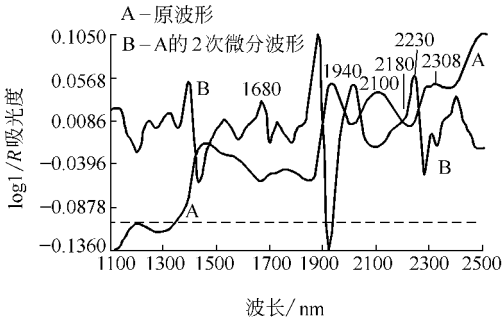


图 17-4 大米光谱

0.2~0.3mm 深处。图 17-4 中的 6 个波长为大米在近红外区域的主要使用波长。

山本制作所开发的大米食味计的食味由蛋白质、水分、黏度值的平衡决定。硬件与静冈相同，采用了干涉滤光片方式的近红外装置 (Dickey John)。将粉碎的精米样品放入样品池，开机测量数秒内显示食味值。

久保田的大米食味计的食味由水分、蛋白质、直链淀粉的平衡决定。硬件部分为独自开发的光栅式分光器。测量整粒精米的透射光谱 (约 $600 \sim 1010\text{nm}$)，由该光谱求出水分、蛋白质、直链淀粉的含量，并将其代入事先开发出食味关系式，求得食味值。

Ketto 科学研究所的食味计主要测量水分、蛋白质、直链淀粉、脂肪酸。硬件部分为独自开发的固定滤光片型分光器。测量完整的糙米、精米的透射光谱 (在波长 $918 \sim 1045\text{nm}$ 范围内的 12 点)，由吸光度值求出上述成分值。

因为上述公司的评价方法各不相同，存在着其测量数值间不能进行直接比较的缺点。

17.2.2 苹果水心病分析

水心病是苹果常见的一种生理失调症状，其发生是无规律的，没有什么办法可以预测得到，而且发病后通常不显示于外表。一般程度的水心病果可以在市场上销售，因其不易

进行贮藏。因此在贮藏前分离严重的水心病果有重要的工业价值。

正常苹果组织有 25%~30% 的体积被空气占据，而水心病果的大量空间充满了液体 (Marlow1984 年)，这些填充细胞间隙的液体改变了苹果组织的光散射特性，光散射因水心病而减少，使通过水果的光学路径更长，有更多的能量透过 (Olson, 1962 年)。Birth 和 Olsen 据此原理研究了一种快速无损伤检测元帅苹果水心病的方法。水心病苹果光谱图见图 17-5，水心病检测装置简图见图 17-6。760nm 波段是水的吸收带，而水心病果的水含量显然高过正常苹果。水心组织在 760nm 处吸收变强，但是水越多，光散射也减少，从而造成 760nm 处散射强度减少；在 800nm 附近苹果有相对很少的物质吸收能量，是测量的最佳波长。测量苹果在 760nm，810nm 的光密度，计算 ΔOD (760~810nm) 是最佳结果。一般而言，优先测量差值，因为许多变量，比如尺寸、形、样品位置，与仪器因素，比如光的稳定性、光度计的稳定性，都会影响光密度的测量值。在做差值时，可降低这些变量的影响。如果要考虑温度对检测的影响需要引入系数 K ，计算公式修正为 $OD_{760nm} - K (OD_{810nm})$ ，当 $K = 0.8$ 时， $R = -0.895$ 。也有报道用 ΔOD (740~805nm) 来检测水心病 (Francis)，检测结果与切开后发病目测的相关系数是 -0.809 。

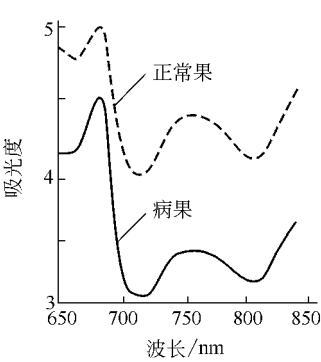


图 17-5 水心病苹果的光谱图

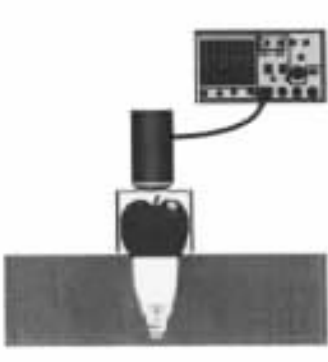


图 17-6 水心病检测装置图

目前最新的研究是利用时间飞行近红外光谱 (TOF-NIRS) 检测水心病，该方法处于初级研究阶段，要想达到最终目的尚需时日。

17.2.3 水果内部褐变分析

水果内部褐变用一般手段难以无损伤检测到，NIR 透射光谱恰好具有提取内部特征信息的特点，且 NIR 应用到在线检测的可行性较大，所以研究者们一直在进行水果内部褐变检测的研究。其中对苹果褐变检测的研究已有一定的基础。

苹果褐变的检测大多与水心病检测相关联，因为水心病是导致褐变因素之一。其检测方法中的光密度差法也类似于水心病检测。Birth 和 Olsen 早在研究水心病检测时就提到为了使水心病检测更加精确，预先计算水心病果的 ΔOD (720~810nm)，这样有利于将内心褐变果剔除掉。Francis 等利用 ΔOD (740~840nm) 正确分离了 91% 的元帅苹果，不过该检测方法在贮藏后期检测才是合理而可靠的。

苹果褐变组织影响其透射光谱，褐变苹果 650~950nm 波长范围的透射率光谱图见图 17-7 (Upchurch 1997 年)，一般的内心褐变果在 720~750nm 处的吸光度比好果的大，褐变果在 750nm 以上的透射率比较高。吸光度的大小取决于组织褐变的程度。透射率的走

势与苹果心部的状况有关，为了减少光路径的影响，将 720nm 处和 810nm 处的透射率之比作为辨别好果和坏果的特征值。当褐变程度增加时，比值减少，并且与褐变的程度相关 ($R^2 = 0.71$)。该方法对 420 个含褐变果的苹果进行检测，结果只有 6.3% 的好果被误判，超过 12.0% 的坏果因为褐变较轻微而被误判为好果。不同因素导致苹果被误判。一些苹果在测量部位有大面积的擦伤时，由于苹果内部的光散射，当擦伤组织与测量位置成 45° 角时，组织的褐变对光谱有轻微的影响。坏果被误判主要是由于较轻微的果肉褐变位置远离测量点，其信号难以检测到。

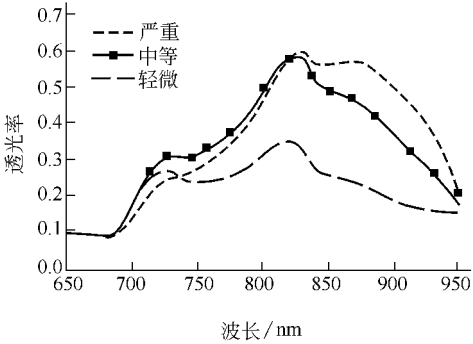


图 17-7 元帅苹果褐变光谱图

C. J. Clark (2003 年) 利用透射近红外检测 Braeburn 苹果的内部褐变和失调，研究了褐变组织在光谱 700~900nm 变化的情况。分别将苹果置于四种不同位置下，应用以下三个模型进行预测：

影响面积 (%) = $\beta_0 + \beta_1 \log\{S(\lambda_1)/S(\lambda_2)\}$ (模型 1)

影响面积 (%) = $\beta_0 + \beta_1 S(\lambda_1) + \beta_2 S(\lambda_2) + \beta_3 S(\lambda_3)$ (模型 2)

影响面积 (%) = $\beta_0 + \sum \beta_i S(\lambda_i), i=1 \sim 100$ (模型 3)

其中： λ 为波长； β 为回归系数； S 为在每个波长光谱预处理之后的离散函数表观光谱值。当果柄花萼果轴平行于检测头，检测头沿着轴中路与光源成 180° 时，模型 (1) $R^2=0.90$, $RMSEP=12.9$, λ 分别为 815nm、823nm。当果柄花萼果轴平行于检测头，检测头位于水果中路且与光源成 90° 时，MLR 模型 $R^2=0.88$, $RMSEP=9.4$ ，所取 3 个波长为 688nm, 697nm, 734nm。该位置下 PLS 模型 $R^2=0.91$, $RMSEP=7.9$ ，波长范围为 697~861nm。

以上所提到的苹果褐变检测和建模的方法虽然在实验中能有相对较好的分级效果，但是褐变检测最终是要在生产中应用，目前的水平还存在大量的问题要解决。单个果测量时间过长，水果尺寸对检测的干扰，褐变在果肉中分布的不对称性等等问题是需要继续深入研究的。

17.2.4 水果表面损伤分析

桃的损伤有擦伤、疤痕、压伤、腐烂、虫害、裂痕等多种形式。在分选过程中，为了有效地识别出损伤果，李晓明等人进行了从可见至近红外区域的光谱特性研究。经研究表明，桃的正常果与损伤果间的置信极限差（正常果和损伤果的置信区间互相接近侧的分光反射率的差）因损伤的种类不同而不同，存在于可见至近红外区域。在不受表面颜色影响的近红外区域可以检测损伤果。腐烂和冲击损伤果在近红外 800~940nm 和 1140~1400nm 区域内的置信极限差为 10% 以上，压伤为 5%，即在此区间可以检测出三种损伤果。而摩擦损伤的置信极限差在 5% 以下，所以检测困难，如图 17-8。

另外，梨在 800~1000nm 区间内，腐烂、擦伤、刺伤的置信极限差为 10% 以上，压伤、碰伤为 5%，可以进行损伤果的检测。腐烂柿在 600~1350nm、伤痕柿 650~900nm

置信极限差为 10% 以上，在 900~1100nm 内为 5% 以上。因此在 800~1100nm 范围内，可以检测腐烂和伤痕柿。柑橘在 500~1300nm 和 1700~2500nm 范围内为 10% 以上，而苹果在近红外区域没有大的置信极限差。

由上述分析可以得出，在 800~900nm 近红外区域内，可以对多数水果进行损伤果的检测。

17.2.5 小麦粉加工适应性的分析

小麦的加工适应性由蛋白质、淀粉、脂肪、水等成分组成，而这些成分在近红外光谱上又产生微妙差异，从这些差异中可以直接分析小麦的加工适应性，通过对 2 次微分光谱的 21 个吸收带数据进行主成分分析，获得图 17-9，所示结果，在第 1 主成分和第 2 主成分散点图上，可以清楚地判明面包、中国面条、糕点及面条用小麦粉。

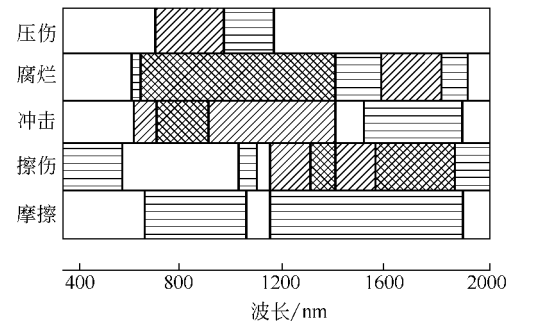


图 17-8 桃损伤果的置信极限差的分布

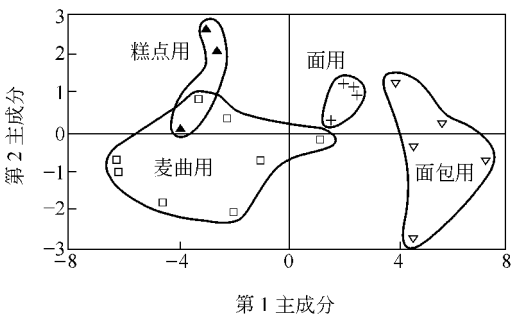


图 17-9 不同加工性能的小麦粉近红外光谱的主成分分析结果

17.2.6 蔗糖分析

长期以来，由于传统分析技术的局限性，按质论价的原料收购体系在我国甘蔗制糖企业一直没有真正建立，甘蔗收购也一直沿用按重量的计价方式，蔗农也因此一味追求高产量，轻质量。低质量甘蔗大大增加了糖厂的生产消耗，以某厂为例，从 1994~1999 年，该厂连续 5 个榨季平均甘蔗糖分仅为 11.28%，制糖成本高居不下，1996~1998 年几个榨季连续亏损。自从 1999 年该厂引进近红外技术以来，甘蔗糖分逐年提高，从 11.50% 提高到 12.05%，大大提高了经济收益。

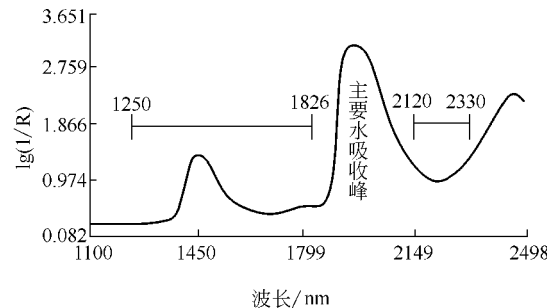


图 17-10 典型蔗汁扫描光谱

陈继红处理蔗糖光谱时，进行了光谱区间的筛选（图 17-10）。由于 1820~2120nm 存在较大吸收水峰，并波动较大，这将严重干扰定标，所以定标中需将此段光谱信息删除。从 2300~2500nm，由于光纤在此波段噪音较大，定标中此波段的信息也将被剔除。为提高光谱的质量，还需对光谱进行数学处理。利用 1100~1820nm，2200~2300nm 波段信息，首先

对选定区段进行散射处理，处理技术为 SNV+Detrend 标准正常化处理技术和去偏异技术；然后对光谱做一阶导数处理，每隔 4 个波点同时做平滑。数学处理后的光谱采用改进最小二乘法 MPLS 建立定标，最佳主成分项由内部交叉验证获得。交叉验证分组数为 4。

近年南非糖业研究所 (SMRI) 和南非糖业协会试验站 (SASEX) 合作进行了以下研究：用 NIRSystem5000 近红外光谱仪测定了 500 个碎蔗样品的 Pol、Brix 和水分。1994 年在一个糖厂开发的定标方程分别用 1992 年和 1995 年两个榨季 4 个糖厂的数据进行了评定。用偏最小二乘方回归分析得到的定标 R 值对于 Brix 和 Pol 都超过 0.97，但对水分较低 (0.945)，对定标的校正表明 NIR 对 Pol、Brix 和干物质的预测 (期望) 值一般是令人满意的， R 值在 0.91~0.96。应当指出，以上结果是用不同 NIR 仪器和实验室在 4 个不同糖厂试验的结果，其后定标方程在商品蔗 (原料蔗) 和清洁的破碎样品的无性系评估中表明对期望值和实验室测定有很好的关联。因此表明 NIR 技术可以充分可靠地用于糖蔗快速分析。

1997 年 SMRI 对 4 个榨季收集的 550 个混合汁样品和超过 900 个 C 糖蜜样品，用 1100~2400nm 波长进行了扫描，混合汁是随机地分成两部分，分别用 NIR 和传统方法检测了 Brix、Pol、蔗糖、葡萄糖、果糖、灰分和纯度，并对定标方程进行了校正。NIR 的定标和期望值对于工厂糖汁中蔗糖、Pol 和 Brix 是最令人满意的 [$SEP < 0.1$, $bias$ (偏差) = 0, 斜率 = 1]，NIR 对蔗糖、Pol 和 Brix 的 SEP 与 CTS 实验室方法非常类似。

日本开发的甘蔗榨汁液品质分析专用设备的自动化程度高，1993 年在日本 21 个生产厂家以附加了网络支持系统的形式推广使用。图 17-11 为甘蔗品质分析网络系统图，根据研究结果，用近红外法测量甘蔗榨汁液中的蔗糖含量可以获得高精度 ($SEP = 0.33^{\circ}Brix$)。

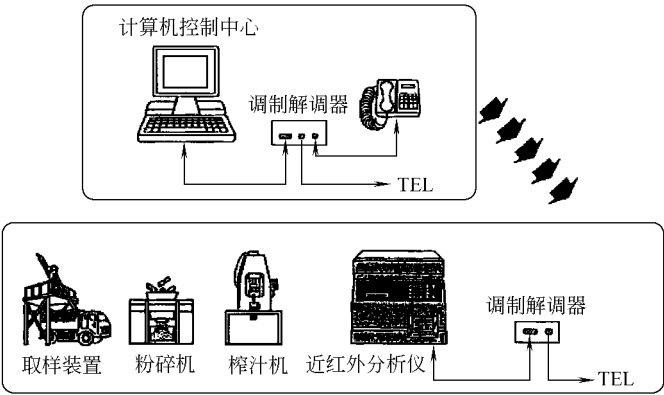


图 17-11 甘蔗品质分析网络系统图

广西大学糖业工业技术研究中心利用 FOSSNIR-System6500 近红外光谱仪对 1998~1999 年榨季 3 个糖厂近 920 个样品的白砂糖色值进行扫描了定标和校正，发现 NIRS 与传统法分析有较好的关联 ($R = 0.911 \sim 0.94$)。

17.2.7 酱油分析

酱油的生产包括原料处理、制曲、压榨、酿造、罐装等工序。在生产过程中，半成品需要多项湿式分析，费时费力。1991 年，日本利用近红外法进行了酱油原料、酱油滤饼、

酱油成分的分析。近红外法对原料大豆、小麦的总氮量、水分的分析与化学法相关程度高，对于酱油滤饼，水分、食盐、粗脂肪、粗纤维、灰分等的测量在实际中已经应用。对于酱油，以总氮量、乙醇、乳酸、葡萄糖、还原糖、谷氨酸为对象进行了分析。从酱油生产过程中取 245 个样品，对光谱数据和化学法分析值进行了回归分析，其结果如表 17-5 所示，除了还原糖以外，与化学方法的相关系数为 0.90~0.99，标准偏差为 0.19 以下。

表 17-5 酱油成分的光谱数据与回归分析结果			
分析项目	分析方法(常规法)	重相关系数	SEE(误差的标准偏差)
食盐	莫尔(mohr)法	0.99	0.08
氮	凯氏定氮法	0.99	0.009
乙醇	气相色谱法	0.99	0.07
乳酸	酶法	0.90	0.06
谷氨酸	酶法	0.95	0.04
还原糖	贝特兰(糖定量)法	0.67	0.33
葡萄糖	酶法	0.95	0.19

基于上述实验结果，近红外法在酱油生产管理中得到了应用。由于各个发酵罐的酱油品质不同，为了生产质量稳定的高品质酱油，必须不断地监视发酵罐内的产品，为此，日本开发了近红外酱油成分自动分析装置。这种装置由近红外仪器、水槽、自动取样器及泵组成。由自动取样器抽取的样品通过水槽中的加热电阻圈时被加热为 20℃，然后被送往近红外仪器进行光谱的自动测量，由数学模型计算出多个成分。1993 年开发出适合多种酱油的检测系统，并作为 JAS（日本农林标准）检测法广为使用。

17.3 结构分析

17.3.1 淀粉糊化

近红外分光分析法可以检测淀粉的糊化程度。

在对不同温度下糊化的淀粉光谱的二阶导数光谱分析中可以看出，1204、1368、1436、1700、1748、1784、1924、2088、2280、2320、2348nm 波长附近的吸光度变化较大，如图 17-12 所示。1924nm 处的吸收为水的 OH 基团，其他峰值主要是淀粉的 OH 基团或 CH 基团的吸收。2100nm 及 2280nm 处的吸收峰随着糊化度的进展，显示出非常复杂的变化。因为在这两个波长中存在着淀粉的吸收特征，所以它们反映了由糊化引起的结构变化信息。

在波长 2100nm 附近，未糊化淀粉在 2070nm 处有最小吸收值，在 2120nm 处也有吸收。当糊化温度上升到 56℃ 时，2070nm 处的吸收逐渐消失，2090nm 处的吸收值为最小。在 56~100℃ 间的糊化淀粉中，2090nm 处的吸收增大，且 2070nm 附近的吸收再次增大，如图 17-13 所示。在淀粉近红外吸收光谱 2100nm 左右，重叠存在着 2090nm 及 2100nm 两个吸收带，上述光谱的复杂变化与这两个吸收变化密切相关。

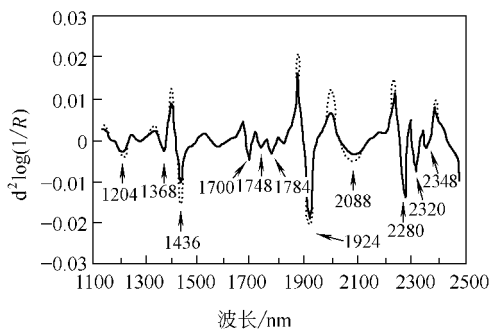


图 17-12 不同温度下大米淀粉二阶导数光谱图

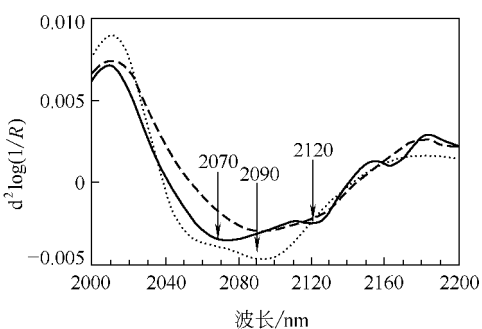


图 17-13 大米淀粉光谱区间图

以糊化度为依变量，吸光度二阶微分值为自变量所进行的回归分析结果如图 17-14 所示。在 1220nm, 1807nm 及 2095nm 处得到了高度正相关，在 1220nm 及 2095nm 左右，可以观察到淀粉的吸收，在 1807nm 处虽然相关系数大 ($R=0.99$)，但没有明确的吸收。由此可以得出，在 1807nm 处获得高相关的原因是与糊化度相关的其他因素，如糊化淀粉的粒度等物理因素。从以糊化淀粉粒度为依变量的回归分析中可以得到与糊化度形状基本相同的相关光谱。另外，糊化度与粒度间存在着较高的相关性 ($R=0.96$)。这意味着通过加水加热的糊化处理，促进了淀粉结晶状态的崩溃，随着糊化度的进展，膨胀后的淀粉粒的粒度变大。因此，可以说，影响光谱 1807nm 处实验结果的原因是粒度等物理因素。

在对波长 2100nm (2040~2144nm) 左右的吸光度二次微分值进行的主成分分析 (PCA) 中可以看出，第 2 主成分得分和糊化度之间有明显的直线关系 ($R=0.98$) 如图 17-15。在主成分分析中，一般由测量对象的物理特性产生的近红外吸收光谱变动大时，这些物理特性、噪声主要包含在第 1 主成分，其他潜在的信息集中在第 2 主成分中。通过进行主成分分析，可以去除伴随近红外吸收光谱淀粉结构变化的间接影响因素，提取出伴随糊化的只与淀粉结构变化直接相关的信息。

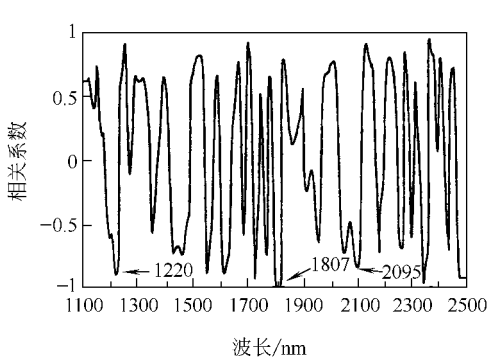


图 17-14 相关系数与糊化度

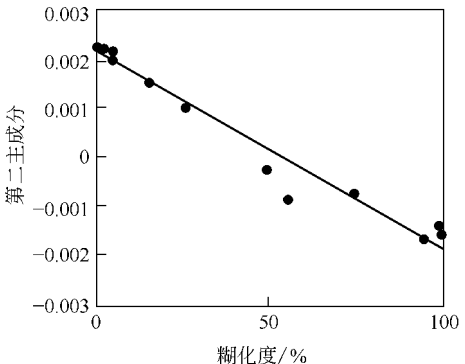


图 17-15 第二主成分与糊化度间的关系

17.3.2 颗粒粒度

颗粒的大小影响到光谱的基线漂移和光谱图的质量。刘建学等人利用微分光谱的二阶导数建立数学模型，有效地消除了大米颗粒对基线漂移的影响。从表 17-3 和表 17-6 中的

预测标准差、平均绝对偏差以及平均相对误差、 R^2 几个评价指标在不同粒度之间的值没有明显变化可以得出，二阶导数谱可以消除粒度不同对测定结果的影响，而且相关系数也在 0.9 附近。所以，根据近红外光谱法对大米蛋白质含量的预测结果，以二阶导数谱作为光谱参数应为首选方法。

表 17-6 二阶导数谱的大米蛋白质含量定标和校正结果

粒度 /mm	样品	定标集容量 /个	校正集容量 /个	主成分数 /个	定标标准差	预测标准差	平均相对误差/%	平均绝对偏差
0.15	精米	60	37	8	0.27	0.44	3.5	0.34
0.15	糙米	60	37	6	0.31	0.49	2.6	0.25
0.20	糙米	60	36	13	0.28	0.44	2.4	0.24
0.28	糙米	60	36	9	0.33	0.52	3.1	0.30

食品质量常表现在以下几个方面：①营养质量，用化学方法进行测定；②感官质量，用感官板进行测定；③工艺质量，用物理方法进行测定；④卫生质量，用微生物方法进行测定；⑤伦理质量，由消费者的观念进行评定。这些评定指标都是应用传统的离线测定方法，常常需要破坏样品，耗费时间，需要装配良好的实验室和经过培训的操作人员，有时还需要具有危险性的化学试剂。因此，这些测定方法是非常昂贵的。为控制一个产品的加工过程和质量，就需要一种快速，费用低和非破坏性的离线或在线的测量方法。因此，需要一种简单的方法来测定几种不同而有关联的质量指标。

NIR 这种方法适合于离线和在线（at- and on-line）测定，它能够完成以上所提到的各种指标的测定。

17.4 奶制品质量分析

17.4.1 概述

经济和法律都要求对奶产品的成分进行严格的控制。例如，牛奶中的脂肪是牛奶中最重要的成分，所有奶产品中的脂肪含量受法律控制或由消费者确定其含量。在荷兰，对农民实行按质付费方案，根据奶中的蛋白和脂肪含量付给不同的价钱，这就强迫牛奶业分析成千上万的鲜奶样品。另外，在世界范围内，竞争变得越来越激烈，奶粉、黄油、奶酪有大量的剩余，特别是在欧洲和新西兰。奶产量定额意味着农场主需要把定额转化成生产含有适宜含量的脂肪、蛋白和酪蛋白的牛奶。牛奶业也供应以奶为基础的婴儿食品和其他奶产品。这些产品中必须包含所要求的成分以满足营养的要求。

官方推荐的分析方法，如水分、蛋白、脂肪、乳糖的测定方法，都需要太长的时间，不能满足加工业对加工过程进行控制的要求。因此急需这样的仪器，它能够对质量和其他一些重要的经济指标进行快速检测。近红外反射光谱为这种需求提供了很大的可能性，多年来一直在实践中应用。

用 NIR 快速分析奶中的脂肪、蛋白和总固形物的含量已经对奶产业产生了巨大的影响，NIR 已是 AOAC 承认的方法。奶粉，奶酪和其他的固体样品在用近红外吸收光谱测定前一般需要溶解在溶液中以使其混合均匀。反射光谱测定的是从样品表面对光反射回来

的密度，因此，消除了上面对样品制备的要求。特别是，NIR 光谱测定运用反射光谱这种原理。酪蛋白、黄油和奶粉的 NIR 光谱已由 Goulden 测定，结果表明可选用几个波长来测定样品的组分。

应用 NIR 测定奶粉的化学成分已经被广泛地研究。Birth 和 Washan 提高了用反射光谱测定蓝奶酪的重复性。Giangiacomo et al 研究表明反射测定与粉状冻干蓝奶酪的成分之间存在显著的相关。Frank 和 Birth 的研究数据表明奶的主成分的吸收谱带之间存在着广泛的重叠。

17.4.1.1 NIRS 在奶牛业中的应用

NIRS 可应用于奶牛业中的不同领域，归纳为 3 个方面：

① 分析收购来的牛奶中的成分，如脂肪、蛋白质和乳糖，以确定付给农场主的费用并按奶中的脂肪、蛋白质和固形物含量把奶进行标准化处理。

② 分析干燥产品中的脂肪、水分、蛋白质的含量以获得制造特殊产品中各成分的适当比例，并分析用于生产奶油、奶酪、酪蛋白和乳清的原材料中的成分。

③ 用于终产品分析进行产品控制，以符合法律规定和消费的特殊要求。

政府或其他官方机构进行牛奶产品质量控制的实验室可应用 NIRS 作为快速分析未知样品成分的一种方法。除 NIRS 外，没有其他指标在过去与化学或物理方法之间存在明显的相关。

17.4.1.2 标准方法

实验室的湿化学分析的标准方法在各国间有很大的变化。然而，表 17-7 中的方法在奶牛业中被接受为标准方法。

表 17-7 奶牛业中湿化学分析标准方法

组 分	方 法	组 分	方 法
水分	Toluene	乳糖	Dye binding
	Oven		Photometric
	Karl Fisher		Reductometric
	Vacuum		Polarimetric
脂肪	Gerber	乳酸和乳酸盐	Titrimetric
	Soxhlet extraction		Enzymatic
	Gravimetric		Photometric
	Rose-Gottlieb		Enzymatric
	Schmid-Bondzynsk	灰分	Furnace
	i-Ratzlaff	盐分	Potentiometric
	extraction		Titrimetric
	Mojonnier		Gravimetric
蛋白质	Babcock	pH	Electrometric
	Kjeldahl		
	凯氏定氮法		

17.4.2 定标程序

NIRS 分析的准确性决定于定标的准确性。所以，定标一台 NIRS 分析仪，需要大量的费用并需要达到很大的准确性。定标就是先获取一套定标样品，通过标准参考方法获得

某些参数值。为获得良好的定标方程最少需要 40 个定标样品。一套定标样品应能覆盖将来要分析的样品范围。

为获得功能良好的定标，需要一段时间采集定标样品。可根据样品的组分、产品生产过程、生产阶段和生产区域收集定标样品。定标样品的数量范围难于准确设定，作为一个基本原则，此范围应不低于主成分的 2%，不超过主成分的 20%。但应知道大部分奶产品成分在一段时间内会发生变化。奶粉能够非常快速地吸收和散发水分，脂肪含量高的产品在数月内会发生脂肪的氧化，乳糖和其他糖含量高的产品会发生结晶现象。氧化和结晶过程能够改变脂肪和碳水化合物的吸收特性。适当贮存可以防止吸收光谱产生很大的变化。最好是用鲜样进行定标，使样品分析过程与光谱收集在同一时间内进行，为了获得良好的定标结果，用于定标的实验室参比方法必须准确可靠。当收集到覆盖一定范围并足够多的样品时，即可根据光谱数据与实验室参比方法的分析数据进行回归分析，并获得一套定标常数。每个产品及每个组分都有它自己的一套定标常数。

NIRS 仪器都带有定标软件，可进行统计分析，并提供为不同定标选择合适滤光片及波长组合的程序。

对某一特定成分，为测定其成分含量需间隔 3~8 个波长进行吸光度的测定。标准差、相关系数和 t -检验是应用多元线性回归获得一个好的定标的最重要的参数。定标方程最终的检验是用一组未参与定标的检测样品，看其实测值与 NIRS 预测值之间的变异程度。如果定标与参考方法获得了很好的一致性，定标方程即可用于常规分析。

重要的是，要经常对定标方程变异性进行检验，以保证定标测定结果的准确性。

17.4.3 奶

奶的近红外吸收光谱与水的很相似，如图 17-16。奶与水的区别是奶中包含了分散于水中的脂肪球和蛋白胶粒。一束电磁辐射穿过这样一个颗粒分布稀疏的介质时，电磁辐射走过的路径大于样品厚度，同时伴有吸收的增强或者吸收波段的增加。小微粒的介入可能带来错误的信息，因消光和透射光（不真实了）部分收集会影响波长。这些问题可通过用积分球收集透射光来克服。

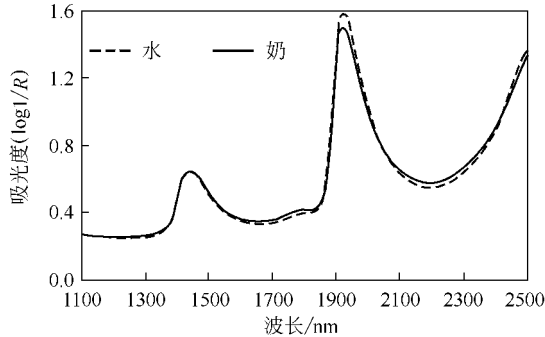


图 17-16 奶和水的近红外反射光谱

另一个常见问题是水的强吸收带覆盖了奶中蛋白质和脂肪的大部分吸收光谱。

这种覆盖作用使建立特殊波段与组分之间的相关变得困难。另外，标准奶中蛋白质和脂肪含量的微小变化只引起 NIR 光谱的微弱变化。因此，有必要对原始数据进行一阶导数和二阶导数处理，以分离和校正奶中被覆盖的蛋白质和脂肪的特殊波段 [图 17-17 (a) 和 17-17 (b)]。用这样的方法可获得可靠而精确结果。

NIRS 配备一个特殊的带有液体室的可推拉装置，就可进行奶及其加工产品中脂肪、蛋白质、乳糖和总固形物含量的测定。它应用透射和反射原理，包含了透射和反射的优点。在这种方式下检测到的反射光包括来源于液体表面和小室后面两方面的光。后一种方

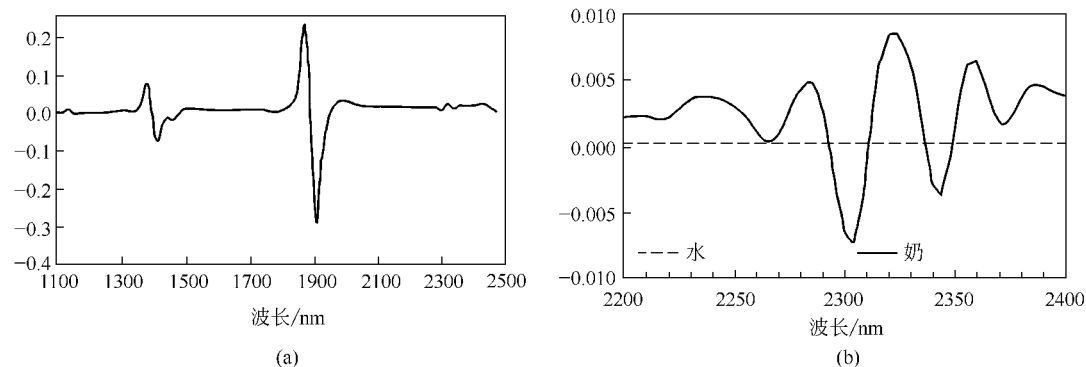


图 17-17 奶和水及其脂肪的二阶导数光谱

(a) 奶和水的二阶导数光谱 (b) 奶和水中脂肪在 2304nm 和 2343nm 的二阶导数光谱

式测定的光是穿透了小室中含有的薄层液体。高压泵和三阶段的均质化处理将使牛奶乳状液均匀到在测微计下可观察到的微粒状态。液体测量装置还提供了温度控制和自动清洗功能。为获得准确的测定结果，样品需在水浴中升温到 45℃ 时放到进样口外进行搅拌，在新鲜状态进行测定。

对 NIRS 测定混合奶（几个来源于不同个体的奶样混合在一起的混合样）中的脂肪、蛋白和糖的可行性研究表明，用 NIRS 的测定结果来进行奶品的标准化处理是可以接受的（表 17-8），对 NIRS 测定山羊奶中的脂肪、蛋白质、乳糖含量的研究表明，由 30 个来自不同个体的检测样品组成的检验样品集的 SEC_s （表 17-9）比用混合奶制成的定标样品的 SEC_s 高 5~10 倍（表 17-10）。

表 17-8 用 InfraAlyaer 500（配有液体样品室和均化器）测定的鲜牛奶（混合样品）
组分值/（g/100g）与参比值的比较

组分	n	SEP	含量范围	要求的精确性
脂肪	50	0.034	4.1~7.5	0.04
蛋白质	50	0.038	2.1~2.8	0.04
乳糖	50	0.043	3.5~4.1	0.05

注： n = 样品数； SEP = 预测标准差；参比方法：脂肪：Gerber；蛋白质：Kjeldahl；乳糖：enzymatic。

表 17-9 鲜山羊奶（个体样品）NIRS 测定的组分值/（g/100g）与参比值的比较

组分	n	SEP	含量范围
脂肪	30	0.226	2.8~7.1
蛋白质	30	0.170	2.8~4.1
乳糖	30	0.281	3.5~5.1

注： n = 样品数； SEP = 预测标准差。

表 17-10 鲜山羊奶（混合样品）NIRS 测定的组分值/（g/100g）与参比值的比较

组分	n	SEP	含量范围
脂肪	20	0.022	3.0~6.3
蛋白质	20	0.028	3.0~3.9
乳糖	20	0.058	4.5~4.9

注： n = 样品数； SEP = 预测标准差。

这种差异不仅是由于样品组成的变异引起的，更主要原因是由于某些山羊的乳房炎而

引起蛋白和脂肪组成的差异。然而通过添加样品和增加定标波长数，可能使定标方程具有更高的精确度（表 17-11）。

表 17-11 鲜山羊奶（混合样品和个体样品）NIRS 测定的组分值/（g/100g）与参比值的比较

组分	<i>n</i>	SEP	含量范围
脂肪	50	0.096	2.8~7.0
蛋白质	50	0.079	2.8~4.1
乳糖	50	0.133	3.3~5.1

注：*n*＝样品数；SEP＝预测标准差。

综上所述，NIRS 测定奶中的脂肪、蛋白质和乳糖含量是成功的。根据我们测定牛奶和山羊奶的研究经验，关键问题是确定具有代表性的样品集，选择具有良好重现性的检测样品方法，并鉴定具有异常光谱的未知样品方法。

Bertrand et al 认为主成分分析（PCA）能够选择具有相关光谱的定标样品（不依赖于化学组成）并且根据异常光谱确定超线样品。

17.4.4 奶粉

17.4.4.1 概述

奶粉因其具有均匀的粒度和形状以及组分的稳定，因此它最适合于 NIRS 分析。奶粉粒度小，不需要粉碎。但重要是，奶粉必须被很好地混合搅拌。粉状物可以简单地被装入样品杯中，表面易于形成水平面并且易于被背板密封。为获得功能强的定标，我们用了两年时间从最重要的奶粉加工 EEC 国家收集奶粉样品。按照组分、加工过程和生产日期选择样品进行定标。样品以两个重复按如下方法进行分析：水分用烘干法测定（102℃ 下烘 2h）；蛋白质用凯氏定氮法（总氮 * 6.38）；脂肪用 Rose-Gottlied 浸提法；灰分含量用干灰分法（500℃ 灼烧 1h）。用酶法测定乳糖和乳酸盐的含量。按照 EEC 方法 L296/10 测定等酪蛋白的酶凝固结果，以此检测结果计算奶牛饲料中脱脂奶粉的含量。当奶粉的温度达到室温后，用 InfraAlyzer400 检测反射光谱值。

用能代表定标样品的一组检验样品来检测奶粉中的所有成分。但在检测期间，获得的检验样品集并不与定标样品集完全一致，因此检验样品集的 *R* 值有时小于定标样品集 *R* 值。

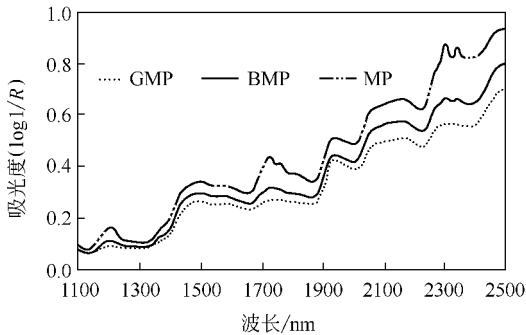


图 17-18 脱脂奶粉、酪奶粉和全奶油奶粉的典型 NIR 反射光谱
GMP—脱脂奶粉 BMP—酪奶粉
MP—全奶油奶粉

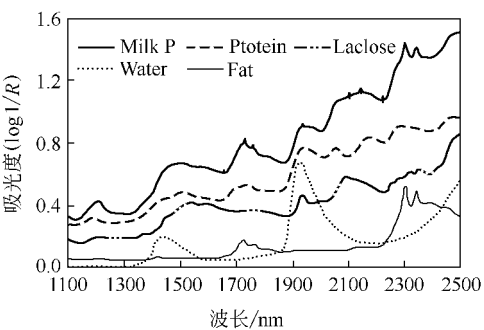


图 17-19 脱脂奶粉和各分离组分的 NIR 反射光谱
Milk P—全奶 Water—水分
Ptolein—蛋白质 Fat—脂肪
Laclose—乳糖

为确保选择正确的定标波长，用 InfraAlyzer500（图 17-18 和图 17-19）扫描了几个不同奶粉样品及提纯的奶蛋白质、脂肪和乳糖的光谱值。测定奶粉中脂肪、蛋白质、水分和乳糖的参考波长列在图 17-19 和表 17-12 中。

表 17-12 奶粉组分的吸收波段

滤光片号 ^①	吸收波段 ^② /nm	在此波段产生吸收的成分	滤光片号 ^①	吸收波段 ^② /nm	在此波段产生吸收的成分
20	1650	水分, 蛋白质, 乳糖参比波长	14	2100	蛋白, 乳糖
16	1940	水分	9	2139	蛋白, 乳糖, 水参比波长
17	1734	蛋白, 脂肪	6	2230	脂肪, 蛋白, 乳糖的参比波长
15	1759	脂肪, 蛋白	4	2310	脂肪, 蛋白

注：① 与给出的波段对应的 Technicon InfraAlyzer 400 的滤光片号；

② 近似的波长值。

17. 4. 4. 2 结果和讨论

(1) 蛋白质

用凯氏定氮法测得的奶中蛋白质含量与 InfraAlyzer 400 测定的光谱值之间存在高度相关（表 17-13）。*R*（相关系数）在 0.97~0.999 之间，定标方程的标准差在 0.18%~0.34%之间，这个结果与参考方法标准差 0.15%相比是可以接受的。

表 17-13 某些奶粉样品蛋白/（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
脱脂奶粉	159	0.97	0.27	32.9~41.2	100	0.97	0.25	33.5~40.6
黄油奶粉	116	0.97	0.21	28.9~34.6	50	0.97	0.21	29.2~34.6
脱脂奶粉+黄油奶粉	274	0.98	0.34	28.9~41.2	150	0.97	0.36	29.2~40.6
变性奶粉	92	0.98	0.34	27.1~40.1	50	0.96	0.42	28.5~39.7
添加非脂肪的奶粉	61	1.00	0.18	17.6~26.0	50	0.98	0.25	18.4~25.4

注：定标参数：*n*=样品数，*SEC*=定标标准差；预测参数：*n*=样品数，*SEP*=预测标准差。

重要的是凯氏定氮法测定的蛋白质含量是粗蛋白质而不是纯蛋白质的含量。粗蛋白质含量或多或少与真蛋白质含量不同，因季节导致非蛋白氮（NPN）含量的波动。也许 NPN 与真蛋白质 NIR 吸收光谱不同，因而导致粗蛋白质定标的标准差高于真蛋白质定标的标准差。

预测样品集的 *SEP* 和 *R* 值与定标样品集结果相同，因而可得出结论认为定标的性能良好而可信。用 4~8 个滤光片获得了好的定标结果，蛋白质吸收的波段在 2319nm 和 1734nm。

(2) 水分

烘干法测得水分与 InfraAlyzer 400 测得水分之间的相关系数为 0.97~0.99。定标的平均标准差为 0.12%，是可以接受的（表 17-14）。检验样品集与定标样品集的结果具有可比性。定标程序选定 3~6 个滤光片，对水分吸收影响最显著的滤光片为 1940nm。

表 17-14

某些奶粉样品水分/（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
脱脂奶粉	159	0.97	0.12	2.9~5.8	100	0.97	0.11	3.1~5.7
黄油奶粉	116	0.99	0.13	2.7~6.3	50	0.99	0.10	2.7~6.3
脱脂奶粉+黄油奶粉	274	0.97	0.13	2.9~6.3	150	0.99	0.14	2.7~6.3
变性奶粉	100	0.97	0.11	2.8~6.3	50	0.96	0.16	3.1~6.2
添加非脂肪的奶粉	64	0.98	0.13	1.3~4.4	50	0.96	0.14	1.5~4.3

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

奶粉水分的测定采用烘干法。烘干法受空气中相对湿度的影响。空气相对湿度影响水分测定的重复性。用烘干法测定的水分结果进行 InfraAlyzer 400 的标定，NIR 的精确性受到参比方法准确性的高度限制。NIR 制定水分的线性关系建立在用大量的从分布不同时间阶段的样品基础上。在这种情况下可消除空气中相对湿度影响。为了验证这一点，24 个具有均匀粒度经喷雾干燥的脱脂奶粉样品被密封贮存起来。在 2 个月的时间里每隔几天用烘干法和 InfraAlyzer 400 同时测定样品中的水分含量。

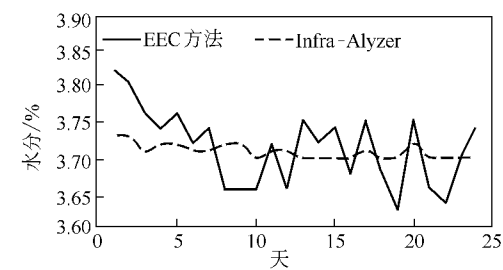


图 17-20 脱脂奶粉在 2 个月的时间
内水分与时间的相关关系

图 17-20 给出了 2 个月时间内水分随时间变化的关系，NIR 的重复性明显地好于烘干法的重复性，但两种方法都给出了与预期相同的平均值。NIR 测出的水分值比烘干法测值更精确。

(3) 脂肪

进行脱脂奶粉脂肪含量的定标不可能得到足够大的脂肪含量范围。因为这种产品已经过标准化处理，所以定标结果不理想。所有定标

的相关系数高于 0.98，定标的标准差在 0.09%~0.31% 之间（表 17-15）。0.31% 的标准差是来自于脱脂奶粉和黄油奶粉的混合物样品定标。混合样品定标的标准差是单个样品的 2 倍，原因是混合样品组分的变异升高。另一方面，不含有奶中脂肪的奶粉样品在脂肪含量上有一个宽的变异范围。其 *SEC* 值是 0.30。把此定标分成两个定标，*SEC* 值降为 0.14。预测方程的 *SEP* 和 *R* 值与定标方程相符。

用 5~7 个滤光片获得的定标结果很好，特别是脂肪的吸收峰为 2310nm 和 1759nm。

表 17-15

某些奶粉样品脂肪/（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
脱脂奶粉	159	0.98	0.09	0.8~2.8	100	0.98	0.10	0.5~1.3
黄油奶粉	116	1.00	0.15	3.3~11.2	50	0.99	0.14	5.3~10.7
脱脂奶粉+黄油奶粉	275	0.99	0.31	0.6~11.2	150	0.97	0.41	0.5~10.7
变性奶粉	92	0.99	0.19	0.6~7.5	50	0.96	0.24	0.8~6.5
添加非脂肪的奶粉(全范围)	62	1.00	0.30	26.2~50.4	50	0.99	0.34	27.2~49.9

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
添加非奶脂肪的奶粉(低范围)	33	1.00	0.15	32.1~41.5	25	0.99	0.14	32.3~40.9
添加非奶脂肪的奶粉(高范围)	29	0.99	0.13	47.4~50.4	25	0.99	0.14	47.9~49.9

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

(4) 乳糖

脱脂奶粉，酪乳粉和含有非奶脂肪的奶粉中乳糖的定标的标准差为 0.38%，0.35%和 0.34%（表 17-16）。考虑到参比方法的标准差变异为 0.25%，上述定标的标准差是可以接受的。脱脂奶粉和黄油奶粉混合样品及变性奶粉定标的标准差太高，不能得到精确的结果。检验样品集中也得到相似的结果。原因是由于样品物理和化学成分的变异及乳糖在整个 NIR 光谱区段有数的特征光谱段而引起的。因此，需要更多滤光片进行定标。大约 10 个滤光片中含有很强的信息。虽然每一个滤光片的相关系数很低，但 10 个滤光片的组合可得到被接受的 *R* 值。

表 17-16 某些奶粉样品乳糖（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
脱脂奶粉	159	0.95	0.38	42.6~51.6	100	0.94	0.44	43.6~50.2
黄油奶粉	116	0.99	0.35	37.0~47.6	50	0.99	0.37	37.1~47.3
脱脂奶粉＋黄油奶粉	275	0.96	0.62	36.6~51.6	150	0.95	0.83	37.1~50.2
变性奶粉	92	0.96	0.62	36.8~49.0	50	0.94	0.75	37.5~48.2
添加非奶脂肪的奶粉	64	1.00	0.34	23.4~37.6	50	0.98	0.38	24.5~36.1

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

(5) 乳酸盐

虽然并不要求分析脱脂奶粉中的乳酸盐的含量，我们仍然作了奶中乳酸盐的定标（表 17-17），因为它能反映出奶中乳糖的含量，乳糖与蛋白质的比率，或者根据乳酸含量添加中性化合物的量。

表 17-17 某些奶粉样品乳酸盐（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
黄油奶粉	116	1.00	0.16	0.0~9.0	50	0.99	0.21	0.1~7.2
脱脂奶粉＋黄油奶粉	275	0.99	0.23	0.0~9.0	150	0.96	0.54	0.0~7.2
变性奶粉	92	0.99	0.10	0.1~4.2	50	0.97	0.19	0.3~4.1

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

脱脂奶中乳酸盐的含量大约为 30~100mg/100g，含量太少，而难于用 NIR 测定。用脱乳奶粉和酪乳粉的混合样品同时定标导致精确度的提高。

为了测定酪乳和变性奶粉中的乳酸盐，定标采用了许多滤光片来获得信息量很强的数据。样品中物理和化学成分的变异和整个 NIR 光谱中乳酸盐的特征光谱区段少而导致需

要多个滤光片。

(6) 灰分

与多数有机化合物相比，大多数无机化合物在 NIR 区段没有特征反射信号。在 500℃ 下测得的灰分含量与 NIR 预测值之间的相关是建立在与有机物或与奶中结合盐分的相关基础上（表 17-18）。

表 17-18 某些奶粉样品灰分/（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
脱脂奶粉	159	0.76	0.08	7.7~8.3	100	0.73	0.11	7.6~8.2
黄油奶粉	116	0.97	0.19	6.9~10.0	50	0.92	0.23	6.9~9.5
脱脂奶粉+黄油奶粉	274	0.85	0.23	6.9~10.0	150	0.85	0.37	6.9~9.5
变性奶粉	100	0.86	0.18	6.6~9.5	50	0.79	0.27	4.0~6.4
添加非奶脂肪的奶粉	63	0.99	0.09	3.9~6.6	50	0.96	0.15	6.7~9.4

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

用 InfraAlyzer 预测灰分含量是按照其与所有有机物的含量和水分的相关进行测定的，因而需要大量的滤光片来获得有意义的信息。

因为脱脂奶粉灰分含量的变异范围很小，所以用 NIR 进行测定的 *SEC* 和 *SEP* 值高而相关系数小，特别是对于预测样品集。对于其他产品的灰分进行数量分析也是可能的，然而分析结果的精度低。

(7) 脱脂奶粉的含量

脱脂奶粉通过添加非奶来源的脂肪可转化成小牛饲料。因为这种产品是 EEC 援助的，测定小牛饲料中脱脂奶粉的含量按 EEC 描述的方法进行测定。测定脱脂奶的含量可通过酪蛋白的酶凝固法进行测定，它是一种很费时间的测定方法。另外，酶法缺乏精确性。NIRS 预测奶粉含量的工作已经进行。

如果脱脂奶粉的含量是基于上述提到的方法，再加上蛋白质、乳糖、脂肪和灰分的含量则预测结果会更好。因为这些原因，InfraAlyzer 400 的定标是基于这些参数进行的。

结果表明，NIR 可定量地测出脱脂奶粉在奶粉和非奶来源脂肪混合物中的含量。（表 17-19）。

表 17-19 含有非脂来源脂肪的奶粉中脱脂奶粉含量/（g/100g）的 NIR 反射光谱定标数据

产品名称	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
添加非奶脂肪的奶粉	63	1.00	0.43	50.3~76.8	50	0.98	0.55	50.2~75.8

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数。

17.4.4.3 重复性和重现性

本研究计算了用 InfraAlyzer 400 测定脱脂奶粉中水分、脂肪、蛋白质、灰分和乳糖含量的重复性和重现性。

24 个脂脱奶粉的次级样品以 2 个重复测定为基础在 24 天的连续工作日中用 InfraA-lyzer 400（表 17-20）和参考方法（表 17-21）进行测定。从表 17-13 和表 17-14 中可总结出 InfraAlyzer400 测定奶中水分、脂肪和蛋白质的含量与标准方法相同或好于标准方法。同时，所有参数用 NIR 测定的重现性与标准方法相同或好于标准方法。

表 17-20 用 InfraAlyzer 400 测定脱脂奶成分的重复性和重现性

	平均值	<i>r</i>	<i>R_W</i>	<i>CV_{R_W}</i>
水分	3.710	0.034	0.034	0.3
脂肪	0.712	0.037	0.042	2.1
蛋白	35.705	0.18	0.18	0.2
灰分	7.964	0.085	0.085	0.4
乳糖	48.779	0.39	0.49	0.4

注：*r*＝重复性＝2.83*S_r*；
R_W＝覆盖整个实验室范围的重现性＝2.83*S_{R_W}*；
CV_{R_W}＝覆盖整个检测范围的变异系数＝*S_{R_W}*/平均数·100％。

表 17-21 用参比方法测定脱脂奶成分的重复性和重现性

	平均值	<i>r</i>	<i>R_W</i>	<i>CV_{R_W}</i>
水分	3.712	0.068	0.15	1.4
脂肪	0.718	0.037	0.042	2.1
蛋白	35.802	0.207	0.327	0.3
灰分	7.952	0.029	0.075	0.4
乳糖	49.021	0.180	0.478	0.4

注：*r*＝重复性＝2.83*S_r*；
R_W＝覆盖整个实验室范围的重现性＝2.83*S_{R_W}*；
CV_{R_W}＝覆盖整个检测范围的变异系数＝*S_{R_W}*/平均数·100％。

17.4.5 酪蛋白和酪蛋白盐

酪蛋白的生产是通过沉淀奶中的蛋白质并把沉淀物干燥到大约 8％的水分而得到的。酪蛋白是一种白色的结晶物质，根据沉淀的过程人工生产两种形式的酪蛋白：酸性酪蛋白和凝乳素酪蛋白。酪蛋白的粒度包括不同的目数和未粉碎的酪蛋白。最普遍的目数是 30 目，60 目，90 目和 100 目。为制备酪蛋白的钠或钙盐，把酪蛋白、水和氢氧化钠或氢氧化钙混合在一起并把混合物喷雾干燥到大约 4％的水分。生产出来的粉状物是非常精细的极易溶于水的白色粉末。因酪蛋白盐具有极小的粒度所以不需要粉碎。60 目、80 目、100 目的酪蛋白不需要粉碎，但 30 目和未粉碎的酪蛋白需要用实验室磨进行短时间的粉碎（大约 20 秒）。制造酪蛋白和酪蛋白盐的

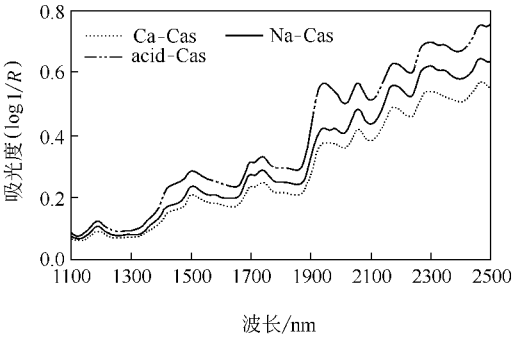


图 17-21 喷雾干燥的酪蛋白钠盐，钙盐和酸性酪蛋白的 NIR 反射光谱

方法不同不仅使产品具有不同的粒度、形状和组成，而且反射光谱特性也有差异（图 17-21）。因此最好对不同类型的产品分别进行定标。因为产品有严格的规格要求，所以很难获得覆盖一定组分含量范围的相关样品。一些分析者因而通过一些方法控制定标样品的组成。为获得样品的一定的水分含量范围，可把样品烘干和把样品暴露于空气中吸收水分来获得。脂肪是酪蛋白和酪蛋白盐中另一个保持恒定的组分。然而，通过混合低含量和高含量的样品得到组分的不同含量范围不能代表样品的加工过程，因此也不能给出精确的结果。另一种获得足够含量范围定标样品的方法是控制酪蛋白和酪蛋白盐加工过程，这是一种相当浪费和昂贵的操作方法。

为了定标 NIR 仪器，我们用了一年的时间从荷兰最重要的酪蛋白和酪蛋白盐加工厂收集酪蛋白和酪蛋白盐样品。然而，特别是对脂肪和灰分，不可能获得具有足够含量范围的定标样品集以建立可靠而功能强大的定标。为获得样品组分更大的变异，试图把酪蛋白和酪蛋白盐放在一起建立一个混合定标，虽然酪蛋白和酪蛋白盐存在上面提到的粒度和反射特性的差异，但定标参数的结果却意外地好（表 17-22）。确定水分可接受的精度为 0.1%，脂肪为 0.05%，蛋白质为 0.3%，灰分为 0.1%，但 NIRS 测定的精确性相对要低一些。然而覆盖组分含量范围广的大量样品的 NIR 与组分含量之间存在线性关系。因此，通过扫描所有类型的酪蛋白和酪蛋白盐可获得可靠而强大的定标，这已被 100 个含有不同组分的检验样品所证实（表 17-22）。

表 17-22 用 InfraAlyzer 500 反射测定的酪蛋白和酪蛋白盐混合定标参数

组 分	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
水分	100	0.99	0.14	3.2~8.9	100	0.99	0.15	3.5~8.7
蛋白质	100	0.95	0.41	85.9~92.9	100	0.93	0.50	86.2~92.8
脂肪	100	0.75	0.10	0.52~1.22	100	0.71	0.11	0.6~1.2
灰分	100	0.97	0.25	0.3~4.9	100	0.96	0.27	0.5~4.7

注：定标参数：*n*＝样品数，*SEC*＝定标标准差；预测参数：*n*＝样品数，*SEP*＝预测标准差。

17.4.6 黄油

黄油中的组分，如水分和非脂固形物含量已由标准进行控制。从制造业的经济学角度要求终产品尽可能与目标值接近。因此在工业中大范围的传统的经验性检验已改进为快速测定这些参数以能够立即对生产过程进行校正。NIRS 技术的优点是其分析的快速性可进行更多的测试，因此可提供更严格的控制。为建立黄油水分和非脂固形物的可靠定标，了解对定标精确性产生负影响的一些因素是很重要的。首先的一个因素是组分的变异。因为产品有严格的规格要求，不容易获得适宜数量的覆盖水分和非脂固形物含量范围大的定标样品。

第 2 个因素是温度。温度相差 5℃ 引起吸收光谱的差异即使经过二阶导数处理也不能消除这种差异，因而对定标精度产生负影响 [图 17-22 (a) 和图 17-22 (b)]。

因此绝对需要标准化进行 NIR 分析的样品温度。第 3 个重要因素是样品的前处理和样品装入样品杯的方法。黄油具有非常高的奶脂肪含量，压制样品时可导致结构的变化而

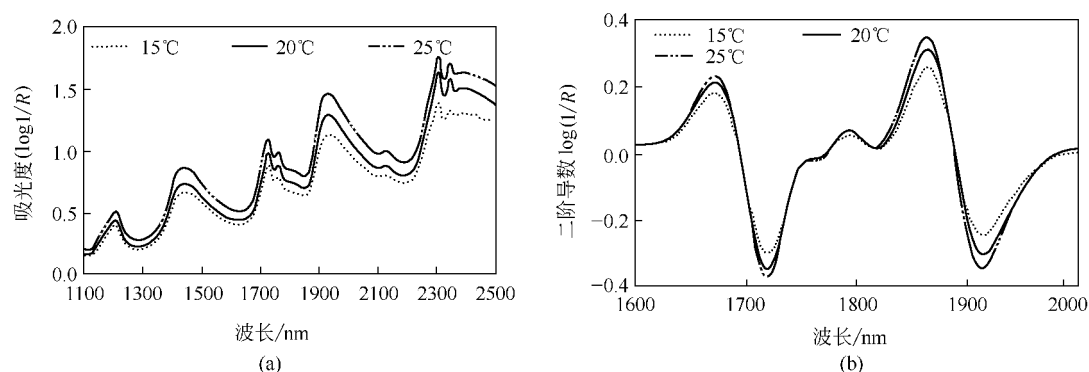


图 17-22 黄油在几种温度下的光谱差异

(a) 黄油在 15℃, 20℃ 和 25℃ 获得的 NIR 反射光谱 (b) 黄油在 15℃, 20℃ 和 25℃ 获得的 NIR 反射光谱的二阶导数光谱

引起反射光谱的变化。因此样品的准备和装入样品杯的方法必须标准化。硬化黄油不经过揉捏很难装进带有石英窗的样品杯, 而且密封的样品杯测定下一个样品前需要更多的时间进行清理, 所以选用敞开的样品杯。一些经验表明可以用小勺把黄油装入敞开的样品杯并产生适合质量和组分分析的表面。

从 4 个不同的工厂收集了 50 个硬化黄油样品 (酸性未着色), 用参考方法和 NIR 方法以 2 个重复的方式对样品进行分析。选择最能代表定标样品的一套检验样品对定标进行检验。定标的水分和非脂固形物标准差为 0.07% (表 17-23), 而参考方法的变异为 0.05%。检验样品的 *SEP* 与定标相符。因而可得出结论: 黄油样品水分和非脂固形物的定标是可以应用的。水分的定标程序表明有 3 个滤光片含有很强的信息, 包括水分特征滤光片 1940nm, 非脂固形物的定标选择 4 个滤光片, 包括脂肪滤光片 1759nm, 水分滤光片 1940nm 和蛋白质滤光片 2100nm。

表 17-23 反射法测定黄油水分和非脂固形物/ (g/100g) 的定标数据

组 分	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
水分	50	0.96	0.07	15.3~16.36	50	0.94	0.08	15.33~16.33
非脂固形物	50	0.94	0.07	1.59~2.41	50	0.88	0.09	1.63~2.37

注: 定标参数: *n*=样品数, *SEC*=定标标准差; 预测参数: *n*=样品数, *SEP*=预测标准差。

17.4.7 奶酪

奶酪是最难于用 NIRS 进行分析的物质之一, 因为它的加工过程不同, 物理特性和化学组成不同, 并且含有高含量的水分和脂肪。因此标准化样品制备和准备程序非常重要, 这样才能获得精确而具有代表性的测定结果。研究的目的是探讨应用 NIRS 测定荷兰最重要的两种奶酪: 伊顿奶酪 (Edam) 和荷兰扁圆奶酪 (Gouda) 的常量和微量成分及成熟时间的可行性。为此, 从荷兰 6 个不同生产厂收集这两种奶酪样品大约 200 个。样品的前处理是采用先去掉干酪皮, 再用装有符合荷兰标准的奶酪装样装置的磨粉碎。样品以两个

重复的方式分析下列指标：水分，盐分和 pH。pH 的测定按荷兰国标进行。脂肪含量按照 Schmid-Bondzynski-Ratzlaff 浸提法进行测定。总氮量按照荷兰国标，用凯氏（Kjeltec）自动分析系统测定。蛋白含量按总氮量乘以 6.38 算得。可溶性氮按 Noomen（30）的方法进行测定。

样品温度差异将影响结果的精确性，脂肪含量高的样品可能引起结构变化而导致反射光谱的差异。因而在装样时应避免使粉碎后的奶酪长时间暴露于空气中。

下面是 RIKIL 为 NIR 分析而进行奶酪样品的准备程序：

- ① 按荷兰国家标准采样；
- ② 按荷兰国标准备样品（注意：如果样品不能及时分析，应贮存在一个低温的地方（8℃左右），在分析前几小时把样品从贮存的地方取出来，使其达到室温（20℃），但不要人为加热，样品应在装样器内用样品勺混合均匀）；
- ③ 用样品勺把样品装入开口的样品杯内，直到装满为止；
- ④ 用样品勺或压舌片把粉碎的样品压到仅低于样品杯的边缘。避免样品散开；
- ⑤ 把样品杯放入 NIR 仪器的样品抽屉内，关上样品抽屉；
- ⑥ 开始分析；
- ⑦ 分析后，移去样品并清扫样品杯；
- ⑧ 预测未知成分含量的样品时，应重复分析三次，重复 3~6 步骤；
- ⑨ 通过定标成分的百分含量关系图计算结果并把结果进行平均（注意：求定标关系图时，每个样品测定一次就够了）。

把三次测定的吸收/反射光谱值平均可降低 SEC 或 SEP 大约 30%。样品温度差异在 5~10℃增加 SEC 或 SEP 10%~50%，这决定于被分析的成分。

（1）常量成分的定标结果

与预想的一样，NIR 测定水分，脂肪和蛋白质（表 17-24）与参比方法之间存在高度相关。NIR 定标的标准差与需要的精度相比是相对比较低的，水分可接受的标准差为 0.2%，脂肪为 0.2%，蛋白质为 0.25%。把一个定标分成两个定标即每种类型的奶酪分别定标效果更好，SEC 降低了大约 30%。从预测样品集得到的结果可以认为样品常量成分的定标是强大而可靠的。为获得理想的定标效果每个指标选择 4 个波长。

表 17-24 用 InfraAlyzer500 反射测定的两种奶酪（Gouda 和 Edam）各常量指标的混合定标参数

组分	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
水分	125	0.993	0.32	34.04~46.04	100	0.991	0.35	43.11~45.98
脂肪	125	0.995	0.26	22.89~43.06	100	0.992	0.28	23.01~34.04
蛋白质	125	0.995	0.37	23.23~28.45	100	0.995	0.36	23.25~28.40

注：定标参数：*n* = 样品数，*SEC* = 定标标准差；预测参数：*n* = 样品数，*SEP* = 预测标准差。

（2）微量成分的定标结果

虽然 NIRS 是一种令人激动的新技术，但它的主要应用是测定常量成分，如所有类型产品中的水分、蛋白质、脂肪和糖类。然而仍然探讨了 NIRS 测定奶酪中微量成分的可能性。研究了 NIRS 测定用于控制的指标：盐分和 pH 的精确性和测定水中可溶性氮或溶于三氯乙酸（TCA）氮的可能性。荷兰国家农业质量控制协会的研究表明水溶性氮和 TCA

溶解氮与荷兰奶酪的成熟时间之间存在高度相关，似乎可通过测定水溶性氮和 TCA 溶解氮来控制年轻奶酪的成熟度。

奶酪各指标的定标参数结果相当好（表 17-25），定标参数包括：食盐，pH，水溶性氮/总氮，TCA 溶解氮/总氮，最初的水溶性胺浓度。二阶导数光谱进行食盐定标效果最好，对食盐的测定是基于水溶性成分而引起的光谱变化。

表 17-25 用 InfraAlyzer500 反射测定 Gouda（48+）和 Edam（40+）各微量成分的混合定标参数

组分	定标参数				预测参数			
	<i>n</i>	<i>R</i>	<i>SEC</i>	含量范围	<i>n</i>	<i>R</i>	<i>SEP</i>	含量范围
盐分(NaCl)	125	0.937	0.15	1.2~3.3	100	0.932	0.16	1.2~3.3
pH	125	0.747	0.08	5.1~5.8	100	0.713	0.08	5.1~5.7
水溶氮/总氮	125	0.916	2.07	11.1~34.7	100	0.892	2.50	11.1~36.0
TCA 溶解氮/总氮	125	0.905	1.34	8.4~22.8	100	0.891	1.51	8.4~24.3
水溶性 初始胺浓度 (μmol/g 蛋白质)	70	0.976	43.1	300~1160	—	—	—	—

Bernal 和 Fowler 的研究表明添加电解质改变了水在谐波波段的光谱。Luck 的研究表明，温度变异引起类似的光谱变化，原因是由氢键的数量变化而引起的。Begley et al. 的研究表明，肉中盐分浓度最好的 NIR 定标产生在这样的波长点，即用扫描型 NIRS 仪可把此点由盐分引起的光谱从其他光谱变异中分离出来。

pH 定标的相关系数低，为 0.747，原因是 pH 的变异范围小，仅为 0.6 个 pH 单位。另一方面，NIRS 测定奶酪 pH 是基于许多非常微弱的重叠谐波。因此用 NIRS 测定 pH 不可能得到高精确度的定标。对奶酪样品 NIR 原始光谱进行数学处理表明下列指标——水溶性氮/总氮，TCA 溶解氮/总氮，水溶性最初胺浓度，与总蛋白之间存在高度相关。这样的结果是可以预见的，因为蛋白质在成熟过程中主要由于酶解作用被降解成肽和氨基酸。因 NIRS 进行蛋白质测定基于酪蛋白分子和各种类型肽和氨基酸对 NIRS 的吸收。很难把蛋白质的吸收波段分解成几个小的吸收波段并把这些波段与产生它们的成分对应起来。仅有的可能性是应用更高次序的数学处理方法。

(3) 测定成熟时间

① 总论 荷兰的法律规定 Gouda 和 Edam 奶酪在出售前在温度低限 12℃ 下至少熟化 28 天。更长的熟化时间法律没有规定。零售的奶酪按熟化时间作如下分类：Young（青年奶酪，至少 28 天），Young-Matured（稍成熟奶酪，至少 2 个月），Matured（成熟奶酪，至少 4 个月），Extra-Matured（特成熟奶酪，至少 7 个月），Old（老奶酪，至少 10 个月），Extra-Old（特老奶酪，至少 1 年）。因同类型奶酪的成熟条件变化不大，可以预见成熟度（成熟的程度）与成熟时间或类型之间存在相关关系。因销售价格由奶酪的类型决定（越年轻价格越便宜），能够区分奶酪的类型是很有意义的，荷兰国家农业控制协会的研究表明，从分析和经济的观点来看，Gouda 和 Edam 奶酪的成熟时间最多与三个独立的指标有关。通过逐步线性回归分析选择的三个指标为：TCA 溶解氮/总氮的比率（它在任何情况下的相关都是最高的），盐分浓度和水分含量。TCA 溶解氮/总氮的比率随时间而增加，但蛋白质的降解过程将随水浓度的降低和盐浓度的升高而降低。因此成熟温度，湿度，酶活性，总的组成等因素的变异都会影响成熟阶段和年龄的关系。各阶段最大的相

关系数在 0.94~0.95 之间。如果区分 Gouda 和 Edam 的 Young（青年奶酪）和 Young-Matured（稍成熟奶酪）95%置信水平的时间是 12~13 天，而区分 Young-Matured（稍成熟奶酪）和 Matured（成熟奶酪）的时间是 18 天。

除了上面提到的化学指标，还研究了 NIRA 测定奶酪成熟阶段的适宜性。用了与进行化学分析同样的样品。从不同的生产商收集了大约 100 个具有不同成熟阶段的样品。奶酪的成熟时间从 25~412 天不等。

② 结果 我们发现总样品集的相关系数为 0.92，标准差（SEC）为 28 天（表 17-26）。SEC 高的原因是这些奶酪样品的成熟时间范围大，为 25~412 天。似乎在这样大的范围内成熟阶段与奶酪年龄之间不存在线性关系（图 17-23）。用一套成熟时间到 160 天的次级样品定标降低了 SEC，相关系数大，为 0.96，SEC 降为 11 天，化学分析中多少也有大于 8 天的情况。更进一步，把次级样品按奶酪类型分成两个样品集：干物质中含 40% 脂肪和干物质中含 48% 脂肪。两种奶酪的定标相关系数高达 0.97，标准差分别为 7.2d（图 17-24）和 10d（表 17-26）。

表 17-26 用 InfraAlyzer 500 反射测定 Gouda 和 Edam 成熟时间的定标参数

奶酪类型	定标参数			
	<i>n</i>	<i>R</i>	SEC	含量范围
40%+48% 脂肪	100	0.92	28	25~412
40%+48% 脂肪	80	0.96	11	25~160
40% 脂肪	40	0.97	7	25~160
48% 脂肪	40	0.97	10	25~160

注：*n* = 样品数，*R* = 相关系数，SEC = 定标标准差。

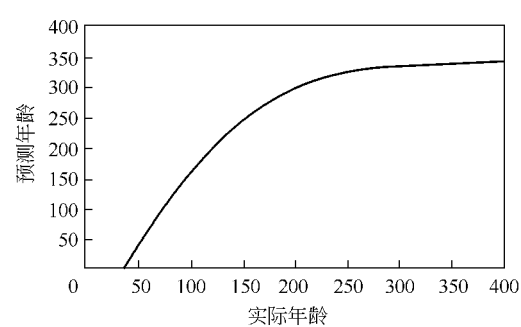


图 17-23 两种类型的荷兰奶酪（25~412d）
实际年龄与预测的年龄相关

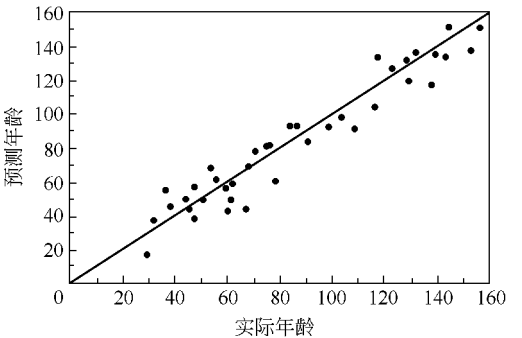


图 17-24 含 40% 脂肪的奶酪样品（25~160d）
实际年龄与预测的年龄相关

令人惊奇的是预测年龄的波长与预测 TCA 溶解氮/总氮、水分百分比和食盐的百分浓度的波长相同。这正好与化学分析中发现的指标一样。因为 Yang-Matured 奶酪是主要产品，用 20 个未知样品对年龄范围在 25~160d 的混合定标（40%+48%）进行验证。结果表明这个定标不足以从数量上区分 Yang-Matured 和 Matured 奶酪。这种奶酪 95% 置信范围的最小耐受限度或拒绝限度为 $1.64 \times$ 残余标准差。对于检测样品 95% 置信范围的 SEP 值大约为 25d ($1.64 \times 15 = 25$)。正像前面提到的一样，在此区域需要的

分析精度和经济准确度为 14d。因此得出结论用 NIRS 分析奶酪的成熟时间是不充分的。

结论：乳牛业对质量控制日益高涨的要求需要发展能快速测定质量指标的仪器。近红外光谱仪具有这种潜能。

17.5 饮料质量分析

17.5.1 概述

NIRA（近红外分析技术）在饮料和营养配方工业中有重要的应用。此技术已应用于测定饮料中的营养成分，包括含酒精的饮料如啤酒、葡萄酒和烈性酒；不含酒精的饮料，如果汁、咖啡、茶和软饮料，及其他产品，如婴儿和成人按营养配方制成的产品。下面将阐述 NIR 在这些方面的一些应用。

17.5.1.1 样品准备

NIRA 测定液体样品与粉状或颗粒状样品在一些关键方面有所不同。液体分析不仅使用真实的光谱反射值，并且更趋向于使用透射或透射和反射的结合值，此结合值有时被称作透射反射方法。液体分析与粉样分析的另一不同点在于样品装入仪器的方法不同。液体样品的前处理包括稀释、悬浮或均化。因此要求一个不同类型的样品槽，一般是某种形状的流动液体槽。

17.5.1.2 液体槽的设计

当设计或选择液体槽时，应考虑以下几个基本因素。检测的方式——透射法或透射反射法将影响检测的结果。其他应考虑的因素是样品的物理性状。液体样品可以是浆状液、悬浮液、溶液或乳浊液。另一个考虑因素是所测样品的使用环境。远端应用将使用流动的液体槽，通过此槽样品可经人工方式注入其中，或经均匀器处理后自动压入其中。在线应用也许将要求样品具有持续流动的能力或在程序控制中，根据需要使槽交替地充满或倒空液体。

设计液体槽经常要考虑温度对测量结果重复性的影响。据报道，在 1980 年早期联邦谷物检测服务试验室就发现，温度作为一个影响因子可使 NIRA 测定蛋白质含量产生不同值。Williams et al 1983 年的研究表明，粉状样品的温度和它们相应的表观蛋白质含量存在负相关。下面给出三种可消除 NIR 蛋白质值差异的建议。一种是选择对温度变化敏感性最小的波长，第二种是使测定样品温度与开发定标方程时温度相同。第三种方法是，通过使用相关关系图以求出一个温度校正因子以消除温度的影响。在液体分析中使用 NIR 技术的工作人员也注意到这种温度敏感性的问题。尽管上述给出解决此问题的建议之一是选择那些对温度不敏感的波长，但大多数液体槽在设计时已具有保持恒温的能力。

17.5.2 酒精饮料

NIRA 被广泛用来分析各种酒精饮料的成分。其中包括啤酒、葡萄酒和白酒。应用范围从直接测定酒精含量到测定啤酒麦芽叶最初浓度（决定酒精的一个功能）。

17.5.2.1 啤酒

人们用 NIRA 测定啤酒中的酒精和其他成分的含量。COVENTRY 和 HUNSTON 已报道用透射方法做出了测定啤酒酒精的定标方程。在各种淡色和黑啤酒中使用最成功的定标方程仅使用一个波长，1672nm。定标的酒精含量范围为 0~11%，标准误差变化范围在 0.2%~0.3%之间。

欧洲酿造业广泛地使用 NIRA 监测酿造程序。NIRA 测定的主要参数包括酒精、啤酒麦芽叶最初浓度（决定酒精的一个功能）、真实浸出物、表观浸出物、糖类以及总固形物。为了保证这些测量数值的重现性，应保持温度恒定。实现此目的另一种方法是，使具有符合温度的水在修改后的固体样品槽内循环。欧洲的科技人员设计了一种可保持恒温的槽。大约可装 20mL 的液体，只需 30 秒钟就可达到需要的温度。表 17-27 是 Dzwinczyk 给出的温度恒定时定标使用的几个关键波长。

表 17-27 测定啤酒成分的定标波长			
组分	波长/nm	组分	波长/nm
酒精(参考波长)	2230	水	1445
酒精	2310	糖分	2100

酿造工业已将 NIRA 技术推广到分析原材料成分方面，如测定大麦和麦芽中的总氨基酸以及啤酒花中的 α -酸。

17.5.2.2 葡萄酒和蒸馏酒

在法国，有 50 多个地方使用 NIR 测定葡萄酒和蒸馏酒中的酒精及糖含量。早在 1976 年 KAFFKA 和 NORRIS 就测定了葡萄酒中的酒精浓度。对于含酒精 11%~17%的葡萄酒，用透射方法测得的定标精确度小于 0.1%。同时也测定了蔗糖含量，对于含量在 0.8%~8.8%的蔗糖水平，定标精确度为 0.2%。他们还检测了葡萄酒酿造工厂感兴趣的另一个成分——酒石酸，浓度范围在 0.5%~8.2%之间，其精确度为 0.1%。

17.5.3 非酒精饮料

NIRA 还定量分析了许多非酒精饮料中的成分，如咖啡中水分和咖啡因。其他应用正向茶，软饮料和果汁的成分分析方面发展。

17.5.3.1 果汁

通过使用透射法和一个 45℃的液体槽，人们测定了果汁的白利度和酸度。表 17-28~表 17-30 列出定标参数以及每个定标所使用的波长。

表 17-28 覆盖整个范围的果汁白利度和酸度的 NIR 定标参数					
	白利度(范围 = 4.4%~19%)	酸度(范围 = 0.3%~2.26%)		白利度(范围 = 4.4%~19%)	酸度(范围 = 0.3%~2.26%)
波长/nm	1445	1828	R	0.999	0.996
	1778	2010	SEE	0.200	0.041
	1818	1576	N	66	66
	2190	2150	F ratio	7348	1779

资料来源：Courtesy of L. P. McDermott, Technicon Industrial Systems, Ref. 9.

表 17-29 果汁白利度不同范围的
NIR 定标参数

波长/nm	白利度含量范围/%		
	4~8	9~14.4	14.5~19
R	1436	1352	1436
	1450	1744	2206
	1506	2220	2388
SEE	0.990	0.995	0.997
N	0.161	0.146	0.127
F ratio	15	29	22
	184	89	1072

资料来源：Courtesy of L. P. McDermott, Technicon Industrial Systems, Ref. 9.

表 17-30 果汁酸度以白利度含量为
基础的 NIR 定标参数

波长/nm	白利度含量范围/%		
	4~8	9~14.4	14.5~19
R	2276	1436	1352
	1954	1828	1534
	2052	1758	1800
SEE	0.999	0.990	0.960
N	0.023	0.038	0.024
F ratio	15	29	22
	4822	405	70

资料来源：Courtesy of L. P. McDermott, Technicon Industrial Systems, Ref. 9.

果汁的碳水化合物浓度是通过 NIRA 透射方法及一个 2.2nm 长的槽进行测定。以这种方式获得定标的应用效果在各种果汁中都非常好，并且在含 9.3%~11.3% 碳水化合物的橘汁中，所求得的相对标准误差为 0.25%。

17.5.3.2 水溶液中的葡萄糖

NIRA 定标现在也应用于测定专为儿童生产的葡萄糖水溶液方面。这些物质一般含有 5%~10% 的葡萄糖。由于从葡萄糖原物质直接配制标准溶液变得十分简单，所以使 NIRA 技术可作为标准 AOAC 方法的替代方法，且其具有十分快速、精确、可再重复的特点。

17.5.4 配方产品

婴儿配方产品和其他满足病人需要的营养配方产品，每年可带给生产厂商几十亿美元的利润。这些液体配方产品是复杂的营养成分悬浮物，其中的营养成分被严格控制。为了监测这些产品，特别是婴儿配方产品，我们需要进行一系列精确的营养成分分析。而这些产品看起来很像奶或奶制品的分析，其实截然不同，它们存在其自身的特殊分析问题。在几次解决这些问题的过程中，我们都借助了光谱学技术，包括 NIRA 在内。

17.5.4.1 历史背景

1970 年后期，人们使用 NIRA 检测婴儿配方产品中的成分含量。最初的目的是测定蛋白质、脂肪、内容物和总固体物质。早期工作随着 NIRA 在奶产品中应用的发展而发展。那时的仪器比较简单，仅使用一个液体槽代替粉样槽即可。样品经预热后，由人工用注射器注入其内。

1980 年早期，应用于其他种类配方产品的 NIRA 技术得到显著发展，已可测定以大豆、奶为基础以及葡萄糖/水配方产品中的组成成分。其中包括在婴儿和医院应用的产品。因此应用在奶和液体方面的仪器得到显著改进。通过使用外部样品准备程序，样品经预热、均化后注入槽内。

在 1986 年，有关 NIRA 配方产品应用的发展扩展到程序处理的环境因素方面。人们还设计出不同程序的定标方程，在不同的程序处理阶段，通过使用台上仪器实验人员就可以检测蛋白质、脂肪和总固体物含量。表 17-31，表 17-32 是这些定标使用的波长及回归

统计数值。NIRA 发展的下一个方向将是实现在线技术方面的上述指标的检测工作。

表 17-31 大豆配方产品中主成分的 NIR 定标参数

	蛋白质	脂肪	固形物
波长/nm	2100	1759	1818
	2180	2270	1982
		2310	2139
		2348	
R^2	0.9933	0.9989	0.9998
SEE	0.0770	0.0529	0.0751
N	31	31	31
F ratio	6316	19378	172592

资料来源：Courtesy of Bristol-Myers Squibb MJNG, Ref. 14.

表 17-32 以奶为主的配方产品中主成分的 NIR 定标参数

	蛋白质	脂肪	固形物
波长/nm	1818	2270	1445
	2100	2310	2100
	2139	2336	1982
	2180		2270
R^2	0.9946	0.9993	0.9995
SEE	0.535	0.0477	0.1446
N	45	45	45
F ratio	5977	62364	59852

资料来源：Courtesy of Bristol-Myers Squibb MJNG, Ref. 14.

17.5.4.2 精确度和准确度

人们已证实使用 NIRA 分析配方产品即省钱又可靠。通过使用一个精细的检测样品程序，确保分析数据的长期精确度和可靠度。因为这些分析数据是有关人员花很长时间，通过 NIRA 技术得到的。表 17-33 给出应用在配方产品的一个精确度和准确度代表性例子。

表 17-33 NIR 分析液体配方产品典型的准确度和精确度

	总固形物	脂肪/油脂	蛋白质
精确度/%RSD	<0.5	<0.3	<0.5
相对准确度/%	±1	±1	±3
N	6	6	6

资料来源：Courtesy of Bristol-Myers Squibb MJNG, Ref. 15.

NIRA 作为一种主要分析技术，应用于特殊液体配方产品的总结见表 17-34。随着食品工业越来越广泛的使用 NIRA 分析技术，所列产品也将增加。

表 17-34 NIRA 在液体配方产品中的应用

应 用	成 分	应 用	成 分
乳浆 酪蛋白婴儿配方产品	蛋白质,脂肪,固形物	完全等渗配方产品	蛋白质,脂肪,固形物
大豆蛋白婴儿配方产品	蛋白质,脂肪,固形物	完全液体食物	蛋白质,脂肪,固形物
葡萄糖水溶液(5%和 10%)	葡萄糖		

资料来源：Courtesy of Bristol-Myers Squibb MJNG

17.5.4.3 结语

NIRA 在分析饮料和婴儿产品上的应用正在快速增加。如果某种产品的销售与市场范围有关，并且因此在质量控制中所需样品的数量也大，这种产品很适合使用 NIRA 技术。投资在仪器、定标软件和操作人员培训方面的费用很快且很容易就会得到回馈。可是至今为止，由于市场竞争激烈，有关于饮料和配方产品使用 NIRA 的资料发表得较少。随着应用和参与使用 NIRA 技术实验室数目的增多，这种情况将得到很好的解决。

17.6 烘烤食品质量分析

NIR 反射和透射技术已成功地应用于大麦的分析方面，并且得到了美国谷物化学家联合会以及国际谷物科学和技术协会的认可。NIR 技术在谷物交易中的成功应用引起了人们极大的兴趣，以期望此技术能推广到更大的食品范围，本节将综述 NIR 反射技术在烘烤业中实际和潜在的应用，包括成分分析（特别是淀粉），加工过程的中产物（生面团）以及终产品（面包、饼干、糕点、专制食品和早餐片）的分析。

17.6.1 面包店产品成分

17.6.1.1 用于烘烤的面粉

面粉几乎是所有烘烤产品中最主要的原料，因此也成为产品质量说明的关键，NIR 在烘烤业最实际的应用就是它对面粉的分析，这一技术在美国和英国已得到普遍应用。NIR 实验操作简单，易于由非技术烘烤人员使用，并且它能瞬时产生结果，使面粉在进入面粉仓之前进行检验。这样我们就可以检查装载的面粉是否和其规格一致，从而也可以防止不合格的面粉进入生产系统，保证烘烤线产品的质量。应用 NIR 检测的面粉烘烤质量参数有：蛋白质、水分、颗粒度、灰分、颜色、淀粉损害度和水分吸收度。

(1) 蛋白质

蛋白质是烘烤面粉中最重要的组成成分，因此也得到 NIR 研究者密切的注视^[1~7]。它也为 NIR 定标方程的发展、合理化、检测提供了最好的例子。蛋白质检测通常用含一定数目（6~19）滤光片或是翻转滤光片（含有可扫描一定范围波长区间的固定波长或滤光片）的 NIR 仪测定。尽管现在已不再生产翻转式滤光设备，但它们对每个样品只记录一种光学数据：两个原始光谱数据的比率，所以其定标方法简单易懂。这样得到的光谱数据与 kieldahl 方法测得的蛋白质值之间通过最小二乘法进行处理。

另一方面，由于粒度影响反射光谱从而产生多共线问题（图 17-25），使应用含固定滤光片的 NIR 仪的定标工作变得困难。多共线现象意味着不同波长处的反射数据（ $\log_1/R$ 或 $\log$ 值）之间的相关比其与蛋白含量之间的相关更密切。此点违反了多元回归分析的一个基本要求即独立的变数（如 $\log_1/R$ ）应该不进行修正。用来开发定标方程的方法通常都基于 $\log_1/R$ ，幸运的是，多元回归可通过在方程中使用参考值来消除粒度的影响。问题在于，要选择一个最合适的测试和参照波长。

从多元回归中选择变量的最普遍做法是一个逐级递进程序，首先选择一个与参照数据相关性最好的变量，然后每次增加一个变量，每一步都使残余标准差最小。当从统计上达到显著时，逐级递进程序即可终止。因为此程序不允许一个被选择的参考值从方程中除去，并且多共线现象使第一个波长的选择具有偶然性，所以此程序对 NIR 波长选择存在负影响，这种影响可通过下面几个方面加以解决：① 通过重复可删去波长，例如，在每次增加之后，一次可删去一个波长并用最好的波长替换，直到每个波长仅能由其自己替换为止；② 开始一次选择两个波长；③ 分析所有可能的组合；④ 在回归前把数据规范化。

典型的固定滤光片型 NIR 仪仅有 6 个可能波长，可计算出包含一个或更多波长的 $2^6 - 1 = 63$ 个可能组合的残余标准差，然后，以残余标准差为标准，比较含有相同数目波

长的等式，从而选出含两个波长的最好几个方程，如此继续，选出含三个、四个波长的最好回归方程等。因为涉及项目越少方程性能越可靠，所以可从基于光谱值的选项中选择一个最好的方程。Osborne et al 使用此方法选择了淀粉蛋白定标的三项方程，包括在 2180nm, 2100nm 和 1680nm 处的光密度。2180nm 对应着谷类蛋白的吸收波，谷类蛋白的吸收波位于淀粉波的肩部，而淀粉最大吸收波为 2100nm（图 17-26），所以上述三项方程比较合理。因此，2180nm 是蛋白测定波长，而 2100nm 可以对 2180nm 处淀粉产生的影响做一个调整。定标方程中包括 2100nm 波长的另一个原因是面粉淀粉和蛋白间存在着高度负相关，它是面粉的一个重要特性。淀粉特征波长对蛋白的定标方程有重要影响。然而必须小心使用包含这种相关关系的定标方程。因为它可能会产生不稳定的数值（面粉中的粗纤维、灰分），1680nm 是一个中性波长，可调整样品粒度大小的差异。

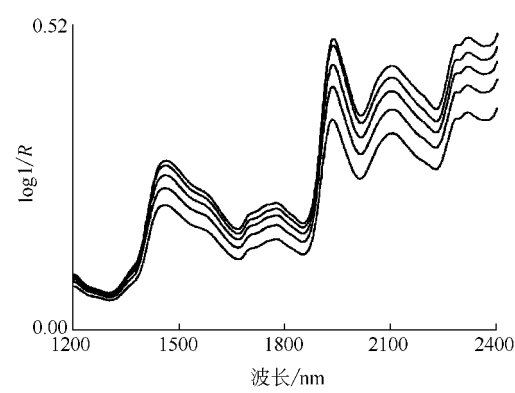


图 17-25 5 种不同粒度面粉的 NIR 光谱

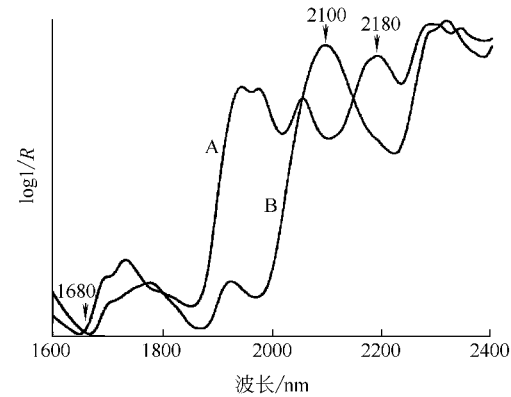


图 17-26 小麦面筋(A)和小麦淀粉(B)的 NIR 光谱

固定滤光片型仪器定标的另一种替代方法是在回归分析前使数据标准化。随微机在滤光片型仪器中应用的增加使实验室的质量控制成为可能。最简单的做法是所有的 log1/R 数据用中性波长即 1680nm 处的数据去除。然而更加满意的做法是用平均光谱值做除数或旋转每个样品的光谱以使其分布在平均光谱的周围。实现后者的计算方法称作附加散射校正（MSC），其过程如下：如果 y_{rw} 代表第三个样品在波长 W 的 log1/R 值， y'_{rw} 代表所有定标样品在 W 波长时的 log1/R 平均值，所有波长的数据可通过最小均方校正得到： $y_{rw} = a + by'_{rw}$ 通过下式计算用于调整波长的 a、b 估测值：

$$y'_{rw} = \frac{y_{rw} - a}{b}$$

虽然不可能完全消除多共线现象，但却被降低了（图 17-27），所以应尽可能使用逐级递进回归方程来建立一个定标方程。用 MSC 算法处理上述面粉蛋白定标数据的结果如下（57 个样品， $w=2310\text{nm}, 2230\text{nm}, 2180\text{nm}, 2100\text{nm}, 1940\text{nm}, 1680\text{nm}$ ）：逐级递进回归方程先选 2180nm 然后是 2100 nm，尽管这种选择比较接近（因为 2100 nm 与蛋白存在高度负相关）。第三个选择波长是 2230 nm，它与那些只含三个变量的定标的波长（1680nm）有所不同，但具有相同的作用，因此 MSC 具有简化方程开发和预测的作用。

当测定与定标样品不同的普通面粉时，定标精确度（用 SEP 表示）为 0.205% [1, 3, 5]，并且发现定标方程可应用于样品类型的一个广泛范围，同时在较长的一段时间内

不需做调整,而保持稳定,由于 NIR 已被用来检测烘烤面粉的等级的一种方法^[2],很显然在 NIR 定标方程的准确性方面,供给者和烘烤者会产生争论。发生争执时,通常的解决方法是用 Kieldahl 标准方法分析有争议的样品,此时所测结果与原 NIR 结果间肯定存在一个差异,可是这并不表明需要重新调整定标方程,通过计算不同时间的 SEP 值,可找到产生问题的原因所在。 $SEP^2 = S_K^2 + S_{NIR}^2 + S_e^2$,其中 S_K 是用 Kieldahl 方法测定值, S_{NIR} 是 NIR 测定值, S_e 是代表定标方程有所缺陷的第三个值。此值可看成是某些面粉的一个随机变化值,但它可表示某个淀粉样品的一个系统误差。因此,只能讨论某种淀粉是定标方程中产生偏差的问题所在,这样讨论也比较合理,并且也只能用此种面粉中具有代表性的样品进行测试。

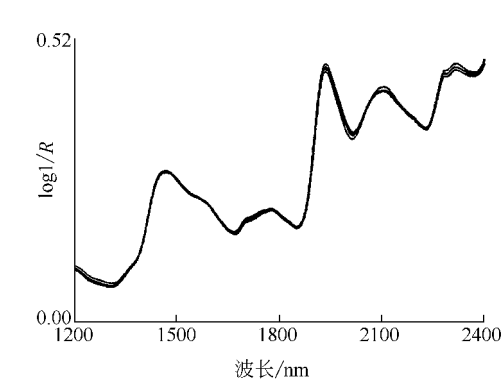


图 17-27 图 17-25 中的样品经倍增分散校正后的 NIR 谱图

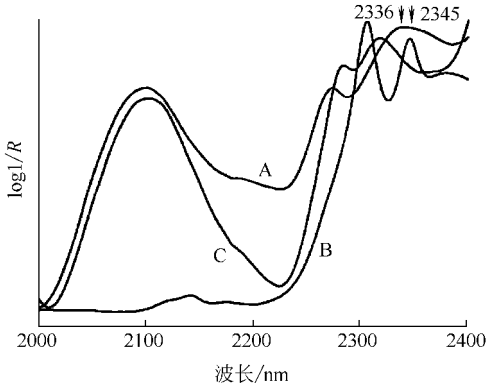


图 17-28 一个面粉样品在烘箱干燥前(A)后(B)的 NIR 光谱

为使 NIR 定标方程得到发展和适于生产,需花大量的时间和精力。将一个定标方程应用在具有相同模式的其他仪器,能给生产者带来很大的收获,因为全部定标的过程不需在每个焙烤厂重复。可是,定标方程也有一个很重要的先决条件,即 NIR 定标能分析应用于面粉厂的所有面粉类型。Osborn 和 Feam 已成功地把手粉蛋白定标常数在 9 个固定滤光设备间转化。滤光片型设备分别放在不同的实验室,每个设备的 F_0 值都用 20 个样品(一套样品)做调整,用第二套 20 个样品评价与 Kjeldahl 方法比较过的仪器测定值。NIR 结果的平均变动范围从 10.92%~11.30%,而 Kjeldahl 测定值平均为 11.03%,NIR 重复性的差异为 0.17, Kjeldahl 为 0.14%。通过样品检测方式,我们可对 NIR 和 Kjeldahl 测定值做常规检测^[8]。

(2) 水分

NIR 技术的另一个特点是,它能因时分析样品的几个成分。测定水分是 NIR 在食品分析方面的首次应用,因此本书中也专门用一章的篇幅来阐述 NIR 在此方面的应用和发展。对于烘烤业的面粉,应测定其水分含量,并且此步工作能和蛋白测定同步完成^[1~7]。水分在 1940nm 处有一个强吸收,并且面粉中的其他含量也不含对此波产生影响,从而水分的定标方程变得简单易行,因为它只需一个参照波长(2310nm)^[9]。

NIR 定标水分 SEP 值的精确度为 0.15% [1, 3, 5],此值与蛋白值相比要好些。实验室间的测定方法结果也很好,与烘干箱烘干法 13.13% 平均值相比,9 个仪器间 20 个样品平均值的变化范围从 13.08%~13.28%,重复性为 0.16%。

(3) 颗粒度

平均颗粒度 (MPS) 对于是饼干 (蛋糕) 粉而言是一个重要的质量指数。因为颗粒度对 $\log 1/R$ 值有一个显著的影响, 同样 NIR 也可对其进行测定, 定标只需一个偏中性的波长^[3], MPS 和 2230nm 的 $\log 1/R$ 呈明显的非线性关系, 但在典型 U. K. 饼干粉的一个小范围内, 可得到线性定标方程。对于经过 75 μ m 筛孔的干粉, 其 57%~74% 样品的 SEP 为 3%。

(4) 灰分

传统检测面粉纯度考虑的是灰分中米糠污染情况。灰分的定义是面粉在高温灰化后其氧化残留物。因为米糠产灰分的量高达 8%, 远高于胚乳 (0.3%), 因此面粉中灰分含量是米糠污染的一个灵敏指标。白面粉中高灰分不仅影响烘烤质量, 而且影响由其制成的面包瓢的颜色。因此, 面粉趋向于获得最低的灰分含量。

灰分没有 NIR 特征吸收波长, 但米糠中的一个重要组成成分是纤维素, 其特征波长为 2336nm (图 17-29)。尽管淀粉和它有一部分波形重叠。可此波长 NIR 测量值仍可作为面粉中米糠含量的一个指标, 因而与灰分含量高度相关。已报道了灰分的经验定标方程, 检测结果也较满意。但是米糠与油和蛋白的胚乳相比, 灰分含量很高, 并且这些灰分定标方程常包括油和蛋白的特征波长。

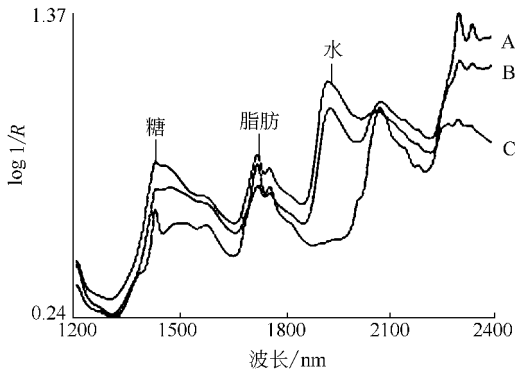


图 17-29 纤维素(A)、淀粉(B)和油(C)的近红外光谱

对于某种谷物, 灰分似乎和颗粒度有关, 因为软质小麦比硬质小麦更易产生灰分高的面粉。Bolling 和 Ewingllbeng 实验结果显示了灰分与颗粒度之间的相关。他们报道, 尽管实际灰分未变, 可是谷物却随蛋白含量的上升, NIR “灰分” 值也在改变。同样, Hilliams et al^[1] 和 Inamoto et al^[11] 也报道灰分定标方程中使用了油和颗粒度的吸收波长, 并且对不同的样品使用这些定标方程产生了严重的错误。Willans et al^[1] 建议在灰分定标中加入 2345nm, 理由是它不会存在油的干扰。可是, Iwamato et al^[11] 表明在脱脂面粉的光谱中, 对 2345nm 的吸收波消失, 此现象与纤维素无关, 而与一种溶剂萃取物有关。图 17-29 给出纤维素和油的光谱, 支持了上述结论。

很明显, 大量研究结果表明应用 NIR 定标方程测定面粉的灰分含量不太合适, 但油有自己的近红外光谱, 可以进行测定。Bertrand et al^[12] 建议用油作为粗粒面粉中米糠的一个指标, 其定标结果不会违反标准方法的结果。Iwamoto et al 也指出 1724nm 作为油的波长, 其与面粉的纯度有关。

很明显, 大量研究结果表明应用 NIR 定标方程测定面粉的灰分含量不太合适, 但油有自己的近红外光谱, 可以进行测定。Bertrand et al^[12] 建议用油作为粗粒面粉中米糠的一个指标, 其定标结果不会违反标准方法的结果。Iwamoto et al 也指出 1724nm 作为油的波长, 其与面粉的纯度有关。

(5) 颜色

颜色测定值可代替灰分作为面粉纯度的一个指标, 因为面粉颜色直接与由它制成面包瓢的颜色有关。至少一个 NIR 设备包括一个可见波 (在 540nm) 滤光片, 这使得可以在测定蛋白质和水分的同时, 测定颜色^[2]。可是面粉颜色是一个复杂的现象, 它包括米糠含量 (级), 黄色 (Agrron) 等级以及颗粒度的共同作用。为了测定与其他因素相分离的米

糠含量（因为它影响面包体积），Ket-Tones 和 Martin 设计了一种颜色梯度，它使测定值与 Agtron 方法有一个相似的波长，但是使用的是面粉和水的浆状物，此测量法称作梯度颜色指数（GCF）。NIR540nm 波长反射测量值与 Hunterlab 干粉值呈高度相关（ $R = 0.95$ ），但与 GCF 相关性很差。可是，如果使用两个波长（540nm 和 1680nm），后者相关系数提高到 0.93，方程的 SEP 为 0.46GCF 单位^[3]。

(6) 淀粉破坏度和水分吸收度

在粉碎程序中，淀粉颗粒受到机械破坏的程度是一重要质量因素，因为现代面包制造业的制造者要求每批淀粉的水分吸收不能变化，尽管淀粉损害不是影响水分吸收度的惟一因素，但它是磨坊主可调节的一个因素。OsbJorne 和 Douglas 用一个扫描单色器，开发了测定淀粉损坏度的定标方程^[13]，然后获得了用于这种测量的合适的滤光片，并把它用在具有固定滤光片的仪器上。此定标方程的 SEP 为 3.2Fannand 单位，与面包粉通常的范围（20~40Farrand 单位）相比，显得很大。更严重的问题是，定标常数是如此之大，以至于所测淀粉损坏度变化很大时，其中方程的变化值都很小。这就使定标对误差因素，如样品温度的变化，非常敏感。通过对 Ref. B 中定标数据用主要成分分析法分析后，发现了原因。主要成分分析法鉴别光谱数据中变化值的各个来源的数目，以及它们波长的依赖性。在此试验中 99.94% 的变化值由前四个主要成分得到（表 17-35）。第一个是颗粒度，第二、三个是水分；第四个主成分与最初定标有相同的特征波长，但它只占了 0.06% 的变化值，并且与淀粉损害度相关性也小。从淀粉损坏度和 1~3 主要成分的关系来看，很明显定标方程大部分依赖于颗粒度和水分。NIR 直接对水分吸收定标方程，其 SEP 值为 1.5%^[5]，所使用的波长与蛋白质、水分以及淀粉损害度的波长相等，而这些都是影响水分吸收的主要因素。可是由于损坏的淀粉对于用湿化学数据预测水分吸收的 Fannand 方程起到明显的作用，所以 NIR 必须建立除蛋白质和水分外的淀粉损坏测定方法，以能成功的测定面粉的水分吸收情况。

表 17-35 淀粉损坏度定标数据的主成分分析

主成分	变异系数%	与淀粉损坏度的相关系数	波长/nm
1	98.65	-0.33	无特异性
2	1.04	0.41	1940
3	0.19	0.63	1940,2100
4	0.06	0.29	1466,1580,1940,2100,2282

17.6.1.2 面粉焙烤质量

既然烘烤者用 NIR 技术检测面粉质量，是否可以用 NIR 检测面包或饼干中的某些可测成分而预测烘烤质量，下面是面粉加工焙烤研究所及 Chorleywood 所得到的一些未发表的资料，设计了两个试验以验证这种可能性。

49 个商业粉碎面粉来自英国和加拿大的工厂，在严格控制每步程序下做成 400g 的面包，除水分吸收外，混合，发酵、模制及烘烤的条件保持一致。用含 19 个滤光片的近红外仪器记录面粉 NIR 数据，用 Kjeldahl 方法（ $N \times 5.7$ ）测定面粉蛋白质含量（8.7%~14.5%），用种子替换法测量面包的体积（1170~1600mL），使用所有 19 个吸光度值的 NIR 数据，建立面包体积的多回归方程；然后选取标准偏差接近的小部分数值作为整个

模型的代表，几个可能波长数值组给其定标标准误差（SEC）范围在 40~50mL 间，其中蛋白波长是主要影响因素，2180nm、2100nm、1680nm 数值的波长组合的 SEC 值为 50mL，可用 Kjeldahl 蛋白质预测相同样品的面包体积，SEC 为 49mL。从上可知，NIR 有预测面包体积的能力，主要在于它能测定蛋白质含量；NIR 数值不比 Kjeldahl 数据好或坏。用扫描单色光谱仪测试大麦的工作也表明了同样的结论。有报道说，NIR 不能测定淀粉酶的活性，表 17-36 中的数据证明了此点，其中 49 个面粉（一种强加拿大）中的一个面粉阶段性加入真菌淀粉酶。这增加了测定的面包体积，但不影响由 NIR 测定的面包体积。

表 17-36 α -淀粉酶活性对测定和 NIR 预测面包体积的影响

实际体积/mL	预测体积/mL	α -淀粉酶/Farrand 单位	实际体积/mL	预测体积/mL	α -淀粉酶/Farrand 单位
1462	1502	2	1568	1505	55
1511	1528	12	1604	1476	102
1544	1502	21			

在 5 年间, 共有 75 个面粉样品用严格控制的程序烘烤为低糖和少糖的饼干, 同时用扫描单色器记录了其 NIR 数据。将 NIR 数值经过二阶导数转换, 与四个饼干性质数据(水分含量、饼干重量、结构和一批面粉密度)进行多元回归计算, 并得到 5 个数据群中的每个样品群不同的 F 。通过每个参数所使用的波长计算出相关系数, 与面粉蛋白质和选取波长的逐级递进得到的回归方程所算出的相关系数相近。

根据现有实验数据,可以说 NIR 不能用于预测面粉烘烤质量,测定灰分、淀粉损坏度和水分吸收度也有所限制,但可用来测定蛋白、水分、颗粒度和颜色,可用于实际生产。

17.6.1.3 小麦面筋

从麦粉中分离出的面筋以一种干粉状态加入面包中作为一种添加成分，有两种原因：①帮助基料面粉携带惰性米糠，以维持由棕色和粗面粉做成的面包体积和其高品质，因为如果不加入面筋，上述一切都不可能实现；②全部或部分地代替面包中由高蛋白质小麦磨出的面粉中的蛋白质。

用 29 个商用样品, 可得到面筋中 NIR 蛋白质和水分的定标方程, 这些样品已用 Kjeldahl 方法 ($N \times 5.7$) 测定其蛋白质含量, 以及在烘箱中, 在 103°C 下烘干 16h 测定其水分含量。用 6 个滤光片的仪器记录了 NIR 数据, 通过研究所有波长组合, 选取的定标方程情况见表 17-37。因为 NIR 可用来测定基料面粉 (P_F) 和面筋 (P_G) 的蛋白质含量。为提高面粉的蛋白质含量到 P_S 所需添加面筋的量 (W_G), 可由下式计算出:

$$W_G = \frac{100(P_S - P_F)}{P_G - P_S} \%$$

表 17-37 NIR 进行面筋分析的统计参数

统计参数	蛋白质	水分	统计参数	蛋白质	水分
波长/nm	2180,2100,2300	1940,2100,2230	相关系数	0.98	0.985
范围/%	62.6~78.2	6.9~9.5	SEC/%	0.73	0.12

表 17-39 给出发表的 NIR 分析焙烤产品的一些应用。

产品	参考数据					
	水分	蛋白质	脂肪	淀粉	糖	纤维
面包	20~23	20~22	20~23	24	22	—
饼干	19	—	19	—	19	—
糕点混合物	—	—	25	—	25	—
早餐谷物	—	—	—	—	28	26~27
面团	29	29	29	—	—	—
快餐食物	—	—	—	31	—	31

17. 6. 3. 1 面包

英国、美国、南非已用 NIR 测定片式面包的蛋白、脂肪和水分含量^[20~23]。采用循环取样，即从 5~6 片面包中取一片面包，并把获得的 NIR 数据进行平均化处理，获得了满意的定标方程，需要注意的是应防止在分析过程中水分的丢失。对面包进行干燥和磨碎不能明显改善 NIR 分析方法的精确度^[20,22]，在不同的研究中，标准误差很一致（表 17-40），对于日常使用也可接受（变异系数 1%~2%）。已证明定标方程在不同的烘烤者之间存在稳定的变异性^[23]，并且定标方程可以在很宽的用不同程序和面粉制作的面包之间进行转移。定标方程对于面包片中总糖分（0.9%~5.6%）的 SEC 值为 0.34%，干燥的面包中淀粉的范围 51%~65%，其 SEC 为 2%。

表 17-40 NIR 预测面包片的精确性

定标方程	SEC/%	SEP/%	Ref.
蛋白质	0.09	0.20	20
	0.24	—	21
	0.23	0.22	22
脂肪	0.16	0.18	20
	0.13	—	21
	0.24	0.20	22
水分	0.46	0.51	20
	0.67	0.62	22
干物质	0.57	—	21

17. 6. 3. 2 饼干

人们研究了用 NIR 检测小块面团饼干产品中脂肪、蔗糖、干粉和水分的可能性^[19]。这次研究中所采用的实验设计，尽量减小主要变异值与广泛用于单色器波长选择方案间的相关关系。人们改变标准配方，为研究四个组成成分的每一个成分提供尽可能大的范围，为了阻止脂肪和蔗糖间的高度相关性，通过改变这两成分间各自的数量，而使配方随机化。然后，计算出每个配方的水分和面粉量，使面团能操作在转模上。整个饼干放在 NIR 扫描单色器上，有字和刺孔的一面对着光柱，就像实际生产线上一样。波长选择方案一次选择两个波长，以找出最好的一对。以这种方法一次选择两个波长可克服多共线问题，并且当一个两项方程即够用时，此程序就应能找到它。如果必要，进一步选择波长可

用逐步回归方法，这样就能得到 NIR 用于测定饼干中脂肪、蔗糖、干面粉和水分的定标方程。可是，尽管定标方程与其他方法相比相关系数高，精确度也高，但只有脂肪的定标方程（图 17-30）才能有足够的精确度以检测商业烘烤业中可能遇到的成分计量错误。

17.6.3.3 面粉甜食

上述面团、面包和饼干的 NIR 定标方程，是在面包厂严格控制条件下研究出来的。不管怎样，NIR 将应用于商业情况，其典型情况是分析干蛋糕混合物^[25]，用 19 个滤光片仪器作定标，然后在烘烤厂具有质量控制的情况下评价其优劣。希望对于一定范围内不同的混合物（面包、茶点小蛋糕、松蛋糕、酥饼和酥皮）都用一个定标方程，这可通过测定脂肪和蔗糖而实现。对于范围在 3%~26% 的脂肪，其定标方程 SEP 为 1.7%；对于范围在 13%~45% 的蔗糖，其定标方程 SEP 值为 2.7%，尽管这些 SEP 值与通过相关参考法重复测定的值（Soxhlet，脂肪 0.38%，高效液相色谱的蔗糖为 1.04%）相比要高，可是考虑到 NIR 结果获得的快速法，所以它们在质量检测中是可以被接受的。

17.6.3.4 各类早餐谷物

Baken et al 研究早餐谷类中性洗涤纤维测定值与 NIR 之间的关系。在初步的研究中，相关因子为 0.985，SEC 为 1.36%，含量范围在 0~31.7% 之间。所用样品为 213 个，基于各种小麦、玉米、燕麦和大米的谷物产品。用两种不同的定标方程都得到了上述结果。其中一个定标方程的 NIR 数据是由 $\log 1/R$ 的一阶导数除以参考波长下的 $\log 1/R$ 得到的。单色器型或倾转滤光仪器都可用来做这些数据的处理。不同的单色器仪器^[27]做进一步的工作，就可得 SEC 为 1.5% 的一个相似方程，用完全独立的样品检测此方程的 SEP 值为 1.8%。纤维定标所用的波长（1352nm，2204nm，2448nm，2450nm）被选择以避免由样品中晶体蔗糖产生的干扰。由于干扰的存在，很明显可用 NIR 测定早餐谷物中总糖含量，在验证这种假设的实验中^[28]，使用了 84 个等级的谷类，总糖含量从 0 变化为 54%。其中 40 个样品用于定标，44 个样品用于预测。用下列波长的一阶导数比率（1432/1822nm 和 1692/1782nm）求出定标方程，与标准气-液色相方法相比，此方程给出定标样品 SEC 值为 2.7%，预测样品 SEP 值为 2.8%。SEP 值与报道的蛋糕混合物^[25]中的值非常接近，并证实 NIR 可用来测定谷类食品中的糖。

17.6.3.5 专制食品

Kaffka ef al，用一个仔细控制实验研究了 NIR 用于检测粉状专制食品中脂肪、蛋白质、水分和蛋的含量^[29,30]。第一个实验中^[29]，专制样品每公斤中加入 0、2、4、或 6 个蛋，用化学分析法测定脂肪和蛋白含量，尽管取得了极好的定标，由于使用了相同的波长，所以说蛋、脂肪和蛋白含量间呈高度相关。虽然蛋白和脂肪的定标适用，然后由其求出面制食品中蛋的含量，如果想成功的应用，应该独立地测验这 2 个成分。因此在第二个

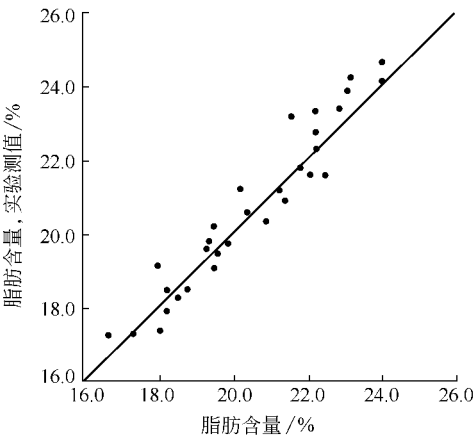


图 17-30 32 个样品脂肪含量 NIR 测定值与常规法的相关图

实验中，向食品中加入不同比例的蛋黄和清蛋白，以使脂肪和蛋白水平随机化。脂肪和蛋白水平间的实际相关为 0.362，但是当同样水平的样品被粉为不同粒度时，相关系数下降，可求出用 $\log 1/R$ 二阶导数比率的简单定标方程（脂肪、蛋白和水分），其 SEC 值分别为 0.11%，0.24%和 0.24%。

17.6.3.6 其他谷类食品

以上所讨论的用 NIR 技术来分析烘烤产品，一般常用于特定种类的产品，如面包或饼干，尽管某些情况下^[25~28]定标方程是同类中的其他产品获得的。可是，在烘烤中使用特定样品的定标方程非常不方便，因为需分析大范围的产品，要用许多定标方程。除简单仪器的内储问题外，这也产生了由于使用错误定标方程而产生更大随机误差的可能性，NIR 研究者开始意识到通用定标方程的可能性和意义。

烤制者用 NIR 测定粉碎、休闲食品（马铃薯片、球片、挤压食品，爆玉米花、薄脆饼干和椒盐卷饼）的纤维和淀粉含量，其 SEP 值分别为 1.4%和 3.5%，这些定标方程中使用了研究成分特征波长 $\log 1/R$ 衍生值和参考波长衍生值间的比例，Osborne 给出测定面包、饼干和早餐食品中纤维的简单定标方程。用 MSC 算法标准化后，使用 1720nm 和 1700nm $\log 1/R$ 数据的定标方程，其 SEP 值在脂肪含量为 1.2%~27.3% 范围内为 1.10%，可是，很有必要在 NIR 分析前，冷冻一干燥面包，粉碎所有样品，因此，尽管能求出一个通用的定标方程，仍需要更多的样品准备工作，并且，尽管蛋糕混合物的 SEP 值较好，可是对于面包来说要差些（表 17-40）。

另一个获得通用定标方程的方法是提取样品，并分析提取物，Menrens et al^[33] 提议 NIR 分析水溶物的一种技术，称作近红外反射干物提取光谱法（DESIR）。它将小量水溶物放在玻璃纤维纸上（一种扩散反射的载体），然后将溶液在微波炉中快速蒸发，再用 NIR 检测玻璃纤维纸。图 17-31 给出从含 0.2%~2.0%蔗糖的 10 个可溶物中得出的二阶导数光谱，蔗糖含量以 0.2% 递增，蒸发到玻璃纤维纸上。图 17-32 给出了蔗糖含量（mg）与其在 2092nm 处的二阶导数值间的关系。在谷物中应用 DESIR 可降低由淀粉、蛋白质、脂肪和纤维产生的干扰。另外，定标方程可用配置标准液的方法得到，代替预先分析样品，如图 17-32。

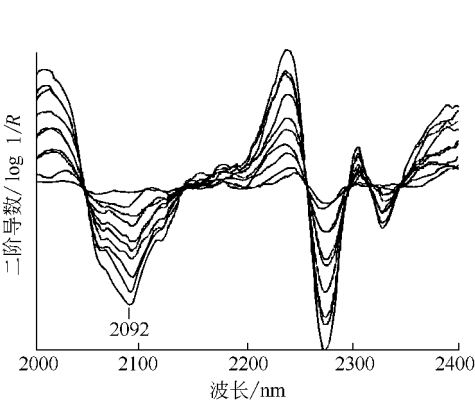


图 17-31 含量为 0.4~4.0mg 蔗糖在玻璃纸上的 NIR 光谱

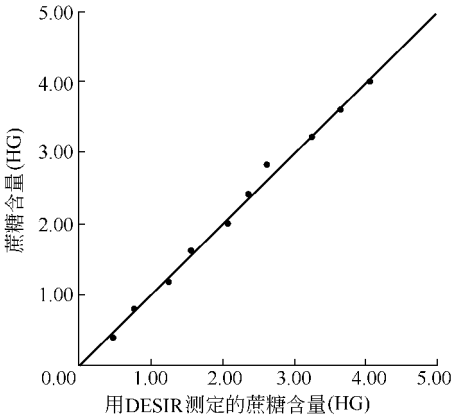


图 17-32 用 DESIR 开发的蔗糖定标方程

结论:

烘烤工业广泛使用 NIR 分析原料成分、程序中间产物和终产品, 本节给出了许多成功的定标方程, 但如果 NIR 技术的优点和缺点能够得到更好的理解后, 会有更多成功的定标方程产生, 而且可从未成功的定标中学到更多的东西。

参考文献

- [1] B L Upchurch, J A Throop, D J Aneshansley. Detecting internal breakdown in apples using inter-actance measurements. *Postharvest Biology and Technology*, 1997 (10): 15~19.
- [2] Bhatti R S. Measurement of oil in whole flaxseed by near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists Society*, 1991, 68 (1): 34~38
- [3] Birth G S, K L Olsen. Nondestructive detection of watercore in ‘delicious’ apples. *Proc Amer Soc Hort Sci*, 1964, (85): 74~85.
- [4] C J Clark, V A McGlone, R B Jordan. Detection of brownheart in ‘Braeburn’ apple by transmission NIR spectroscopy. 2003, (28): 87~96.
- [5] Che Man, Y B. Determination of free fatty acids in palm oil by near-infrared reflectance spectroscopy. *JAOCS*, 1998, 75 (5): 557~562.
- [6] Chiba Minoru, et al. Principal component analysis applied to near infrared spectra for classifying wheat flours. 1995 (42): 796, 801
- [7] Delwiche S R, McKenzie K S, Webb B D. Quality characteristics in rice by near Infrared reflectance analysis of whole grain milled samples. *J Cereal Chem*, 1996 (73): 257~263
- [8] K H Norris, W R Hruschka, M M Bean, et al. A Definition of wheat hardness using near infrared spectroscopy. *Cereal Foods World*, 1989 (34): 696, 705
- [9] Li Hui van de Voort, F R Ismail, A. A et al. Discrimination of edible oil products and quantitative determination of their iodine value by Fourier transform near-infrared spectroscopy. *JAOCS*, 2000, 77 (1): 29~36
- [10] Marlow G C, W H Loescher. Watercore. *Hortic Reviews*, 1984 (6): 189~250
- [11] Moh M H. Determination of peroxide value in thermally oxidized crude palm oil by near infrared spectroscopy. *JAOCS*, 1999, 76 (1): 19~23
- [12] Olsen K L, H A Schomer, G S Birth. Detection and evaluation of watercore in apples by light transmittance. *Wash State Hort Assoc Proc*, 1962 (58): 195~197
- [13] Onda takumi Application of near infrared spectroscopy to quality control of foods. **日本食品保蔵科学会誌**, 2002, 28 (6)
- [14] Sumio Kawano, et al. Determination of Sugar Content in Intact Peaches with Fiber Optics in Inter-actance Mode. *J Japan Soc Hort Sci*, 1992, 61 (2): 445~451
- [15] Sumio Kawano, et al. Nondestructive Determination of Sugar Content in Satsuma Mandarin using Near Infrared Transmittance. *J Japan Soc Hort Sci*, 1993, 62 (2): 465~470
- [16] Yildiz G. Method for determining oxidation of vegetable oils by near-infrared spectroscopy. *JAOCS*, 2001, 78 (5): 495~502
- [17] 陈继红, 赵武善. 采用近红外技术分析糖汁用于甘蔗的收购计价. **广西蔗糖**, 2002 (4)
- [18] 恩田匠. 食品の品質管理への近赤外分光分析法の応用に関する研究. **日本食品保蔵科学会誌**, 2002, 28 (6)
- [19] 恩田匠等. 近赤外分光分析法によるスモモ果実の糖度、酸度および硬度の非破壊計測. **日本食品科学工学会誌**, 1994, 41 (12): 908~912

- [20] 顾伟珠等. 多元线性回归分析油菜籽含油量的近红外光谱数据. 中国粮油学报, 1995, 10 (27): 57~64
- [21] 河野澄夫. 近赤外分光法による食品分析. 月刊フードケミカル, 1998 (3)
- [22] 李晓明等. 损伤水果图像处理检测系统 (1). 日本农业机械学会志, 1993, 55 (4): 91~98
- [23] 林晓鹰. 近红外水分仪的研制. 中国仪器仪表, 2001 (2)
- [24] 刘建学等. 大米直链淀粉含量的近红外光谱分析. 农业工程学报, 2000, 16 (3)
- [25] 刘建学等. 近红外光谱法快速检测大米蛋白质含量. 农业机械学报, 2001, 32 (3)
- [26] 卢家炯, 黎庆涛. 近红外光谱技术在制糖工业应用动态. 广西蔗糖, 2000 (1)
- [27] 千葉実等. 近赤外スペクトルの主成分分析による用途特性の異なる小麦の分類. 日本食品科学工学会誌, 1995, 42 (10): 796~801
- [28] 舒庆尧等. 用近红外反射光谱技术测定精米粉样品表观直链淀粉含量的研究. 中国水稻科学, 1999, 13 (3): 189~192
- [29] 汪旭升. 近红外光谱分析测定棉籽粉中油含量的研究. 浙江农业学报, 2001, 13 (4): 218~222
- [30] 徐晓峰. 用 GQA-31 型近红外光谱仪测试食品油的方法. 生物技术, 1996, 6 (1): 44~46
- [31] 岩元陸夫等. 近赤外分光法入門. 幸書房, 1994
- [32] 郑咏梅等. 小麦近红外特征波长提取及蛋白质含量测定. 激光与红外, 2003, 33 (2)

18. 近红外光谱分析在石油化工工业分析上的应用

18.1 概述

近红外光谱分析技术最早应用于农业领域^[1~3]。进入 20 世纪 80 年代后期，近红外光谱分析技术获得了长足的发展，主要得益于两点：①在这期间，在中红外光谱仪器硬件技术成熟发展基础上，近红外光谱仪器硬件技术有了迅速发展；②计算机技术快速发展促进了化学计量学学科的迅速发展和应用^[4~8]，为近红外光谱测量技术提供了大量行之有效的定量和定性方法与软件^[9,10]。值得提及，美国华盛顿大学化学系的过程分析中心（CPAC）在这期间做了大量将化学计量学、传感器和光谱分析结合起来用于油品性质分析的开拓性工作^[11]，从此改变了近红外光谱仅在农业领域中应用的局面。尤其是，近红外光谱分析辛烷值在时间和成本上的效益特别突出，可用于在线控制分析，使得当时各大石油公司竞相研究开发与应用这种技术，所以，在后来的十多年中，它在石油化工领域中的质量检验和生产控制方面获得了迅速发展和广泛应用^[12~26]。

作为传统工业的石化工业仍然是各国国民经济支柱产业之一，因此，发达国家都特别注重采用现代信息技术提升石化生产过程的精确监测和控制技术^[15~26]，目的是提高产品质量，不断满足世界日益增长的环境保护严格要求，降低生产成本，提高经济效益。与其他产业相比，在石化工业生产过程中需要分析的项目特别多，所涉及技术几乎囊括了分析学科的各类现有技术。如汽油主要分析项目有：辛烷值（包括：研究法辛烷值 RON，马达法辛烷值 MON，和抗爆指数）、烯烃、芳烃、苯含量、含氧化合物、蒸汽压、密度、馏程、色度、氧化安定性、硫含量等。各种项目需要不同的分析设备，分析时间也较长。如汽油辛烷值采用标准方法 ASTM D2699^[27]，需要一台爆震引擎辛烷值测定仪（通常采用进口的 WAUKSHA 设备，价格为 200 多万元人民币），一个辛烷值（RON 或 MON）测量时间至少为 40min，同时对该测定仪标定还需要进口或配制标准燃料，这项费用也很高。因此，当用于中间控制分析（特点频次高，样品多）时，不仅仅是成本高，主要是不能及时分析，往往漏检不合格的原料、中间产物和产品，造成生产上的巨大浪费（因为加工规模巨大的特点，稍有偏差引起的实际价值波动就会很大）。传统方法解决这些问题的途径，一是增加取样频次，其结果增大了工作量，耗费了巨大的人力物力和时间，但仍然不能有效地解决其分析滞后的问题；另一种方法是控制生产条件使质量指标留有余地，以保证产品合格。但是代价为产品的实际质量指标明显高于产品规格，超出规格要求的质量指标部分是生产厂家“白送（Giveaway）”的；控制条件留有余地方法会造成能源及设备的巨大浪费，产生的损失也是惊人的^[15]。

近红外分析技术已成功地替代许多石化产品质量指标的标准分析方法，在化验室和在线方面使用^[15~26]。与现行标准方法比较，近红外分析方法具有高效、快速和方便的特

点,“高效”是指通过对样品一次光谱测量,可提供样品的多种性质参数;而“快速”是指完成分析时间在几秒到几分钟范围内。近红外在大多数石化产品的分析中,一般不需要样品的预处理,操作方便,无化学污染,又被称为绿色分析技术。近红外光可以通过光纤传输,能方便地进行在线控制分析,成功地解决了过去依靠标准分析方法存在的分析严重滞后问题,在生产上可实现“卡边”控制,避免了上述“质量白送”和浪费现象的产生,由此产生的巨大社会效益和经济效益令人注目^[15,26,28]。

近红外分析技术在国外石化工业早已获得了较广泛的成功应用。发达国家几乎所有炼油厂在油品调和、重整、原油蒸馏等各种炼油与化工工艺^[24]中先后相继采用了在线近红外光谱分析技术,成为炼油厂保障产品质量、降低生产成本和提高经济效益所必须依靠的技术手段之一,而且在与传统在线仪表、气相色谱、核磁等技术并存和竞争中脱颖而出,其应用处于主导地位。国内石化领域中,早在1994年,国外近红外光谱仪器代理商开始宣传与推广,至今其进展仍甚小。因为初期(20世纪90年代初)国内对这项新技术需要有一个认识过程。但在经过一段认识过程之后,人们发现,最适合应用的近红外光谱分析技术是化验室中间控制分析和装置在线分析。并且近红外光谱分析技术应用能够奏效的前提是:近红外光谱仪、化学计量学软件和各种校正模型的三位有机结合为一体。进口仪器所带的模型大多不适应国内油品(主要原因是我国炼油工艺和原料与国外的差别很大)。事实上,国际近红外光谱分析技术应用的十多年实践已经表明该技术的应用成功极大地取决于技术服务^[21]。所以模型与技术服务是目前国外近红外光谱仪器在国内石化领域中应用受阻的主要原因,另一方面,进口仪器价格高(台式仪器为5~7万美元)。还有仪器硬件的售后服务不仅费用高,而且送货不及时。

近年来,国内也在近红外光谱分析技术的研究和应用方面做了大量工作^[23~25,29~50],例如石油化工科学研究院成功地开发了CCD近红外光谱仪系列产品(台式光谱仪包括NIR-2000和NIR-3000,在线仪器NIR-6000和便携式油品测量仪等),化学计量学分析软件和各种油品性质的校正模型等组成的成套技术,并已在国内30多家炼油厂生产控制分析中进行了较长期的成功应用^[50],为提高我国炼油厂化验室中间控制分析效率发挥了积极作用。结合我国石化分析,研究建立稳定模型和模型传递等理论问题,对在石化工业推广该技术提供了技术保障。还有部分厂家也引进了汽油调和用在线分析仪。预计在今后的几年内,近红外快速分析技术在我国石油化工生产领域内将有大的发展。

18.2 近红外光谱分析技术对石化工业分析的适用性

石油主要由复杂烃类化合物组成。组成石化产品的主要结构官能团有C—H、O—H、N—H和S—H等,它们都含有H原子,其分子振动吸收光谱的倍频和组合频正好位于近红外光谱范围(700~2500 nm)内,表18-1列出了各化学功能团的谱带归属^[5,51]。而其他化学官能团在该波长范围的干扰相对较少。因此,从信息选择性上讲,近红外分析非常适合石化产品。

	芳 C—H	甲基 C—H	亚甲基 C—H	N—H	O—H
合频	2150,2460	2250,2360	2290~2450	—	
一级倍频	1680	1695,1705	1725,1765	1540	1450
合频	1420~1450	1360,1435	1395,1415		
二级倍频	1145	1190	1210	1040	960
合频		1015	1053		
三级倍频	875	913	934	785	730
四级倍频	714	746	762	—	—

近红外光谱分析技术在石化应用发展过程中，曾有过各种物理方法如采用中红外光谱、拉曼光谱、核磁等分析技术以及近红外光谱技术的报道^[52,53]。但是发展至今，从世界范围内石化工业实际应用情况看，在油品性质分析方面，虽然有少量中红外仪器在应用，但仍以近红外光谱仪器为主；在在线分析方面，除很少量的在线核磁仪器在应用外，近红外光谱技术占主导地位，这是因为近红外光谱有以下技术特点：

石油产品的中红外光谱和近红外光谱都属于分子振动光谱。其中，中红外光谱为振动基频，而近红外为振动倍频与合频，其产生几率远比中红外基频的小，中红外和不同的倍频谱带的强度有显著差别（见表 18-2），倍频级数每增加 1，其吸收强度降低 1 个数量级。近红外吸收系数较小，虽不灵敏，但由于“不灵敏”，近红外光谱可以使用长光程（光通过样品的距离），如短波近红外（Short Wavelength Near Infrared, SW-NIR）可使用 2~100 mm 的光程，与中红外（光程为 0.02mm）相比，不仅对光程精度的要求显著下降，而且对痕量物质的干扰影响也不明显，因此，近红外光谱测量的抗干扰性强。石油产品如汽油、航煤和柴油等大多是透明或颜色较浅的液体，不经稀释，也不做任何预处理，使用比色皿或光纤探头就可对其直接测量，十分方便与快捷。通常，中红外油品分析仪器在分析不同油品，如汽油和柴油，需要不同光程的样品池，而近红外油品分析仪不论什么油品，使用同一光程即可。由于近红外使用长光程，也使测量池有清洗方便和拆卸后再安装时重现精度高的特点。

表 18-2

C—H 红外不同倍频谱带的强度对比^[25]

谱带	波长/nm	相对强度	需用的光程 (对液体烃样品)/cm
基频	3380~3510	1	0.01~0.4
一级倍频	1690~1755	0.01	0.1~2
二级倍频	1127~1170	0.001	0.5~5
三级倍频	845~878	0.0001	5~10
四级倍频		0.00005	10~20

中红外光无法通过普通的石英或玻璃光纤，近红外光则很容易穿过。因此，近红外光谱技术可以使用光纤进行远距离（300~1000m）测量，能够在装置上直接对物料性质进行测量。虽然过去在石化生产装置上有许多种类的在线分析仪，如气相色谱、质谱、馏程、密度等，但是这些装置通常需要较强现场维护，经常由于仪器硬件故障给生产控制带来不便。而近红外光谱仪在结构上，需维护的硬件较少，大大地减少了现场维护的强度。与其他测量手段相比，不仅近红外测量速度快，而且能同时提供多种质量分析参数，一台

仪器可替代多种在线分析仪使用，节省大量设备投资。在线近红外光谱仪器通过多路光纤转换器可方便实现对不同物流点的切换监测，而且安装也很方便。因此，近红外光谱进行在线测量更方便、更快捷、更可靠和成本更低。

与油品的传统标准分析方法相比，近红外光谱方法测量速度极快，测量精度高，这对于石化生产控制分析尤为重要；由于是非破坏性测量，也不使用任何化学试剂，不造成环境污染，极大地改善分析人员工作环境，也称绿色分析技术；近红外分析样品用量（0.2mL 到十几 mL）很少，可以适合一些研究项目分析，如小型装置产物分析。因此，近红外光谱分析技术在石化工业分析中非常受欢迎。

但是近红外光谱技术也存在以下困难，即必须事先准备好计算模型，石化产品的变化范围大，模型的建立也就需要较多投入，同时如以上提到的近红外光谱的灵敏度低，虽然可以应用较长的光程进行主要成分的检测，但对含量低于 0.1% 的微量物质，在一般情况下，不便检测。

18.3 方法原理与特点

18.3.1 方法原理

方法原理与其他应用领域的原理基本相同，如图 18-1 所示。被测样品近红外光谱含有样品组成与结构的信息，性质与样品组成结构又是相关的。收集一定数量的样品，其组成或性质是已知的（称为训练集）。使用数学方法将近红外光谱与组成或性质进行关联，可确立这两者间的定量或定性关系，即校正（或称建立模型）。在分析（即预测）中，用校正模型和待测样品近红外光谱计算待测性质。因此，近红外光谱分析方法是一种间接测量方法，其中，校正模型是方法的核心之一。建立模型工作主要包括创建训练集（Training Set）、模型关联（Calibration）和模型验证（Validation）等。这些工作程序应严格遵循标准方法 ASTM E1655^[54] 规定进行。

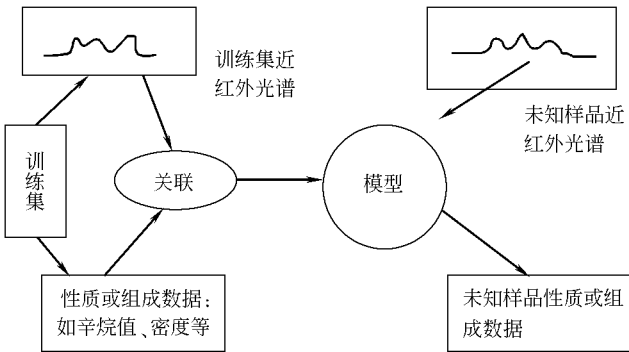


图 18-1 近红外光谱分析流程图

18.3.2 训练集

训练集的样品应具备代表性，即样品集中样本组成或性质的变化范围能够涵盖待测样

品的变化范围。根据样品的特点，确定最佳的光谱测量条件。所有校正集的样本光谱都须在同一测量条件下逐一测量。这些测量条件也将在预测未知样品时使用。近红外光谱仪的工作状态、测量环境等因素对光谱测量精度均有一定影响，近红外光谱仪硬件性能必须十分稳定，对光谱测量操作规程应当规范化。这样可降低近红外光谱测量的误差。训练集中样本性质参数一般是使用传统的标准方法测定得到的，这些数据的测量精度常受仪器状态和操作人员经验的影响，而这些数据的测量精度将直接关系到所建模型的性能，因此，对训练集的性质数据测量精度，应予特别重视。

18.3.3 校正方法

在近红外光谱分析石化产品常用的多元校正方法^[26]，包括多元线性回归（multi-varate linear regrression, MLR）、主成分分析（principle component analysis, PCA）、主成分回归（principle component regression, PCR）、偏最小二乘法（partial least squire, PLS）、拓扑学方法和人工神经网络（ANN）等。MLR, PCR 和 PLS 方法主要在样本光谱与被测性质之间具有较好的线性相关关系情况下应用，比如汽油辛烷值。图 18-2 为重整汽油短波近红外光谱-辛烷值（RON）相关系数曲线，可看出，在 850~1000nm 区间，相关系数接近于 1，具有很好的线性相关关系。而拓扑学方法和 ANN 方法等常用于非线性相关关系的情况，如油品的馏程，与汽油辛烷值相比，其相关性较差。实际工作中应根据具体应用情况的特点选用不同的方法。各种校正方法的特点如下：

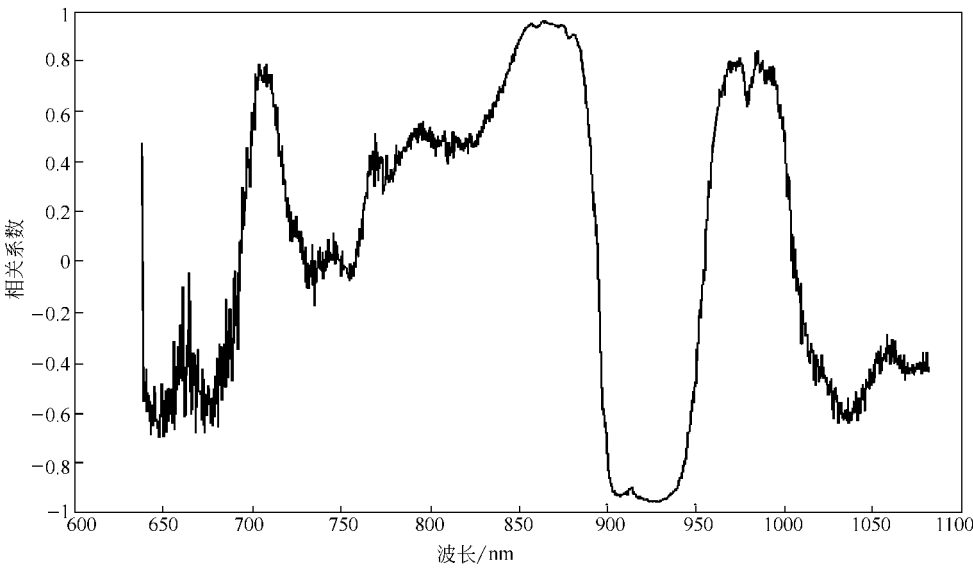


图 18-2 重整汽油短波近红外光谱-辛烷值（RON）相关系数曲线

MLR 方法计算简单，物理意义明确，易于理解，但参加关联变量数目有限制，这样常常丢掉许多光谱信息，要求训练集样品的个数多于变量数目，即使这样它仍对许多性质测量（如辛烷值等）均可获得较好的结果，这种方法在早期近红外光谱定量分析软件中常采用。

PLS 是近红外石化产品分析中使用最多和效果最好的一种方法。它是将因子分析和

回归分析结合的方法。一张近红外光谱通常包括几百到几千个波长处的吸收或透过等数据，它们既包含了待分析组分化学信息，又有干扰组分或其他因素（如仪器和环境等）的信息。PLS 通过因子分析将光谱（多维空间数据，维数相当于波长数目）压缩为较低维空间数据，其方法是光谱数据向协方差最大方向投影。将原近红外光谱分解为多种主成分光谱，不同近红外光谱的主成分分别代表不同组分和因素对光谱的贡献，通过对主成分的合理选取，去掉干扰组分和干扰因素的主成分，仅选取有用的主成分参与模型的回归计算。PLS 对光谱数据的利用可随意选取，可以是全谱也可以是部分数据，与 MLR 相比，更能充分利用光谱的信息，得到更好的关联结果，所不足的是建立模型需要的计算时间稍长，现在的微机（586 以上）计算速度已解决了这个问题。目前，已有许多商品软件使用了这种方法。采用 PLS 建立校正模型，其中很重要的工作是如何对所建模型的质量进行评判，评判方法比较多，使用最多的还是交互验证（Cross Validation）方法^[26]。如何确定奇异点（Outlier）常常影响到模型的质量，也可以说，PLS 方法常容易受奇异点的干扰。采用稳健偏最小二乘法（Robust PLS），在一定程度上能有效抑制奇异点的干扰。化学计量学光谱分析软件版本 2（石油化工科学研究院研制）中也有这一算法。

使用这些方法建立的模型在应用中遇到的一个重要问题就是模型的适应性。如果未知样品不在模型所涵盖的范围内时，预测结果将产生较大偏差。解决问题方法之一是进行模式识别^[9]，即在预测前，首先对未知样品进行类型判别，选择正确的校正模型，或者给予操作者警示。常用识别方法有模糊聚类、马氏距离、和 SIMCA 等^[26]。

近红外光谱仪的状态、测量环境等因素对光谱的影响大多属于非线性的，还有一些性质和光谱的相关关系也是非线性的。解决这类问题方法除对仪器光谱测量进行标准化处理（Absolute Virtue Instrument, AVI）外，常采用拓扑方法和人工神经网络方法。但从现有的商品软件和应用方面来看，这些方法在石化产品分析尚不普及。常用的人工神经网络算法是 BP 算法和自组织。对预测与光谱呈线性相关的性质，人工神经网络并不比 PLS 明显优越，而且人工神经网络还具有不确定性；人工神经网络的外推能力较差，对于超出模型界外的样品，PLS 给出更合理解释的结果。人工神经网络在解决非线性问题和模型抗干扰性方面比 PLS 方法优越^[41,55,56]。使用人工神经网络的重要限制是输入神经元数不能过多，一般是不超过 6~7 个。因此，输入的近红外光谱数据的选取对人工神经网络学习和运算时间有着重要影响。石油化工科学研究院对此做了大量工作^[55,56]，一般，先对近红外光谱数据通过主成分分析进行压缩，选取 3~4 个主成分的得分作为输入参数。一般使用两层的网络结构。

18.3.4 光谱预处理

样品的状态如油品的颜色和聚合物颗粒形状等、仪器外部工作条件如室温和样品温度等都会直接影响着光谱测量精度，引起基线漂移或光谱变形等。通常在校正或预测前，采用数学方法对光谱进行预处理以消除或降低这些影响。常用的方法包括微分、平滑、MSC 和扣减等^[26]。

另一个重要预处理是波长区域的选取，通常根据训练集光谱方差图或光谱-性质相关系数图进行选择^[26,55~57]。

18.3.5 方法精确度判别

根据在实践中的发现,初学者往往在精确度方面产生一些混淆认识,为此指出,通常方法精确度定义规定:使用同一种仪器或同一种方法,检验重复性、再现性和与标准物对照的结果。这些定义同样适合近红外光谱分析方法,除此之外,近红外光谱分析方法还有其特殊性。由于近红外光谱分析是一种间接测量方法,因此,近红外光谱测量结果准确性判别方法是与建立模型所使用的参考方法对照。对于油品性质,参考方法为标准分析方法,如 GBXXXX 或 ASTM DXXXX 方法。常用于近红外光谱分析精确度分析的参数有交互验证标准偏差 (SEC)、预测标准偏差 (SEP) 和近红外方法-参考方法结果的相关系数等。 SEC 反映的是对训练集而言的校正效果; SEP 则指该模型对所选验证集样品而言的预测能力。使用 SEC 和 SEP 评价校正效果一定要考虑所选训练集和验证集的代表性,这种代表性越接近实际,应用要求才越有意义。习惯上,常将近红外方法-参考方法结果的相关系数作为衡量近红外方法结果的准确性,其实,更确切的应使用这两种方法测定结果间的偏差大小。还有许多初学者把这个偏差称之为再现性也是不对的,因为再现性定义是对同一种方法在不同实验室实验结果而言。正确表述应当是:如果两种方法测定结果间的偏差落在参考方法的再现性误差要求范围内,就认为近红外分析方法结果是准确的,反之亦然。

由于 SEC 和 SEP 是评价模型代表性的关键指标,因此,也预示着近红外光谱分析方法在使用过程中,需要不断地进行验证和不断地补充训练集样品变化范围。通常仪器厂家在仪器安装时有一个初始模型,用户在使用后再根据样品的变化情况,对该模型做适当补充。

18.4 汽油分析

1989 年,Jeffrey J. K. 等建立了使用短波近红外光谱技术测定无铅汽油的辛烷值方法^[11],这是一项开创性工作,将近红外光谱分析技术应用引入到石化领域。从此近红外光谱分析技术在石油化工领域获得了非常广泛的应用。可以说,近红外光谱在石化领域中最成功的应用是汽油辛烷值测量,这项应用给石化工业带来巨大经济效益,因而备受关注。除此之外,近红外光谱还可用来测定汽油的烯烃、芳烃、苯、蒸汽压、馏程、碳数分布和密度等项目。

18.4.1 辛烷值分析

辛烷值是表示汽油抗爆性能的重要参数。车用汽油牌号是按辛烷值等级划分的,国内汽油牌号包括 70 号,90 号,93 号和 97 号等,标号越高,品质越好。因此,辛烷值是在汽油生产和使用中规定必须检测的项目之一。汽油辛烷值测定的标准方法为 ASTM D-2699^[56],我国采用与之等价的国家标准方法 GB/T 5487—1995 (研究法)和 GB/T 503—1995 (马达法)。这些标准方法规定辛烷值测定是在标准操作条件下,采用标准引擎试验机 (ASTM-CFR),比较标准校验燃料和待测汽油的爆震强度,通过内插或压缩比方法计算待测汽油的辛烷值。这些标准方法在应用中存在的主要问题是分析周期长,难以满足生产控制和在线汽油调和工艺等对辛烷值快速分析的要求;消耗样品量 (1000 毫升/次) 较

大，无法用于研究中的中型、微反装置产生少量样品的辛烷值测定；所需试验设备投入高（200 多万人民币/台试验机），标准校验燃料消耗和设备维护费都很高；此外，设备操作复杂、污染大。对于炼厂汽油生产的中控分析，成品油流通领域中的质量监督，以及国防野战情况下汽油品质分析工作，要求分析样品数量多、分析速度快，标准方法无法满足其

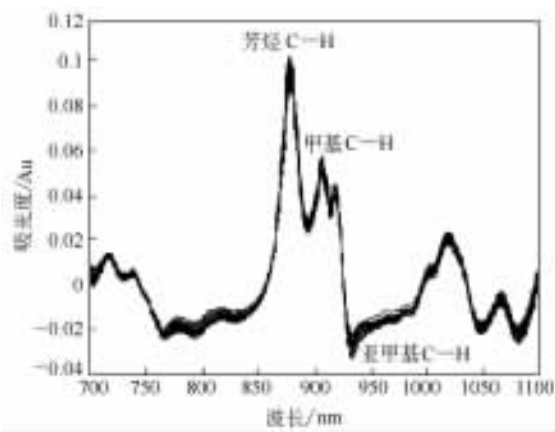


图 18-3 一批重整汽油光谱

应用要求。因此，为解决上述问题，迫切需要建立一种更简便和更快速的辛烷值分析方法。多年来，为此展开了采用核磁共振、气相色谱、红外、拉曼和近红外光谱等方法测量汽油辛烷值的大量研究，经长期实践表明，近红外光谱分析方法是具有实用性和发展前途的方法。

汽油是由各种烃类混合物组成的。烃类主要官能团包括甲基、亚甲基、次甲基、烯烃、芳烃等。它们在近红外光谱区域均有吸收，如图 18-3，图 18-4 所示，有关某些谱带解析列于表 18-3。可以看出不同样品的近红外光谱图形状很相似，不同官能团的吸收谱带之间的重叠严重，但彼此之间也有微小差别。如果对这些光谱做方差处理，可以得到突出谱图变化的方差图。

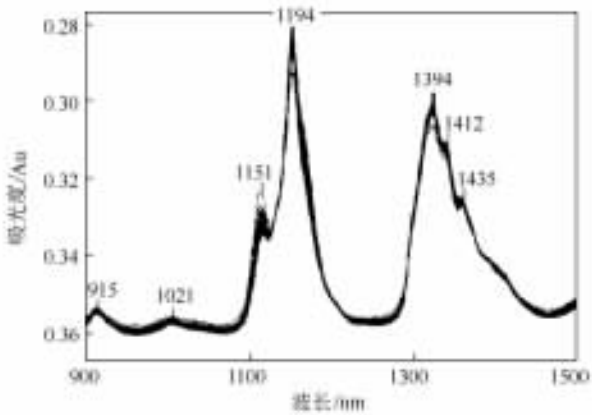


图 18-4 一批成品油光谱

表 18-3 汽油官能团光谱解析

波长/nm	归 属	波长/nm	归 属
915	CH ₂ 伸缩振动三级倍频	1394	CH ₂ 的合频
1021	CH ₂ 与 CH ₃ 的合频	1412	芳烃和 CH ₂ 的合频
1151	芳烃和 CH ₃ 伸缩振动二级倍频	1435	芳烃和 CH 的合频
1194	CH ₃ 伸缩振动二级倍频		

图 18-5 为 30 个重整汽油光谱方差。它表明，在 875nm、910nm 和 935nm 处有明显的光度变化（最大变化为 8×10^{-3} 吸光度），对应于芳烃 C—H 的三级倍频。主要因样品

芳烃含量（60％～75％ m/m）变化引起的。910nm 和 935nm 处的峰分别对应于甲基和亚甲基的三级倍频吸光度的变化。说明近红外光谱能反映汽油在组成和结构上的变化差别，可以作为关联计算辛烷值的变量。

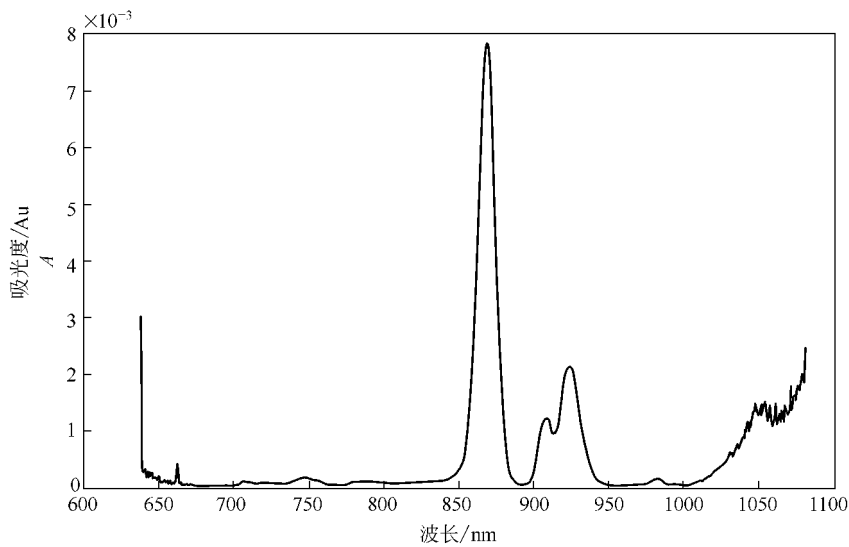


图 18-5 30 个重整汽油的光谱方差图

从已发表的众多工作看，近红外光谱测定汽油辛烷值使用的校正方法以 PLS 方法为主。我们就近红外光谱测量重整汽油研究法辛烷值，比较了 MLR、PCR 和 PLS 方法的关联效果（如表 18-4 所示）。根据表中的平均偏差，说明 PLS 校正方法的校正效果最优。

表 18-4 MLR、PCR 和 PLS 方法校正效果对比							
序号	GB*	MLR 预测	MLR 偏差	PCR 预测	PCR 偏差	PLS 预测	PLS 偏差
1	103.3	103.4	0.1	103.6	0.2	103.5	0.1
2	99.7	98.1	1.1	99.0	0.5	99.7	0.0
3	102.4	101.5	0.6	102.8	0.3	102.5	0.1
4	98.1	97.5	0.4	98.2	0.0	98.1	0.0
...	...	...	...	...	...	...	...
24	100.0	101.7	1.2	100.2	0.2	99.9	0.0
25	100.9	102.6	1.2	100.5	0.3	100.7	0.1
26	102.9	104.3	1.0	103.0	0.1	102.9	0.0
27	103.5	104.1	0.4	104.7	0.8	103.5	0.0
28	99.4	99.5	0.1	99.2	0.1	99.3	0.1
29	93.9	92.6	0.9	93.2	0.5	93.7	0.1
		平均标准偏差＝	0.53	平均标准偏差＝	0.34	平均标准偏差＝	0.26

* GB 为标准方法，即爆震引擎试验机方法。

近红外光谱分析方法测定汽油辛烷值的应用非常广泛，概括起来主要为在自动汽油调和辛烷值在线分析、炼油厂汽油生产中间控制分析、成品汽油分析和国防用燃料的辛烷值测定等方面。有关在线分析将在本章 18.9 中介绍，以下分别举例介绍后者的应用。为了说明近红外光谱测定辛烷值方法的精确度，首先介绍标准方法的精确度要求，如表 18-5 所示。

表 18-5		标准方法再现性			
性质	试验方法	再现性	性质	试验方法	再现性
研究法(RON)	GB/T 5487	0.7(90)	马达法(MON)	GB/T 503	1.2(80)

18.4.1.1 炼油厂分析应用

对于燃料型的炼油厂，辛烷值分析是汽油生产的一项重要工作，例如中国石化集团公司沧州分公司（年加工原油 300 万吨左右，属国内中型炼油厂），每天多次分析各生产装置汽油馏出口和汽油成品罐的辛烷值。其中，涉及的生产装置包括两套催化装置和两套精制装置。分析的汽油包括两种稳定汽油和汽脱汽油（简称汽脱）。过去全厂依靠 1 台进口的标准引擎试验机（ASTM-CFR）测定辛烷值，不仅设备维护费用高，更主要的问题是不能及时为生产提供辛烷值数据，严重影响汽油生产质量和效益的有效控制。

沧州炼油厂是国内最早开始将近红外光谱分析技术用于汽油辛烷值分析的单位。自 1997 年，在化验室内使用 NIR-2000 型 CCD 近红外光谱仪（英贤仪器公司制造，石油化工科学研究院监制）和化学计量学光谱分析软件（石油化工科学研究院研制）进行汽油辛烷值常规分析工作，测量条件：分辨率为 0.5nm，玻璃比色池（光程 50mm），室温测量。

殷宗玲和李克忠采用近红外光谱方法和 GB/T 方法同步测量辛烷值结果做了对比，被测量的汽油辛烷值变化范围为 89~96，两种结果间的偏差均小于 0.2 个辛烷值单位，也小于 GB/T 5487 的再现性（0.7）要求。在 CCD 近红外光谱仪上，在 1 天内对一个一稳汽油研究法辛烷值进行了 10 次重复测定，测量间隔为 5 分钟，最大偏差为 0.1 辛烷值单位，标准偏差为 0.042，优于 GB/T 5487 的重复性要求，0.3（90）。

虽然沧州炼油厂生产规模不大，但辛烷值中间控制分析状况与国内其他炼油厂的类似，因此，具有代表性。采用 GB 方法，单个样品汽油辛烷值分析需约 40min。通常，从取样到数据反馈的周期为 4h，分析滞后生产要求。当不合格产品漏检进入成品罐时，必须再加工处理，将造成大的经济损失。使用近红外光谱分析技术后，大大缩短分析时间（5min/样品）。可及时监测生产装置产品的质量，有效地解决了上述问题。例如 1998 年 5 月 25 日和 6 月 15 日，近红外光谱分析发现二稳装置汽油的研究法辛烷值为 90.6 和 88.8，出现不符合 90 号汽油质量要求的情况。由于监测及时，对工艺条件做了及时调整，如采取降低直汽回炼量，提高反应温度等，使产品质量迅速恢复到正常情况。显然，避免一次加工再处理带来的较大经济损失。

根据沧州炼油厂化验室成本数字统计，采用 GB 方法，辛烷值分析成本费用（包括标准燃料与马达维护）约为 300 元/个。根据分析频次为 324 个样品/半年，推测辛烷值单项分析费用： $300 \times 324 = 9.72$ 万元，一年则需 19.44 万元。而 CCD 近红外光谱仪测定辛烷值除正常使用少量电费外，其他开支很少。所以，近红外光谱分析技术测定辛烷值将节省大量化验费用。

目前，国内已经有几十家炼油厂的化验室成功地使用近红外光谱分析技术（主要以 NIR-2000 型和 NIR-3000 型仪器为主）测定汽油辛烷值^[59~64]。以下列举几个实例说明国内近红外测定汽油辛烷值应用情况。

上海石化股份有限公司炼化部蒋健采用 NIR-3000 近红外光谱仪测定了催化裂化汽油，成品油（90 号，92 号，93 号和 95 号）的辛烷值^[59]，用近红外光谱快速分析技术代替标准方法在汽油调和试验中进行控制分析。表 18-6 列出近红外方法与标准方法结果对比。并用成对 t 检验判断这两种方法是否存在着显著性差异。给定显著性水平 $\alpha=0.05$ ，查表得 $t_{(0.05,19)}=2.09$ ，可看出， t 检验结果 $|t|$ 均小于 2.09，说明这两种分析方法测定结果是一致的。

表 18-6 近红外方法与标准方法测定结果的对比统计结果				
项目	样本数	范围	差值 ^①	$ t $
RON	20	91.2~94.5	-0.17	1.38
MON	20	80.1~83.3	0.24	0.95

注：① 两种方法结果偏差平均值。

王京华，邵波和黄小英等人^[61]在 NIR-2000 近红外光谱仪上，建立了成品罐汽油辛烷值模型，经过一年多时间的反复验证，实现了在 5min 内快速准确测定成品油的辛烷值，较大限度地避免油品质量不合格或质量指标过度富余的情况出现，对汽油产品的调和出厂起着积极的作用。表 18-7 列出了预测准确性数据。

表 18-7 成品罐汽油验证试样集对比统计结果				
测定项目	样品数量	性质范围	SEP	标准方法再现性要求
RON	20	88.2~94.2	0.36	0.8~0.6

张新志，余新铭等人^[63]将 NIR-2000 近红外光谱仪用于汽油辛烷值测定，建立模型的样品包括 90 号成品汽油，催化装置馏出口汽油和混合油（含石油和 MTBE）。表 18-8 列出与 GB 方法分析结果对照情况，说明 GB 方法和近红外方法结果的偏差小于 GB 方法的再现性要求。按照每周对馏出口检测 2 次辛烷值计算，一年化验室因此节省费用 5.76 万元人民币。

表 18-8 近红外光谱法测定汽油辛烷值数据对照表						
样品名称	RON			MON		
	实测	NIR	偏差	实测	NIR	偏差
催 0327	92.7	92.5	0.2		79.8	
联 3027	93.0	92.8	0.2	81	80.7	0.3
装 90 号 0325	92.3	92.2	0.1	80.1	80.2	-0.1
装 90 号 0326	92.5	92.4	0.1	80.1	80.0	0.1
装 90 号 0327	92.6	92.5	0.1	80.5	80.3	0.2
201 号 0321	93.0	93.2	-0.2	80.5	80.3	0.2
201 号 0326	92.4	92.5	-0.1	80.0	80.2	-0.2
202 号 0204	93.6	94.1	-0.5	80.9	80.6	0.3
202 号 0325	92.4	92.4	0.0	80.0	80.3	-0.3
204 号 0328	92.2	92.4	-0.2	80.4	80.0	0.4
202 号 0329	92.3	92.7	-0.4	80.5	80.2	0.3

18.4.1.2 军用汽油辛烷值分析

早在 1991 年，美军在巴拿马与代号为沙漠盾牌和沙漠风暴战争中发现野外军需燃料的品质快速评价具有重要意义。为此，美国西南研究院在 20 世纪 90 年代初期开展红外光谱以及近红外光谱快速分析军用汽油辛烷值及其他燃料性质的研究工作^[65]。这种军用需求和后来研究开发工作导致很多便携式近红外油品分析仪产品的面世，如 FOX 公司红外燃料分析仪，Zeltex 公司的滤光片式便携式近红外燃料分析仪，ASD 公司的 GS—500 便携式近红外分析仪等。这些产品携带方便符合军用要求，应用在分析模型（原因之一是国内燃料性质与国外的差别较大）适用范围上受到一定的限制。

我国也很重视这一先进技术的应用，刘慧颖，熊春华和韦锐等人并一直对此进行研究^[66]。在近几年来做了大量工作，开发了便携式近红外燃料分析仪，适合车载或便携用途。从全国各地收集了大量车用燃料样品，组成混合校正集，分别在 NIR-2000 近红外光谱仪和便携式仪器上建立了各种燃料性质模型^[67~69]。表 18-9 列出了汽油辛烷值测定结果，说明近红外光谱测定车用燃料的结果与 GB 方法测定结果的一致性较好。

表 18-9 汽油辛烷值实测值和近红外光谱预测值的结果

性质 编号	RON			MON		
	实测值	预测值	偏差	实测值	预测值	偏差
	GB/T 5487	NIR-2000		GB/T 5487	NIR-2000	
1	91.9	91.9	0.0	79.3	80.1	0.4
2	91.9	92.7	0.8	79.3	80.1	0.8
3	—	—	—	79.1	79.8	0.7
4	91.2	91.2	0	—	—	—
5	91.6	91.7	0.1	79.4	80.0	0.6
6	94.8	95.0	0.2	82.8	82.6	—0.2
7	91.1	91.1	0	79.1	79.1	0
8	92.6	93.3	0.7	80.8	8.05	—0.3
9	91.4	91.6	0.2	79.0	79.2	0.2
10	91.6	91.8	0.2	79.8	79.8	0
11	91.0	91.4	0.4	79.8	79.6	—0.2
12	92.6	92.8	0.2	79.9	80.0	0.1
13	89.2	89.5	0.3	77.8	77.9	0.1
14	92.4	92.0	—0.4	80.7	79.9	—0.8
15	88.1	88.9	0.8	77.1	78.0	0.9
16	92.0	91.6	—0.4	80.7	80.1	—0.6
17	92.6	93.2	0.6	81.0	81.4	0.4
18	93.0	92.3	—0.7	80.2	80.2	0
19	91.8	92.7	0.9	79.8	80.2	0.4
20	92.0	92.4	0.4	79.2	79.8	0.6
21	92.6	92.8	0.2	81.1	81.3	0.2
22	91.1	90.7	—0.4	78.7	78.9	0.2
23	91.3	90.4	—0.9	79.8	79.9	0.1
24	91.1	90.2	—0.9	79.3	79.1	—0.2
25	92.0	92.3	0.3	80.3	80.2	—0.1
26	91.8	92.7	0.9	—	—	—
27	91.5	91.5	0	80.6	81.0	0.4
28	91.6	91.5	—0.1	81.1	81.0	—0.1

续表

性质 编号	RON			MON		
	实测值	预测值	偏差	实测值	预测值	偏差
	GB/T 5487	NIR-2000		GB/T 5487	NIR-2000	
29	91.4	91.2	−0.2	80.4	80.7	0.3
30	91.2	91.5	0.3	80.7	80.8	0.1
31	91.2	91.5	0.3	79.4	79.9	0.5
32	—	—	—	80.7	81.4	0.7
33	93.3	93.1	−0.2	—	—	—
34	93.0	92.4	−0.6	80.7	80.4	−0.3
35	91.6	92.0	0.4	80.2	80.2	0
36	—	—	—	80.7	80.3	−0.4
37	—	—	—	80.6	80.2	−0.4
38	93.0	92.1	−0.9	80.6	79.9	−0.7
39	93.0	92.7	−0.3	80.0	79.9	−0.1
40	93.5	93.5	0	82.5	82.1	−0.4
41	—	—	—	79.7	79.0	−0.7
42	93.1	92.6	−0.5	81.8	81.2	−0.6
43	90.8	91.0	0.2	80.1	81.1	0
44	91.6	91.4	−0.2	80.2	79.6	−0.6
45	90.9	90.9	0	79.9	79.2	−0.7
46	90.4	91.0	0.6	82.5	83.3	0.8

此外，GB 方法测定辛烷值需要样品量约为 500mL，显然，对科研中一些量很少的样品辛烷值测量是不适用的。近红外光谱测量需要的样品量很少（几毫升），适合一些微量产物辛烷值的快速测定，杨玉蕊等在这方面做了大量有成效的工作，使得近红外光谱分析技术在催化剂和工艺研究与评价中发挥了积极作用^[70]。

总之，近红外光谱分析技术非常适合用于汽油生产和科研工作中的辛烷值检测，具有方便、快速和准确的优点，而且还可为化验室节省大量分析费用。

18.4.2 其他性质分析

除辛烷值外，汽油重要性质还包括蒸汽压、馏程和组成等。近年来各国相继出台有关空气洁净的环保法规，其中，对汽油中烯烃、芳烃、苯和硫含量的限制越来越严格。因此，在生产和使用中对这些性质的测量越来越受到重视。由于有关现有标准方法具有操作复杂和速度慢等缺点，不能适应目前清洁燃料生产和使用中对分析频次与精度的要求。而近红外光谱分析技术特点和优势恰好顺应了这一发展的新要求，因此，在石油炼制工业中备受关注。从近年来近红外光谱在国内炼油厂应用情况看，炼油厂特别注重近红外光谱用于汽油组成、馏程以及蒸汽压等性质测量。

近红外光谱测定汽油组成、蒸汽压和馏程的原理与它测定辛烷值的原理是相同的。虽然近红外光谱测量这些性质均能够获得满意结果，但是各种性质的测量精度有所不同。笔者根据长期实践经验认为，近红外光谱测定燃料组成的精确度比测量其他性质的更好，相对来讲，馏程性质，尤其终馏点和初馏点的测量较困难。

18.4.2.1 二甲苯异构体的分析

褚小立在 NIR-3000 近红外光谱仪上测得的邻-二甲苯、对-二甲苯和间-二甲苯的近红

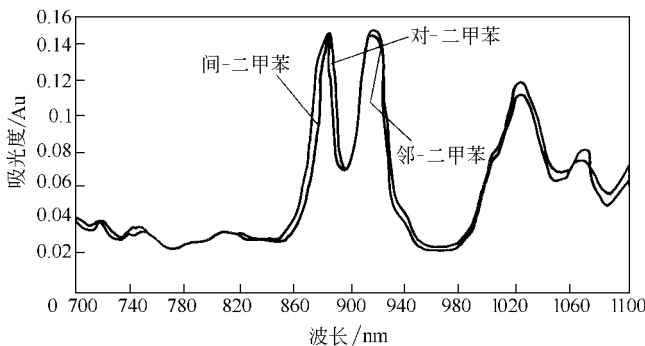


图 18-6 邻-二甲苯、对-二甲苯和间-二甲苯的近红外光谱图

外光谱图（如图 18-6 所示）^[71]。可以看出，二甲苯异构体芳环 C—H 特征吸收存在差异（对-二甲苯：880.8nm，间-二甲苯：878.4nm，间-二甲苯：877.0nm）由于相邻甲基的振动偶合，使邻-二甲苯的甲基 C—H 有两个特征吸收峰（907.4nm，918.6nm），与对-二甲苯和间-二甲苯的甲基吸收存在较大的差异。从光谱特征上可以推断，虽然三者之间有较严重的干扰和重叠。为了研究近红外光谱测量二甲苯三种异构体含量，使用化学纯试剂，配制 40 个不同二甲苯异构体浓度（0.5v%~15v%）的正己烷溶液。用 K-S 方法^[72]选出 28 个样品作校正集，剩余 12 个样品验证集。分析体系的微分光谱-性质相关系数曲线表明，尽管 3 种组分的光谱相互干扰大，但在光谱-性质之间还是具有一定的相关关系，其中，邻-二甲苯的相关关系最强。采用 PLS 方法关联，得到非常满意的定量结果（如图 18-7）。

列举这个实例的目的是为了说明近红外光谱与组成之间具有更强的线性相关性，近红外光谱方法可以成功解决一些在结构上非常相似的成分测定难题，那么对一般组成测量则比较容易，比如近红外光谱非常适合测定汽油组成。

18.4.2.2 汽油组成分析

在炼油厂化验室，近红外光谱分析组成技术主要用来替代常规分析烯烃、芳烃含量方法（FIA，荧光指示剂色谱法）。以下是几个汽油组成分析应用实例。

刘莎，在 2000~2002 年期间，共收集了 220 多个汽油样品，取自催化裂化装置馏出口，采用 GB/T 11132—1989 方法（即 FIA 方法）测定了汽油的饱和烃，烯烃和芳烃作为近红外光谱建立模型的参考数据。在 NIR—2000 近红外光谱仪建立了组成校正模型，用于中间控制分析^[73]。表 18-10 比较 NIR 方法和 FIA 方法测量结果，说明两种方法结果一致。表 18-11 是近红外光谱方法测量精密度的，明显优于 FIA 方法。

表 18-10 NIR 方法与 FIA 方法测定汽油族组成结果对比^[68]

项 目	饱和烃/v%	烯烃/v%	芳烃/v %
FIA 再现性要求	5.00~5.50	8.40~8.60	2.00~2.50
<i>d</i>	0.34	-0.37	0.18
SEP	0.94	1.18	0.61
<i>t</i>	1.73	1.50	1.42
<i>t</i> (0.05,23)		2.07	

注：*d* 为两分析方法之间差值的平均值；SEP 为两分析方法之间差值的标准偏差。

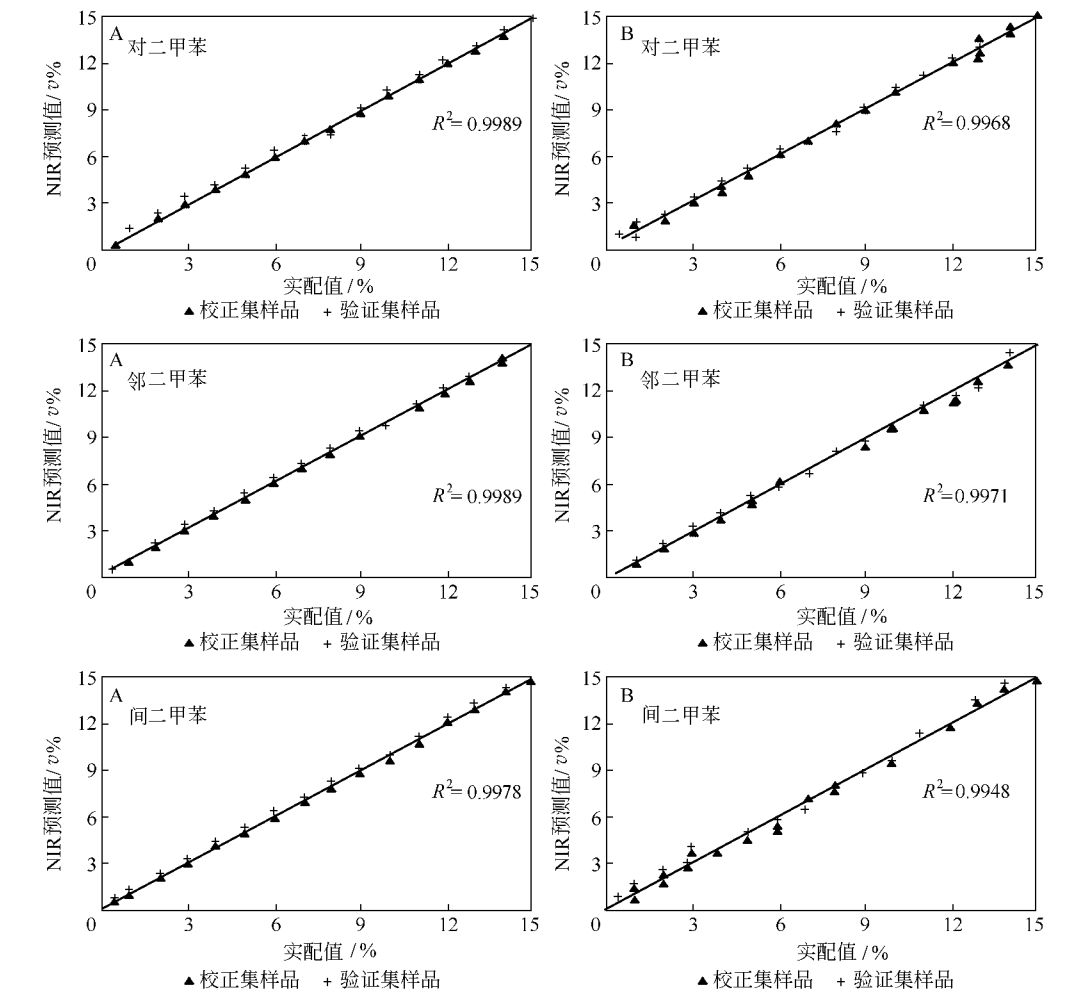


图 18-7 二甲苯三种异构体 NIR 预测值与实配值的对比结果

A—实验室 NIR-3000 B—在线 NIR-6000

表 18-11 NIR 方法重复性试验结果 ^[73]						
测试次数	三催化汽油			二催化汽油		
	饱和烃/v%	烯烃/v%	芳烃/v%	饱和烃/v%	烯烃/v%	芳烃/v%
1	40.24	45.43	15.26	33.52	53.92	12.42
2	39.92	45.80	15.29	33.54	53.91	12.49
3	39.89	45.68	15.30	33.47	54.04	12.40
4	39.70	45.83	15.29	33.76	53.61	12.53
5	39.79	45.58	15.34	33.73	53.61	12.68
FIA 重复性要求/v%	1.70	2.00	1.20	1.70	2.10	1.20
NIR 重复性/v%	0.20	0.16	0.03	0.12	0.20	0.11
NIR 平均值/v%	39.90	45.66	15.30	33.60	53.82	12.50
NIR 相对标准偏差/%	0.50	0.35	0.20	0.36	0.37	0.88

中国石化集团股份有限公司沧州分公司自 1999 年开始，在汽油调和生产控制分析中，采用近红外光谱方法替代 GB/T 11132—1989 方法测定汽油组成^[74]。汽油调和成分为催

化裂化汽油、重整汽油和 MTBE 等。殷宗玲和李克忠建立了有关 90 号和 93 号汽油的芳烃和烯烃共 4 个模型（其参数如表 18-12 所示）。表 18-13 是分析清洁汽油的结果，说明与 FIA 方法结果一致。几年来，该方法一直在汽油调和和生产中发挥着重要作用。

表 18-12		调和汽油模型 ^[74]		
模型种类	样本数	范围/%(v/v)	相关系数	平均偏差
90 号烯烃模型	40	21.2~47.8	0.9858	0.7866
90 号芳烃模型	40	17.8~31.0	0.9654	0.8753
93 号烯烃模型	40	22.4~36.8	0.9519	0.7057
93 号芳烃模型	40	27.0~39.0	0.9269	0.7377

表 18-13		近红外分析与 FIA 测定数据比较 ^[74]				
样品	NIR 测定芳烃/%v	FIA 测定芳烃/%v	偏差	NIR 测定烯烃/%v	FIA 测定烯烃/%v	偏差
93 号 1	32.8	33.8	−1.0	31.0	30.8	+0.2
2	35.6	37.6	−2.0	28.9	27.0	+1.9
3	39.3	39.0	+0.3	23.4	24.2	−0.8
4	31.2	32.5	−1.3	34.0	32.5	+1.5
5	32.0	31.2	+0.8	32.2	33.7	−1.5
6	30.7	31.1	−0.4	31.0	30.7	+0.3
7	34.1	35.2	−1.1	33.9	34.2	−0.3
8	27.9	28.3	−0.4	30.2	29.3	+0.9
9	29.5	27.8	+1.7	38.1	37.7	+0.4
	极差	2.0		极差	1.9	
	SEP	1.2		SEP	1.1	
90 号 1	27.9	27.2	+0.7	38.5	39.6	−1.1
2	23.3	24.0	−0.7	42.5	41.7	+0.8
3	21.8	23.0	−1.2	43.5	43.2	+0.3
4	23.1	22.1	+1.0	44.9	44.3	+0.6
5	21.6	20.8	+0.8	43.9	45.0	−1.1
6	24.1	24.1	0	45.6	46.5	−0.9
7	23.6	22.5	+1.1	41.1	39.6	+1.5
8	29.9	28.5	+1.4	30.3	30.5	−0.2
9	31.0	30.3	+0.7	32.7	32.0	+0.7
	极差	1.4		极差	1.5	
	SEP	1.0		SEP	0.9	

对重整汽油的烷烃、环烷烃和芳烃（PNA）组成，目前经典的分析方法是采用毛细管色谱法，由于汽油的组成较复杂，要把 C₃~C₁₀⁺的各烷烃、环烷烃和芳烃分开，必需采用一套分析软件对 120 个色谱峰进行辨别计算，分析时间需要 2.5h，这往往不能及时为生产装置提供分析数据。黄小英、邵波等^[75]人采用近红外光谱方法用于重整汽油分析^[73]，建立各碳数的校正模型并在 3min 内一次性预测未知样品 C₃~C₁₀⁺的 PNA

组成，预测结果与色谱法比，准确性及重复性满足要求。共收集了 110 个重整样品（90 个为校正集，剩余 20 个作验证集），采用 Q/SH016.03.175—1995 方法测定建立模型的参考数据。使用 NIR-2000 近红外光谱仪测定其近红外光谱。使用 PLS 方法建立一系列模型（模型参数见表 18-14），表 18-15 为与 Q/SH016.03.175—1995 方法测定结果的对照。任取一个重整汽油，用近红外光谱重复测定 6 次，并用已建立的模型预测各碳数组成，结果表明近红外光谱测定重整汽油族组成的标准偏差优于标准方法的标准偏差。

表 18-14 重整汽油 C₃~C₁₀⁺ PNA 各组成建模所选参数和结果

组成	样品数	含量范围/w%	主因子数	相关系数	SEC/w%
C ₃ P	88	0.2~0.7	6	0.8024	0.16
C ₄ P	85	0.3~1.0	7	0.8712	0.16
C ₅ P	87	11.2~17.9	7	0.9226	0.40
C ₆ P	88	1.3~3.9	3	0.7994	0.75
C ₇ P	90	8.4~11.6	9	0.9329	0.29
C ₈ P	90	3.1~6.5	7	0.8758	0.32
C ₉ P	90	0.5~1.6	6	0.9231	0.16
C ₁₀ ⁺ P	91	0.2~0.8	6	0.7685	0.10
C ₅ N	82	0.3~0.9	8	0.8569	0.19
C ₆ N	85	0.6~2.5	5	0.9719	0.20
C ₇ N	88	0.2~0.5	7	0.8395	0.06
C ₈ ⁺ N	89	0.06~0.18	7	0.8154	0.03
C ₆ A	87	4.1~9.4	7	0.9924	0.26
C ₇ A	90	13.9~18.0	7	0.9343	0.41
C ₈ A	90	19.1~23.3	8	0.9443	0.51
C ₉ A	90	14.1~18.6	7	0.9300	0.86
C ₁₀ ⁺ A	88	1.4~6.3	6	0.9144	0.31

注：所有光谱预处理都采用二阶微分，光谱区间为 850~950nm，SEC 为校正标准偏差。

表 18-15 重整汽油 C₃~C₁₀⁺ PNA 各组成近红外与标准方法分析结果对比

序号	C ₃ P		C ₄ P		C ₅ P		C ₆ P		C ₇ P		C ₈ P		C ₉ P		C ₁₀ ⁺ P		C ₅ N	
	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH
1	0.40	0.36	0.58	0.56	2.51	2.43	13.55	14.53	9.63	9.83	4.16	4.33	0.91	0.92	0.32	0.28	0.78	0.75
2	0.15	0.14	0.66	0.55	2.46	2.39	13.83	14.12	10.08	9.99	4.83	5.24	1.28	1.25	0.36	0.39	0.32	0.33
3	0.22	0.15	0.48	0.32	2.27	2.09	15.77	16.09	10.26	10.27	4.72	4.97	0.98	1.05	0.28	0.29	0.34	0.40
4	0.52	0.59	0.73	0.69	2.50	2.17	13.86	14.34	9.63	9.73	4.25	4.53	0.96	1.08	0.32	0.28	0.57	0.57
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16	0.67	0.72	0.86	0.94	2.40	2.86	12.91	13.80	9.85	9.80	4.71	4.21	1.15	0.84	0.25	0.22	0.44	0.51
17	0.78	0.85	0.83	0.93	2.71	2.87	13.62	14.51	9.64	9.36	3.91	3.41	0.71	0.58	0.13	0.15	0.45	0.45
18	0.63	0.68	0.83	0.79	2.65	2.33	14.23	14.48	9.01	9.10	3.64	3.77	0.83	0.83	0.19	0.22	0.41	0.38
19	0.66	0.74	0.69	0.88	2.83	2.70	12.78	13.66	9.27	8.83	3.80	3.53	0.80	0.72	0.24	0.20	0.21	0.14
20	0.44	0.53	0.47	0.60	1.96	2.09	11.80	11.74	9.55	8.91	4.46	4.07	1.02	1.02	0.22	0.26	0.78	0.68
SEP	0.09		0.10		0.28		0.70		0.56		0.31		0.15		0.04		0.05	
t	1.79		1.47		1.19		0.81		0.64		1.21		1.10		0.18		1.58	

续表

序号	C ₆ N		C ₇ N		C ₈ ⁺ N		C ₆ A		C ₇ A		C ₈ A		C ₉ A		C ₁₀ ⁺ A	
	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH	NIR	Q/SH
1	1.61	1.41	0.38	0.35	0.24	0.21	6.17	6.51	15.90	16.85	3.69	3.64	0.91	0.92	0.32	0.28
2	2.30	1.86	0.41	0.30	0.22	0.23	7.36	7.75	16.80	16.51	2.74	2.54	1.28	1.25	0.36	0.39
3	1.55	1.76	0.31	0.33	0.18	0.20	6.77	7.04	15.97	15.94	4.08	3.33	0.98	1.05	0.28	0.29
4	1.64	1.86	0.33	0.36	0.18	0.18	6.78	6.79	16.40	16.11	4.49	4.15	0.96	1.08	0.32	0.28
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
16	0.99	0.95	0.32	0.29	0.13	0.09	5.62	5.63	16.10	16.43	3.67	3.99	1.15	0.84	0.25	0.22
17	0.96	0.88	0.29	0.25	0.12	0.11	6.52	6.48	17.94	17.83	2.82	2.82	0.71	0.58	0.13	0.15
18	1.10	0.89	0.43	0.58	0.11	0.11	5.66	6.99	17.64	18.84	2.91	3.58	0.83	0.83	0.19	0.22
19	1.00	0.96	0.28	0.30	0.11	0.09	6.10	6.04	16.79	16.75	3.48	3.28	0.80	0.72	0.24	0.20
20	1.50	2.03	0.36	0.35	0.12	0.14	8.15	8.09	16.24	16.44	4.63	4.35	1.02	1.02	0.22	0.26
SEP	0.22		0.06		0.05		0.28		0.89		1.33		1.05		0.31	
t	1.41		0.32		0.98		1.07		0.69		0.50		0.67		0.12	

18.4.2.3 馏程分析

汽油馏程是炼油厂化验室分析任务中最繁重的工作之一，为此，对使用近红外光谱方法尝试解决汽油馏程快速分析的研究工作始终没有停止过。但是，与辛烷值和组成相比，馏程测定相对比较困难。经验表明，对于特定的工艺和加工原料变化不大的情况下，通常，近红外光谱方法可以用于测定 10%，50%和 90%的馏出温度，但用于测定终馏点和初馏点的误差比较大。其原因是决定初馏点和终馏点的组分在样品中的含量很低，而且在组成上与样品本体的很接近，导致它们的光谱信息很弱。因此，汽油馏程的近红外分析模型适用范围和外延能力远不如分析辛烷值和组成。以下是炼油厂应用近红外光谱测定汽油馏程的实例。

中国石化集团股份有限公司济南分公司^[76]的常减压、重油催化裂化、催化重整和加氢精制等装置都生产多种汽油馏分，因此，汽油馏程分析项目特别多，频次繁，工作量大。采用标准方法 GB/T 6536—1997 测定，需要十几台恩氏馏程测定仪，三人同时操作才能完成，分析时间也较长，不能起到及时指导生产的作用。为此，杨晶使用 NIR-2000 近红外光谱仪，建立了近红外光谱测定汽油馏程分析方法^[76]。所测定的汽油组分包括 800kt/a 重油催化裂化粗汽油和稳定汽油，1.40Mt/a 重油催化裂化粗汽油和稳定汽油，常减压直馏汽油和溶剂（汽）油，催化重整原料油、预精制油（C-101↓），精制油（C-102↓），重整生成油等。参考方法为 GB/T 6536—1997 《石油产品蒸馏测定法》。表 18-16 为各模型的性质参数；表 18-17 为与 GB/T 6536—1997 方法测定的对照。

表 18-16 部分汽油样品的校正模型参数

项目	催化裂化粗汽油						催化裂化稳定汽油					直 馏 汽 油			
	HK	10%	50%	90%	KK	HK	10%	50%	90%	KK	HK	10%	50%	90%	KK
主因子数	10	11	10	10	11	11	11	11	11	11	10	10	9	9	9
相关系数(R)	0.90	0.95	0.96	0.94	0.96	0.95	0.88	0.95	0.94	0.86	0.97	0.98	0.96	0.94	0.96
标准偏差(SEC)	0.82	1.12	0.80	1.19	1.38	0.5	0.68	0.46	1.05	1.42	0.20	0.47	0.72	0.66	0.69

18.5 航煤分析

航煤是重要燃料之一。我国采用规格标准 GB 6537—1994 基本与国际通用的 Jet A-1 标准一致，规定了各项性质的标准试验方法（如表 18-19 所示），有关试验方法的再现性要求见表 18-20。但是，这些标准试验方法操作费时，为此人们在试验方法的快速和自动化方面做了大量的研究工作。法国 ISL 公司^[77]遵循 ASTM 标准方法，发明了一系列性质（如冰点、闪点、凝点、馏程等）自动测定仪，降低了操作强度，提高了工作效率。奥地利 Anton Paar 公司生产了密度/重度测定仪。但是，在许多情况下还是不能满足对喷气燃料性质的快速检测要求，比如野外喷气燃料质量检验、生产装置产品在线检验等。S. Garrigues^[78]研究了 FT-IR 光谱方法预测喷气燃料的七种性质（密度、冰点、闪点、芳烃含量、初馏点、终馏点、黏度），其重复性和再现性均在 ASTM 要求的范围之内。20 世纪 90 年代后，美军为了寻找能满足在靠近战场条件下实现快速、准确鉴别燃料质量的技术途径，美国西南研究院进行了用光谱方法代替传统化验方法的大量研究，近红外光谱方法是其中之一。与中红外光谱方法相比，近红外测量操作更方便。

表 18-19 喷气燃料规格规定的标准试验方法					
性 质	国家标准试验方法	ASTM 试验方法	性 质	国家标准试验方法	ASTM 试验方法
冰 点	GB 2430	ASTM D2386	密 度	GB 1884 或 1885	ASTM D1298
芳烃含量	GB 11132	ASTM D1319	馏 程	GB 6536	ASTM D86

表 18-20 喷气燃料质量性质项目、试验方法及再现性要求		
性 质	试验方法	再现性要求
馏程 /℃	GB/T 6536	
10%		6.0
20%		6.0
50%		6.0
90%		8.0
终馏点		10.5
密度(20℃)/(g/cm³)	GB/T 1884 或 SH/T 0604	0.001
冰点/℃		2.7
运动黏度/(mm²/s)		2.2%
闪点(闭杯)/℃		4.0

20 世纪 90 年代初期，后勤工程学院与北京分析仪器厂曾联合研制过“喷气燃料质量指标测定仪”，在近红外光谱方法快速分析喷气燃料性质研究方面做了一些工作。1997 年袁洪福，龙义成和陆婉珍研制出了 CCD 近红外光谱仪^[29]，随之也展开对喷气燃料性质快速测定的研究。近红外光谱可以测定的航煤性质包括：冰点、芳烃含量、馏程、密度和含氧化合物等。

冰点是用来表示喷气燃料的低温使用性能。西欧、美国及国际空运协会采用 ASTM D2386 和 ISO 3013，我国采用其等效标准方法 GB 2430 方法。该方法采用制冷剂如干冰

对油品降温处理，以温度计确定冰点，操作复杂和费时。分析一个样品至少需要 1h。笔者和岳俊奇分别使用短波近红外光谱（700~1100nm）和长波近红外（1100~2500nm）光谱对航煤冰点快速测定进行了研究。结果表明，近红外分析能够替代 GB 方法使用^[30]，具有快速和简便的特点。图 18-8 和图 18-9 分别为实际值与 GB 2430 方法测定结果的对比。GB 2430 方法规定，在不同冰点仪上测定重现误差不大于 2.7℃。因此，近红外方法测定结果符合 GB 2430 方法的要求。从预测样品集中任选 6 号样品，连续测定 10 次，其测定结果的标准偏差为 0.5℃，优于标准方法的精度（0.7℃）。一个航煤冰点的近红外光谱测量全过程仅需 3 分钟或更短。

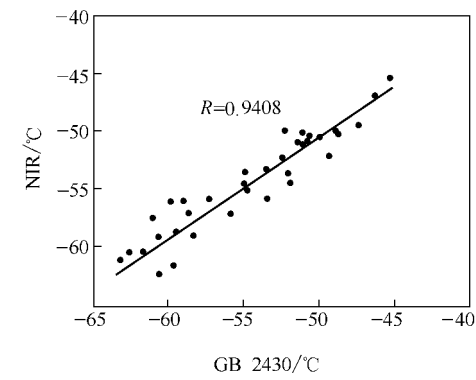


图 18-8 短波近红外光谱测定冰点

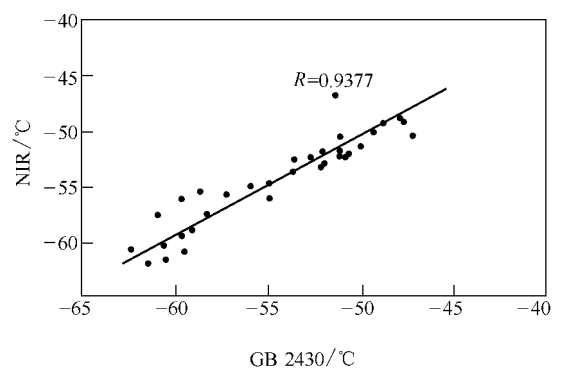


图 18-9 长波近红外光谱测定冰点

芳烃含量是影响航煤燃烧性能的重要质量指标之一。我国规定航煤中芳烃不超过 20%（v/v），RP-5 航煤的芳烃含量不超过 25%（v/v）。常用的测定方法为 GB 11132，即荧光色谱方法。不仅分析速度慢，且重复性差。岳俊奇使用近红外光谱和 MLR 和 PLS 方法测定航煤芳烃的统计结果列于表 18-21。图 18-10 为近红外法测定结果对 GB 11132 方法测定结果作图，预测平均偏差为 0.7%（v）。任选一个样品，连续 10 次测定，其平均偏差为 0.2%（v），优于 GB 11132 对重复性的要求。

表 18-21 近红外测定芳烃总含量结果				
算法	相关系数	平均偏差/%（v）	波长数	主因子数
MLR	0.9027	0.9	5	—
PLS1	0.9600	0.7	全谱	5

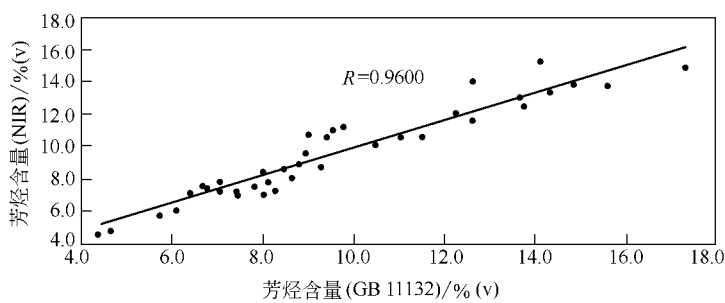


图 18-10 近红外方法与 GB 11132 方法结果对照

近红外光谱方法还可量测航煤的密度、馏程和闪点等性质^[78] 近红外方法测定的准确度和精度均符合标准分析方法的要求。

刘慧颖等人近期对近红外光谱测定喷气燃料性质做了大量工作^[68]。表 18-22 列出所建立的模型参数。工作表明, 除闪点性质外, 其他性质测量结果均与标准方法测量结果相一致。

表 18-22 **喷气燃料校正模型数据结果**

性质	样品数量	性质范围	R^2	SEC
馏程/℃				
初馏点	53	156~167	0.7836	3.86
10%	54	164~184	0.7433	2.58
20%	57	169~188	0.8159	2.30
50%	56	180~201	0.7173	2.52
90%	49	195~238	0.6963	3.30
终馏点	47	211~259	0.8201	4.65
密度 20℃/(g/cm ³)	52	0.7781~0.7974	0.9476	0.0014
黏度 20℃/(mm ² /s)	50	1.39~1.70	0.5680	0.024
闪点/℃	56	42~51	0.8692	1.21
冰点/℃	23	-48.5 ~ -60.0	0.9410	1.20

18.6 柴油分析

近红外光谱分析技术也非常适用于柴油组成和性质的快速分析。主要分析项目为柴油组成和性质包括十六烷值、馏程、密度、黏度、闪点以及各种低温性质如凝点、冷滤点等。其标准试验方法及其再现性要求见表 18-23。

表 18-23 柴油性质参数及标准试验方法的再现性要求

性 质	试验方法	再现性要求
十六烷值:	GB/T 386	2.5~3.3
馏程/℃	GB/T 6536	
50%		6~8
90%		9
密度(20℃)/(g/cm ³)	GB/T 1884 或 SH/T 0604	0.001
凝点/℃	GB/T 510	4
运动黏度/(mm ² /s)	GB/T 265	2.2%
闪点 ℃	GB/T 261	4

18.6.1 十六烷值分析

十六烷值是柴油抗爆性质指标。标准分析方法为 GB/T 386—1991，分析周期长，使用的爆震引擎试验机投资高（200 万人民币左右/台），开机和维护费用也较高。

目前国内外柴油需求正在呈上升趋势，促使各炼油厂多产柴油。用重质原料生产柴油是提高产量的重要途径之一，但由此带来柴油十六烷值变差的问题。一般，采用调和方法

使柴油十六烷值达标。调和过程需要频繁检测十六烷值，显然，GB/T 386—1991 在分析速度和费用上均不适合调和过程的控制分析。

与测定汽油辛烷值原理相似，近红外光谱方法也非常适合测定十六烷值。徐广通等人将 CCD 近红外光谱仪用于济南炼油厂柴油调和控制分析^[35]。表 18-24 列出近红外光谱方法与标准方法的结果对照。两种结果间的平均偏差为 0.78，小于 GB 11139 再现性误差要求（±2 单位）。

表 18-24 柴油十六烷值分析结果对比			
编 号	GB/T 386—1991	近 红 外	偏 差
1	50.5	49.9	—0.6
2	47.2	47.6	0.4
3	44.1	45.4	1.3
10	41	41.7	0.7
...		...	
32	44.2	46	1.8
33	45.8	47.2	1.4
			0.78

18.6.2 馏程分析

馏程是汽油和柴油等石油产品的重要质量指标，是工艺计算的重要基础数据，在生产中常用作控制操作条件的依据。目前，我国石油产品的馏程测定方法是 GB 6536，但该方法操作较复杂，且费时费力。在汽油分析中已经讨论过，近红外光谱方法测定馏程比测定其他性质困难。尽管如此，近红外光谱用于柴油馏程测量仍然对柴油生产非常有帮助。

李桂莲和徐新举等人将近红外光谱用于催化裂化柴油，和常三线柴油的馏程测量，包括初馏点，10%，50%，90%，95%^[79]。以 95%点为例，2002 年 4 月 5 日至 2002 年 4 月 9 日监测常三线柴油，连续监测 10 个点，结果如表 18-25 所示。可以看出，两种方法的最大差值为 5℃，在 GB 6536 规定的精密度允许范围内。

表 18-25 同一样品 NIR 结果与 GB 6536 结果比较										
取样时间	4.5 8 : 00	4.5 16 : 00	4.6 8 : 00	4.8 8 : 00	4.8 16 : 00	4.9 8 : 00	4.9 16 : 00	4.10 8 : 00	4.10 16 : 00	4.11 8 : 00
NIR 结果	347	347	358	344	345	350	336	327	340	352
BG 6536 结果	348	348	359	347	347	353	341	329	339	350
偏差	1	1	1	3	2	3	5	2	—1	—2

18.6.3 闪点分析

王艳斌在 NIR—2000 型近红外光谱仪上，研究了多元校正方法测定柴油闪点方法^[41]。其中人工神经网络校正方法的效果最佳，说明近红外光谱与柴油闪点的关系具有一定的非线性。样品为沧州炼油厂的一线、二线、三线直馏柴油，其样品性质范围见表 18-26。

表 18-26		样品性质范围	
性质	一线柴油	二线柴油	三线柴油
初馏点/℃	141-170	194-230	221-248
终馏点/℃	176-224	260-342	337-383
闪点/℃	41-62	70-97	95-111

表 18-27		不同校正方法对柴油闪点预测结果			
样品号	GB/T 261/℃	近红外/℃			
		ANN	LWR	PCR	PLS
1	41	44	38	32	31
2	42	44	45	39	38
3	47	44	48	47	46
4	49	51	51	61	60
5	55	53	55	62	61
6	62	62	61	67	66
7	70	73	69	73	72
8	76	77	75	78	77
9	106	100	104	98	98
10	106	103	109	107	107
11	109	104	109	109	110
相关系数		0.9956	0.9972	0.9729	0.9746

由表 18-27 可以看出，神经网络（ANN）、局部线性回归（LWR）预测准确性优于主成分回归（PCR）、偏最小二乘（PLS），这是由于一、二、三线柴油的馏程不同（见表 18-26），它们的组成基质也不同，这种较宽的性质变化范围使光谱与闪点之间的数学关系偏离了线性，因而对不同馏程柴油统一建模时，应采用适于关联非线性响应的校正方法。局部权重回归方法虽然准确性好于人工神经网络方法，但是它对未知样品光谱测量中的微小变化十分敏感，模型抗干扰性差。在样品性质范围较宽，化学量测值与光谱数据不呈线性时，人工神经网络方法更为准确和抗干扰，是一种比较好的校正方法。

18.6.4 芳烃分析

柴油中芳烃含量的高低，不仅直接影响其燃烧和抗爆性能，而且柴油燃烧产物对大气产生不同程度的污染。柴油中芳烃、尤其多环芳烃的存在对环境非常有害。为此，一些国家和地区组织相继制定了严格限制柴油中芳烃含量的新规定。

标准分析方法主要有荧光指示剂吸附法（ASTM D1319—1984）、质谱法（ASTM D2425—1988）和超临界色谱法（ASTM D5186—1996）。超临界色谱仪器昂贵且操作复杂，其应用实例较少。ASTM D1319—1984 是一种经典的柱色谱方法，它是利用硅胶对不同烃族组分的吸附能力的差异使油品中的不同烃类分离，并根据色谱的不同颜色进行测定。但该方法要求油品 95%的馏出温度不超过 315℃，而我国成品柴油的 95%馏出温度大约为 334～368℃ 之间，超出了荧光指示剂吸附法的使用范围，而且我国柴油颜色较深，不易判别馏分的颜色界线。因此该方法不适宜分析我国柴油的芳烃含量。质谱法是目前使用最多的标准方法，国内 SH/T 0606—1994 等效于 ASTM D2425—1988 方法。该方法能得出柴油详细的烃族组成。但需要昂贵的进口仪器和较高标准的实验室，而且还需要柱色

谱预分离柴油样品的饱和组分和芳烃组分，操作复杂、速度慢、溶剂使用量大，因此限制了质谱的广泛推广使用。显然，不适合柴油生产过程中的控制分析。

为此，杨玉蕊和李文新等人建立了近红外光谱快速测定柴油的芳烃方法，用于柴油生产过程中芳烃控制分析，分析项目包括：单环芳烃、双环芳烃、多环芳烃和总芳烃。该方法的参考方法为 SH/T 0606—1994。SH/T 0606—1994 方法再现性要求见表 18-28。

表 18-28 SH/T 0606 再现性误差要求

	单环芳烃/%	双环芳烃/%	三环芳烃/%	总芳烃/%
SH/T 0606 再现性	1.7	1.6	0.4	2.3

表 18-29 模型建立主要参数及指标

	单环芳烃/%	双环芳烃/%	三环芳烃/%	总芳烃/%
校正集样本数	53	54	54	51
浓度范围/%	7.5~23.5	4.4~13.9	0.1~2.8	12.0~37.8
预处理方法	二阶微分 (21 点)	二阶微分 (21 点)	二阶微分 (21 点)	二阶微分 (21 点)
波长范围/nm	850~960	850~960	850~960	850~960
主因子数	5	4	4	5
SEC	1.05	0.54	0.22	1.10
R ²	0.9228	0.9422	0.9039	0.9737

表 18-30 近红外光谱法与 SH/T 0606 对照

序号	单环芳烃含量/%			双环芳烃含量/%			三环芳烃含量/%			总芳烃含量/%		
	SH/T	NIR	残差	SH/T	NIR	残差	SH/T	NIR	残差	SH/T	NIR	残差
1	11.4	11.5	0.1	7.6	8.3	0.7	1.5	1.4	-0.1	20.5	20.8	0.3
2	16.7	17.4	0.7	10.7	9.8	-0.9	2.1	1.7	-0.4	29.5	29.9	0.4
3	8.0	9.2	1.2	4.8	4.8	0.0	0.3	0.3	0.0	13.1	12.8	-0.3
4	10.5	11.7	1.2	6.6	6.5	-0.1	0.6	0.6	0.0	17.6	19.5	1.9
...	...	...	...	...	...	...	...	...	...	...	...	...
22	18.0	18.9	0.9	9.9	10.0	0.1	2.1	1.8	-0.3	29.9	31.8	1.9
23	21.4	20.2	-1.2	10.1	10.0	-0.1	1.7	1.7	0.0	33.3	31.3	-2.0
24	17.5	17.5	0.0	9.7	9.5	-0.2	2.0	1.7	-0.3	29.3	29.9	0.6
SEP ^①				0.80			0.38			0.16		
$\bar{d}$ ^②				-0.2			0			0		
SDV ^③				0.80			0.41			0.26		
t				1.00			0.51			0.87		
$t_{0.05}^{③}$				2.07			2.07			2.07		

注：① 验证集标准偏差；② 两种分析方法测定值之间差值的平均值；③ 两种分析方法测定值之间差值（即残差）的标准偏差。

使用 PLS 方法建立模型参数与性能如表 18-29 所示。表 18-30 是两种方法结果对照，各族芳烃的 SEP 表明两种分析方法测定值偏差不大于 SH/T 0606 再现性误差要求；各族芳烃的 t 值表明两种分析方法结果之间无显著性差异。

目前这项工作仍在继续研究，以期使其成为标准方法。

18.6.5 便携式近红外光谱仪测定柴油性质

北京总后油料研究所研制了 NIR3000ZH 型便携式近红外光谱仪，目前该仪器可快

速、实时检测多种燃料的多项性质参数。并使用该仪器对柴油性质测定做了大量工作^[69]，其中包括十六烷值、馏程、凝点、闪点、黏度、密度等。通过在北京、南京和广州地区完成的使用试验情况，便携式油料质量分析仪能够较好的适应外场或野外作业环境条件，可以代替多台（套）传统石油产品测定仪器，在特殊环境条件下，实时、准确地进行油料的质量监控，满足军队油料质量的实时保障需要。

校正集样品取自于东北、华北、华东、华南等地区 16 个炼油厂和油库的—10 号车用柴油 52 个；其余使用油品：轻柴油 10 个，3 号喷气燃料 10 个。验证集样品为不同地区的—10 号车用柴油 7 个。表 18-31 和表 18-32 分别为模型性能和与标准方法结果的对照情况。

表 18-31 柴油的校正模型

性 质	样品数量	性质范围	主因子数	R^2	SEP
十六烷值：	50	48.4~73.5	3	0.9308	2.0
馏程/℃					
50%	52	249~272	8	0.8412	3.3
90%	52	270~316	8	0.8255	3.8
密度 20℃/(g/cm ³)	52	0.7913~0.8369	5	0.9561	0.002
黏度 20℃/(mm ² /s)	52	2.11~4.86	7	0.7416	0.13
闪点/℃	52	49~113	8	0.8642	3.7
凝点/℃	52	-11~-19	5	0.8404	2.1

表 18-32 标准方法结果与近红外结果的比较

样品 编号	十六烷值			密度/(g/cm ³)			馏程 50%/℃			馏程 90%/℃		
	GB/T 386	NIR	偏差	GB/T 1884	NIR	偏差	GB/T 6536	NIR	偏差	GB/T 6536	NIR	偏差
1	55.0	56.5	1.5	0.8166	0.8150	-0.0016	255	254	-1	285	284	-1
2	55.5	54.4	-1.1	0.8142	0.8124	-0.0018	253	252	-1	284	284	0
3	54.0	53.4	-0.6	0.8231	0.8224	-0.0007	263	262	-1	291	291	0
4	58.0	58.2	0.2	0.8117	0.8113	-0.0004	260	260	0	292	292	0
5	59.0	58.4	-0.6	0.8060	0.8075	0.0015	270	269	-1	303	303	0
6	63.0	63.1	0.1	0.8090	0.8091	0.0001	264	264	0	311	310	-1
7	57.0	56.8	-0.2	0.8170	0.8179	0.0009	275	277	2	316	314	-2
SEP	0.8			0.0013			1.2			1		
再现性 要求	2.5			0.0010			6			9		

样品 编号	凝点/℃			20℃黏度/(mm ² /s)			闪点/℃		
	GB/T 261	NIR	偏差	GB/T 510	NIR	偏差	GB/T 265	NIR	偏差
1	-18	-18	0	3.78	3.80	0.02	91	90	-1
2	-19	-18	1	3.65	3.67	0.02	92	93	1
3	-17	-18	-1	3.97	3.95	-0.02	79	79	0
4	-14	-15	-1	3.83	3.83	0	95	94	-1
5	-14	-12	2	4.30	4.32	0.02	78	79	1
6	-14	-16	-2	3.97	3.99	0.02	78	79	1
7	-16	-16	0	4.67	4.67	0	73	73	0
SEP	1.4			0.018			0.9		
再现性 要求	4			0.06			4		

18.7 重质油料分析

与轻质油品相比，近红外光谱分析重质油品性质的难度较大，因为样品颜色深，黏稠及固体，流动性和均匀性差，取样困难等，光谱测量不如轻质油品的方便；建立模型使用的参考数据（如 4 组分 SARA）的取得也不易。

尽管有关重油的实验室近红外分析详细报道较少^[43,57,80,81]，近红外光谱快速测量重油性质对于炼油工业具有重要意义。国内对近红外光谱测定原油、润滑油、渣油和沥青等性质课题也做了大量研究工作^[55,57]，这些工作展现了近红外光谱在重质油料性质快速分析方面具有很光明的应用前景。但如果将其用于工业生产，尚需要在样品收集和校正基础研究等方面做大量而艰苦的工作。

18.7.1 原油性质分析

市场和原料的变化使炼油厂常采用多种原油及混兑原油作为原料。另外，在原油蒸馏过程中，原油的切换也使原油在管道中自然混合。由于传统原油性质分析方法速度慢，不能及时监测原油性质，而原油的频繁更换及混合使得原油蒸馏装置的有效操作很困难。操作人员不得不在进行原油切换时留有一定的安全余地，使处理量降低，产品产率减少，操作费用增加。因此，生产的发展要求开发快速测定原油性质的方法。

近红外光谱可以用于测定原油的密度和馏程，对指导原油加工过程有着十分重要的意义。例如法国一家炼油厂采用近红外光谱在线监控原油的密度和实沸点，根据进料性质及时调整操作参数，可最大限度地发挥装置的加工能力，带来了可观的经济效益^[18]。新加坡 Hidajat 等人报道了采用近红外光谱和 PLS 方法测定原油实沸点，共建立了 14 个模型^[82]。原油来源十分复杂，各自的组成和性质差异很大。王艳斌针对国内部分炼油厂原料（如表 18-33 所示），使用近红外光谱和化学计量学方法对原油馏程快速测定方法进行了研究^[55]。参考数据测定方法为 ASTM D5307（气相色谱模拟蒸馏方法），其精密度要求见表 18-34。

表 18-33 原油样品表

样 品 名 称	原油基属	特性因数	样品来源
东河塘原油	含硫石蜡-中间基	12.1	兰州炼油化工总厂
解放渠东原油	含硫石蜡-中间基	12.2	兰州炼油化工总厂
吐哈原油	低硫-石蜡基	—	兰州炼油化工总厂
塔中混合原油	含硫-中间基	12.1	兰州炼油化工总厂
桑塔木原油	含硫石蜡-中间基	11.9	兰州炼油化工总厂
巴喀原油	低硫-中间基	12.3	兰州炼油化工总厂
轮南原油	含硫石蜡-中间基	12.2	兰州炼油化工总厂
采南原油	低硫-石蜡基	12.4	兰州炼油化工总厂
丘陵原油	低硫中间-石蜡基	13	兰州炼油化工总厂
北疆原油	低硫-中间基	—	兰州炼油化工总厂
白克原油	低硫-石蜡基	12.5	兰州炼油化工总厂
库勒勒原油	低硫-中间基	12.2	兰州炼油化工总厂
沙参 7 井原油	低硫-石蜡基	12.2	乌石化总厂
沙参 45 井原油	低硫-石蜡基	12.4	乌石化总厂
雅克 1 井原油	低硫-石蜡基	12.2	乌石化总厂
大庆原油	低硫-石蜡基	12.3	燕化

续表

样 品 名 称	原油基属	特性因数	样品来源
长岭管输油	—	—	长岭炼油厂
西江原油	—	—	石油化工科学研究院 18 室
伊朗也门马西拉原油	—	11.9	茂名石化公司
阿联酋穆尔班原油	—	12.2	茂名石化公司
伊朗轻油	含硫-中间基	11.8	石油化工科学研究院 1 室 101 组
伊朗重油	含硫-中间基	11.9	石油化工科学研究院 1 室 101 组
伊拉克原油	—	11.9	石油化工科学研究院 1 室 101 组
沙特原油	含硫-中间基	11.8	石油化工科学研究院 1 室 101 组
哈萨克斯坦原油	含硫-中间基	12.3	石油化工科学研究院 1 室 101 组

表 18-34

ASTM D5307-97 精密度

收率 /m%	馏出温度/℃		收率 /m%	馏出温度/℃	
	重复性	再现性		重复性	再现性
初馏点	3.7	10.6	60	11.8	25.9
5	4.6	14.8	70	17.6	39.2
10	6.9	11.3	80	24.8	38.8
20	6.8	15.4	85	18.8	38.8
30	7.6	20.4	90	20.7	44.9
40	9.3	24.6	残余/m%	2.6	8.1
50	10.6	30.3			

表 18-35

模型主要参数及校正结果

项 目		馏出温度/℃							
		初馏点	5%	10%	20%	30%	40%	50%	60%
校正集 统计数 据	样品数	57	57	57	57	57	57	57	57
	最小值	9	55	87	123	162	202	242	282
	最大值	180	220	251	304	343	399	456	522
验证集 统计数 据	样品数	14	14	14	14	14	14	14	14
	最小值	29	73	100	140	190	231	279	323
	最大值	172	206	231	276	316	353	402	454
PCR	主因子数	9	7	7	7	7	7	7	7
	SEC	21. 25	13. 84	12. 49	11. 31	9. 68	11. 62	15. 62	19. 51
	SEP	12. 52	6. 80	5. 89	6. 70	6. 55	9. 17	11. 62	14. 82
	R ²	0. 8410	0. 9149	0. 9275	0. 9431	0. 9560	0. 9269	0. 8742	0. 8289
PLS	主因子数	5	5	5	5	5	6	8	8
	SEC	20. 67	12. 56	10. 74	8. 60	8. 01	9. 01	10. 85	14. 19
	SEP	13. 33	7. 24	6. 09	6. 65	6. 14	9. 08	10. 51	13. 99
	R ²	0. 8397	0. 9294	0. 9454	0. 9678	0. 9698	0. 9566	0. 9407	0. 9097
PCA-A NN	结构	7-2-1	6-2-1	7-2-1	6-2-1	7-2-1	6-2-1	7-2-1	8-2-1
	学习率	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002
	SEC	6. 17	5. 22	4. 93	6. 71	5. 68	10. 51	14. 01	13. 89
	SEP	4. 73	4. 66	4. 63	5. 23	4. 89	7. 30	10. 18	13. 51
	R ²	0. 9881	0. 9887	0. 9893	0. 9806	0. 9853	0. 9410	0. 9011	0. 9200
PLS-A NN	结构	4-3-1	5-2-1	5-2-1	4-2-1	5-2-1	6-2-1	6-2-1	8-2-1
	学习率	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002	0. 002
	SEC	6. 55	5. 07	4. 81	5. 27	5. 03	5. 65	8. 88	9. 00
	SEP	4. 93	3. 77	3. 80	5. 62	4. 60	8. 82	10. 66	12. 71
	R ²	0. 9865	0. 9893	0. 9898	0. 9881	0. 9885	0. 9836	0. 9625	0. 9679
ASTM D5307 再现性		10. 6	14. 8	11. 3	15. 4	20. 4	24. 6	30. 3	25. 9

研究中研制了塑料袋和特制板夹测量部件，有效解决了原油近红外光谱测量问题。分别使用四种多元校正方法建立了初馏点，5%，10%，20%，30%，40%，50%和 60%的校正模型，其校正参数和结果如表 18-35 所示，人工神经网络方法（PLS-ANN，PCA-ANN）的校正结果好于偏最小二乘（PLS）及主成分回归（PCR）方法的结果。与 PCA-ANN 相比较，PLS-ANN 具有较好的准确性及较小的网络规模，其 SEC 及 SEP 值均在 ASTM D5307-97 再现性范围内。

表 18-36				验证集测定结果								
样品 序号	初馏点/℃			5%馏出温度/℃			10%馏出温度/℃			20%馏出温度/℃		
	GC	NIR	差值	GC	NIR	差值	GC	NIR	差值	GC	NIR	差值
1	30	39	−9	83	85	−2	113	113	0	147	153	−6
2	29	35	−6	93	90	3	125	121	4	175	170	5
3	168	162	6	202	197	5	230	224	6	275	272	3
4	158	163	−5	195	196	−1	222	223	−1	266	270	−4
5	170	166	4	206	210	−4	231	233	−2	271	280	−9
6	158	161	−3	192	197	−5	219	222	−3	261	270	−9
7	166	163	3	202	203	−1	224	228	−4	268	275	−7
8	160	155	5	198	191	7	223	219	4	270	263	7
9	163	162	1	200	198	2	225	225	0	263	264	−1
10	172	164	8	206	200	6	231	226	5	276	274	2
11	166	165	1	201	202	−1	228	227	1	272	274	−2
12	32	28	4	73	77	−4	100	105	−5	140	147	−7
13	36	34	2	81	81	0	110	109	1	153	153	0
14	159	159	0	200	198	2	229	224	5	267	263	4
ASTM D5307 再现性			10. 6				14. 8				11. 3	15. 4
SEP			4. 93				3. 77				3. 80	5. 62
$\overline{d}$			0. 79				0. 50				0. 79	−1. 71
S _d			4. 92				3. 80				3. 58	5. 40
t			0. 60				0. 49				0. 83	1. 18

样品 序号	30%馏出温度/℃			40%馏出温度/℃			50%馏出温度/℃			60%馏出温度/℃		
	GC	NIR	差值	GC	NIR	差值	GC	NIR	差值	GC	NIR	差值
1	190	187	3	249	269	−20	294	311	−17	336	331	5
2	223	226	−3	297	295	2	345	339	6	404	409	−5
3	316	311	5	353	348	5	402	394	8	454	443	11
4	305	306	−1	341	342	−1	385	384	1	434	440	−6
5	308	314	−6	341	349	−8	384	395	−11	430	444	−14
6	302	308	−6	336	345	−9	378	393	−15	427	451	−24
7	307	311	−4	340	351	−11	381	395	−14	429	445	−16
8	312	305	7	350	339	11	398	384	14	451	440	11
9	298	304	−6	325	329	−4	352	364	−12	390	390	0
10	315	311	4	350	348	2	399	393	6	452	437	15
11	311	310	1	345	348	−3	392	393	−1	444	442	2
12	196	190	6	231	242	−11	279	289	−10	323	338	−15
13	200	198	2	252	259	−7	301	308	−7	343	357	−14
14	305	301	4	332	331	1	366	364	2	409	398	11
ASTM D5307 再现性			20. 4				24. 6				30. 3	25. 9
SEP			4. 60				8. 82				10. 66	12. 71
$\overline{d}$			0. 43				−3. 78				−3. 57	−2. 78
S _d			4. 70				7. 95				9. 94	12. 46
t			0. 34				1. 78				1. 34	0. 83

采用 PLS-ANN 模型对原油馏程性质进行预测，与标准方法结果对照见表 18-36。成对 t 检验结果表明两种方法结果无显著性差别；两种结果的差值小于气相色谱模拟蒸馏的再现性要求，因此可认为在所研究样品范围内，近红外光谱方法测定原油馏程性质是准确的。

18.7.2 SARA 分析

渣油 SARA 是重油加工工艺所需要的重要数据。目前仍然依靠经典的洗脱色谱法 (EC) 分析，方法使用有毒溶剂、分析周期为几天，测量精度低。因此，使用近红外光谱测定渣油 SARA 对石油加工研究和生产均有重要意义。褚小立在紫外光谱测定渣油 SARA 方法研究的同时，也对近红外光谱测定渣油 SARA 方法进行初步探讨^[57]。收集和配制了具有代表性的 21 个常压渣油、27 个减压渣油和 27 个加氢渣油组成样品集。SARA 组成参考测定方法为 RIPP 10—90 方法 (EC 方法)，与 ASTM D—3279 标准方法 (EC) 等效。其方法精密度规定：单个试样 2 次平均测定的 SARA 结果之差不应大于 2%。

表 18-37 常压渣油饱和烃 NIR-PLS、UV-PLS 预测结果与 EC 实测值 单位：%

序号	渣油名称	EC 实测值	NIR-PLS		UV-PLS	
			预测值	偏差	预测值	偏差
1	伊朗重质>350℃	37.17	39.60	2.43	36.89	—0.28
2	冀东>350℃	60.46	54.38	—5.63	54.98	—10.64
3	西江>350℃	65.62	52.41	—3.21	51.64	—8.82
4	二连>350℃	51.52	49.40	—2.12	50.31	—1.21
5	Heglig>350℃	58.58	62.32	3.47	57.60	—1.25
6	伊朗混合>350℃	38.93	42.33	3.40	37.91	—1.02
7	沙特中质>350℃	35.38	36.88	1.50	36.49	1.11
8	轮南>400℃	56.05	49.89	—7.16	43.88	—12.17
9	曙光>350℃	40.84	38.63	—2.21	40.59	—0.25
10	塔里木混合常渣	56.34	57.32	0.98	44.52	—11.81
预测标准偏差			3.88		7.33	

表 18-38 常压渣油芳烃 NIR-PLS、UV-PLS 预测结果与 EC 实测值 单位：%

序号	渣油名称	EC 实测值	NIR-PLS		UV-PLS	
			预测值	偏差	预测值	偏差
1	伊朗重质>350℃	39.09	38.52	—0.57	39.16	0.07
2	渤海常渣	29.49	30.78	1.29	30.13	0.64
3	乌东>350℃	22.30	19.70	—2.60	22.65	0.35
4	胜利>350℃	22.71	23.49	0.78	24.11	1.40
5	吉林英台>400℃	25.40	21.55	—3.85	18.79	—6.61
6	孤岛常渣	29.95	31.37	1.42	28.85	1.10
7	吉林抚余>400℃	28.50	24.01	—4.49	19.78	—8.72
8	冀东>350℃	20.88	22.78	1.90	20.39	—0.49
9	哈萨克>350℃	25.12	27.81	2.69	24.05	—1.07
10	Heglig>350℃	18.43	19.16	0.73	17.61	—0.82
预测标准偏差			2.53		3.73	

表 18-39 常压渣油胶质 NIR-PLS、UV-PLS 预测结果与 EC 实测值			单位: %			
序号	渣油名称	EC 实测值	NIR-PLS		UV-PLS	
			预测值	偏差	预测值	偏差
1	伊拉克>350℃	13. 18	19. 63	6. 45	14. 00	0. 82
2	伊朗重质>350℃	18. 34	19. 04	0. 70	19. 74	1. 40
3	渤海常渣	31. 60	30. 14	—1. 46	31. 93	0. 33
4	乌东>350℃	25. 87	23. 90	—1. 97	21. 87	—4. 00
5	中原四混常渣	20. 93	18. 47	—2. 46	16. 82	1. 15
6	管输常渣	26. 61	25. 99	—0. 62	26. 86	0. 25
7	克拉玛依>400℃	24. 72	17. 51	—7. 21	24. 46	—0. 26
8	吉林抚余>400℃	16. 12	20. 39	4. 27	23. 62	7. 50
9	沙特中质>350℃	15. 56	17. 77	2. 21	14. 68	—0. 88
10	轮南>400℃	13. 46	28. 63	15. 17	27. 59	14. 13
预测标准偏差				6. 32	5. 54	

表 18-40 常压渣油沥青质 NIR-PLS、UV-PLS 预测结果与沥青质仪实测值			单位: %			
序号	渣油名称	沥青质仪 实测值	NIR-PLS		UV-PLS	
			预测值	偏差	预测值	偏差
1	秦皇岛>350℃	2. 20	3. 50	1. 30	2. 85	0. 65
2	沙特中质>350℃	3. 10	2. 53	—0. 57	2. 87	—0. 23
3	伊朗重质>350℃	5. 40	4. 19	—1. 21	4. 85	0. 55
4	伊朗常渣	3. 60	3. 77	0. 17	3. 85	0. 25
5	孤岛常渣	4. 00	2. 68	—1. 32	4. 77	0. 77
6	管输常渣	3. 70	2. 65	—1. 05	3. 40	—0. 30
7	二连>350℃	1. 90	0. 76	—1. 14	1. 99	0. 09
8	塔里木混合常渣	2. 00	4. 33	2. 33	2. 04	0. 04
9	冀东>350℃	2. 30	1. 93	—0. 37	2. 56	0. 26
10	吉林英台>350℃	3. 10	1. 70	—1. 40	3. 05	0. 05
预测标准偏差				1. 30	0. 42	

结合近红外光谱、紫外光谱和 PLS 方法，分别建立了近红外光谱测定常压渣油 SARA 校正模型，和紫外光谱测定常压渣油 SARA 校正模型。表 18-37～表 18-40 分别列出了近红外光谱方法、紫外光谱方法与 EC 方法的比较情况。研究发现：对于渣油的饱和烃和芳烃，近红外光谱方法结果优于紫外光谱方法结果；对于渣油的胶质和沥青质，紫外光谱方法结果优于近红外光谱方法结果。根据渣油的光谱信息可解释其现象，因为饱和烃和芳烃的近红外光谱信息远比其紫外光谱信息强，而胶质和沥青质含量比较低，在近红外光谱上的信息较弱。因此，上述预测结果的准确度变化规律与光谱信息变化规律是一致的。对于大多数样品，近红外光谱方法分析结果与 EC 分析结果的偏差小于 2%，不大于 EC 方法的精度要求。对于少量样品偏差大于 2% 的原因应归咎于所收集的样品集范围限制。对于组成差别很大的渣油（比如来自不同工艺等）的校正需要分别建立校正模型，褚小立就渣油分类进行了模糊聚类研究，不仅用于聚类建立模型，而且用于预测中对未知样品的识别。近红外光谱测定渣油 SARA 的研究工作尚属初步探讨，希望随着今后工作的继续开展，使该方法未来能够直接服务于生产。

挪威 Narve Aske^[80]使用红外光谱，近红外光谱和高效液相色谱（HPLC）研究了 16 个原油。通过液相色谱将样品分离为饱和烃、芳烃、胶质和沥青质。探讨了分别用红外光谱，近红外光谱结合 PCR 和 PLS 预测原油的 SARA 组成。结果表明，对所研究样品，红

外光谱预测饱和烃、芳烃、胶质和沥青质的精度分别为 2.5wt%，2.2wt%，1.4wt%和 1.3 wt %；近红外光谱预测饱和烃、芳烃、胶质和沥青质的精度分别为 2.8wt%，2.4wt%，1.4wt%和 1.0 wt %；光谱分析精度与 HPLC 的相当；并认为近红外光谱具有更适合快速分析的特点。这项工作仅限于方法研究探索，因为 HPLC 方法分析在定量上有很多不确定性，也不属于标准方法。如果面向工业，则必须采用标准方法（即 EC）测定 SARA 数据作为校正的参考数据。

18.7.3 沥青蜡含量分析^[55]

随着高速公路的发展对高等级道路沥青的需求量越来越大，目前，我国主要靠进口，因此研制和生产这种沥青已成为炼油工业的重要任务之一。其中，沥青蜡含量是关键性质之一。目前采用 SH/T 0425 方法，操作复杂，周期长（几天），直接影响沥青的研究与生产。该方法分析精密度列于表 18-41。为解决这项分析难题，王艳斌研究了近红外光谱快速测定沥青蜡含量的分析方法。由于高等级沥青种类有限和收集困难，所收集的沥青样本不多（共 30 多个样品），但研究结果表明，近红外分析方法与标准分析方法（SH/T 0425）的测定结果一致。近红外分析方法具有重现性好、分析速度快、使用样品量少、测试过程简单等特点，是一种较好的沥青蜡含量快速测量方法。表 18-42 为近红外光谱方法与 SH/T 0425 方法的对照。

表 18-41		SH/T 0425 标准方法精密度			
蜡含量 <i>t</i> /%	重复性	再现性	蜡含量 <i>t</i> /%	重复性	再现性
0.0~1.0	0.1	0.3	>3.0	0.5	1.5
>1.0~3.0	0.3	1.0			

表 18-42		验证集测定结果比较	
样品编号	蜡 含 量 $t/\%$		偏差
	SH/T 0425	近红外	
1 #	3.2	3.5	—0.3
2 #	2.2	2.1	0.1
3 #	1.6	2.0	—0.4
4 #	2.0	2.1	—0.1
5 #	1.9	2.2	—0.3
6 #	3.8	3.5	0.3
7 #	1.7	1.8	—0.1
8 #	2.4	2.0	0.4
9 #	2.4	2.1	0.3
10 #	2.8	3.2	—0.4
11 #	2.0	1.9	0.1
验证标准偏差		0.297	
$t(\alpha=0.05)$		$ t $ (0.41)< $t_{(0.05,10)}$ (2.23)	

18.7.4 润滑油基础油性质分析^[55]

根据润滑油基础油分类的新标准，即 API 分类^[83]，硫含量，饱和烃含量和黏度指数是分类的基本依据。其中，饱和烃含量、黏度和黏度指数的标准分析方法分别为 EC 方

法，GB/T 265 和 GB/T 1995，缺点是操作复杂和分析周期长。因此，建立其快速分析方法对润滑油基础油的生产过程控制及润滑油的研制、调配具有重要的指导意义。王艳斌对近红外分析方法测定润滑油基础油的化学族组成、黏度和黏度指数进行了研究，并对光谱与性质之间的关系进行了解释。

试验样品基本状况列于表 18-43。样品的饱和烃质量分数范围 67.3%~99.6%，芳烃质量分数范围 0.29%~30.7%，胶质质量分数范围 0.1%~2.1%，黏度指数范围 64~134。从样品的来源、工艺、牌号及各性质分布来看，试验样品覆盖范围较广，具有一定的代表性。

表 18-43 试验样品情况	
项目	样 品 情 况
原油	伊朗、新疆、大庆、南阳、独子山及其他不明原油来源油样
工艺	白土精制、加氢精制、加氢处理、加氢异构化脱蜡、基础油原料油
牌号	200N,150ZN,300ZN,750ZN,100SN,150SN,500SN,650SN,150BS,混兑油样及部分未知牌号油样

不同工艺和原料的润滑油基础油具有明显不同的近红外光谱特征，可根据主成分分析将样品进行分类（如图 18-11 所示）。加氢处理油、ZN 基础油、SN 基础油和加氢原料油四种不同来源的原料可以清楚地区分。

18.7.4.1 黏度分析

分别采用 PLS 和人工神经网络进行校正，研究结果表明，在比较窄的样品范围内，采用 PLS 方法建立基础油的近红外光谱-黏度模型是可行的，在较宽的范围采用人工神经网络方法建立校正模型可以得到较好的预测效果。表 18-44 为采用 ANN 模型对验证集 17 个样品的预测结果及其成对 *t* 检验结果。ANN 模型所得的 *t* 值小于显著性水平为 0.05 时的 *t* 分布值，说明 ANN 模型的预测值与实际结果之间没有显著性差异，同时其预测值对实际值的相对残差都小于标准的再现性要求（2.2%），说明采用 ANN 模型得到的预测结果是准确的。

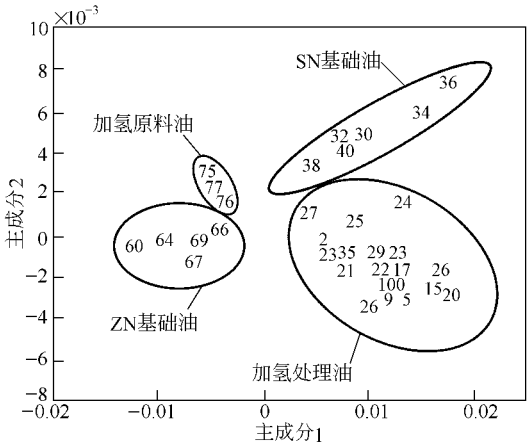


图 18-11 润滑油基础油的主成分分类图

表 18-44 验证集预测结果				
样品序号	GB/T 265 /mm ² · s ⁻¹	近红外 /mm ² · s ⁻¹	差值 /mm ² · s ⁻¹	相对残差/%
1	30.30	30.83	-0.53	-1.74
2	46.50	46.19	0.31	0.66
3	47.00	46.20	0.80	1.70
...	...	...	...	...
15	144.70	146.23	-1.53	-1.06
16	142.50	143.24	-0.74	-0.52
17	148.90	150.82	-1.92	-1.29
RSEP			1.59	
$\bar{d}$			0.37	
<i>S_d</i>			1.30	
<i>t</i>			1.17	
<i>t</i> (0.05,16)			2.12	

18.7.4.2 黏度指数

黏度指数是表征润滑油黏温性能的重要参数，及时了解润滑油基础油的黏度指数对润滑油的生产质量监控、润滑油的调配、研制以及高等级润滑油基础油生产方法的开发具有重要的实际意义。目前黏度指数是通过测定 40℃ 和 100℃ 黏度计算得来，方法比较复杂。

GB/T1995 精密度取决于用来运算的 40℃ 黏度和 100℃ 黏度。表 18-45 为两个样品黏度测定中出现 1％（重复性）和 2.2％（再现性）误差时计算所得的黏度指数差值，对黏度指数影响较大的是 100℃ 黏度的测定。由于黏度误差的影响，黏度指数的误差为 5～7。

表 18-45 黏度测定误差对黏度指数的影响

	测定值	计算值	误差	计算值	误差	计算值	误差
40℃ 黏度/mm ² ·s ⁻¹	142.6	141.2	－1％	142.6	0	139.5	－2.2％
100℃ 黏度/mm ² ·s ⁻¹	12.14	12.26	1％	12.41	2.2％	12.41	0
黏度指数	66	70	4	71	5	69	3
40℃ 黏度/mm ² ·s ⁻¹	38.44	38.05	－1％	38.44	0	37.59	－2.2％
100℃ 黏度/mm ² ·s ⁻¹	6.211	6.273	1％	6.348	2.2％	6.211	0
黏度指数	108	113	5	115	7	113	5

采用 PLS 建立校正模型，校正集样品包括加氢处理基础油、加氢处理原料油、石蜡基精制基础油和中间基精制基础油，覆盖范围较宽。表 18-46 为模型参数及结果，表 18-47 列出与标准方法的对照结果。说明近红外光谱测定润滑油基础油黏度指数方法与 GB/T 1995 方法一致。

表 18-46 润滑油基础油黏度指数模型校正参数及结果

	长波近红外	短波近红外
校正样品数	30	
黏度指数范围	65～132	
预处理方式	二阶微分	二阶微分+平滑
主因子数	4	3
SEC	1.57	1.47
R ²	0.9948	0.9956

表 18-47 验证集测定结果

序号	GB/T 1995	短波近红外		长波近红外	
		预测值	差值	预测值	差值
1	100	102	－2	104	－4
2	108	107	1	108	0
3	104	102	2	103	1
4	122	119	3	120	2
5	112	109	3	110	2
6	104	102	2	102	2
7	105	105	0	104	1
8	107	106	1	106	1
9	111	111	0	110	1
10	125	127	－2	126	－1
11	134	139	－5	137	－3
12	128	129	－1	127	1
13	64	68	－4	67	－3
14	70	70	0	70	0
15	67	68	－1	67	0
16	68	69	－1	67	1
17	66	68	－2	68	－2

续表					
序号	GB/T 1995	短波近红外		长波近红外	
		预测值	差值	预测值	差值
18	67	66	1	67	0
19	68	68	0	66	2
20	70	68	2	68	2
21	73	74	−1	72	1
22	76	75	1	74	2
23	66	69	−3	69	−3
SEP			2.13	1.89	
$\bar{d}$			−0.26	0.13	
S_d			2.11	1.89	
$ t $			0.59	0.33	
$t_{(0.05,22)}$				2.07	

18.8 化工过程中的应用

今天，近红外光谱方法已经广泛应用于化工领域各个分支，主要用来快速测定原料、中间产物和产品的组成和性质。石油化工科学研究院将近红外光谱分析技术应用于聚丙烯树脂性质分析、己内酰胺工艺分析、聚异丁烯高活性端基含量及分子量分析和强酸以及混胺的分析等。

18.8.1 聚丙烯树脂性质分析^[84,85]

聚丙烯树脂为 5 大通用塑料产品之一，在国民经济和生活中占有重要地位。聚丙烯树脂在生产过程中需要测量的性质很多，表 18-48 列出一些主要性质及其试验方法，其缺点如下：

表 18-48 聚丙烯树脂主要分析项目及试验方法	
分析项目	目前的分析方法
相对分子质量及其分布	GPC、拉曼光谱、核磁共振谱
熔体流动速率	GB/T 3682
清洁度	分检法
等规指数	GB/T 2412(二甲苯或正己烷作溶剂,抽提 12h)
颗粒度分布	标准筛筛分
堆积密度	重量分析
粉末灰分	GB/T 9345(重量分析法)
聚合物中游离单体及水含量	重量分析法,卡尔费休法,GC
鱼眼	GB/T 6595
力学性能 拉伸屈服强度	GB/T 1040
化学组成 共聚单体含量及共混比例	拉曼光谱、核磁共振谱、红外光谱
添加剂含量	紫外光谱、核磁共振谱

① 标准分析方法速度慢，难以满足生产控制及产品出厂的要求。如拉伸屈服强度分析周期约为 30h；熔体流动速率、等规指数测定方法也存在类似的问题。

② 分析设备投资太大，如测定聚丙烯树脂拉伸屈服强度的设备投资高达 500 万元；对七万吨/年聚丙烯装置，分析设备投资就高达 1300 余万元。核磁共振、拉曼光谱及分子量测定用的 GPC 等仪器投资均很大。许多生产厂家无力购置以上设备，缺少必要的分析手段。

③ 分析方法准确度低，如 GPC 分析聚丙烯树脂的相对分子质量，方法的再现性仅为 25％。

④ GPC，等规度及添加剂含量等性质分析方法使用了有毒溶剂，不利于分析人员的健康。

⑤ 核磁共振、拉曼光谱、注塑、压膜及 GPC 等分析方法的技术含量高，不仅对操作人员的技术素质要求较高，而且对实验室及配套设备要求高、投资大。

因此，研究开发快速测定聚丙烯树脂性质的方法对工业生产具有重要意义。吴艳萍采用漫反射近红外光谱快速测定聚丙烯树脂性质方法进行了详细研究^[56]，研究测定的主要性质包括熔融指数（MFR）、等规质数、拉伸屈服强度、平均分子量等。测量的聚丙烯树脂形态包括粒料、粉料和板材。

与油品测量方式（透射）不同，聚丙烯树脂在生产过程的主要形态为粒料和粉料，因此，采用漫反射测量方式。由于颗粒表面的镜反射特别强（尤其粒料）以及颗粒的不均匀性，由此引入近红外漫反射光谱中更多与性质无关的噪声。与油品光谱相比，聚丙烯树脂近红外光谱与被测量性质之间常为非线性关系，因此，聚丙烯树脂的近红外光谱校正更复杂，研究结果表明以聚丙烯树脂粒料熔融指数的校正难度最大，其主要原因是除颗粒形态引起较大光谱干扰影响外，标准方法测定熔融指数的误差随熔融指数变化比较大。在详细分析影响光谱测量的各种因素的基础上，研究了 PLS，局部权重回归，和人工神经网络校正规律，提出了根据光谱模式识别与分段人工神经网络校正的方法，对粒料聚丙烯树脂熔融指数校正，取得了满意结果。表 18-49 和表 18-50 分别为 MFR<20g/10min 粒料聚丙烯树脂和 MFR>20g/10min 粒料聚丙烯树脂的人工神经网络校正模型参数与结果。表 18-51 为漫反射近红外光谱测定 MFR>20g/10min 聚丙烯粒料熔融指数结果与标准方法结果对照。结果表明，漫反射近红外光谱预测结果与 GB 方法测定结果是一致的，且两种方法测定结果间相对误差在参照方法再现性范围内。

表 18-49 MFR<20g/10min ANNPLS 模型参数及结果

预处理方法	ANN 结构	区间/cm ⁻¹	校正集		验证集 1	
			SEC	R ²	SEP1	R ² 1
二阶微分(51) 样品数	6-4-1	6215~4286	0.42	0.9946	1.58	0.9148
			40		44	

表 18-50 MFR>20g/10min ANNPLS 模型参数及结果

预处理方法	ANN 结构	区间/cm ⁻¹	校正集		验证集 1		验证集 2	
			SEC	R ²	SEP1	R ² 1	SEP2	R ² 2
二阶微分(51) 样品数	6-4-1	6215~4286	1.34	0.9193	2.37	0.7969	2.87	0.8453
			53		38		56	

序号	GB/ (g/10min)	NIR/ (g/10min)	残差/ (g/10min)	相对误差/%
1	27.36	26.10	-1.26	-4.61
2	29.40	24.15	-5.25	-17.86
3	34.07	35.25	1.18	3.47
4	22.34	22.82	0.48	2.15
...	...	...	...	...
35	28.25	25.29	-2.96	-10.48
36	26.39	23.46	-2.93	-11.10
37	36.87	32.02	-4.85	-13.16
38	25.71	25.53	-0.18	-0.70
<i>t</i>			2.12	
<i>t</i> _(0.005,37)			2.42	
标准偏差			2.23	

与粒料相比，粉料和板材的熔融指数的近红外光谱测量更容易，其准确性也越高；与熔融指数测量相比，聚丙烯树脂的等规指数、拉伸屈服强度和相对分子质量的测定更容易些，读者可参考文献 [56，83，84]。

18.8.2 己内酰胺工艺分析

鹰山石油化工总厂在己内酰胺生产控制中，使用 NIR-2000 近红外光谱仪测定不同体系中己内酰胺含量和羰基等指标。分析样品取自羟胺肟化装置中的肟化塔、重排反应槽、萃取塔、离子交换塔和三效蒸发塔等。

样品体系：①无机工艺液体体系含 NH₃OH⁺、H₂PO₄⁻、NH₄⁺、环己酮肟、甲苯等，总羰基含量为 1.0%~2.0%。②己内酰胺水溶液体系含水、己内酰胺、硫酸、SO₃、环己酮肟等，己内酰胺含量为 0.1%~0.5%。③苯萃取溶液体系含苯、己内酰胺、水、硫酸等，其中己内酰胺含量为 0.1%~0.5%和 3.5%~4.5%。

18.8.2.1 无机工艺液中总羰基分析

磷酸羟胺与环己酮在肟化塔反应生成环己酮肟后从塔底出来的无机工艺液中含磷酸羟胺、甲苯、环己酮、环己酮肟等。总羰基含量是指环己酮和环己酮肟的总量，样品代号为 AT，控制指标为 1.1%左右。

采用 PLS 建立模型，AT 训练集含有 75 个样品，总羰基变化范围为 1.0%~1.64%。交互验证结果表明，预测值与参考数据间的相关系数 (*R*²) 为 0.8984，两者之间的平均残差为 0.02%。使用该模型对 30 个样品进行预测，标准偏差为 0.11% 小于过程分析要求 (0.3%)。成对 *t* 检验结果为 1.47，小于 |*t*|_{30,95%} = 2.04，说明近红外分析方法与现行标准方法之间无显著差异。

18.8.2.2 水体系中己内酰胺含量分析

环己酮肟经过重排、中和、离子交换、加氢、蒸发等过程生成己内酰胺。这些过程中其己内酰胺含量不同，成分有些差异（见表 18-52）。

表 18-52

体系来源、组成与指标要求

代号	样品来源	主要成分	指标要求
BJ	重排	己内酰胺,水,硫胺,硫酸,SO ₃ 等	70
B2J	中和	己内酰胺,水,硫胺,硫酸,SO ₃ 等	68
CJ	离子交换	水,己内酰胺,少量不饱和和杂质,OH-	30~31
C2J	加氢	水,己内酰胺,少量不饱和和杂质,OH-	30~31
DJ	蒸发	水,己内酰胺	88~90

BJ 的 PLS 模型中,训练集含有 99 个样品,指标变化范围为 68.8%~72.8%。交互验证结果表明,预测值与参考数据间的相关系数 (R^2) 为 0.9799,两者之间的平均残差为 0.16%。使用该模型对 27 个样品进行预测,结果列于表 18-53,标准偏差为 0.88% 小于过程分析要求的 2.5%。成对 t 检验结果为 0.53,小于 $|t|_{27,95\%} = 2.06$ 。CJ 和 DJPLS 模型训练集含样品共 125 个,其结果也列于表 18-53,标准偏差分别为 0.45%和 0.64%,均小于过程分析要求的 1.5%和 2.5%。 t 分别为 2.00 和 0.29,均小于 $|t|_{27,95\%} = 2.06$ 。说明近红外分析方法与现行标准方法之间无显著差异。

表 18-53 BJ, CJ 和 DJ 的预测结果与标准方法结果的对照

BJ			CJ			DJ		
近红外	标准	残差	近红外	标准	残差	近红外	标准	残差
69.99	70.23	-0.24	30.30	29.73	0.57	86.72	87.63	-0.91
70.37	70.09	0.28	30.13	29.53	0.60	86.51	87.35	-0.84
71.26	70.96	0.30	29.37	28.42	0.95	87.65	87.88	-0.23
72.53	70.40	2.13	30.25	30.26	-0.01	86.55	86.84	-0.29
71.65	71.27	0.38	29.31	29.82	-0.51	87.60	88.01	-0.41
...	...	...	...	...	...	...	...	...
70.60	70.72	-0.12	28.88	28.75	0.13	88.24	86.98	1.26
69.31	69.49	-0.18	29.68	29.43	0.25	87.93	88.51	-0.58
69.32	69.82	-0.50	29.53	28.49	1.04	87.49	87.79	-0.30
69.38	69.50	-0.12	30.03	29.75	0.28	88.99	89.37	-0.38
69.47	69.44	0.03	31.47	31.47	0.00	89.74	90.46	-0.72
69.60	67.90	1.70	29.20	29.26	-0.06	87.60	88.34	-0.74
69.64	69.96	-0.32	29.93	30.25	-0.32	88.52	88.83	-0.31
69.71	71.11	-1.40	30.61	30.54	0.07	88.73	87.95	0.78
70.38	71.03	-0.65	29.83	30.38	-0.55	88.26	88.86	-0.60
69.55	71.08	-1.53	30.98	30.99	-0.01	88.65	88.57	0.08
69.25	69.75	-0.50	31.07	30.44	0.63	87.40	87.18	0.22
—	—	—	31.89	32.01	-0.12	—	—	—
平均残差		0.66	平均残差		0.37	平均残差		0.55
标准偏差		0.88	标准偏差		0.45	标准偏差		0.64
t		0.53	t		2.00	t		0.29

18.8.2.3 苯体系中己内酰胺含量分析

苯体系中的主要成分为苯和己内酰胺, EJ 为己内酰胺萃取塔顶, 硫胺溶液被苯萃取后的产物, 还含有水和硫胺, 控制指标为 3.5%~4.5%。FJ 为萃取塔顶水萃取苯中己内

酰胺的产物，控制指标为 0.1%。分别建立了 EJ 和 FJ 的 PLS 模型，表 18-54 列出了预测结果和标准方法结果的对照情况，方法准确度符合标准方法的误差要求，*t* 检验结果表明近红外方法与标准方法无显著性差别。

表 18-54 EJ 和 FJ 预测结果与标准方法结果的对照

EJ			FJ		
近红外光谱法	标准方法	残差	近红外光谱法	标准方法	残差
4.30	4.72	−0.42	0.17	0.14	0.03
4.04	4.62	−0.58	0.19	0.13	0.06
3.90	4.39	−0.49	0.20	0.15	0.05
4.18	5.03	−0.85	0.22	0.21	0.01
4.13	3.71	0.42	0.16	0.16	0.00
3.78	3.80	−0.02	0.17	0.15	0.02
3.86	4.17	−0.31	0.18	0.21	−0.03
3.66	3.58	0.08	0.19	0.19	0.00
4.21	4.21	0.00	0.20	0.14	0.06
4.10	4.05	0.05	0.17	0.17	0.00
4.14	4.16	−0.02	0.13	0.15	−0.02
4.61	4.62	−0.01	0.14	0.10	0.04
4.39	4.40	−0.01	0.12	0.08	0.04
4.13	4.08	0.05	0.15	0.19	−0.04
4.32	3.92	0.40	0.17	0.11	0.06
4.27	4.31	−0.04	0.21	0.07	0.14
4.68	4.70	−0.02	0.18	0.15	0.03
4.19	4.18	0.01	0.20	0.17	0.03
4.53	4.52	0.01	0.12	0.09	0.03
4.31	4.26	0.05	0.12	0.08	0.04
5.29	4.93	0.36	0.17	0.20	−0.03
3.98	3.44	0.54	0.20	0.35	−0.15
3.58	4.10	−0.52	0.12	0.12	0.00
—	—	—	0.12	0.11	0.01
—	—	—	0.16	0.13	0.03
—	—	—	0.16	0.16	0.00
—	—	—	0.18	0.18	0.00
—	—	—	0.18	0.18	0.00
—	—	—	0.12	0.18	−0.06
—	—	—	0.09	0.08	0.01
—	—	—	0.11	0.09	0.02
平均残差		0.23	平均残差		0.03
标准偏差		0.34	标准偏差		0.05
<i>t</i>		0.81	<i>t</i>		1.42

18.8.3 聚异丁烯高活性端基含量及相对分子质量测定^[44]

带有活性端基的聚异丁烯聚合物是理想的热加成反应原料，通过衍生可以得到各种重要化工产品（比如用来生产聚异丁烯基丁二酰亚胺无灰润滑油分散剂），具有广泛用

途和很高的经济价值。在合成过程中得到产物是包括主要三种（ α 型、 β 型和 T 型）异构体的混合物，其中 α 型烯烃端基的活性最高， T 型的基本无活性。活性端基的含量与相对分子质量是其最重要的质量参数，作为生产控制和产品检验的主要技术指标。聚异丁烯分子相对分子质量在 1000 左右，端基外分子结构很类似，而活性端基在分子中占的部分很小，目前依靠核磁方法区分端基结构的定性与定量。而其相对分子质量测定方法为蒸汽压渗透法。这些方法特点是仪器投资高，操作复杂，且分析周期很长，不适合做控制分析。

为解决上述问题，作者提出了使用近红外光谱分析聚异丁烯活性端基含量与相对分子质量的快速测定方法，建立了测定聚异丁烯聚合物相对分子质量和 α 、 β 烯烃活性端基含量的校正模型，并考察了方法的精密度，并与 VPO 和 NMR 方法进行了对比。获得满意研究结果，如用于工业仍需进一步收集样品，扩展模型范围。

18.8.3.1 聚异丁烯 α 活性端基含量测定

使用 PLS 方法将聚异丁烯的近红外光谱和核磁测定活性端基含量进行关联，建立校正模型。表 18-55 为交互验证的结果。成对 t 检验方法 $|t| < |t_{17, \alpha=95\%}| = 2.11$ ，说明在 95% 的置信度下，两种方法测定结果无显著性差异。

表 18-55 核磁和近红外测定聚异丁烯活性端基含量结果比较

样品编号	核磁	近红外	样品编号	核磁	近红外
10	79.52	83.84	31	86.93	84.66
12	84.37	84.12	33	80.28	80.4
15	86.75	85.13	34	85.18	88.12
17	91.45	87.08	40	86.65	89.1
24	87.1	86.17	48	90.78	84.55
25	87.61	87.5	50	82.28	82.76
28	86.42	86.37	51	80.24	84.85
29	87.23	87.63	54	87.37	86.59
30	82.95	85.57	$t = -0.26$		

18.8.3.2 聚异丁烯 β 烯键含量测定

样品 β 烯键含量都在 8%~17% 之间。使用上述同样方法得到聚异丁烯 β 烯键含量结果如表 18-56 所示，成对检验结果为： $|t| = 0.21 < |t_{20, 95\%}| = 2.09$ ，说明核磁方法和近红外方法测定结果之间无显著性差异。

表 18-56 核磁和近红外测定聚异丁烯 β 烯键含量结果比较

样品编号	核磁	近红外	样品编号	核磁	近红外
2	17.70	15.85	29	12.77	10.98
10	11.70	14.50	30	14.03	12.73
12	15.63	15.50	31	10.82	12.06
15	12.99	12.59	33	15.68	17.16
17	11.81	11.91	34	14.82	15.79
24	8.55	10.75	48	8.36	10.02
25	12.90	11.64	50	12.37	12.61
26	12.39	10.99	51	14.45	14.34
28	12.03	8.91	$t = 0.21$		

18.8.3.3 相对分子质量测定

相对分子质量是聚异丁烯聚合物的重要性质之一。VPO 方法使用有毒溶剂，操作费时。对溶剂纯度及测量条件有着严格的要求。通常这种方法的精度在 10％左右。使用 PLS 方法将近红外微分光谱与相对分子质量关联建立校正模型，交互验证对照结果如表 18-57 所示，成对 *t* 检验结果为： $|t|=0.29|<|t_{20,95\%}|=2.09$ ，说明这两种测定结果无显著性差异。

表 18-57 VPO 和近红外光谱测定聚异丁烯相对分子质量对照

样品编号	VPO	近红外	样品编号	VPO	近红外
2	836	959	30	879	898
5	547	416	31	767	685
10	1110	1082	33	975	1106
12	714	684	34	1405	1314
15	757	778	35	1013	1023
16	1322	1365	40	697	705
17	748	782	48	949	864
25	1118	1091	51	933	979
26	1295	1383	53	1513	1379
28	1122	960	54	754	774
29	994	1108	<i>t</i> =0.29		

18.8.4 红烟硝酸分析

红烟硝酸是一种重要氧化剂，其标准试验方法特点具有操作复杂、费时（3 天，2 人/样品全分析）和有毒性的缺点。其中某些主要分析项目与方法列于表 18-58。邢志娜^[87]等建立近红外光谱测定了红烟硝酸密度、氧化氮、水分、磷酸和氢氟酸含量的 PLS 校正模型。表 18-59 列出了与标准方法测定结果的对照情况，其近红外光谱方法的预测结果偏差与标准方法的再现性误差相当。说明近红外光谱方法预测红烟硝酸组成是可行的，所得结果基本满足实际测量时对误差的要求；与标准方法相比，近红外方法分析速度快，将原先一周的化验时间缩短为 10min，将两个人以上的工作变为 1 个人的工作；重现性好；分析成本低；样品用量少；便于实现现场在线分析。

表 18-58 分析项目与方法

项 目	分 析 方 法	项 目	分 析 方 法
密度(15℃)/(g/cm³)	密度计法	磷酸含量/(% (m/m))	酸碱滴定法
四氧化二氮含量/(% (m/m))	氧化还原滴定法	氢氟酸含量/(% (m/m))	电极法
水分含量/(% (m/m))	浓差原电池电位法		

表 18-59 验证集的 NIR 方法与标准方法测定值一览表

编号	密度/(g/cm³)			水分/(% (m/m))			N ₂ O ₄ /(% (m/m))			磷酸/(% (m/m))			氢氟酸/(% (m/m))		
	标准方法	CCD-NIR	偏差	标准方法	CCD-NIR	偏差	标准方法	CCD-NIR	偏差	标准方法	CCD-NIR	偏差	标准方法	CCD-NIR	偏差
Y003	1.597	1.596	—0.001	1.69	1.72	0.03	20.41	20.55	0.14	1.12	1.13	0.01	0.504	0.504	0
Y005	1.595	1.595	0	2.25	1.81	—0.44	20.24	20.33	0.09	1.15	1.14	—0.01	0.526	0.503	—0.023
Y011	1.594	1.591	—0.003	2.08	2.32	0.24	21.07	20.54	—0.53	1.08	1.15	0.07	0.530	0.526	—0.004
Y017	1.593	1.595	0.002	1.81	1.62	—0.19	20.57	21.19	0.62	1.16	1.09	—0.07	0.520	0.493	—0.027
Y025	1.602	1.601	—0.001	1.31	1.36	0.05	21.50	21.50	0.00	1.14	1.10	—0.04	0.492	0.527	0.035
预测标准偏差			0.002	0.27			0.416			0.054			0.025		

18.9 在线近红外光谱过程分析

石油炼制工业生产非常需要能够对原料、中间产品和最终产品的质量指标参数进行实时监测，以便及时确定加工方案、调整各装置单元的操作条件，使生产过程达到最优化和降低能耗。近年来国际上主要大型现代化石油炼制企业都纷纷采用了 NIR 在线分析技术来监控炼油工艺，优化了生产过程，取得了巨大经济效益^[26,88]。以下列举几个应用实例。

18.9.1 应用实例

【例 1】 法国 Lavera 炼油厂将在线近红外光谱分析技术用于产量为 100 万吨/年的在线汽油调和^[89]。在线近红外光谱应用前辛烷值测量精确度 0.6~1.0，应用后，辛烷值测量精确度仅为 0.3。仅节约辛烷值带来每年收益为 200 万美元。该厂用作汽油的调和组分有 8 种：丁烷、MTBE、裂化汽油、蒸汽裂解汽油、异构化产品、C-5 馏分、烷基化油及重整产品。

【例 2】 法国拉贝拉子公司在蒸汽裂解装置上安装了在线红外分析仪^[90]。该工厂有 24 台裂解炉，年产 67 万吨乙烯。在线近红外光谱分析在 1 min 内可提供石脑油性质：密度、化学组成、平均相对分子质量、甲烷、乙烯、C3 和 C4 不饱和化合物的潜含量，PI-ONA（链烷，异构烷烃，环烷烃，烯烃和芳烃），结焦指数；提供粗汽油性质：密度、各切割馏分段的化学组成、潜含量、潜含辛烷值和蒸馏曲线、PINOA、苯、甲苯，各种二甲苯和二烯烃含量等。根据这些分析数据和 15 年的装置操作经验，对裂解炉实时操作进行自动优化。通过在线近红外光谱分析的生焦指数对裂解炉辐射管和输油管换热器清焦工序操作。由此带来效益达 100 万美元/年。

【例 3】 西班牙炼油厂，原来是用 RON、MON、蒸汽压及馏程数据四种在线分析仪分别测定以上各性质，改用 NIR 在线仪后可同时测得四种性质。因而大量降低了维护和操作费用。

【例 4】 以色列 Haifa，HIRAM 工艺控制工程有限公司和炼油有限公司在汽油调和、重整和 FCC 三个装置安装了 NIR 光谱分析仪^[17]，在线测量汽油性质（如表 18-60 所示），该分析系统测量精度列于表 18-61。该分析系统在 30s 内完成 13 种性质的测定。

表 18-60 近红外光谱分析测定的汽油性质

化 学 组 成	物 理 性 质	
芳烃总含量/v%	MON 辛烷值	冰点
苯含量/v%	RON 辛烷值	倾点
MTBE/v%	雷氏蒸汽压	闪点
总烯烃含量/v%	黏度	馏程
	API 重度	浊点

表 18-61

HIRAM 近红外光谱分析系统的准确度数据

性质	单位	范围	均方预测误差	相关系数
RON	辛烷值单位	90~100	0.27	0.996
MON	辛烷值单位	80~90	0.23	0.998
雷氏蒸汽压	psi	5~11	0.15	0.99
总芳烃含量	v%	5~45	1.48	0.98
总烯烃含量	v%	0~30	0.94	0.99
苯	v%	0~1.2	0.06	0.95
10%的馏程	F	100~180	3.32	0.97
50%的馏程	F	150~260	4.5	0.98
邻二甲苯	v%	0~100	0.98	0.999
间二甲苯	v%	0~100	0.33	0.999
对二甲苯	v%	0~100	1.06	0.998
己基苯	v%	0~100	0.32	0.999
PDEB	v%	0~100	0.86	0.999

【例 5】 在线近红外光谱分析技术优化原油蒸馏装置^[18]：现代化炼油企业面对的原油来源，原油组成和价格随着时间变化很大。装置需要有灵活的操作方式来适应这种变化，以降低能耗，提高产品质量和产率，才能获得好的效益。在 NIR 在线分析技术出现前，为了保证装置的正常运行，采用保守的操作方法，即通常装置的处理量远低于装置的设计处理量，以防止由于进料组成变化而引起的超出装置设计处理量而发生事故，其结果就是降低了产率，增加了消耗。如当进料原油的品质由重变轻时，蒸馏塔为了防止溢塔和换热器过热事故发生，和防止拔头（即轻产品）产品产量过高，会使下游气体厂过载，因此，在进料更换过渡期间，一般采用逐渐降低进料量的方式，不能有效地利用装置的加工能力，降低了装置的处理量。

在线近红外光谱技术实时检测原油蒸馏装置进料的组成和性质（馏程、相对密度、硫含量、苯胺点、黏度、折光指数等），据此实时控制装置的操作条件，使装置发挥最大安全的处理能力。仅这一项应用就给法国南海岸一家中规模炼油厂增加效益 30 万~50 万美元/年。该厂使用在线近红外光谱分析技术控制调整原油蒸馏装置各产物流出比，降低高附加值产品损失，由此获得效益 30 万美元/年。

【例 6】 在催化裂化装置操作优化中的应用^[18]：在众多的石油炼制工业工艺中，流动床催化裂化（FCC）工艺是效益最好的工艺之一，装置操作难度也是最大的，工艺进料变化对装置操作影响较大，以往离线的实验室分析的速度慢不能满足 FCC 装置操作的快速变化要求（目前国内炼油厂依然靠有实际操作经验的专家承担该车间的主管责任），而在线近红外光谱分析技术能较好地解决了这个问题。在线近红外光谱监测进料的质量和反应器和再生器间影响生焦率主要因素变化情况，及时调整操作条件，以保持最佳生焦量和在最优化装置负荷下多出汽油产品。湿气压缩器（WGC）常常是限制 FCC 装置处理能力的一个处理单元，液体进料质量的变化对 WGC 产生很大影响，需要给装置操作员足够的时间进行操作条件的调整，为了安全生产，通常给予 WGC 操作有较大余地，造成装置不能满负荷生产，在线 NIR 分析技术实时检测提供进料的组成和性质参数，根据这些参数及时调整操作

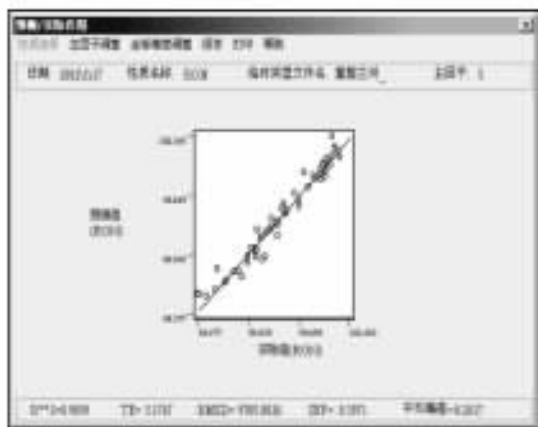


图 18-12 研究法辛烷值模型验证结果

比较情况见图 18-13，其平均偏差为 0.35 个辛烷值单位，满足辛烷值标准方法对分析结果的要求。

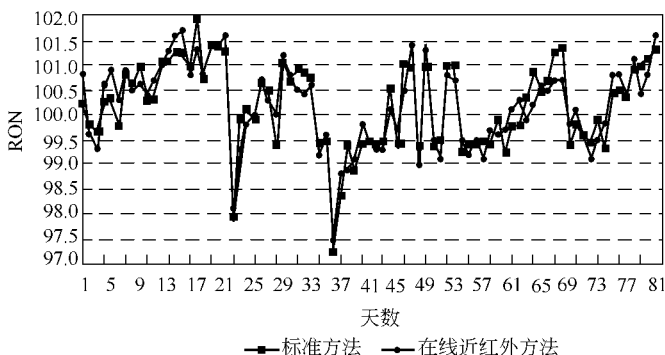


图 18-13 在线近红外测定生成油辛烷值

条件，能使装置处于平稳和高处理量状态。

【例 7】 中国石油化工科学研究院和北京英贤仪器公司合作研制了新型在线 CCD 近红外光谱分析仪^[49]，将其应用在炼油厂重整装置生成油的在线辛烷值检测，根据实时检测数据优化生产装置操作，为企业节省经济效益达 196.74 万元人民币/年。

图 18-12 为 61 个重整汽油的近红外光谱法结果与标准方法结果对照情况；新型在线 CCD 近红外光谱分析仪安装于兰州炼油厂重整装置上，在 2001 年 9 月 13 日至 2001 年 12 月 13 日期间，将在线测量结果与离线化验室标准方法结果进行对比，比

18.9.2 在线近红外光谱在若干炼油工艺应用中的主要分析项目

(1) 原油蒸馏 (Crude Distillation)

原油	汽油	石脑油	煤油	瓦斯油(柴油)
馏程	RON	PIONA 组成	PIONA 组成	Cloud Pt.
密度	MON	ASTM 馏程	ASTM 馏程	ASTM 馏程
	ASTM 馏程	相对密度	相对密度	相对密度
	相对密度		冰点	闪点
			闪点	

(2) 催化裂化 (FCC)

① FCC 进料

相对密度
ASTM 蒸馏曲线
黏度 @ 100℃
折光指数 (Ref. Index)

② FCC 汽油

MON
RON
调和指数 (Blending Indices)

平均馏程
苯胺点
K(UOP)
相对密度
康氏残碳

③FCC 循环油(cycle oil)
闪点
黏度
碱性氮
芳烃
ASTM 蒸馏曲线

(3) 汽油调和 (Gasoline Blending)

调和原料：
 调和指数 (Blending Indices)
 RON, MON, 蒸汽压, ASTM 蒸馏曲线
调和汽油产品：
 RON, MON, 蒸汽压
 蒸馏曲线
 % 苯, %MTBE, % 氧化 物, %硫,
被表征的调和原料：
isomerase, penexate, butane/butene
reformate, TBA, Cat Cracked spirit
alkylate, MTBE, pyrolysis gasoline, 和其他

(4) 瓦斯油调和

调和料 (Blendstocks)
调和指数
十六烷值, 十六烷指数
Cloud point
闪点
瓦斯油调和料的种类包括：
煤油, 低硫的轻瓦斯油, 重瓦斯油 (ex CDU)
高硫的重瓦斯油 (ex VDU), LCO (ex FCC)
加氢裂解瓦斯油

调和瓦斯油产品
十六烷值, 十六烷指数
加和响应
闪点
Cloud Point,

(5) 乙烯生产

石脑油进料
PIONA 组成
ASTM 蒸馏曲线
潜在裂解产量
密度
结焦趋势
相对分子质量

裂解汽油
PIONA 组成
ASTM 蒸馏曲线
苯、甲苯和二甲苯 (BTX) 含量
二烯烃 (% diene)
RON
MON

18.9.3 在其他化工领域中的应用

由于近红外光谱主要提供含氢基团的信息，所以被广泛地用于制药、化学、化妆品、油漆、塑料橡胶、香料等工业领域，进行产品纯度检验和已知杂质分析。可以测定羟值

(OH Value), 皂化值 (Saponification Value), 碘值 (Iodine Value), 过氧化价 (Peroxide Value), 异氰酸脂含量 (Isocyanate Content), 顺/反比率 (Cis/Trans Ratio)、含水量 (Moisture Content)、烷基磺酸含量 (Alkyl Sulfonate Content) 和胺值 (Amine Value), 高分子化合物的相对分子质量、结构序列、单体含量、熔融指数等物化性质^[91,92], 这些性质的测定传统方法大多费时和费力, 近红外光谱方法可在生产控制中替代它们使用, 具有方便和快速的特点。

18.10 小结

近红外光谱分析技术借助计算机技术和化学计量学分析手段, 具有快速和高效的特点; 近红外光谱包含有机物中含氢基团的丰富化学信息, 非常适合石化样品质量参数的分析; 近红外用于石化分析, 不需要化学试剂, 也不破坏样品, 操作方便, 因此, 也称为绿色分析技术。它的应用能显著提高炼油厂化验室的工作效率, 及时提供原料、中间产品和产品的质量分析数据, 为生产装置平稳运行和优化提供了准确数据信息。在线近红外油品质量分析技术和先进过程控制技术配套使用, 实现原油加工方案、各种炼油装置生产的质量卡边控制、安全生产和全厂生产的优化, 是提高整体炼油技术水平和增强企业的市场竞争力的强有力的先进技术。具备该技术与否也是衡量现代炼油企业技术水平的一个重要标志。该技术的广泛应用为石化工业取得了丰厚的经济效益。

参考文献

- [1] Norris K H, et al. *Agricult Eng*, 1962, (43): 54
- [2] Norris K H, et al. *J Food Sci*, 1968, (33): 64
- [3] Stark E, et al. *Appl Spectrosc Rev*, 1986, (22): 335
- [4] Kowalski B, et al. *Chemical Systems under Indirect Observation*. In: Joreskog K, Wold H. *Systems under Indirect Observation*. North-Holland, Amsterarn, 1982: 191~209
- [5] Wold S. In: Martens H, et al. *Food Research and Data Applied Science Publishers*. London, 1983
- [6] Mardia K, et al. *Multivariate Analysis*. Academic Press, London
- [7] 俞汝勤. 化学计量学导论. 长沙: 湖南教育出版社, 1991
- [8] 梁逸曾. 白灰黑复杂多组分分析体系及其化学计量学算法. 长沙: 湖南科学技术出版社
- [9] 可人. 现代科学仪器. BCEIA 金奖产品近红外光谱仪和化学计量学光谱分析软件, 1999 (5)
- [10] 袁洪福, 陆婉珍. 现代光谱分析中常用的化学计量学方法. *现代科学仪器*, 1998 (5): 6
- [11] Kelly JJ, Barlow CH, Jinguji TM, et al. *Anal Chem*, 1989, 61 (4): 313~320
- [12] Jeffrey J Kelly, James B Callis. Non-Destructive Analytical Procedure for Simultaneous Wstimation of Major Classes of Hydrocarbon Constituents of Finished Gasolines. *Anal Chem*, 1990, (62): 1444~1451
- [13] Rocco DiFoggio, Maya Sadhukhan, Martha L Ranc, et al. The Application of NIR Spectroscopy for the Prediction of Properties of Australian Refined Reformate. *Applied Spectroscopy*, 1991, 45 (7): 1153~1157
- [14] Swarin, Stephen J, Drumm Charlene A. Predicting Gasoline Properties Using Near-IR Spectroscopy. *Spectroscopy*, 1992, 7 (7): 42~49

- [15] Espinosa A, et al. On-line NIR Analysis and Advance Control improve Gasoline Blending. Oil and Gas Journal, Oct 17, 1994
- [16] Diaz A, et al. Hydrocarbon Processing, 1996 (2): 71
- [17] I Zilberman, J Bigman, I Sela, et al. Use telecommunications for real-time process control. Hydrocarbon Processing, May, 1996: 91~97
- [18] Al Kabia. NPRA Annual Meeting, San Antonio, Texas, 1997
- [19] Aaron Ranson Lopez, Adriano Parisi. A Venezuelan Technology for Measuring Fuel Quality. 1998, 5 (2): 109~116
- [20] Andrej Popkowski, Ara Barsamian. New gasoline gasoline blending unit started at Poland refinery. oil&gas, 1999, 97 (27): 64~70
- [21] A Barsamain. Get the most out of your NIR analyzers. Hydrocarbon Processing, January 2001: 69~72
- [22] Mullins O C, Daigle T, Crowell C, et al. Gas-Oil Ratio of Live Crude Oils Determined by Near-Infrared Spectroscopy, Applied Spectroscopy. 2001, 55 (2)
- [23] 袁洪福, 陆婉珍, 石油炼制与化工, 1998, 29 (9): 47~50
- [24] 陆婉珍. 用现代信息技术改造传统炼油工艺, 中国石化集团公司科学技术委员会论文选, 2002
- [25] 陆婉珍, 袁洪福, 许广通等. 现代近红外光谱分析技术, 北京: 中国石化出版社, 2000
- [26] www. nirplus. com
- [27] Annual Book of ASTM Standards, (1985) Vol. 05. 04.
- [28] Barsmian A. Inline Blending Minimizes Tankage, Inventory, Giveaway, Boosts Client Satisfaction and Bottom Line. World Refining, Nov-Dec, 1999. 46
- [29] 袁洪福, 龙义成, 陆婉珍. 近红外组成分析仪, ZL 96 2 18361. X
- [30] 袁洪福, 岳俊奇, 陆婉珍. CCD 近红外光谱快速测定喷气燃料的冰点. 分析化学, 1998, 26 (5): 603~606
- [31] 王宗明等. 石油炼制与化工, 1997, 28 (1): 22
- [32] 袁洪福, 陆婉珍. 现代科学仪器, 1998, 5: 6~9
- [33] 徐广通, 袁洪福, 陆婉珍. CCD 近红外光谱快速测定柴油中的芳烃含量. 分析化学, 1999, 27 (1): 29~33
- [34] 徐广通, 陆婉珍, 袁洪福等. CCD 近红外光谱仪在柴油生产控制分析中的应用. 石油炼制与化工, 1999, 30 (9): 57
- [35] 徐广通, 袁洪福, 陆婉珍. 近红外光谱法测定柴油十六烷值. 石油学报, 1999, 15 (4): 62
- [36] 徐广通, 陆婉珍, 袁洪福. 近红外光谱法测定柴油中的芳烃含量. 石油化工, 1998, 28 (4): 253
- [37] 袁洪福, 龙义成, 陆婉珍. 近红外光谱仪的研制. 分析化学, 1999, 27 (5): 608
- [38] 袁洪福, 陆婉珍. 近红外光谱分析技术正在掀起一场分析效率革命. 现代仪器
- [39] 袁洪福. 近红外光谱 (NIR) 分析技术在中国的发展. 中国石油化工. 北京: 中国石化出版社, 2000. 148
- [40] 褚小立, 袁洪福, 陆婉珍. 光谱结合主成分分析和模糊聚类方法的样品聚类与识别. 分析化学, 2000 (4)
- [41] 王艳斌, 袁洪福, 陆婉珍. 人工神经网络用于柴油近红外光谱的多元非线性校正. 分析化学, 2000 (28)
- [42] 袁洪福. 近红外 (NIR) 光谱分析技术优化催化裂化工艺操作. 2000 年催化裂化年会, 上海, 2000
- [43] 王艳斌, 袁洪福, 陆婉珍. 近红外分析方法测定润滑油基础油黏度指数. 润滑油, 12 (6): 53~55

- [44] 袁洪福, 褚小立, 王京等. 聚异丁烯高活性端基含量及其相对分子质量测定方法的研究. 光谱学与光谱分析, 2002
- [45] 杨玉蕊, 袁洪福等. 近红外快速测定催化裂化汽油重馏分加氢脱硫产物辛烷值的研究. 国外分析仪器技术与应用, 2002
- [46] 许育鹏, 袁洪福, 陆婉珍. 采用相关分析解析石油产品近红外光谱. 现代仪器, 2002
- [47] 田高友, 袁洪福, 陆婉珍等. 小波变换在近红外光谱分析中的应用进展. 光谱学与光谱分析, 2002
- [48] 褚小立, 袁洪福, 陆婉珍. 近红外稳健分析模型的建立. 光谱学与光谱分析, 2002
- [49] 袁洪福, 褚小立, 陆婉珍等. 一种新型 CCD 在线近红外光谱分析仪的研制. 分析化学, 2003
- [50] 石油和石油化工系统第六届光谱分析技术报告会文集. 中国石油学会石油炼制分会 & 中国石油学会石油科技装备委员会, 北京, 2002
- [51] Jerome J W. Applied Spectroscopy Reviews, 1996, 31 (3): 251~320
- [52] Chung H, Ku M. Comparison of Near-Infrared, Infrared, and Raman Spectroscopy for the Analysis of Heavy Petroleum Products. Applied Spectroscopy, 2000, 54 (2)
- [53] John B Cooper, Kent L Wise, William T Welch, et al. Determination of Weight Percent Oxygen in Commercial Gasoline: Comparison between FT-Raman, Ft-IR, and Dispersive Near-IR Spectroscopy. Applied Spectroscopy, 1996, 50 (7): 917~921
- [54] ASTM E 1655
- [55] 王艳斌. 人工神经网络在近红外分析方法中的应用及深色油品的分析. 石油化工科学研究院博士学位论文, 2000
- [56] 吴艳萍. 漫反射近红外光谱测定聚丙烯树脂性质的方法研究. [学位论文]. 北京: 石油化工科学研究院, 2003
- [57] 褚小立. UV 光谱快速测定渣油 SARA 四组分的方法研究. [学位论文]. 北京: 石油化工科学研究院, 1999
- [58] Annual Book of ASTM Standards, (1985) Vol. 05. 04.
- [59] 蒋健, 张金华, 冯莲英. 近红外光谱法在调和汽油控制分析中的应用. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002. 92
- [60] 陶敬华. CCD 近红外光谱法快速测定汽油辛烷值. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 111
- [61] 邵波, 黄小英, 陈月嫦. 近红外光谱法快速测定成品罐汽油的有关性质. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 132
- [62] 王艳丽, 李述忠. NIR-2000 近红外光谱仪测定汽油辛烷值. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 146
- [63] 余新铭. 近红外光谱仪在汽油质量过程控制上的应用. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 149
- [64] 田高友, 褚小立等. 小波变换与微分方法结合用于重整汽油辛烷值近红外分析. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 69
- [65] GEFodor. Analysis of Petroleum Fuels by Midband Infrared Spectroscopy. SAE TECHNICAL PAPER SERIES, 1994 (94): 10~19
- [66] 刘慧颖, 戴百龄. 近红外光谱分析技术用于石油燃料质量检测的探讨. 中国石油协会石油炼制分会第三届年会论文集, 济南, 1997. 641
- [67] 刘慧颖, 韦锐, 熊春华. 用近红外光谱仪测定车用汽油有关质量性质的研究. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 75

- [68] 刘慧颖, 韦锐, 熊春华. 用近红外光谱仪测定喷气燃料有关物理性质的研究. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 82
- [69] 韦锐, 刘慧颖, 熊春华. 用 NIR3000ZH 型便携式近红外光谱仪测定军用柴油的性质. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 88
- [70] 杨玉蕊, 董金辉, 袁洪福. 近红外光谱快速测定催化裂化汽油重馏分选择性加氢脱硫产物的辛烷值. 国外分析仪器技术与应用, 2002 (2): 68~72
- [71] 褚小立, 袁洪福, 谢智彬. 近红外光谱方法同时测定异构体的含量. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 51
- [72] R W Kennard, L A Stone. Technometrics, 1969. 137
- [73] 刘莎, 朱虹. 近红外光谱在汽油组成分析中的应用. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 122
- [74] 李克忠, 殷宗玲. 近红外光谱快速测定调和汽油中芳烃、烯烃的含量. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 119
- [75] 黄小英, 邵波, 林楚堂, 等. 近红外光谱法快速测定重整汽油族组成. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 128
- [76] 杨晶. 近红外光谱法快速测定生产过程汽油馏程. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 115
- [77] <http://www.isl-france.com/>
- [78] Garrigues S, Andrade J M, Prada D. Anal Chim Acta, 1995 (317): 95
- [79] 李桂莲, 徐新举等. 近红外法测定石油产品馏程在生产中的应用. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 143
- [80] Narve Aske, Harald Kallevik, Johan Sjo blom. Determination of Saturate, Aromatic, Resin, and Asphaltenic (SARA) Components in Crude Oils by Means of Infrared and Near-Infrared Spectroscopy. Energy & Fuels, 2001 (15): 1304~1312
- [81] 褚小立, 袁洪福, 陆婉珍. 紫外-可见光谱-偏最小二乘法测定渣油四组分含量. 分析化学, 2000, 28 (12): 1457~1461
- [82] Hidajat K, Chong S M. J Near Infrared Spectrosc, 2000 (8): 53~59
- [83] 关子杰. 润滑油, 1998, 13 (5): 1~6
- [84] 吴艳萍, 陆婉珍, 袁洪福等. 聚丙烯树脂拉伸屈服强度的测定方法. 中国专利, 02125849. X, 2002-7
- [85] 吴艳萍, 陆婉珍, 袁洪福等. 测定聚丙烯树脂等规指数的方法. 中国专利, 02125851. 1, 2002-7
- [86] 邢志娜, 王举香, 申刚. 近红外光谱在推进剂领域中的应用. 石油和石油化工系统第六届光谱分析技术报告会文集, 北京, 2002 (10). 164
- [87] Finch Peter, Measurement Control, May, 1994, V. 27
- [88] Knott D. J Oil & Gas, Mar, 1997. 39
- [89] Lambert D, et al. Hydrocarbon Processing, Dec, 1995, 103
- [90] BUHLER 公司近红外光谱仪广告资料
- [91] Classon R. Chemical Engineering, April 1993, 103

19. 近红外光谱分析在制药工业中的应用

近红外光谱分析是近年来迅速发展和推广应用的一种分析技术^[1~2]。NIR 光谱分析的最大特点是对样品无侵害性与破坏性；操作简便，分析迅速，一次分析仅需几十秒到几分钟；仪器价格便宜，使用成本低；测量信号可以远距离传输和分析；特别是与计算机技术和光导纤维技术相结合，采用 NIR 透射、散射、漫反射光学检测方法，可不使用化学试剂，不进行样品预处理，直接对颗粒状、固体状、糊状等不透明的样品进行分析。

20 世纪 80 年代以来近红外光谱技术取得了重要进展，特别是在药学领域，已有大量文献介绍了近红外光谱技术在这些方面的应用。

19.1 药物制剂分析

19.1.1 早期应用

近红外光谱最早用于药物分析是在 20 世纪 60 年代后期^[3]，Sinsheimer 和 Keuhnelian 利用近红外对包含有铵盐/氯化钾混合物的丸剂和只含铵盐的丸剂样品进行了分析。在 1050~2800nm 区域采集样品的近红外光谱，利用光谱峰的信息进行定性分析。随后，Oi 和 Inaba 在日本杂志上发表了定量分析制剂中烯丙异丙基乙酰甘胺酰化物和那非西丁的文章。样品用氯仿（在近红外区内基本透明）溶解，分别使用 1983nm 和 2019nm 处的吸收值测定烯丙异丙基乙酰甘胺酰化物和那非西丁的浓度，可以排除制剂中其他成分的干扰。

美国食品药品监督管理局（FDA）的 Sherken 最早使用近红外对片剂药物进行含量测定。在这项研究中，作者建立了对片剂和市售制剂中眠尔通（甲丙胺酯）片含量测定的近红外光谱分析法，使用近红外对一系列的眠尔通标准溶液进行测定，并用以建立校正方程。近红外法的准确度与对照方法的准确度相当。这一新方法比较简单，因此被推荐作为眠尔通片的法定含量检测方法。

Allen 使用近红外光谱法对片剂中的卡立普多（异丙眠尔通）、那非西丁和咖啡因进行含量测定。每种样品取 20 片研碎并溶解于氯仿中，制备一系列卡立普多、那非西丁和咖啡因标准溶液，在 2750~3000nm 之间进行扫描，在所测定的浓度范围内均与比尔定律相符（最大浓度值为 200mg/mL）。分别使用 2820nm 和 2910nm 处的吸收值可以对卡立普多和那非西丁进行同时测定。在计算药物浓度前，扣除了每个波长处的基线吸收值。咖啡因的浓度在 3390nm 处测定，但是在这个波长处，氯仿有强吸收，所以必须使用四氯化碳代替氯仿作溶剂。结果表明，近红外方法能够准确地进行定量分析，对这 3 个药物，其变异系数（CV）均在 1.4% 以下。

在 1977 年发表的文章中，Zappala 和 Post 考察了近红外对片剂和缓释胶囊中眠尔通的分析。在这篇文章发表以前，USP XIX 版采用了比色法作为片剂中眠尔通的含量测定方法，比色法比以前的方法快捷，但试剂的 pH 对结果有较大影响，因此，近红外法仍然

是制剂中眠尔通含量测定的最具吸引力的方法，因为它采用了眠尔通在 1958nm 处的主要合频带，这就避免了 2915nm 处吸收峰所受的干扰，尽管 1958nm 处的吸收较弱，但已足以建立校正方程。这种方法要求取至少 20 粒片剂或胶囊研碎，将相当于 300~400mg 眠尔通的样品溶于氯仿中。为检验所建立的校正方程，选取了 4 个不同生产厂家的 9 种市售品进行测定，所得变异系数片剂为 0.7%，胶囊为 1.3%，均好于对照方法。

19.1.2 原料药、辅料等的质量评价

对药品生产原料的严格质量控制是确保药品质量的重要前提。传统的质量控制方法是从生产车间中取原料到分析中心或化验室进行定性定量分析，需要取样，按标准方法分析，并作权威评价，然后反馈到生产车间。这种分析耗时较长，达不到及时控制质量的目的。

现代的近红外分析方法可以克服上述局限性，通过光导纤维将传感探头与 NIR 分析仪相连，可以在原料进入生产车间时，立刻检测生产原料的质量，全部的定性定量分析可以在很短的时间内完成，真正实现了严格的质量控制，达到及时、快速分析的目的^[4]，从而可用以确认和评价药品生产过程中的各种原料。

19.1.2.1 原料药鉴别测定

Shah NK. 等用近红外光谱法对不同厂家生产的制药原料进行了检测，包括三种不同类型的微晶纤维素、乳糖、磺胺甲基异恶唑（SMZ）以及一种专有成分。作者在 1100~2500nm 范围内以 4nm 为间隔测定了近红外光谱。为对合格与不合格的原料进行分类，使用了两种分类方法处理近红外光谱数据，一种方法是首先计算出近红外全谱区吸收的主成分，然后用马氏距离法进行分类。这种方法使用样品的近红外全谱数据，可以确保光谱所包含的信息均能反映到分类结果中。另一种方法是 SIMCA 法，该法是基于主成分分析建立分类模型。作者发现这两种方法具有一定的互补性，并指出通过一种模式识别技术对原料的光谱进行分类时，不一定能够正确地识别出不能接受的原料，而联合使用几种独立的分类技术能够给出更为可靠的分类结果。

Lunardelli L. 等用近红外光谱在称重时鉴别不同的原料药。其结果表明，近红外光谱系统具有可靠、重现性高等特点。这种装置操作简单，耗时少，可用于分析固体混合物、半固体、液体、颗粒以及压缩材料。Ren Y. L. 等用漫反射近红外光谱结合偏最小二乘法建模测定了粉末中的磺胺甲基异恶唑含量。Jacobsson S. P. 等^[8]用反射光谱结合偏最小二乘法定量测定了水杨酸偶氮磺胺吡啶，结果可与液相色谱相媲美。Smola N. 等^[9]用近红外光谱结合主成分分析、偏最小二乘法、主成分回归等化学计量学方法建立了四环素原料药分析的特定参数，并用这些参数提出一种定性定量测定四环素原料药的方法，作者将其与卡尔-费休测水法、测热法、傅里叶变换红外光谱法鉴别四环素分别做了比较，近红外方法操作简便，适于用做常规的入库原料测定。Ren Y. L. 等^[10]用人工神经网络校正方法测定了甲硝唑粉末的含量。作者以真实数据建立了校正模型，考察了神经网络的拟和程度，并讨论了过拟合的情况。由于人工神经网络具有有效的多元非线性校正能力，测定结果准确可靠。

19.1.2.2 辅料类型鉴别测定

生产药物的辅料必须经过可靠的鉴别才能使用。欧洲药典采用的湿化学鉴别纤维素酯的方法既耗时又比较昂贵。Kramer K 等^[11]研究了用近红外光谱鉴别不同的药物辅料，加

速了这种鉴别过程，减少了在生产过程中的延迟，并运用因子分析和 SIMCA 方法成功地鉴别了微晶和粉末纤维素以及纤维素和纤维素酯。这种鉴别方法在统计上是非常显著的，因而可用来快速鉴别大量不同生产商不同性质的样品。

Candolfi A. 等^[12]将 SIMCA 法应用于近红外光谱，成功地鉴别了几种制药中常用的辅料，包括无水磷酸二钙、无水乳糖、淀粉甘醇酸钠、硬脂酸镁、羟丙基甲基纤维素、聚乙烯吡咯烷酮、十二烷基硫酸钠、淀粉以及微晶纤维素。

Ulmschneider M 等^[13]用近红外光谱结合相应的化学计量学校正方法快速非破坏地鉴别了 11 种固体药用辅料。这种方法在车间、生产线以及实验室中都可使用，而且几乎不需要任何预处理。测量时使用一根 2 米长的光纤以漫反射模式照射样品表面，将光谱数据存储下来，然后应用主成分回归和聚类分析方法来鉴别不同类型的辅料。Berthold Jesper 等^[14]用近红外光谱测定羧甲基纤维素中极性基团对水的吸附。水分子可以吸附到含羟基、羧基等极性基团的多糖中，作者通过近红外光谱和多元校正方法研究了不同程度取代的羧甲基纤维素（CMC-Na，CMC-Ca）与水分子之间的相互作用。水分子的吸附很大程度上受反离子的影响。主成分分析表明近红外光谱的差异主要与反离子的性质、取代程度、相对湿度以及水分子吸附的极性基团有关。偏最小二乘方法从光谱数据中获得的信息与吸附在羟基和羧基上的非凝固态水有关。研究表明，在相对湿度从 0~100% 的情况下，水分子可特异性地吸附在某一极性基团上，其性质则取决于极性基团的属性。

19.1.3 物化性质测定

原料药的结晶状态、粒径和密度在制剂生产和活性成分的控制过程中非常重要。近红外反射光谱可用于对制药工业中所用的不同物理性质原料的检测，可以区分出物质的不同结晶状态、粒径和密度。Dreassi 等人利用近红外反射光谱对药物生产中所用原料的物理状态进行了检测。待测粉末放置在聚乙烯样品池中，盖上显微玻片，在 1100~2500nm 波段进行测定，所检测的样品及其物理性质如表 19-1 所示。

表 19-1 原料药物理性质检测结果

化合物	所区分性质	化合物	所区分性质
咪喃妥因 扑热息痛 吉非贝齐	大晶体和微晶 两种不同密度的粉末 从不同溶剂中结晶	鹅去氧胆酸 布洛芬 四咪唑	四种晶型 高/低密度 左旋体和外消旋体

通过线性判别分析对原始吸收数据进行处理，结果证实了近红外反射光谱法可以作为分析原料药物理参数的一种简单、快捷的工具，具有非侵入性，不破坏样品的特点。尽管使用近红外光谱法进行定性分析时所需做的校正工作相当复杂，但校正模型一旦确立，便能简便快速地获得制剂配方中成分的全面信息（定性定量），因此确保了产品的质量。

Norris T 等运用近红外确定了甲磺酸曲伐沙星两种晶型转化的终点。Seyer JJ 等测定了吡喹酮和蔗糖的结晶度，并对其无定形和结晶状态进行了研究，结果发现其区别主要在于两者转化过程中，近红外光谱振动产生了分裂。Patel A. D. 等用近红外光谱定量测定了含磺胺甲基异恶唑的二元和多元混合物中的多形体，评价了仪器的重现性、方法误差、精确度、检测限等，结果显示近红外技术可在二元和多元混合物中测出 2% 的多形

体。与其他测定方法相比，近红外技术是一种测定固体中多形体的准确度、精度都较高的方法，而且使用方便。Luner Pe 等用近红外光谱测定了固体二元粉末中的晶体成分。作者定性定量分析了磺胺甲基异恶唑、磺胺噻唑、乳糖和氨苄青霉素等不同的晶体成分。近红外光谱数据可直接透过玻璃瓶获得。作者还以磺胺甲基异恶唑和磺胺噻唑为例将近红外方法与 X 射线衍射法进行比较，证明了近红外方法的准确性。作者用磺胺甲基异恶唑和氨苄青霉素进行了浓度测定范围的研究，结果表明完全可以测定到 1% 以下。该实验表明，近红外技术作为一种物理鉴别方法可用来测定较低含量的晶体。ONeil AJ. 等应用近红外光谱结合多元线性回归的方法测定了多种药物和辅料的粒径。这些药物和辅料具有中等大小的粒径，主要有阿司匹林、无水咖啡因、扑热息痛、一水乳糖和微晶纤维素。作者用近红外反射模式获得光谱数据，以其他方法测得的数据建立校正模型，试验表明这是一种快速有效的测定方法。

Blanco M. 等用近红外光谱研究了药物的多形性。多形物不同的物理化学性质在制药工业上具有很大的意义。作者用近红外方法鉴别了米欧卡霉素的晶型和非晶型物，基于这两者光谱的不同，作者建立了定性定量分析这种多形物的模型，以多元回归进行定量分析，逻辑回归进行定性鉴别。这种方法可检测出低含量晶型物 ($<10\%$)，是一种控制多形物纯度的有效方法。Buckton G. 等用近红外光谱测定了室温下无定型和晶型乳糖的含量。

Seyer JJ. 等研究了近红外技术测定吡哆美辛晶体和非晶体粉末中的晶体含量。光谱数据可透过玻璃瓶直接获得，用 X 射线衍射作参考方法。试验显示近红外光谱可准确测定晶体含量。Dziki W. 等^[24]研究了用近红外来测定沙氟沙星晶格中水的流动性。近红外光谱可成功地区分出合格品和不合格品，而 X 射线衍射、中红外等方法都无法区分。近红外光谱可测定出晶格中水分子的位置和取向。

19.1.4 包装材料的质量控制

由于药政管理部门对药品生产的管理越来越严格，作为药品生产原料之一的包装材料的质量也不容忽视。用于药品包装的多聚物能够通过近红外光谱进行鉴别。Ciurczak 等^[25]做了这方面的尝试，用判别分析法处理包装材料样品的近红外光谱数据，成功地鉴别了塑料瓶、PVC 等药品包装材料。

19.1.5 片剂溶出度预测

对于片剂溶出速率的预测是近红外光谱法在药学领域的又一应用。Zaunikos 等人在 1991 年的一项研究中，将贮存在同样高湿度条件下不同厂家的卡马西平的溶出情况做了比较，测定药物置于 USP 溶出仪 II 中 1h 后溶液中药物的浓度，并据此建立校正模型。用光谱仪采集完整片剂的近红外光谱，然后测定片剂的溶出速率。在 5 天高湿度贮存期间，一般仿制片剂的近红外光谱和溶出行为几乎没有改变，而品牌片剂的光谱和溶出行为变化显著。对于品牌片剂，使用主成分分析和 BEAST (Bootstrap Error-Adjusted Single-sample Technique) 算法建立了校正模型，校正的相关系数为 0.99，对 1h 后进一步的溶出进行预测，标准预测误差 (SEP) 为 0.8%。进一步延长片剂的贮存时间对溶出行为和近红外光谱无明显影响，但在 5d 后，近红外光谱上开始出现另一个吸收峰，这被认为是片剂

的降解产物峰。

Drennen 等继续进行了这方面的研究, 于 1991 年发表了一篇文章, 比较了改进的 BEAST 算法和马氏距离法在对卡马西平片的溶出行为分析中的应用。由于马氏距离法假定与校正集和测试集相关的光谱变化是随机的, 这在复杂的药物混合物中不一定成立, 而 BEAST 算法是一种非参数检验方法, 能够用于几乎所有的光谱分析。在这项研究中, 作者分别使用马氏距离法和 BEAST 算法处理了全谱 (710 个波长点) 和主成分光谱, 从而对高湿度条件下的卡马西平的溶出行为做了分类。多次试验结果表明, BEAST 算法给出的定性结果的准确度比马氏距离法好。在一次试验中, 用 9 片溶出速率低的片剂作训练集, 用 21 片不同溶出速率的片剂对模型进行检测, 改进的 BEAST 算法正确地识别出了所有溶出速率比训练集快的片剂, 而马氏距离法对这类片剂的识别错误率达 58%。在几乎所有的实验中, 改进的 BEAST 算法均给出了比马氏距离法更精确、更准确、更快速的结果。

19.1.6 片剂硬度测定

Morisseau K. M. 等用近红外评估和定量分析压力对片剂硬度的影响, 获得了比实验方法更加准确的结果。Kirsch JD. 等运用近红外测定了西米替丁片剂的硬度, 并比较了主成分分析、主成分回归算法和新的最适应算法, 结果相当, 但后者更易推广。

19.1.7 制剂颗粒的质量控制

在药品生产过程中, 颗粒可以作为最终制剂形式, 也可以作为压片或胶囊填充过程的中间体。1987 年, Chasseur 使用近红外光谱分析了西米替丁颗粒的含量。制备了一系列西米替丁含量为标示量的 70%~130% 的样品颗粒, 并采集其近红外光谱。选用一个或两个波长点, 对一阶导数和二阶导数光谱进行分析, 以建立校正模型。结果表明, 使用二阶导数光谱的双波长模型得出的结果最佳, 标准误差为 1.75%。研究人员还比较了近红外法与 UV 对照法在西米替丁含量测定中的应用, 制备了药物含量为标示量的 100% 的样品, 并用两种方法分别进行测定。近红外法的标准误差为 2.73%, UV 法的标准误差为 2.79%, 两者基本一致。

Oneil AJ. 等运用近红外测定赋形剂微晶纤维素颗粒大小分布情况。其利用三波长多元线性回归和主成分回归取得了较好的结果。Oneil AJ.^[31]等利用多元线性回归测定了阿司匹林、无水咖啡因、对乙酰氨基酚等活性成分和乳糖、微晶纤维素颗粒的大小。

19.1.8 药物的定性定量分析

有机化合物在近红外光谱区均有吸收, 其特点是吸收较弱, 样品不需处理就可测量, 适于组分的常量分析, 易于实现快速的非破坏性分析。但近红外漫反射光谱峰较宽, 样品中各组分谱带重叠严重, 这给直接进行非破坏性分析带来了困难。然而化学计量学技术可有效的解决这个问题。目前国际上应用于固态样品的近红外漫反射光谱法中较多的有主成分分析和偏最小二乘法。任玉林等^[32]运用偏最小二乘-近红外漫反射光谱法分析扑热息痛。对常规光谱法和一阶导数光谱法, 分别在其最佳波长间隔、主成分数下, 建立偏最小二乘法的数学校正模型。用所建模型预测 44 个校正集样品中扑热息痛的含量, 得到比较

满意的结果。

Blanco M. 运用近红外光谱法分别测定了混合药品、素片和包衣片剂中活性成分吉非贝齐的含量，并测定了不同原料来源、制粒法、颗粒大小分布等条件下药物的含量。同时他们^[34]还对两种完整片剂中相关成分进行了鉴定并对其中活性成分进行了定量分析，从而可以用于生产过程的控制。Merckle P. 等运用近红外透射光谱法测定了泡腾片中乙酰水杨酸的含量。单方中含量为 14.9%，与维生素 C 组成的复方中含量分别为 17.4% 和 8.7%。用偏最小二乘法进行数据处理，相对标准偏差分别是 1.20%~2.01%，1.91%~2.21%，2.41%~4.50%，并用高效液相色谱法作为参照，但其与 HPLC 法相比，具有快速、简洁、所需样品量小、非破坏性分析且无有机溶剂对环境的污染的特点。

Thosar SS. 等通过透射光谱测定了不同含量组成的片剂中茶碱的含量。通过混合纤维素 R PH-101 和 0.5% (w/w) 硬脂酸镁以及不同浓度的无水茶碱 (0, 1%, 2%, 5%, 10%, 20%, 40%) 获得各种配比的片剂，并发现运用偏最小二乘法可以获得最小预报残差。

Eustaquio A. 等利用近红外光谱法测定了扑热息痛完整片剂中活性成分的含量，可用于生产过程的即时监控。

Guo JH. 等将近红外用于固体制剂的测定，其分别测定了片剂中活性成分（对乙酰氨基酚或茶碱）、黏合剂（羟乙基纤维素）、添加剂（乳糖、硫酸钙、磷酸氢二钙、微晶纤维素）以及滑润剂（硬脂酸镁）的含量。

任玉林等^[6]将偏最小二乘法同近红外漫反射光谱法相结合，分析粉末状磺胺甲基异恶唑，讨论了波长间隔对偏最小二乘法进行磺胺甲基异恶唑含量定量预测的影响。用已建的校正数学模型，预测对照品和试样中的磺胺甲基异恶唑含量，RSD 分别为 0.11% 和 0.56%，结果令人满意。

Takuo Y. 等运用多元线性回归方法测定了抗凝血药中葡萄糖和柠檬酸的含量，为抗凝血药生产过程中的监控提供了一种方法。Kreft K. 等^[41]运用近红外光谱对聚乙烯吡咯烷酮进行了定性分析。

Buckton G. 等运用近红外光谱法测定了无定型和结晶乳糖晶型的转变。Bauer H. 等分别运用扫描测热、X-单晶衍射和近红外方法对喷雾干燥后的甘露醇、山梨糖醇混合物的多形性进行了测定。近红外比其他两种方法更易接近被测样品，使测量更加方便。Ulmschneider M. 等运用近红外光谱法对 11 种固体辅料进行了鉴别。Candolfi A. 等^[45]运用 SIMCA 法对 10 多种药用赋形剂的近红外光谱进行了鉴别。

屈凌波等^[13]利用近红外光谱分析技术，直接测定了头孢类片剂的含量，对不同的分析方法和不同分析对象进行了探讨，考察了近红外光谱分析在制剂生产各个环节中的应用可能，如生产原料的鉴定、粉体样品的确认、微球样品和固体片剂的含量测定。作者对所采集到的近红外光谱数据首先采用数据标准化处理，对三次测量的结果取其平均值，然后利用判别分析，主成分判别分析和聚类分析等进行处理，都可以得到相似的结果，利用判别分析方法来处理光谱数据的二阶导数，可以使药物颗粒大小对图谱的影响较小，尽可能地缩小了散射不均匀对测定造成的影响。作者用多元回归法和偏最小二乘法建立标准曲线，定量测定药物和水分的含量，在预测被测物质活性含量时，偏最小二乘法可以得到较小的标准偏差。

19.1.9 制剂中水分的测定

由于水分子在近红外区有一些特征性很强的合频吸收带，而其他各种分子的合频和倍频吸收相对较弱，这使近红外光谱能够较为方便的测定药物和其他化学物质中水分的含量。Luukkonen P. 等对硅化微晶纤维素和 α -乳糖湿团中的水分进行了测定，并与 MTR (Mixer Torque Rheometry) 进行了比较，同时还给出了水分不同结合状态对谱图的影响。Sanchez MS. 等对三水合苄苄青霉素的水分进行了测定。Berntsson O. 等将漫反射近红外光谱用于在线测定硬胶囊中的水分。利用多元线性回归和偏最小二乘法对数据进行了处理，无需样品预处理且分析时间仅需 1~2min，检测浓度范围是 5.6%~18.0% (w/w)，预报残差为 0.1%。Derksen MWJ. 等运用近红外光谱法分别测定了 8℃、50℃、60℃ 保存两个月的冻干制剂中残留水分的含量。

生理盐水中 NaCl 浓度的测定是药物检验中的一个重要内容，常用的方法是滴定法。该法不但要破坏待检溶液，而且消耗的试剂昂贵，分析成本高。随着计算机的普及和化学计量学软件的开发，近红外光谱法已成为测定水溶液中电解质的简便方法。该法的基础是利用电解质对近红外光谱中水的吸收带的干扰作用。其特点是直接测定样品的近红外光谱，通过计算机的运算，就能预报电解质的浓度。方法简便快速，适于大批量的现场分析。任玉林等^[8]研究了 NaCl 对水的近红外光谱的吸收带的干扰作用，探讨了非破坏分析生理盐水中 NaCl 浓度的可能性。结果表明，同温度下测定可用线性回归预报 NaCl 的浓度。若在不同温度下测定，必须考虑温度对测定的影响，只有采用多变量统计技术中的主成分回归方法才能得到较小的预报标准误差。

19.1.10 其他应用

任玉林等^[6~12]依据药品的近红外反射一阶导数光谱，应用系统聚类分析、逐步聚类分析、主成分分析和逐步判别分析等多变量统计分类技术，分别进行了甲硝唑、磺胺脒、芦丁、甲氧苄胺嘧啶等粉末药品的质量控制方法的研究。该法能将合格药、劣药和假药正确地区分开来。

Berntsson O. 等应用近红外漫反射光谱测量了几种药物和辅料的有效厚度。受检的粉末包括聚苯乙烯、微球、微晶纤维素、薄膜包衣球以及不同粒径的无包衣球，实验表明，光谱吸收信号与粉末厚度呈指数关系。作者还介绍了该实验在生产在线监控中的应用情况。

1986 年，Ciurczak EW. 使用近红外光谱对药物制剂混合物进行了分析，并对三种数据处理方法——光谱的减除、光谱的再现和判别分析进行了考察。制备的混合物样品中包含三种活性成分阿司匹林、布他比妥和咖啡因，这些样品中有些可能缺失某种活性成分，也有些可能其成分含量在标示量的 90%~110% 范围变化。将所有样品放置于粉末研磨机中以确保混合均匀性，然后置于样品池中扫描。首先考察光谱的减除，将安慰剂的光谱从完整制剂光谱中减去，产生的光谱与所缺失的活性成分的光谱非常接近。这一技术可用于对片剂中组成成分的鉴别。接着进行的也是对制剂组成的定性鉴别，配置一系列已知浓度的混合物，利用改进的多元线性回归技术，找出光谱变化与药物浓度变化的相关性。这样如果所鉴别的药物存在于样品混合物中，即可对药物的光谱进行再现。第三个实验是使用

判别分析对样品进行分类。在一系列混合物中，咖啡因、布他比妥和阿司匹林的浓度分别在标示量的 90%~110% 之间变化；在另一系列混合物中，三种药物中的一种不存在于混合物中，其他两种的浓度在标示量的 90%~110% 之间变化。通过马氏距离的计算成功地将这些制剂样品分成各自类似的几类。这一技术同样用于完整制剂（三种药物的含量均为标示量的 100%）、临界制剂以及缺味制剂的分析。尽管所有的样品均得到了正确的分类，作者仍然注意到一个问题，即几乎无法通过视觉对光谱进行区分。这使得分析工作者无法靠肉眼观察光谱差异来进行鉴别，而必须依赖软件的帮助。

19.2 体内药物、生物分子分析

19.2.1 体内药物分析

Hsiao T C. 等用近红外光谱测定了水和全血中葡萄糖的浓度。作者用偏最小二乘回归法压缩光谱数据并建立了回归模型，采用交互证实的方法评价偏最小二乘回归法，在光谱范围内选择回归因子。为了获得最优的控制变量，采用回归因子映射（Regression Factor Map, RFM）和误差表面映射（Error Surface Map, ESM）来确定选择的变量。对水溶液来说，最优的控制变量范围是 0.915~1.335mm，偏最小二乘回归因子数是 9，全血中相应是 1.105~1.335mm 和 10mm。交互证实的平均百分误差（Mean-Percentage Error of Cross-Validation, MPECV）分别是 7.36% 和 9.20%。

Hazen KH. 等用近红外透射光谱测定了未经稀释的人血液中总蛋白、白蛋白、球蛋白、三甘油三酯、胆固醇、尿素、葡萄糖和乳酸的浓度。作者在 4000~5000cm⁻¹ 的光谱范围内测定了 242 份血液。通过数字滤波和偏最小二乘回归为每种分析物建立校正模型。每个分析模型被分别优化，考虑各自的波长范围、因子数以及用作数字傅里叶滤波的高斯型响应函数的宽度和位置，对每种分析物都进行了精确测定，只有乳酸的含量在实验条件下低于检测限。葡萄糖的标准预测误差是 23.3mg/dL（1.29mM）。作者考察了葡萄糖校正模型的相关性和稳定性，在获得校正数据集 19 个月 after，通过一台校正过的光度计对 50 个人血液样本的预测值进行了评估。结果显示预测值存在很小的正向偏差，相应表明预测模型有很小的不稳定性。这种不稳定性可能来自于实验仪器的变化。无论如何，预测值和实际值之间很强的相关性证明校正模型是建立在特异性的葡萄糖信息之上的。

Pezzaniti JL. 等用近红外光谱测定了尿中尿素、肌苷、葡萄糖、蛋白质和酮体的含量。作者用多项式平滑方法处理样品光谱，通过对正常和非正常样品进行部分因子实验设计获得光谱数据。结果显示尿素、肌苷、葡萄糖、蛋白质和酮体在 1350~1800nm 和 2050~2375nm 范围内有很强的吸收，适于用作精确测定。通过实验可得出结论，近红外方法测定尿素和肌苷的准确性只略小于参考方法；测定葡萄糖、蛋白质和酮体具有足够的准确度，可用作筛选工具。近红外方法的优越性在于：不需试剂，样品制备方便，快速测定，可同时测定多种成分。

19.2.2 蛋白质结构分析

Robert P. 等应用近红外技术研究了蛋白质的二级结构。作者应用近红外光谱结合中

红外光谱测定了 12 个固体模型蛋白。肌红蛋白、 β -乳球蛋白和 β -酪蛋白在近红外区显示了特异的吸收带。肌红蛋白的吸收带在 2055nm, 2170nm, 2289nm 和 2350nm, β -乳球蛋白在 2203nm, 2267nm 和 2300nm 处有吸收峰, β -酪蛋白的二阶导数光谱在 2269nm 处有很强的吸收峰。通过通用标准相关分析可以得出近红外光谱与二级结构之间的关系: 2172nm 和 2289nm 代表 α -螺旋结构, 2205nm, 2264nm 和 2313nm 代表 β -折叠, 2265nm 代表无规卷曲。

Marc F. 应用近红外多光谱成像技术考察了固相多肽合成过程。这种成像仪使用声-光可调滤波器和 NIR InGAs 聚焦平面阵列照相机, 具有快速扫描和高灵敏度的特点。这种近红外成像仪具有传统近红外光度仪的全部优点, 可用来非破坏地监控反应进行, 在甘氨酸、丙氨酸、缬氨酸合成多肽反应中鉴别产物。反应可以通过监控氨基产生的 1529nm 峰的减弱和酰胺键产生的 1483nm 处峰的加强来进行。比传统 NIR 光度仪优越的是这种成像仪可以测定同一样品不同位置的光谱, 从而可以同时监控三种不同的固相肽合成。

19.2.3 其他应用

Menno J. B. 等运用近红外荧光检测器成功地完成了六种血管紧张素 I 多肽的毛细管电泳分离。Clesberg J. T. 等运用近红外光谱法测定了蛋白质晶体生长过程中蛋白质的浓度。

封闭的细胞系统与典型的悬浮细胞培养方式相比有很多优点, 其中之一是可获得较高的细胞密度, 而这种给细胞提供营养的方式经常成为限制产量的因素。Frazier BL. 等^[65]应用近红外方法研究了在固定细胞材料中测定营养物质的有效扩散率的方法。近红外光谱被用来在一个特殊设计的扩散腔中从 10 个空间位置和多个时间点非破坏的测定样品中营养物质的含量。作者用这种方法测定了不含细胞的 3% (w/w) 琼脂凝胶中谷酰胺的有效扩散率 ($D_{\text{eff}} = 6.46 \times 10^{-10} \text{ m}^2/\text{s}$), 与理论值一致。

19.3 制药过程控制及实时分析

制药过程控制分析是药物分析的一个重要研究内容。过程分析的特点就是要求分析检测仪器能够在生产操作的现场原位、无损、快速地提供分析对象的定性定量信息。这就要求分析仪器能够快速地对分析测试信号进行响应, 自动处理和存储所测数据和结果, 并且快速地反馈测试结果, 为生产质量控制提供依据。近年来各国药政管理部门对药品生产提出了越来越严格的质量要求, 传统的离线的实验室质控分析已无法适应快速生产过程控制。

近红外光谱分析的最大特点是无需进行样品预处理, 可直接对颗粒状、固体状、糊状、不透明的样品进行分析, 操作简便、快速, 可不破坏样品进行原位测定, 特别是与计算机技术和光导纤维技术结合, 测量信号可以远距离传输和分析。这些特点使得近红外分析技术特别适宜于在线的过程控制和分析。

19.3.1 干燥过程控制

近红外光谱法还被用于冷冻干燥、微波真空干燥、流化床干燥过程等的监控。对干燥

过程的监控主要是对残留水分的检测，以确定是否达到干燥终点。

Morris, K. R. 等进行了流化床干燥过程的监控，结果发现快速干燥过程不仅降低了制粒过程中干燥的时间而且与低温干燥的效果是相同的。Rantanen J. 等运用近红外光谱三个波长点进行了安慰剂制粒过程在线水分测定，1990nm 用于测量水分，1745nm、2145nm 用于进行制粒过程基线矫正。

Leperson S. 等通过测定水分、溶剂和活性成分的含量来测量药物薄膜生产过程中的干燥速率，从而控制整个干燥过程，并用近红外测定了成品中活性成分的分布均匀度和晶型，其对药物贮存和透皮吸收有着重要的影响。

Harris SC. 等运用近红外方法对干燥过程中溶剂蒸发实现了实时监控，并可进一步通过光谱的差异间接测定过程中的温度和压力。

19.3.2 粉末混合过程控制

一般情况下，一种药物制剂中含有五种以上的成分，其中包括主要活性成分（主药）和各种赋形剂。加入赋形剂可以改变药物的硬度、溶出度、外观、口感等，使得药物更为有效，并且更适合临床使用。将主药与赋形剂进行混合是制剂工艺中的基本工序之一。传统的混合均匀性试验方法是将各种成分混合一段时间后，分别取出混合粉末堆的上部、中部和底部具有代表性的样品，送到实验室，用传统分析方法如紫外-可见分光光度法或高效液相色谱法进行测试。如果测试结果不合格，则需继续进行混合。在这个过程中，实际混合操作的时间并不长，而离线分析过程耗时太多。在取样时，样品易黏附在取样器上，从而影响所取样品的代表性。另外，传统的实验方法只对混合物中的主要活性成分进行检测，而假设赋形剂的混合均匀情况与活性成分相一致，这也与实际情况不符。

Berntsson O. 等利用漫反射近红外光谱对乳糖、水杨酸混合物和微晶纤维素、包膜药丸混合物进行了定量测定。Wargo DJ. 等^[73]对粉末混合过程进行了分析，定量地测定了氢氯噻嗪、乳糖、交联羟甲基纤维素、硬脂酸镁混合的均匀度。

Sekulic S. 等报告了用近红外光谱监控粉末混合过程，并研究了几种不同的模型，比较了不同模型间的参数。运用近红外光谱对粉末混合过程进行在线监控，其中活性成分含量为 3.5% (w/w)，微晶纤维素为 62%，磷酸氢钙 31.5%，淀粉羟乙酸钠 2%，硬脂酸镁 1%，并用 DT (Detrending)、SNV (Standard Normal Variates) 和偏最小二乘法对数据进行了处理。结果表明，近红外光谱技术作为一种对药物混合均匀性的“实时”的非侵入式分析方法是可行的、有效的，实现了粉末混合过程的在线控制。

19.3.3 包装过程监控及实时质量监测

Herkert T. 等运用近红外光谱实现了包装线上成品质量的在线控制。其分别对胶囊 A (含 300mg 对乙酰氨基酚和 250 mg 氯唑沙宗) 和胶囊 B (含 500 mg 对乙酰氨基酚和 30 mg 磷酸可待因) 进行了测定，为药物生产提供了很好的非破坏性的质量控制方法。

19.3.4 发酵过程的质量控制

Kazuhiro N. 等利用近红外光谱法对糖脂类发酵过程中赤藓醇脂和豆油进行了测定，为发酵过程中原料和产品的检测提供了一种方法。Forbes RA. 等对莫能星发酵物效力进

行了测定，并利用近红外对其中的脂质体进行了测定。

19.3.5 包衣过程监控

片剂的薄膜包衣常用纤维素多聚物，可以控制释药的速率或改善药物制剂的外观和口感，另外还可以改善光敏性药物的稳定性。近红外光谱可对包衣层成分乙基纤维素或羟丙基纤维素的厚度进行非破坏性测定，且可以继续进行溶出度或生物利用度的测定，以考察与包衣厚度的相关性，并据以进行生产工艺的监控。Kirsch 等发现片剂样品近红外光谱的变化与包衣的厚度之间存在相关性。他们对近红外光谱法在片剂的包衣过程监控中的应用做了进一步的考察。在用包衣材料乙基纤维素或羟丙基纤维素进行包衣的过程中，按一定时间间隔取样，并测定片剂样品在 1200~2400nm 范围内的近红外光谱。为消除样品定位差异引起的光谱基线漂移，采用二阶导数和多元散射校正两种方法对光谱进行处理，然后用主成分分析建立计算包衣厚度的校正模型。由于近红外光谱法具有非破坏性，可以进一步测定样品的溶出度，考察包衣厚度与溶出度的相关性，从而更好的控制包衣制剂的质量。

为了控制药物活性成分的释放，研究人员正在研究一种以包衣技术为核心的制剂新工艺，即在缓释药物片芯外包上一层含有快速释放药物的包衣。这需要一种能够对外层包衣中药物活性成分进行快速的、非破坏性的定量分析方法，对这种高精度要求的包衣过程进行监控。Bunchanan BR. 等对含有活性成分的包衣层进行了定量测定，光谱采集的波长范围是 1100~2500nm。分别用主成分分析和多元散射校正对光谱数据进行了预处理后，用多元线性回归和偏最小二乘法建立了定量校对方程。用近红外光谱法分析所得结果与 HPLC 分析结果一致。这表明能够用近红外光谱法更加快速有效地对新的包衣工艺进行质量评价。

Pierini N. 等将近红外用于丁咯地尔控释片剂包膜层厚度的测定。Dredan J. 等^[81]研究了粒径和包衣程度对近红外光谱的影响，从而可以将近红外用于粉末颗粒大小和包衣层的测定。

Andersson M. 等运用近红外漫反射光谱法定量测定了片剂的两侧包衣质量。并用主成分分析、偏最小二乘法和多元信号校正（Multiplicative Signal Correction, MSC）对数据进行了处理，可以将其用于在线分析控制。Pierini N. 等^[83]运用近红外光谱对缓释片剂包膜的厚度进行了测定，由于缓释片剂活性成分的释放依赖于片剂包膜的厚度，因此可以运用近红外快速、非破坏性对缓释片剂包衣过程进行质量控制。

19.3.6 片剂生产过程监控

Rantanen J. 等完成了流化床制粒过程水分的在线测定，并测定了黏合剂和颗粒大小对测量的影响。Miwa A. 等^[85]对湿法制粒过程中加入水量进行了预测。为了确定加入水的量，其首先对辅料中水分进行了测定，而后再测定制粒过程中加入的水分量。

近红外光谱法被用于药物生产过程的全面质量控制。Plugge W. 以三水化氨苄青霉素为例，利用近红外光谱进行生产过程的质量控制：原料鉴定、晶型鉴定、水含量测定、三水化氨苄青霉素测定、无水氨苄青霉素测定、残留试剂测定、残留有机溶剂测定和残留原料的测定等。

对片剂生产过程进行实时在线的质量控制是非常必要的。Dreassi 等将近红外漫反射分析应用于抗生素头孢呋新酯片的生产过程分析。样品放置在 0.5cm 深的聚乙烯样品池中,用显微玻片覆盖,片剂和片芯在分析前进行了研磨。通过对原始近红外吸收数据或其主成分分析结果的判别分析和聚类分析进行定性,并通过用多元线性回归和偏最小二乘法处理原始吸收数据及导数光谱数据进行定量。考虑到样品(原始颗粒、研磨过的片芯和片剂)的粒径差异较大,将判别分析用于对二阶导数光谱数据的处理,以减少各种头孢呋新酯样品间的光谱差异,并尽可能的减少色散的影响。

制剂生产的每一个步骤,如制粒、压片、包衣等,都会影响最终成品的质量。为了对制药生产的各个中间步骤进行全面的质量控制,有必要分别建立对不同原料(主药、赋形剂)、准备用于压片的颗粒、片芯和成品片剂进行定性分析的校正模型。使用得到的定性模型对片剂中主药含量和含水量正确的定量校正,可以在生产过程的每一步对产品质量进行检查。

Blanco M. 等将近红外光谱法用于药物不同生产过程的监控。运用偏最小二乘法对 1100~2200nm 波长数据进行了分析,实现了对混合药物、囊心物和包衣片等的测定。利用光纤技术实现了在线、快速测定。并对原材料、颗粒分布、紧密度以及制粒等进行了分析,相对标准偏差低于 1%。

Dreassi E. 等对盐酸雷尼替丁片剂的生产过程进行了分析,包括雷尼替丁的鉴定、混合、囊心物、包衣以及其在水溶液中的浓度等各方面的测定。

19.3.7 其他应用

Bauer JF. 等运用近红外对二水红霉素片剂溶解进行了分析,并成功地得出二水化是溶解失败的主要原因,且预测出潮解可以使这种结合作用逆转并提高药物溶解速率。

Morris DL Jr. 等运用近红外光谱进行了 4-甲基盐霉素提取过程的含量测定,并进行了定量控制。

19.4 中药分析

中药材品种繁多,为了保证中药材的质量、用药安全有效,传统的方法是根据中药材的外观性状直接利用感官达到鉴别目的,其对貌似实异的药材难以下结论。近红外光谱的信息源是分子内部原子间振动的倍频与合频,具有信息量丰富、信息强度低、信息分布广的特点。近年来的研究结果表明,从复杂、变动、重叠的近红外光谱中提取弱信息,已经形成了一套完整的理论与技术,几乎可以用于所有与含氢基团有关的样品的化学性质与物理性质分析,又由于吸收强度低,近红外光谱分析不需对样品做任何化学处理,即可取得样品深处的物质信息,是一种较为理想的复杂样品分析方法。

19.4.1 中药分类

伞形科植物具有一定的药用、经济价值,我国伞形科阿魏亚族(Ferulineae Drude)共 8 属 70 余种,是一个自然植物类群,其中川明参、伊犁芹、簇花芹和球根阿魏 4 属为单种属;弓翅芹、胀果芹属为少种属;阿魏属和前胡属为多种属。属间界线比较模糊,不

同作者在划分时常有交叉现象，形态分类与化学分类常存在不能完全对应的现象。另外，该亚族有许多重要的药用植物，如阿魏（*Ferulassp*）、前胡（*Peucedanum ssp.*）、川明参（*Chuanmingshen sp.*）等。因此，研究该亚族的分类，不仅能作为经典分类方法的补充和改进，而且对有效利用该类群药用植物资源有重要意义。李睿等^[17]探讨了近红外光谱分析技术在中国伞形科阿魏亚族植物分类中的应用。文中选用中国阿魏亚族的 14 种植物作为研究对象，植物样本的名称和编号见表 19-2。

表 19-2 供试阿魏亚族植物种类和产地		
编号 No.	种名 Species	产地 Locality
1	弓翅芹 <i>A rcuatop terus filipedicellus</i> Sheh et Shan	云南宾川
2	新疆阿魏 <i>Ferula sinkiangensis</i> K. M. Shen	新疆
3	阜康阿魏 <i>F. fukanensis</i> K. M. Shen	新疆阜康
4	多石阿魏 <i>F. lapidosa</i> Korov.	新疆肖尔布拉克
5	大果阿魏 <i>F. lehmannii</i> Boiss.	新疆玛纳斯
6	球根阿魏 <i>Schumannia turcomanica</i> Kuntze	新疆霍城
7	簇花芹 <i>Soranthus meyeri</i> L edeb.	新疆玛纳斯
8	胀果芹 <i>Phlojodicarpus sibiricus</i> (Steph. Ex Spreng.)K. -Pol.	内蒙古满洲里
9	台湾前胡 <i>Peucedanum formosanum</i> Hayata	广西容县天堂
10	南岭前胡 <i>P. longshengense</i> Shan et Sheh	广西全州
11	广西前胡 <i>P. guangxiense</i> Shan et Sheh	广西武鸣
12	天竺山前胡 <i>P. ampliatum</i> K. T. Fu	陕西
13	川明参 <i>Chuanmingshen violaceum</i> Sheh et Shan	湖北当阳
14	伊犁芹 <i>Talassia transiliensis</i> (Herd.)Korov	新疆

作者使用样本近红外光谱之间的夹角正切值作为聚类统计量，用最短距离法对以上 14 种阿魏亚族植物样本的近红外光谱数据进行聚类分析，结果见图 19-1。近红外光谱分类结果与传统植物学分类相比基本一致。其中，8 胀果芹与 9 台湾前胡近红外光谱特征极为相似，历史上也曾有作者将其置于前胡属，今近红外光谱分析的结果证实了此两者之间的确存在着较近的亲缘关系。

中药蛇床子为伞形科蛇床属植物蛇床（*Cnidium monnieri* (L.) Cusson）的果实，具有散寒、祛风、燥湿、杀虫、止痒和壮阳等功效，全国各地都有生长，其中江苏、安徽等地数量较多。不同居群的蛇床子在主要化学成分及药用价值上存在一定的差异。蛇床子的化学成分比较复杂，不同种属和产地的蛇床子样本中主要化合物的种类及含量会有差别，因而其近红外光谱也不尽相同。刘国林等^[18]

对 42 种不同产地的蛇床子样品进行了近红外光谱分类研究，将样品不经预处理直接置于载样器中，用近红外光谱仪扫描 18 次，以其平均值作为样品近红外信号数据。取各样品

在 $4400\sim 4800\text{cm}^{-1}$ 区间 100 个波长点处的近红外吸收值为变量，用非线性映射法进行处理，得到各样本在二维平面上直观的分布图。结果见图 19-2。

从各样品在非线性映射图上的分布情况可以看出，属于不同种的样本（样本 20 和 36）在图上远离其他的样本，表明不同种蛇床子在近红外光谱上存在重大差异。同时还可以看到，采集于长江以南的几个样本（样本 1，14，16，19，39，40）较为集中地分布于图的左边，显示出不同地域的蛇床子在近红外光谱上也存在一定的差异性，这也为将近红外光谱技术进一步应用于中药产地的识别提供了可能。

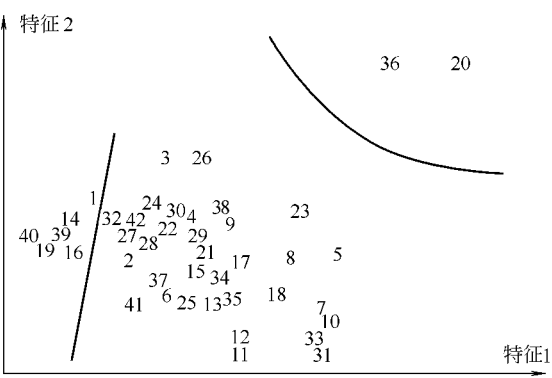


图 19-2 各实验样品在非线性映射图上的分布情况

前胡族是伞形科又一类重要的类群，根据 Drude 系统，前胡族分为三个亚族^[18]，但是族间族内分类界限存在争议，吴拥军等^[19]选用中国 21 种分属三亚族的有代表性植物进行分析。按一定次序对样本进行排序，将植物叶用研钵碾成粉末并过 40 目筛，装入载样器，用不锈钢铲将其抹平，调节光纤升降杆，使其与样本表面保持一定距离，开启恒流泵，带动载样器按一定转速旋转，样本用近红外光谱仪扫描 18 次，并对其信号进行平均，得到样本的近红外光谱数据。用欧氏距离和类平均法对数据进行处理，得到聚类图（图 19-3）。结果表明，本方法应用于前胡族中植物分类与传统植物学分类基本一致。

白芷是一类重要的常用中药，历代本草均有记载。目前，商品药材为栽培品，分为川（杭）白芷和祁（禹）白芷两大类，每年总产量达 2 万吨以上，在中药材市场上占重要地位。对药材白芷的原植物种类存在着多种不同意见，一些中药学和植物学的权威著作的记载也不一致。吴拥军等^[20]采用近红外光谱技术对白芷类中药进行分析，通过非侵入方式获得植物的内在成分信息，结合模式识别技术对其进行分类。作者选用 8 种白芷及其近缘植物的叶进行分析，扫描波长 $3999.778\sim 11999.335\text{cm}^{-1}$ 测定样本的近红外光谱。图 19-4 所示为 8 种样本的近红外光谱。

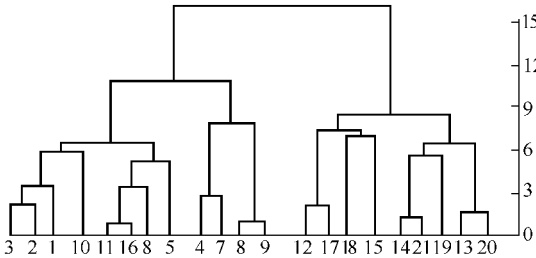


图 19-3 系统聚类分析的谱系

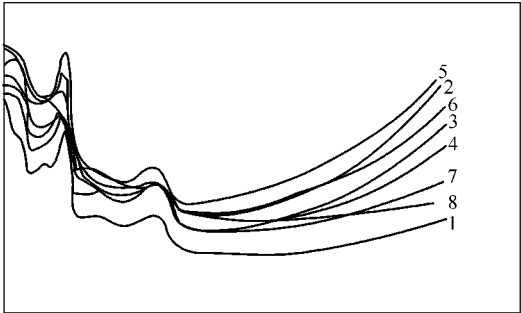


图 19-4 近红外漫反射光谱

试验中以 3.885cm^{-1} 间隔采集光谱数据每个样本选取 2075 个波长点的吸光度数据进行非线性映射，将上述吸光度数据，用欧式距离和类平均法计算得谱系图，如图 19-5、图 19-6 所示。

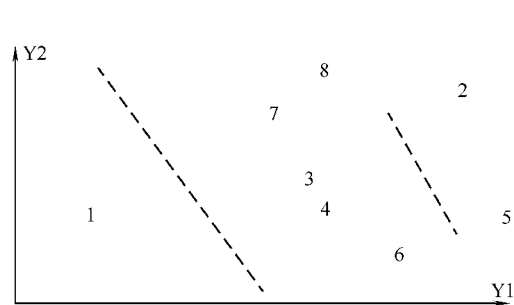


图 19-5 非线性映射

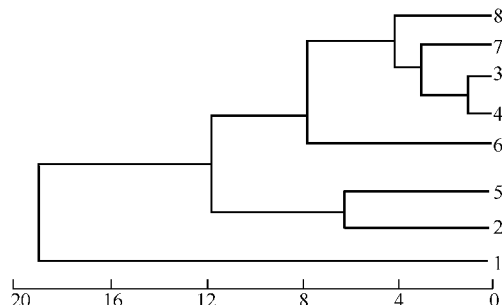


图 19-6 系统聚类分析的谱系图

非线性映射及系统聚类的结果显示，这两种不同的分类方法获得了相同的分类结果，表明其结果是比较可靠的。

羊蹄为民间常用草药，源于蓼科酸模属多种植物的根。其中羊蹄 (*Rumex japonicus* Hoult)、巴天酸模 (*R. patientia* L.)、红丝酸模 (*R. chalepensis* Mill. 金不换)、齿果酸模 (*R. dentatus* L.) 都作为“羊蹄”在各地使用，称为“羊蹄类生药”，有清热解毒、止血凉血、通便利尿、消痈肿、杀虫之功效。临床主要用于治疗各种出血症、炎症及真菌性皮肤病等。刘荔荔等^[21]采用近红外漫反射光谱结合聚类分析法和判别分析法对蓼科酸模属的五种药用植物及三个产地的羊蹄进行鉴别研究，结果与形态分析基本一致。作者将采自南京、余山、上海第二军医大学药圃的多棵羊蹄与采自长春(巴天酸模)、余山(酸模、齿果酸模)、上海第二军医大学药圃(红丝酸模)的多棵属内植物聚类进行分析，结果较为满意。

19.4.2 中药成分定性定量分析

洋地黄中的洋地黄糖甙广泛地用于治疗各种心脏疾病。Kudo M. 等研究了用近红外光谱快速鉴别洋地黄糖甙的方法。与传统的破坏性、耗时的鉴别方法相比，这种鉴别技术具有很多优点。它不仅可以分辨来自不同植物的提取物，还可以鉴别出同一植物中不同部位的提取物。这种方法还可以应用于许多其他天然药物。Molt K. 等用漫反射近红外光谱直接测定了番泻叶中的番泻叶甙含量。作者以 HPLC 作为参考方法，建立了近红外分析的校正模型。

何淑华^[22]等采用近红外光谱法对吉林人参进行定性分类，结果令人满意。作者对吉林 11 个产地的红参、生晒参和西洋参进行了测定。近红外漫反射光谱测试条件：狭缝宽度 20nm，扫描范围 1000~2500nm，以硫酸钡为空白，自动采集光谱数据。一阶导数光谱测试条件： $\Delta\lambda$ 设定为 10nm，其他条件同上。红参、生晒参和西洋参的近红外漫反射光谱重叠严重，而一阶导数光谱特征非常明显。但是三者谱图仍很相似，故采用多变量统计分析技术进行数据处理，结果均准确进入与其相对应的训练集样品分类谱系图的相关区域。

刘国林等^[23]将近红外光谱分析技术应用于中药的定量分析，为元胡止痛散建立一种快速有效的定量分析方法。元胡止痛散是临床较为常用的中药，具有理气、活血、止痛等功效，临床上常用于治疗气滞血瘀的胃痛、肋痛、头痛及月经痛等。对元胡止痛散的鉴别分析一般仅限于定性鉴别，目前尚未见有对其进行定量分析的法定方法或相关报道。本研

究将近红外光谱分析与 BP 神经网络或偏最小二乘法相结合, 对元胡止痛散模拟样本及实际样本中的两种组分进行了定量, 得到了令人满意的结果。作者采用 BRUKERVECTOR22 型近红外光谱仪 (德国, BRUKER), 分支光纤 (德国, BRUKER), 参照处方, 配制不同含量的 25 个元胡止痛散模拟样本, 与三批实际样品一起作为实验对象, 组成各模拟样品。

作者在 $4000 \sim 8000\text{cm}^{-1}$ 波数范围内的 200 个波长点处采集元胡、白芷、实际样品及 25 个模拟样本的近红外光谱数据, 元胡和白芷的近红外光谱图见图 19-7。从 25 个模拟样本中随机挑出 7 个样本组成预示样本集, 其余的 18 个样本作为训练集, 分别采用偏最小二乘法及 BP 神经网络法对样本数据进行处理。其中在 BP 神经网络法中, 为了减少

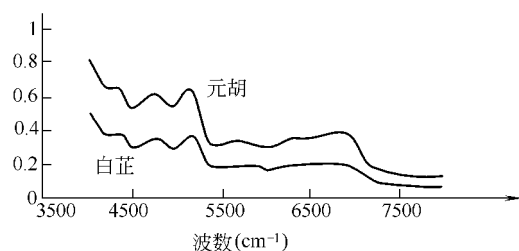


图 19-7 元胡和白芷的近红外光谱图

网络的运算量, 先采用主成分分析法将样本变量压缩为 10 个, 网络的隐含节点数选为 7 个, 学习速率为 1.2, 训练的终止条件为平均训练误差小于 0.001, 经过 9295 次训练, 网络收敛。精密度试验: 取某一预示样本 (样本 7), 重复 7 次测量其近红外光谱数据, 并分别将这 7 次所测量的数据代入由 BP 网络及偏最小二乘法建立起来的校正模型, 计算其中白芷和元胡的含量, 以考察方法的精密度, RSD 均小于 4%, 说明方法的精密度较好。将 7 个预示样本的近红外光谱数据代入由 BP 网络及偏最小二乘法建立好的校正模型中, 计算各预示样本中白芷与元胡的含量, 并与实际加入的白芷及元胡含量比较, 计算回收率。两种方法计算的回收率为: 元胡, $(101.1 \pm 1.4)\%$, $(100.2 \pm 3.1)\%$; 白芷, $(98.0 \pm 3.0)\%$, $(100.8 \pm 5.2)\%$ 。

BP 神经网络及偏最小二乘法对模拟样本及实际样本的预示结果都较好, 对于模拟样本, 平均相对偏差都小于 5%, 对于实际样本, 标示量的百分含量都在 95%~105% 之间, 说明将近红外光谱分析技术应用于元胡止痛散的定量分析是成功的。

中药散剂组成复杂, 若采用常规的方法对其中的主要中药组分进行定量, 必须先对其进行繁琐的预处理, 然后通过测定特定组分的量来确定其中主要中药成分的含量。而近红外不需进行复杂的预处理, 可直接对散剂进行测定, 通过适量的校正样本建立起数学模型之后, 可准确快速地确定未知样本的含量, 尤其适合于对大量重复性样品的快速分析, 为近红外光谱技术应用于中药的定量分析及中药生产中的在线质量控制提供了可能。

19.4.3 中药材产地鉴别

Corti P. 等应用近红外反射光谱建立了一种快速简便地鉴别人参产地的方法。Van Der Vlies C. 等用近红外指纹图谱鉴定了不同来源的药物原料。指纹图谱可区分出不同产地药物在物理或化学性质上的细微差别, 是一种鉴别药物来源的理想工具。Woo Y. A. 等用近红外光谱和模式识别技术区分出了不同产地的药材。中药材在韩国、日本、中国等亚洲国家的临床治疗中具有很重要的地位。而中药材即使是同一物种, 也会因生长环境、气候、土地等的不同在质量上有很大的差异。近红外反射光谱结合模式识别技术可用来鉴别不同产地的药材。作者用这种技术鉴别了紫云英 (Astragali Ra-

dix, AR)、灵芝 (Ganoderma) 等药材。结果表明, 不同产地的药材的二阶导数光谱、主成分数等都有很大不同, 可以用来鉴别药材产地。用相似的方法作者还鉴别出了不同产地的人参。

刘荔荔等^[24]利用近红外光谱技术对不同产地的丹参和不同时间的大青叶等进行了鉴别, 结果较好。Guixing R. 等^[110]运用近红外光谱法完成了美国人参根部总人参皂苷和主要成分的测定。

19.5 药品的无损分析

用近红外反射光谱法解决药品的无损分析, 从而实现药品的质量控制, 目前在国际上正成为热门课题。随着计算机各类软件的开发, 多变量统计分类技术的发展, 信噪比已达 10^4 的近红外光谱仪的出现, 使对固态样品进行非破坏性无损分析得以实现。

第一篇将近红外光谱应用于完整药物制剂分析的报道出现在 20 世纪 80 年代初期和中期美国发生的混有氰化物的胶囊致死事件之后。在这些中毒事件发生后, 美国食品药品监督管理局使用许多种方法分析了 200 万粒胶囊。1987 年, Lodder RA. 等发表了一篇具有里程碑意义的论文, 提出了基于近红外反射光谱分析检测完整胶囊的方法。为达到快速、简便测定胶囊的目的, 设计了一种特殊的反射器, 胶囊剂无需打开, 即可恰好放入这一反射器, 通过胶囊壁直接对胶囊剂进行测定, 采集了 10~13 个未掺假胶囊在 18 个波长点处的光谱, 用作训练集, 并用相同数量的未掺假胶囊对模型进行验证。将氰化钾、氰化钠、铝屑、三氧化二砷以及其他污染物混入乙酰氨基酚胶囊中, 对模型进行进一步的测试。它们用 BEAST 法对所得的近红外发射光谱数据做处理, 建立了测定氰化钾胶囊中含量的校正模型。对于平均质量为 670mg 的乙酰氨基酚胶囊, 氰化钾的检测限为 2.6mg (小于胶囊质量的 0.4%)。如果加入 9mg 氰化钾到胶囊中, 这粒胶囊即被作为不合格品识别出来。

Wang 和 DeJesus 在第 9 届近红外光谱会议上的报告表明, 在使用近红外光谱法分析包衣制剂时, 无需在分析前除去包衣。虽然近红外漫反射光谱已经成为制药工业中快速可靠的 QA/QC 工具, 但是一般的漫反射样品技术不适用于分析覆盖有较厚包衣材料的片剂或胶囊, 而近红外光谱在短波长区 (800~1400nm) 具有低的吸收系数, 能够在几秒内获得药物片剂或胶囊的高质量的发射光谱。因此在这项研究中, 对于有厚胶囊壁的善胃得 (雷尼替丁) 胶囊, 作者考察了近红外发射光谱在分析胶囊囊心的活性成分中的应用, 同时测定了漫反射光谱作为对照。使用 Bruker IFS 28/N 傅里叶变换近红外光谱仪测定 68 粒善胃得胶囊的发射光谱和漫反射光谱, 善胃得胶囊的厚度约为 5mm, 其中胶囊壁的厚度约为 0.5mm, 每粒胶囊的平均总质量为 600mg, 善胃得胶囊中的活性成分的质量在 115.1~162.9mg 之间, 用偏最小二乘法处理漫反射光谱数据, 无法得到胶囊的漫反射光谱与囊心活性成分浓度的相关性。因此, 通过测量近红外发射光谱来进行测定, 使用偏最小二乘法对所得发射光谱数据进行分析, 其中, 浓度在 115.1~162.9mg 的 42 粒胶囊的光谱作为校正集, 另 26 粒浓度在 128.9~153.7mg 之间的胶囊作为独立的验证集。同时进行了交互验证和外部验证, 交互验证的平均根方误是 2.8%, 预示的平均根方误是 2.1%, 使用 HPLC 分析的精度约为 2%。发射光谱的预示精度与 HPLC 相当, 且能为制药企业所接受。因此, 近红外发射光谱分析可作为对药物片剂和胶囊剂的片心和囊心活性

成分分析的又一有效手段。

Dempster MA. 等开发了一种非侵入式近红外反射分析法，用于识别成分相同，但包衣材料不同的片剂。作者采用光纤传感器穿过包装材料直接对样品进行测定，获得了薄膜包衣和非薄膜包衣水泡眼片剂的近红外光谱，并用三种模式识别方法——波长距离法、SIMCA 残差变量法及马氏距离法对所获得的近红外二阶导数光谱进行处理，成功地鉴别出了片剂的类别。

Lodder RA. 等还研究了近红外光谱对片剂的无损分析。他们对载样装置做了改进，在原有的主反射直角锥体的顶部加上一较小的直角锥体，方向与之相对。插入的小锥体能够使光线照到片剂背向光源的一面。使用这种改进的装置在 18 个波长点处对两个厂家的市售阿司匹林片剂进行分析，首先对得到的光谱数据进行主成分分析，然后用 BEAST 判别分析法做进一步的处理。结果表明，使用改进过的载样器所得光谱的分类比将片剂粉碎后置于传统载样器中测定所得的结果好得多。为进一步检验双反射锥形载样器的用途，在片剂的一侧挖了一个孔，装入 10mg 氯化钾粉末，当样品置于单反射锥体样品池中进行扫描时，样品被分到了未掺假的一类。而当使用双反射锥形载样器时，该样品被分到未掺假的一类之外，这显示了双反射锥形装置的优越性。

任玉林等^[25]将偏最小二乘法同近红外漫反射光谱法结合，非破坏地分析了粉末状安体舒通，讨论波长对偏最小二乘法定量预测能力的影响。校正样品和预测样品预测结果的相对标准偏差分别为 0.3202% 和 1.3096%，结果令人满意。

Brashear RL. 等用近红外光谱测定了用作植入物的聚合物基质中的洛美沙星和聚乙二醇含量，通过对二阶导数光谱进行线性回归和偏最小二乘回归得到洛美沙星和聚乙二醇的校正模型。另外洛美沙星通过可靠的方法提取后用紫外测定，对二阶导数光谱的特定波长点处进行线性回归得到洛美沙星和聚乙二醇的含量。实验表明，两者的含量与理论值和紫外测定值非常一致。

Ulmschneider M. 等应用近红外光谱和相应的化学计量学建模方法测定五种密闭玻璃瓶中的药物：无菌头孢曲松二钠盐，沙奎那韦甲磺酸盐，沙奎那韦碱，那非那韦甲磺酸盐和神经酰胺酶抑制剂。这种方法专性强，快速准确，不破坏样品，也不需要预处理样品或清洁操作。

Corti P. 等研究了用近红外技术快速准确分析凝胶和粉末剂型中的酮洛芬的方法。Gold T. B. 等^[118]研究了用近红外光谱测定凝胶胶囊中甲醛含量的方法，应用了主成分分析、多元线性回归等化学计量学方法。

Broad NW. 等研究了用近红外技术直接无损测定琥珀色塑料瓶中乙醇、丙二醇和水的含量。作者用偏最小二乘法和多元线性回归建立了相应模型。Scheiwe M. W. 等^[120]研究了用近红外技术检测二氯酚酸包衣片的方法。使用近红外技术可以实现统一的实时在线监控产品质量，可以全自动的进行质量控制。

Cho C. H. 等用近红外技术测定了制剂中的氨基乙磺酸的含量。作者用二阶导数光谱结合偏最小二乘法建立模型，其相对标准预测误差不超过 1.04%。Robert A. F. 等运用近红外反射光谱测定了劳拉卡比（loracarbef）中残留异丙醇的含量。

Merckle P. 等用近红外研究了三种不同的乙酰水杨酸泡腾片。每种泡腾片都含有 7 种辅料，分别用漫反射和透射方法获得光谱，以 HPLC 为参考方法。与 HPLC 相比，近

红外简单快速，不需样品制备，也不需要使用对环境有害的试剂。

周学秋等^[26]利用傅里叶变换近红外光导纤维探头对塑料包装中的罗红霉素药片有效成分进行非接触、非破坏和无污染分析，取得了令人满意的效果。用近红外光纤探头隔着塑料包装直接测定罗红霉素，测试条件：分辨率 8cm^{-1} ，光谱范围： $4000\sim12000\text{cm}^{-1}$ 。对 16 个样品进行内部交叉验证，根据软件推荐的最佳主成分的维数，计算各样品近红外交叉验证的预测值与高效液相色谱分析数据的相关系数和均方差，结果满意。

19.6 其他应用

Coffey C. 等用化学计量学方法和配备光纤探头的近红外光谱实时的监控化学反应进行。在获得样品光谱数据后立即进行 HPLC 分析，将峰面积与光谱数据比较以建立校正集，用偏最小二乘法建立模型，并用于一个 30g 规模的反应，这个模型成功的预测出了产物的合成过程。这种方法的优点是可同时监控反应中的产物和副产物，可直接从光谱数据中计算出产物含量，相关系数超过 0.9，只需用一个因子。

Ogasawara T. 等用时分辨近红外光谱 (Time-Resolved Near IR) 和二维近红外相关谱 (2D Near-IR Correlation Spectra) 研究了由水分子引起的 3-氨基丙基三乙氧基硅烷聚合过程。在对时分辨近红外光谱去卷积后，组分在 5189cm^{-1} ， 5265cm^{-1} ， 5300cm^{-1} 的吸收带在光谱中 $5000\sim5400\text{cm}^{-1}$ 的区域得到证实，其强度随着反应的进行显著降低。在 5189cm^{-1} 的吸收带是由水分子引起的， 5265cm^{-1} 和 5300cm^{-1} 的吸收带分别是最强和最弱的硅醇基引起的。在二维近红外相关谱中这种聚合过程也找到了明显的证据。

Howard WW. 等用近红外方法在线测定原料药物氢化反应终点。该反应使用 5% 的钨-碳催化剂，以四氢呋喃为溶剂，传统的监控终点的方法是气相色谱法，从取样开始大约需要 60 分钟才能得到结果。使用在线近红外设备，可以每 2min 进行一次取样，从而极大的缩短了分析时间，作者用偏最小二乘法建模成功地确定了反应终点，防止了由于超过最佳反应终点而产生不希望的副产品。

19.7 小结

近年来，近红外光谱在药学领域中的应用取得了很大的进步。就在近二十年前，近红外光谱法与其他分析方法相比还没有什么优越性可言，在进行药物含量分析前，仍须用有机溶剂提取。随着仪器、软件以及样品处理技术的发展，使用近红外光谱技术对药物制剂进行快速的非破坏性分析已成为可能。制药工业企业也已开始发展近红外方法对药物生产过程的各个环节进行监控。有了近红外光谱，制药工业已向“零缺陷”质量控制目标迈进了一大步。

近红外光谱技术已得到药品管理部门的重视。美国 FDA 已经开始认识到近红外光谱的价值，并且已经批准将近红外方法取代传统方法作为氨苄青霉素三水合物的含水量测定、鉴别和含量测定方法。目前，FDA 正与制药企业合作，进一步开发其他的近红外方法。加拿大的药政管理机构卫生部 (Health Protection Branch, HPB) 也已批准将近红外光谱法用于原料和包装材料的鉴别。由此可见，近红外光谱技术受到了越来越多的关注。随着近红外光谱分析技术的不断提高和化学计量学的发展，近红外光谱技术在药学领域中

的应用将越来越广泛。

参 考 文 献

- [1] 安登魁等. 现代药物分析选论药物分析, 1999, 3: 103~108
- [2] 屈凌波, 陈国广, 盛龙生等. 近红外光谱技术及其在药物分析中的应用. 药物分析杂志, 1996, 3: 204~206
- [3] John DK, et al. Appl spectrosc Rev, 1995, 30 (3): 139~174
- [4] Merckle P, Kovar K A. Assay of effervescent tablets by near-infrared spectroscopy in transmittance and reflectance mode: Acetylsalicylic acid in mono and combination formulations. Journal of Pharmaceutical and Biomedical Analysis, 1998, 17 (3): 365~374
- [5] 任玉林, 李伟, 郭晔等. PLS-近红外漫反射光谱法分析扑热息痛. Journal of Mathematical Medicine, 2000, 13 (5): 452~453
- [6] 任玉林, 李伟, 任瑞雪等. PLS-近红外漫反射光谱法分析磺胺甲基异恶唑. 药物分析杂志, 1998, 18 (1): 30~32
- [7] 屈凌波, 相秉仁, 陈国广等. 近红外反射光谱分析在抗生素片剂质量控制中的应用. 第8届中国化学会计算机化学分会会议论文集, 黄山, 2001
- [8] 任玉林, 张滨, 郭晔等. 用主成分回归进行生理盐水的非破坏分析. 数理医药学杂志, 1998, 11 (2): 165~167
- [9] 任玉林, 李伟, 任瑞雪等. 多变量统计分类技术进行甲硝唑质量控制的研究. 数理医药学杂志, 1998, 11 (1): 43~45
- [10] 任玉林, 李伟, 任瑞雪等. 应用多变量统计分类方法对磺胺脒的质量分类. 中国卫生统计, 1998, 15 (2): 38~41
- [11] 任瑞雪, 李伟, 任玉林等. 多变量统计分类技术进行芦丁质量控制的研究. 数理医药学杂志, 2000, 13 (1): 75~77
- [12] 任玉林, 李伟, 郭晔等. 多变量统计分类技术进行甲氧苄胺嘧啶质量控制的研究. 数理医药学杂志, 1999, 12 (3): 269~271
- [13] 屈凌波, 相秉仁, 安登魁等. 近红外光谱在制药过程控制分析中的应用. 中国医药工业杂志, 1996, 27 (11): 526~528
- [14] 严衍禄, 赵龙莲, 李会军. 近红外光谱分析的信息处理技术. 傅里叶红外光谱仪技术及应用论文集, 武汉, 2000. 1~6
- [15] 单人骅, 余孟兰. 中国植物志 (第55卷第3分册). 北京: 科学出版社, 1992, 55 (3): 80~178
- [16] 谢宗万. 中药材品种论述. 上海: 上海科学技术出版社, 1990
- [17] 李睿, 李伟, 白云飞等. 近红外光谱分析在中国伞形科阿魏亚族植物分类中的应用. 西北植物学报, 2000, 20 (4): 666~670
- [18] 刘国林, 蔡金娜, 李伟等. 近红外光谱技术在中药蛇床子分类中的应用. 计算机与应用化学, 2000, 17 (2): 109~110
- [19] 吴拥军, 李伟, 相秉仁等. 光纤近红外漫反射光谱技术在前胡族植物分类中的应用探讨. 计算机与应用化学, 2000, 17 (2): 111~112
- [20] 吴拥军, 李伟, 相秉仁等. 近红外光谱技术用于白芷类中药的鉴定研究. 中药材, 2001, 24 (1): 26~28
- [21] 刘荔荔, 原源, 陈万生等. 近红外漫反射光谱法在羊蹄类生药分类中的应用. 中草药, 2001, 32

(11): 1024~1026

- [22] 何淑华, 孙瑞岩, 任玉秋等. 近红外漫反射光谱法对吉林人参的分类探讨. 吉林大学自然科学学报, 2001, 1: 96~98
- [23] 刘国林, 陈国广, 相秉仁等. 近红外光谱技术在元胡止痛散定量分析中的初步应用研究. 中国现代应用药学杂志, 2000, 17 (5): 383~385
- [24] 刘荔荔, 吴玉田. 中药材的近红外漫反射光谱峰位鉴别法. 光谱仪器与分析, 2001, 2: 23~29
- [25] 任玉林, 李伟, 郭晔等. PLS-近红外漫反射光谱法分析安体舒通. 数理医药学杂志, 1998, 11 (3): 273~275
- [26] 周学秋. 用傅里叶变换近红外光导纤维探头对包装中的罗红霉素有效成分进行非接触式定量分析. 布鲁克公司

20. 近红外光谱分析在烟草与纺织行业中的应用

20.1 近红外光谱分析在烟草行业中的应用

国际上 NIR 技术在烟草行业的应用研究,从 20 世纪 70 年代开始出现,20 世纪 80 年代研究报道不断增加,特别是进入 20 世纪 90 年代后,随着 NIR 软硬件技术的发展,研究对象更是不断扩展。目前 NIR 技术在国际烟草界及国内外的各行各业得到了日益广泛的应用,但在国内烟草行业的研究及应用还处于起步阶段。本章主要针对国际上 NIR 烟草的研究及应用情况,并结合我们自己的一些研究进展,对烟草 NIR 的研究、应用及发展展望做一些介绍。

20.1.1 NIR 定量分析的原理及对烟草的适用性

20.1.1.1 NIR 定量分析的原理及特点

NIR 定量分析,是利用物质中 C—H、N—H、O—H 及 C=O 等基团对近红外光谱吸收较强的特点,根据物质 NIR 光谱吸收信息对相关特征(如物理、化学、生物学性质)进行定量测量的技术。烟草中糖、氮、尼古丁、焦油及其他一些质量特征,与以上基团密切相关,因此可能使用 NIR 对这些特征进行定量测量。

NIR 定量分析具有以下特点:①测试速度快,比较适合于大批量测试;②测试过程及操作简单;③测试过程中无化学反应,不消耗化学试剂是无污染测试;④NIR 是无损检测,样品可以重复使用;⑤可以使用于生产线等在线检测。

20.1.1.2 NIR 定量分析的流程与步骤

NIR 定量分析的流程分为两步:①建立数学模型;②应用模型定量分析。

建立数学模型:使用 NIR 分析之前,首先必须收集一批具有代表性、含量或性质已知的标准样品,测定其光谱,选择性地使用全谱区中的有关光谱信息,使用化学计量学方法,建立光谱与特征间的数学关系,经方法筛选最后建立相关的数学模型。

应用模型测量:模型建立后,首先测定待测的 NIR 光谱,使用模型即可定量计算有关特征值。

由于 NIR 定量分析是居于数学关系模型的二次分析,因此建立稳定的数学模型非常关键。

20.1.1.3 NIR 定量分析数学模型的建立

建立数学模型的基本步骤分以下四步:

① **优选校正样品集:**选择有代表性的样品,选择样品时要尽可能地使标样的待测特征分布范围较大,并包含可能的各种典型样品。

② **建立模型:**应用化学计量学法,从图谱中提取对待测特征敏感的波段,建立光谱信息与待测特征间的关系模型,使用严格的统计方法,通过样品集内部交叉证实与外部证

实法校验数学模型。在模型的建立过程中，要全面考虑背景对模型的影响，对可控因素要尽量保持一致，例如样品的粒度、装样条件尽量一致，使所建模型尽量免受背景波动的影响；而对不可控因素，要尽量有意识地控制标样的这些因素水平，因而使模型对这些因素的变化不敏感，例如烟叶的水分如果无法精确控制的话，则在校正样品集中有意识地使样品集的水分在一定范围内变动。

③ 优化模型：反复重复前两步骤，并剔除其中不好的标样，提高模型的预测精度。

④ 模型修正与维护：当样品的测定时间或空间条件发生变化时，必须用新样品检验模型，如果模型预测效果降低，则需往校正样品集中增加新的校正样品，并重新按照上述步骤建立和优化模型。这是一个没有止境、不断进行的动态过程。

为了能建立理想的数学模型，除了要有优秀的算法及有经验的 NIR 专业人员外，特别重要的是选择并建立有代表性的标准样品集（如品种、产地、等级等），建立规范、严格的测试方法，准确测定校正样品集的特征值和 NIR 光谱数据。数学模型要对应一定的样品范围、背景状况、仪器条件、测试方法、光谱预处理方法和化学计量学方法。为保证模型的“匹配性”，数学模型必须注意不断地维护与校正。

NIR 模型的精度一般使用以下参数加以评价：相关系数 R 、模型校验标准误差 SEC 、模型检测标准误差 SEP 及变异系数 CV 。

20.1.1.4 NIR 对烟草的适用性

一方面，烟草作为天然生长的植物，含有大量对 NIR 比较敏感的 C—H、N—H、O—H 键，因此应用 NIR 研究开发烟草的检测技术应该很有潜力；另一方面，烟草中的糖、氮、尼古丁等化学指标是设计配方和监控质量必不可少的因素，通常需花大量的时间、精力及经费进行分析，特别是在视质量为生命的市场竞争中及自动化程度日益提高的今天，更是迫切需要快速、高效、简便的分析方法为烟草工业服务。因此近红外光谱分析技术（NIR）有望且正在发展成为烟草分析的重要技术。另外，由于 NIR 还具有无损和多特征同时分析的特点，使原料和产品的在线检测和实时控制成为可能，特别适合烟草自动化生产和控制的需求。

20.1.1.5 烟草 NIR 检测需注意的因素

根据 NIR 的特点，在建立烟草 NIR 定量分析模型时，必须考虑以下方面的因素。首先，卷烟方面，要考虑可能含有烤烟、香料烟、白肋烟以及再造薄片等多种原料因素；其次，烟叶方面，则存在品种、地理环境、气候、培植、施肥、烘烤等影响因素；第三，即使同一厂家的产品，不同品牌的烟叶配方结构、香精香料、添加剂数量种类各不相同。这些因素的综合影响，给 NIR 的定量分析带来了极大的挑战，但这些问题不是无法克服的。

尽管开发具有一定适应能力和精度的 NIR 模型需花较多时间和资金，但由于具有快速、使用简便等优点，NIR 仍然不失为一种低投入、高产出的分析方法。因此 NIR 对烟草来说具有广阔的应用前景。

下面分几个方面，介绍 NIR 在烟草方面的研究及应用情况。

20.1.2 NIR 在烟草常规分析化学成分检测中的研究及应用

20.1.2.1 烟草生物碱

（1）烟草生物碱的重要性

烟草生物碱是烟草有别于其他植物的关键物质，其中的尼古丁是烟草中含量最高，且最为重要的一种生物碱。烟草尼古丁含量与烟草的种类、品种、产地、培植方法、气候条件、成熟度等密切相关，另外烟草不同组织（根，茎，叶）、烟叶不同生长部位（上，中，下）、烟叶不同组织（叶片，烟梗），甚至一片烟叶的不同区域，尼古丁含量各不相同，且有一定的规律。品种的影响：马来烟和土耳其烟的尼古丁含量较低，烤烟中等，白肋烟较高；成熟度的影响：尼古丁含量随成熟度提高而提高；烟草组织的影响：烟叶尼古丁最高，根次之，茎最低；烟叶生长部位的影响：上部烟叶尼古丁最高，中部次之，下部烟叶尼古丁最低；烟叶组织的影响：叶片尼古丁高，烟梗的尼古丁低；烟叶上区域的影响：叶尖和叶的外缘尼古丁较高，内部区域较低。

虽然烟草含有数千种化学成分，并由它们的综合作用才形成烟草最终的质量特征，如香气、杂气等，但烟草生物碱特别是尼古丁，为烟草中最重要的化学成分，因为它是烟叶质量判别、烟叶选择、产品设计及产品质量控制的主要因素。因此烟草生物碱的检测对烟草工业来说必不可少的。

（2）烟草生物碱的标准测量方法

最早的烟草生物碱测量方法由 Schloesing 于 1846 年发明，目前经典的方法是采用碱性溶液蒸馏萃取，并用分光光度法定量，如我国烟草生物碱测量的标准方法 YC/T 34—1996 即为此法，该方法测得的是生物碱总量。随着技术的进展，目前已逐渐采用流动注射分析法。另外也可采用气相或液相方法测量烟草生物碱中的某些成分，如尼古丁。

需要注意的是，在用不同方法校正 NIR 模型时，NIR 测定结果与建模时所采用标样的结果信息是一致的，如用分光光度法标定 NIR 后测定的结果是总生物碱，用气相法标定 NIR 后测定的结果是尼古丁。

（3）烟草生物碱的 NIR 研究进展

应用 NIR 进行烟草生物碱的研究最早报道为 1978 年的《烟草科学》（Tobacco Science），由 Pandeya 及其合作者完成。研究对象为烤烟，仪器为 6 滤光片 InfraAlyer，滤光片波长为 1680nm、1940nm、2100nm、2180nm、2230nm 和 2310nm，标样尼古丁范围为 0.02%~4.25%，标样数为 75 个。Pandeya 对颗粒度（20 目至 40 目）进行了详细的研究，得出的结论是为稳定测试精度，测试与建模时的粒度应一致。

Pandeya 使用的是滤光片式 NIR 仪器，1978 年 Hamid^[1]等也公开了使用扫描式单色仪进行烟草生物碱的研究结果，标样数为 68 个，粒度为 40 目，生物碱含量范围为 0.78%~6.1%。研究了 3 种数学变换方式： $\lg(1/R)$ 、一阶导数、二阶导数，波段由计算机计算选择。根据 SEP（预测标准误差），二阶导数的结果最优，一阶导数次之， $\lg(1/R)$ 的最差。最后优选了 10 波段二阶导数的模型。

在 NIR 实际测量过程中，经常发现由于样品粒度变化导致检测结果波动，Long 等人研究表明可以在标定时直接加进波动因素，让模型自动补偿粒度变化的影响^[2]。Cooper 研究了年份变化对关系模型的影响及处理办法^[3]，结果表明：使用早年烟叶建立尼古丁关系模型后，直接测试新年份烟叶时，精度会有所降低，但模型中适当加入新年份烟叶标样后，检测精度即可得到保证。Diffie 等人研究了单色扫描仪与滤光片式仪器间的模型转换^[4]，发现通过必要的误差校正，模型转换是可以实现的、精度也可以得到保证。

（4）烟草生物碱的 NIR 测试结果

笔者及合作者正在进行 NIR 检测方面的研究，所使用的仪器为 VERCTOR22/N，烟叶样品的粒度为 80 目，测试对象为复烤烟叶，来自国内多个产地（以下同）。

表 20-1 为文献^[5]与作者对烟草生物碱的检测结果，资料数据来自 Handbook of Near-Infrared Analysis（by Donald A. Burns，1992）。

表 20-1 烟草生物碱的 NIR 检测结果

尼古丁	相关系数 R^2	SEC	SEP	含量范围
资料	0.988	—	0.09%	1.7%~4.5%
张建平（作者）	0.992	0.038%	0.201%	0.91%~4.13%

20.1.2.2 烟草总糖及还原糖

（1）烟草总糖及还原糖的作用

烟草中的糖对烟气质量具有有利作用，它改善烟叶的吸湿性和控制燃烧速度，因此通常在烟叶调制加工过程中加入一定量的糖。烟叶中含有葡萄糖、果糖、蔗糖等天然糖分，不同品种烟叶的糖含量差异较大，烤烟与东方烟叶的糖含量最高，白肋烟的糖含量较低。还原糖的含量范围为 0.2%~25%，总糖的含量范围为 0.2%~35%。

烟草糖类的检测有多种方法，我国烟草行业的标准方法为芒森·沃克法（YC/T 32—1996）。

（2）烟草总糖及还原糖的 NIR 研究进展

McClure 及其合作者于 1977 年首次公开发表了 NIR 测定烟草还原糖的文章^[6]，其仪器为单色扫描仪，波长范围为 800~2600nm，152 个建模样品（其中 129 个烤烟、16 个白肋烟、7 个东方烟），还原糖的范围为 0~30%，样品烘烤到 2%~3% 水分并磨成 40 目，波长范围 1400~2400nm，光谱二阶导数回归分析的结果最好，糖预测模型最后选择的波长为 1458nm、1914nm、1790nm、2028nm、1516nm、1592nm、1476nm 及 2178nm。对 72 个样品的实测结果：SEP 为 1.36%， R^2 为 0.986。作者认为，使用二阶导数比 lg (1/R) 好，另外如果在样品粒度、采样均匀性、装样稳定性及扫描次数方面做进一步改进的话，结果精度将可以得到进一步改善。

同样，在糖的 NIR 建模和检测过程中，为达到稳定的精度，最好使样品的粒度统一。Pandeya 等人的定量研究表明^[7]，如果建模时使用 20 目而测试时使用标样 40 目样品的话，还原糖的 NIR 测试值将比实际可能高出 2%~3%。试验结果表明，测试 300 个样品，NIR 需 1 天，而通常的化学分析方法则需 8 天。

Tonini 等人利用 14 个滤波片的 InfraAlyzer 400，对烤烟和东方烟两种类型的烟叶进行 NIR 建模^[8]，得到的结果分别为，烤烟：SEC 为 0.802%，SEP 为 1.208%；东方烟：SEC 为 0.883%，SEP 为 0.976%。不同种类烟叶分开与混合建模相比，分开建模 NIR 模型的精度会更高些。

（3）烟草总糖及还原糖的 NIR 检测结果

表 20-2 为文献^[5]与笔者对糖的检测结果。

20.1.2.3 总氮分析

（1）烟草总氮的重要性

据 I. Schmeltz 等人 1977 年对烟草和烟气化学成分的综述^[9]，在当时烟草及烟气中检测出的 2200 种化学成分中，有 30% 的为含氮化合物，其中含有许多重要的烟草化学物

表 20-2

糖的 NIR 检测结果

总糖与还原糖		相关系数 R^2	SEC	SEP	含量范围
资料报道		0.99	0.802%	0.976%	0.2%~35.0%
张建平	总糖	0.99	0.457%	1.59%	7.4%~32.18%
	还原糖	0.99	0.561%	1.59%	7.0%~29.61%

质类，如胺类物质、氨基化合物、亚硝酸盐、氨基酸、硝基苯胺、嘧啶、吡嗪、吡咯，由于时间和资金方面的限制，不可能对这些重要含氮化合物进行全数测量，因此通常测量总含氮量，以便进行粗略的估计。

研究表明，不同烟草的总氮量随种类有较大的变化，一般烤烟的总氮最低，白肋烟的总氮最高。

(2) 烟草总氮的标准测量方法

总氮的经典测试方法是 Kjeldahl 法，通过对 Kjeldahl 法进行适当的修改，也可用流动注射分析法进行总氮的测量。

测定总氮消化样品时，烟草中有大量的难溶性化合物。使用 Kjeldahl 法可以消化许多含氮化合物，但有机物中的氮较难转换，特别是尼古丁及其他生物碱。消化溶液的澄清并不表明含氮化合物向氨转换已完成，根据经验，消化溶液的澄清后样品最好继续消化 1~1.5 个 h，以保证氮的完全转换。当为 NIR 提供标样总氮数值时，必须确保所测数据确实是总氮，否则所建立的数学模型预测的并不是总氮。

(3) 烟草总氮的 NIR 测试结果

有关 NIR 烟草总氮测定的专题文章较少，在文献^[5]中公布了 Cooper 等人未公开发表的总氮数据，国内学者王文真等人 1995 年公开了他们 NIR 进行总氮测定的有关研究^[10]。表 20-3 为 Cooper、王文真等及笔者有关总氮的 NIR 数据。

表 20-3

总氮的 NIR 检测结果

总氮	相关系数 R^2	SEC	SEP	含量范围
Copper 等	—	—	0.078%	1.89%~3.47%
王文真等	0.984	0.232%	0.277%	—
张建平	0.982	0.031%	0.418%	1.45%~3.06%

20.1.2.4 水分分析

(1) 烟草水分测量的重要性及其方法

在烟叶的初烤、复烤过程中，水分是最重要的工艺参数，它直接影响烟叶的初加工质量；在贮存过程中，水分的大小将影响烟叶的可存放时间和贮存养护质量；在制丝过程中，水分能否合理控制将影响烟丝的质量；卷接包的水分直接影响烟支的物理结构及成品质量；成品水分既影响产品储放时间，对卷烟产品吸味也有重大影响。

因此对烟叶初加工、卷烟生产、产品质量控制及产品贮存全过程来说，水分是非常重要的参数，必须频繁检测和严密控制。

水分检测的经典方法是烘箱法：即将测试样品在设定的温度下烘烤规定的时间，根据样品的失重计算其水分。更精确的方法则需要使用 Karl Fischer 法，但这两种方法由于需

要较长的时间，一般只能用于实验室，水分的生产线在线检测与控制目前一般采用 NIR 方法。

(2) 烟草水分的 NIR 检测

1961 年 NIR 水分检测技术被首次应用于烟草工业，当时 Crowell 等人应用 NIR 技术测试湿焦油中的水分。他们使用丙酮/CCl₄ 溶液萃取剑桥滤片上的焦油，研究萃取液的 NIR 吸收光谱，发现当水分含量处于 0.2~1.2mg/mL 范围时，水分与 1980nm 处的吸收系数成正比。1983 年 Long 公开其研究成果表明，水分的重要吸收波段为 1940nm、1840nm、2100nm。同年 Tonini 等人使用 InfraAlyzer 应用 NIR 进行水分检测，其结果为 SEC0.639%，SEP0.646%^[5]。

由于水分测试的重要性及在线检测的需要，有许多厂商生产了许多专用的 NIR 水分检测仪，目前这些仪器被非常广泛地应用于烟草工业中。一些新型号的 NIR 水分检测仪正在克服易受颜色等因素影响的缺陷，精度达到 0.1%^[11]。

20.1.2.5 无机元素的分析

(1) 无机元素的重要性及 NIR 检测的理论依据

一些无机元素对烟草的生长发育十分重要，众所周知氮、磷、钾是重要的植物营养成分，但以下元素也被认为是烟草必要的营养元素：硼、钙、铜、铁、锰、镁、钼、硫、锌等。

通常认为无机元素对 NIR 光谱没有明显的吸收，应该无法使用 NIR 检测。但研究发现碳酸根在 NIR 波段的 1900nm、2000nm、2160nm 及 2350nm 有吸收，有人据此建立了相关模型，检测地质样品中的碳酸盐含量，进而发现无机盐与 NIR 光谱间具有衍生的复杂关系，显示这方面的研究具有使用价值。

早期无机元素的 NIR 检测研究成果有限，这主要是受当时的软硬件水平限制而造成。依据到目前为止的研究进展和认识，由于无机元素会形成盐类物质或与有机物形成螯合物，所以有些无机元素是可以进行间接测量的，例如 NaCl 因为影响水对 NIR 的吸收，因而可以使用 NIR 检测。

(2) 无机元素的 NIR 检测结果

1986 年 McClure 等人公布了应用 NIR 进行无机元素检测的研究结果见表 20-4^[12]。

表 20-4 鲜烟叶无机元素 NIR 检测模型参数

无机元素	R ²	SEC	CV/%	范围/%
硼	0.9365mg/kg	1.292	9.0	10~24
铜	0.9153mg/kg	1.369	1.0	10~90
铁	0.9073mg/kg	8.993	14.7	13~74
锰	0.9515mg/kg	6.112	7.8	16~94
钠	0.9842mg/kg	19.78	6.8	96~387
锌	0.9403mg/kg	4.596	7.5	13~74
钙	0.9865%	0.122	4.2	0.57~3.49
氯	0.9801%	0.071	10.7	0.07~0.73
镁	0.9757%	0.034	6.5	0.31~0.83

无机元素	R^2	SEC	CV/%	范围/%
硝酸盐	0.9088%	0.028	9.3	0.03~0.33
磷	0.9454%	0.016	7.6	0.23~0.44
钾	0.9578%	0.157	6.6	1.81~4.20
硫	0.9515%	0.014	6.7	0~0.21

该研究使用光谱的二阶导数建立模型，由于样品有限（25 个），没有实测验证数据，但作者认为当变异系数（CV）<10%的话，所建模型应该是有效的。这个研究表明应用 NIR 检测无机元素是可能的。

20.1.3 NIR 烟气分析研究

20.1.3.1 卷烟烟气测量的重要性及其方法

对消费者来说，比较关注的是卷烟燃烧后烟气的吸食感觉，如烟气的香气、劲头、余味、杂气等，这些只有通过感官才能判别的特征统称为卷烟的感官质量，它们是卷烟消费者真正关心的质量因素，并将影响消费者吸烟的欲望及对品牌的选择。但卷烟这些感官质量的判别，到目前为止，除了专家的经验外尚无真正有效与公认的仪器。

虽然如此，对烟气的检测并不是一无办法，除了烟气的香味特征方面无法用统一的仪器测量外，目前国际上有通用的烟气测量办法，即在统一的环境条件（温度、湿度）和抽吸方式下，检测卷烟燃烧后烟气中的焦油、尼古丁及 CO 的含量，这是因为除了烟气的香型特征外，这些指标能比较有效地反映烟气的浓度、劲头及对人体有害的 CO 含量。近年来，随着吸烟有害健康呼声的提高，认为烟气焦油中含有一定致癌或诱癌的化学成分，因此各国政府已逐渐规定在烟盒标注焦油、尼古丁和 CO 中部分或全部指标。

每支香烟燃烧后所产生烟气中的焦油、尼古丁和 CO，除了受烟支本身的物理、化学性质影响外，也受抽吸的环境条件和抽吸方式的影响。譬如，即使是同一支卷烟，假如抽吸时的温度、湿度不同，抽吸时的力量大小不同，所产生的焦油及 CO 会有很大的变化；另外即使同一配方的卷烟，由于卷制的充填密度不同，会因燃烧条件及烟草含量的变化，单支香烟燃烧后产生的焦油、尼古丁和 CO 也随之变化。同时，这些条件的变化也会导致吸烟者吸烟感觉的变化。

因此，烟气分析的标准方法规定：烟支必须在温度为 $(22\pm1)^{\circ}\text{C}$ 、湿度为 $(60\pm2)\%$ 环境下静态平衡 96h 或动态平衡 48h；抽烟必须使用专用的吸烟机，其吸烟环境为 $(22\pm2)^{\circ}\text{C}$ 、湿度 $(60\pm5)\%$ ；并规定标准的抽吸方式，包括吸烟的时间、抽吸间隔、抽吸压力的曲线等；再根据标准的方法对烟气中的焦油、尼古丁和 CO 进行计算或分析。

因此烟气分析是一项费时、费力、昂贵而乏味的检测工作。但由于烟气的质量数据对烟草的质量判别与控制极其重要，因此对任何卷烟生产商来说，烟气分析是必须重复进行的一项工作。

20.1.3.2 NIR 的烟气测量方法

Luzio^[13]等应用 NIR 方法研究了烟气的检测方法，所使用的仪器为 INFRAMAT-IC8620，为使所建立的模型能适应各种品牌的卷烟，建立模型所用的标准样品为来自世界

各地的 70 多个卷烟产品。建模所用的数据来自标准的烟气测试方法，吸烟机为 FILTRO-NA SM-350，用剑桥滤片收集的主流烟气，湿焦油使用称重法，水分使用 Karl Fischer 法，尼古丁使用流动注射分析仪测试。NIR 测试的是剑桥滤片，为防止剑桥滤片上水分的挥发，滤片密封在一个特殊的密封装置中。NIR 建模时波长的选择由计算机优选，优选结果为：尼古丁—2310、2230、1818、1759、1722nm；水分—2270、1940、1445nm；湿焦油—2345、2336、1940、1818、1734nm。

Luzio 等人的研究由于是针对卷烟产品，由于不同产品的辅助材料及规格各不相同，所以必须使用吸烟机将烟气收集在剑桥滤片上，而 NIR 所检测的是剑桥滤片。假如要测试烟叶原料焦油产生量的话，只要在统一的辅助材料的基础上，可以使用 NIR 直接测试烟叶原料，从而大大降低测试的前处理工作量与难度（无需使用吸烟机），直接得到烟叶的焦油与尼古丁产生量，该方法对新产品开发、按设定焦油目标设计卷烟配方、稳定产品的质量很有意义。Williamson 等人即使用这种方法进行了烟叶焦油产生量的研究^[14]。

20.1.3.3 NIR 烟气测量结果

Luzio 与 Williamson 的 NIR 烟气研究结果见表 20-5。

应用 NIR 模型的烟气测试结果				
研究者	烟气测试值	被测样品数/个	相关系数 R	预测误差 SEP
Luzio	尼古丁	61	0.973	0.076%
Luzio	焦油	178	0.991	0.84%
Luzio	水分	61	0.979	0.12%
Williamson	焦油	100	0.91	3.4%(SEC=1.6%)

20.1.3.4 NIR 与其他方法的比较

经典方法与 NIR 方法的特点比较见表 20-6。

NIR 与经典方法比较						
烟气测试值	测试方法	速度	精确性	操作简便性	预处理	仪器
焦油	称重法	很快	很高	很简便	无	天平
	NIR	中等	中等	很简便	无	NIR 仪
水分	Karl Fischer	慢	高	繁	有	K. F. 滴定仪
	NIR	中等	中等	很简便	无	NIR 仪
尼古丁	蒸馏法	快	高	繁	无	蒸馏及光度仪
	连续流动法	快	高	繁	有	自动分析仪
	气相法	很快	很高	繁	有	气相色谱仪
	NIR	中等	中等	很简便	无	NIR 仪

研究表明：为使模型的适用性加强，必须广泛收集标样；如要提高模型的精度，则最好有针对性地根据待测样品建立模型。

由于 NIR 具有快速、简便、无污染及多特征同时测定的优点，可以大量而快速地分析更多的样品，作为产品质量检测和 control 的有效手段。在烟气分析中使用该方法，可以减

少 80% 日常工作量, 但经典方法还是不能抛弃的, 在 NIR 建模及测试结果有疑问时都需使用经典方法进行核对。

20.1.4 NIR 在烟草其他分析中的研究与应用

前面介绍了常规化学成分及烟气的 NIR 研究情况, 本节介绍 NIR 其他一些研究情况^[8]。

20.1.4.1 其他化学成分的分析

(1) 总挥发碱

Mosely 等人于 1951 年首先提出, 总挥发碱对白肋烟的烟气动头有重要影响, 可以作为烟气的的一个质量指标。随后的研究发现总挥发碱同样可以用于其他烟叶的质量判别。经典的总挥发碱检测需要蒸馏、滴定等一系列复杂的过程。

研究表明测定总挥发碱的 NIR 波段主要为 2230nm、2180 nm、1982nm 及 1445nm, 其中 2230nm 与 2180nm 主要反映尼古丁与蛋白质的信息, 1982nm 与 1445nm 主要反映 N—H 键的信息。T. Bense 等人在 1983 年公布的研究结果 $SEP=0.094\%$ 。

(2) 多酚

在烟草中发现的多酚包括 Chlorogenic Acid、芸香苷、Scoplin、Scopoletin 等, 影响多酚的因素包括烟草品种、烟叶生长的部位、光照条件、温度、成熟度及烘烤技术, 烤烟中含有大量的多酚。多酚对烟气的香气质和香气量有好的作用, 因此一般优质的烟叶含有较多的 Chlorogenic Acid 和芸香苷, 而白肋烟、马来烟及其他凉制烟一般多酚含量较少 ($0\sim0.5\%$)。

烟叶的棕色化反应主要就是多酚在酶的作用下氧化成醌的结果, 在凉制过程中多酚物很容易氧化, 因此可以用来检验烟叶处理过程中的进程。在烤制过程中多酚物中的部分成分不变, 部分会氧化, 因此在烤烟中常含有多酚物及其氧化物。

目前为止虽然有一系列检测多酚中有关物质的方法, 但液相色谱是检验多酚的较佳选择。

20 世纪 70 年代开始有应用 NIR 进行多酚检测的报道, 总的来说这方面的研究报道不多。1980 年 Tobacco Chemistry 报道了 McClure 等人的研究成果, $SEC=0.70\%$, $SEP=1.089\%$ 。

(3) 薄荷分析

在卷烟制品中加入薄荷会给人清凉的感觉, 因此深受消费者的喜欢。薄荷烟的最早专利出现于 1925 年, 现在全球生产量最大的薄荷烟 Kool 始生产于 1933 年。薄荷烟在初期并不被人们重视, 一直到 20 世纪 60、70 年代薄荷烟开始盛行, 1986 年时薄荷烟占世界烟草总产量的 27.7%。

通常薄荷的实验室分析非常繁复, 首先需将薄荷从卷烟中蒸馏萃取, 然后使用气相色谱仪分析。

John T. Diffie 等使用 NIR 进行薄荷的检测研究。他们收集 135 个各种牌号的薄荷烟, 薄荷含量为 $0\sim1\%$, $SEC=0.091\%$, 对 5 个牌号 100 多样品的测试, $SEP=0.093\%$ 。如分牌号建立 NIR 模型的话, 对一些较好的 NIR 模型, SEP 可以达到 $0.06\%\sim0.08\%$ 。

除了以上成分的检测分析外, NIR 可以检测的化学成分很多, 只要这些成分与 C—H、N—H、O—H、C=O 键之间有直接或间接的关系, 另外是含量不能太微。

20.1.4.2 保润剂分析

甘油、乙二醇等有关化合物常被保润剂, 加入到卷烟制品中, 以保持和稳定水分, 以延长产品的储存周期和使用寿命。保润剂的添加量一般在 1.5%~4.5%。实验室分析保润剂的通行办法一般是气相色谱仪。

Diffie 等人首先开始使用 NIR 进行保润剂的检测。他们收集了 52 个牌号的 104 个样品作为建模的标准样品, 甘油的含量为 0.81%~3.80%, 丙烯乙二醇的含量为 0.23%~1.59%。所得到的结果为: $SEC=0.11\%\sim0.15\%$, $SEP=0.14\%\sim0.15\%$ 。

通常卷烟产品中保润剂的含量聚集在一个很小的范围内, 而从建模方面考虑, 一般认为标样含量范围越宽、分布越均匀, 所建的模型稳定性越好, 因此为了得到较好的模型, 必须在实验室人为配置一些样品, 使标样的保润剂含量均匀分布且含量范围较大, 这样可以大大改善模型的精确度。

20.1.4.3 三醋酸甘油酯分析

在滤棒制造过程中, 三醋酸甘油酯通常作为固化剂, 因此三醋酸甘油酯的含量常作为滤棒质量的一个主要指标。目前滤棒中三醋酸甘油酯的分析, 一般是首先将滤棒中的三醋酸甘油酯萃取出来, 然后使用气相色谱仪进行分析。

NIR 同样可应用于三醋酸甘油酯的分析, 而且不需要将滤棒中的三醋酸甘油酯萃取分离出来。应用 NIR 进行三醋酸甘油酯的分析时, 所使用的波段为 2104nm、2282nm 和 1264nm。据 Diffie 提供的数据, $SEC=0.334\%$ 。

另外据介绍, 红外工程公司 (Infrared Engineering) 已有可以在线检测三醋酸甘油酯的 NIR 仪器。

20.1.4.4 卷烟结构分析

(1) 烟梗烟叶比例分析

NIR 一个很有趣的应用是烟丝中梗叶比例的测定。该研究由 C. R. Downs 等人公布于 1980 年出版的《国际烟草科学进展》上。该研究的独特性在于它要区别两种既相似又不同物质的物理混合物, 事实上到目前为止在实验室尚无有效的办法可以分析烟叶与烟梗的比例, 所以在建立 NIR 的模型时, 可以通过称重的办法调配出一系列不同烟叶烟梗比例的样品。

建模时所使用的波长为 2270nm、1840nm、2310nm 和 1680nm。分析表明 2270nm 与 1840nm 与梗含量呈正相关性, 估计这是由于烟梗纤维素中的 C—C、O—H 在 2270nm 处有较强的吸收, 纤维素中的 C=O 与 O—H 在 1840nm 处吸收较强; 而 2310nm 和 1680nm 与梗含量呈负相关性, 因为烟叶中油、蜡物中的 CH_2 键和蛋白质中的 $CONH_2$ 键对 2130nm 吸收较多, 而芳香族化合物中的 CH 键和胺类物质中的 NH 键在 1680nm 处吸收较多。因此梗叶比例的检测主要依据是烟梗中富含的纤维素, 及烟叶中富含的油、蜡、蛋白质及芳香物。

(2) 烟叶配方研究

卷烟产品实际上是由不同烟草及添加剂混合而成的混合物, 由于烟草的品种、产地、等级及其性质的复杂多样性, 要对卷烟性质进行定量分析是非常困难的。对卷烟工业的一

个主要挑战是如何分析原料及产品的质量，一个很容易涌现的问题是，卷烟产品中加入了何种烟叶及加入的比例是多少？

Su-Chin Lo 于 1995 年公布了其最新研究成果：应用 NIR 和神经网络对烟草混合物进行快速分类和配方分析^[15]。其结果表明，采用 NIR 和神经网络分析法结合的办法，在合理的误差范围内可以对烟草混合物进行分类和定量分析，烟叶的混合比例的估测误差平均为 2.5%，该方法经进一步研究开发后，有可能应用于烟丝掺配过程中混合比例的检测和控制。

当烟叶的年份、产地、等级等发生变化时判断及维持产品的质量，或者根据目标产品开发新产品时，往往依靠感官、物理、化学等方面提供的信息开展这方面的工作，Bense T. 等人的研究认为，NIR 可以对这方面的工作提供有益的技术支持^[16]。

20.1.5 NIR 的应用展望及其他应用研究

20.1.5.1 在线检测

目前 NIR 广泛应用于水分的生产线实时检测和自动控制，新型 NIR 仪器对烟草颜色、种类及流量敏感性也逐渐减小，检测精度不断提高。

同时一些新型 NIR 仪器可进行烟碱和糖的在线检测，该功能使用在打叶复烤线上，可以将主要化学信息直接附在包装好的烟包上；也可以用来监视烟叶的混配过程，为混配或回掺提供预警。随着这方面仪器的成熟，制丝生产线的化学成分最终也将像水分一样进行在线检测，产品的重要化学成分也许会像水分一样进行适当的在线控制。

另外已经开发出的 NIR 仪器可以进行以下方面的检测^[17]：在线不接触测量卷烟烟盒外包装的薄膜厚度，在线检测包装材料纸张、纸板的厚度、水分与颜色，在线检测条包纸复合层内施加的胶、漆和涂层的厚度，在线检测丝束和三醋酸甘油酯的一致性。可以检测香精浓度的专用仪器也正在开发中。因此有人称 NIR 为“测量整个制品的技术”。

可以预见，这些仪器成熟和广泛使用后，对卷烟工业将有革命的促进作用，可以大大地提高烟草制品的质量、产品稳定性和生产效率。

20.1.5.2 其他应用研究

另外根据理论分析，除了以上所讨论过的研究及应用外，随着 NIR 软硬件的发展，NIR 还有望应用于其他许多方面，并曾有不少学者在早期已进行过有关探索研究^[5]。

NIR 可能应用于总粒相物、烟叶重量、烟草质量指标等方面的定量分析。Paradeya 等人早在 1978 年曾进行过这方面的研究，但限于当时的软硬件条件限制，该研究没有取得重大进展。

淀粉、石油醚提取物及 Holocellulose 的检测，McClure 等人于 1986 年公布了这方面的 NIR 研究结果，结果比较乐观，但灰分的结果不尽如人意。

研究烟草香型与香气方面，Solanesol 是烟草中十分丰富的 Terpenoid，与烟草的香型与香气有密切关系，Williamson 等人（1987 年）对低 Solanesol 烟草品种进行的 NIR 研究，结果还不算太理想。

如能开发便携式 NIR 仪器的话，在烟草采购过程中，有望对烟叶的主要成分进行现场测试，从而为烟叶的采购提供更为可靠的依据。

NIR 还可望在烟草生物、生化、栽培与施肥等方面得到应用。

20.1.6 小结

本章所综述的大多是 20 世纪 70~80 年代的一些 NIR 研究成果,而当时 NIR 技术在软硬件方面存在以下方面的不足:①NIR 仪器大多是滤光片式的,因此波段的选择具有很大的局限性,而现在已推出各种型号和规格的扫描式 NIR 仪器,精度大大提高,波段的选择具有更大的自由度,对 NIR 的研究者提供了极大的便利;②早期的计算机在速度、计算能力和储存容量非常有限,研究开发过程中往往需要根据计算机硬件条件来开发适合的算法。

而现在计算机的发展已基本可以不受这些方面的限制,随着应用数学和计算机软件的发展,已有许多现成的计算软件可供 NIR 研究者使用,对研究者来说这方面的要求大为降低,从而使更多的学者可以轻易地进入 NIR 的研究领域。

另外一些专业化的 NIR 仪器不断出现和性能不断提高,更为 NIR 应用扩展提供坚实的技术基础。

因此 NIR 技术已渐渐成为一种大众化的研究手段,而其自身所具有的众多优点及非常的可扩展性,必将使这种技术以更快的速度扩展到各行各业。近年来,近红外技术在我国烟草领域的应用已逐步得到认可,引进的傅里叶变换型近红外仪器就有数十台,并发表了一批相关的研究论文^[18~19]。NIR 在烟草行业的应用前景是非常诱人的,可以预见不久的将来,NIR 将成为烟草行业“全能”的检测技术和研究手段。

20.2 近红外光谱分析在纺织工业中的应用

在中红外谱区,对纺织工业中纺织材料、纺织化学品、织物上的整理剂的分析以及反应过程、反应机理的研究都已成功应用在实际生产过程中。随着近红外光谱技术的发展,及其分析速度快、无需前处理、分析过程中无需化学试剂属于绿色分析技术等优点,纺织工业中应用近红外光谱技术进行质量监控和定量、定性分析织物的组分及物理参数等越来越广泛。下面将近红外光谱分析技术在纺织工业中的应用做一简单介绍。

近红外光谱在纺织工业中的定性分析主要结合模式识别方法可以区别和鉴定不同聚合物的形态和性状差别。如对尼龙 66 和尼龙 6 的鉴别,不同厂家聚酯纤维的鉴别。Warren J. Jasper 和 Eva T. Kovacs 用人工神经网络定性鉴别过 390 个样品中 17 种不同类型的织物^[20],结果证明此法是一种快速可靠的织物定性鉴别方法。

近红外光谱在纺织工业中的定量分析主要应用在以下几个方面。

20.2.1 聚酯纤维和棉花混纺织品中成分的定量测定

纺织品加工厂常常将几种不同属性的材料混合在一起使成品具有某种特性。如,棉花中混有聚酯纤维(PET)可以增加成品的耐久性并且使织物容易保养。为保证成品的质量,在整个生产过程中必须使不同材料的混合比例保持一致。通常聚酯纤维和棉花混纺织品的化学分析需要在 70% 硫酸溶液中溶解,此方法需耗时 8h 才能得到最终结果。而用近红外光谱漫反射光谱分析法只需 2min 就可以得到最终结果。由于聚酯纤维和棉花在近红外谱区都有特征吸收带。聚酯纤维和棉花的近红外光谱见图 20-1,由图可以看出两者在主要吸收带有很大差别。

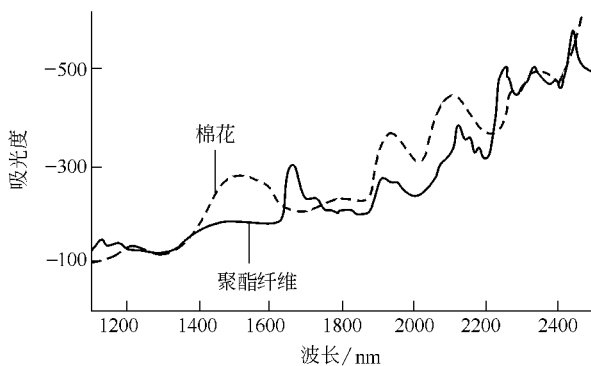


图 20-1 棉花和多元酯纤维的近红外光谱

通常选用 3~4 个波长进行回归就可以准确地测定两者的比例，即混合物的组成。Wayne 和 Tincher 等人^[21]对此做过研究并取得满意的成果，在实际的棉含量和近红外分析结果之间进行相关分析，相关系数达 0.99，预测标准差为 0.85，符合实际要求。

20.2.2 定量分析混纺毛织品中羊毛的质量分数

羊毛织品中毛的质量分数是一个重要的品质指标，但感官测定需要经验，理化分析又相当困难。中国农业大学近红外课题组在傅里叶变换型近红外光谱仪上用偏最小二乘法定量分析羊毛的质量分数方面做过尝试性研究，结果表明该方法有望成为一种快速、非破坏性的定量分析混纺织物的手段^[22]。用近红外光谱分析混纺毛织品中毛的质量分数必须研究两个问题：近红外光谱分析必须首先建立数学模型，建立模型必须具备一批含量已知的标准样品，但收集毛质量分数已知且质量分数范围较宽的毛混纺织品又有一定的困难。这里用混纺毛织品的原料（纯羊毛、黏胶、涤纶）按不同的比例人工混合后作为标准样品，纯羊毛的添加量即为该标准样品中毛的质量分数，依此质量分数和标准样品的近红外光谱建立数学模型，研究预测混纺毛织品中毛的质量分数；各种毛织品用的染料不同，而且纺织技术也不同，这些因素是否会影响数学模型的应用。

在 BRUKER 公司的 22N 型傅里叶变换近红外光谱仪上用直径 3mm 的微量漫反射池取得的样品光谱如图 20-2 所示。

由图 20-2 可见粘胶的近红外光谱中只有 C—H 与 O—H 的吸收峰，羊毛中有 C—H、O—H 与 N—H 的吸收峰，涤纶为人工合成样品，因此一些细节的吸收在图中也能显示出来，这些光谱结构总体的差别是进行近红外定量分析的基础。

将纯羊毛、黏胶和涤纶 3 种纤维原料剪碎后按不同比例混合建立 3 个模型。模型 1：纯羊毛和涤纶混合成比例不同的 30 种样品；模型 2：纯羊毛和无色粘胶混合成比例不同的 30 种样品；模型 3：纯羊毛和有色粘胶混合成比例不同的 30 种样品。模型的性能见表 20-7。

表 20-7 不同纺织纤维原料混合建模结果

模型	样品数	混合原料	相关系数	标准差	相对标准差
1	30	纯羊毛、涤纶	0.978	2.74	4.77
2	30	纯羊毛、无色粘胶	0.890	2.59	2.99
3	30	纯羊毛、有色粘胶	0.982	2.97	4.26

注：1 号模型和 2 号模型的样品是用光纤扫描的，3 号模型的样品是用旋转样品池扫描的。

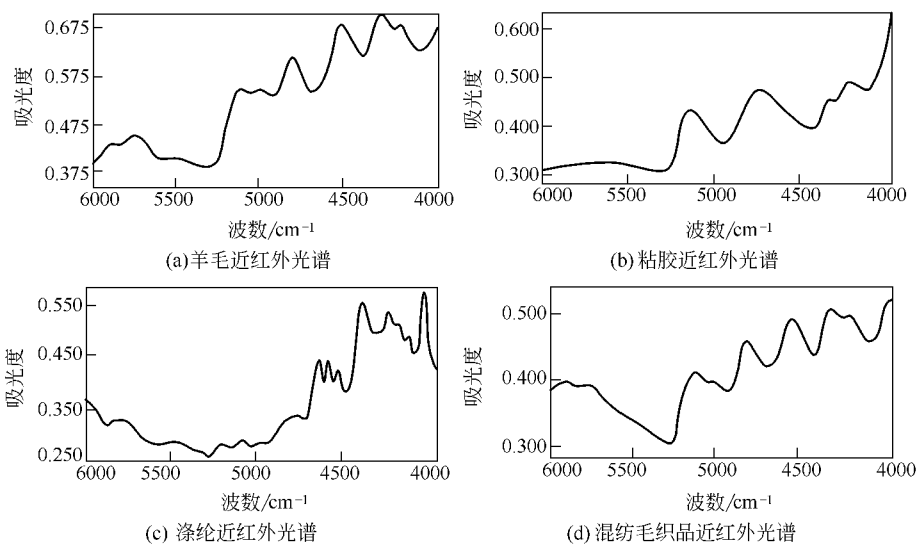


图 20-2 纯羊毛、粘胶、涤纶和混纺毛织品的近红外光谱

用建好的模型预测羊毛质量分数已知的混纺成品（羊毛和粘胶）共 6 种，羊毛质量分数均为 70%，颜色不同，这样可以有效地检测模型的预测能力。结果如表 20-8。

表 20-8 3 个模型对羊毛质量分数已知混纺成品的预测结果

样品号	1 号模型预测结果	2 号模型预测结果	3 号模型预测结果	样品号	1 号模型预测结果	2 号模型预测结果	3 号模型预测结果
1	98.8	71.9	66.63	4	98.7		66.43
2	96.9	73.1	71.6	5	96.35		74.4
3	97.5	69.8	70.4	6	98.35		70.27

由预测结果可见 3 号模型预测效果比 1 号模型好，其原因显然是 1 号模型是用羊毛和涤纶混合样品建立的，而未知样品则由羊毛和粘胶混纺而成，涤纶和粘胶的近红外光谱差别很大，因此 1 号模型的预测效果较差。其他两个模型均是用混纺毛织品的原料（羊毛和粘胶）建立的，预测效果较好。从模型的散点图看，2 号模型比较离散，其原因可能是用光纤扫描未经旋转造成样品比较离散。而 3 号模型使用旋转样品池，减少了样品不均引起的离散，对未知样品的预测值的变异系数为 4.1%，结果令人满意。表明建立近红外模型的模型背景要与待测样品背景一致；选择合适的模型可以降低颜色对预测结果的影响；近红外漫反射光谱技术可以通过导数等光谱处理措施校正样品状态（纤维的人工混合状态与纺织品状态）的差别。

20.2.3 纺织品丝光度的测定

在丝光工序中为增加织物染色时的亲和力及色彩的稳定性等，通常要将棉花放在 20% 的碱溶液（NaOH）中处理。未经处理天然状态棉的纤维主要是纤维结构 I，当用碱溶液处理时由纤维结构 I 转变成结构 II。传统判定丝光程度（Barium Activity Number，简称 BAN）的方法需要 6h 才能取得结果，远不能满足对丝光工序的过程监控。Nikitin^[23]对未经丝化和经过丝化的棉花的近红外光谱进行研究，发现丝光引起光谱峰值的

漂移并导致某些吸收峰的消失。利用两者差别，Gjonnes 和 Norman^[24] 选用 3 个波长进行回归，化学分析的 BAN 和近红外分析值之间存在很好的相关，散点图如图 20-3 所示。

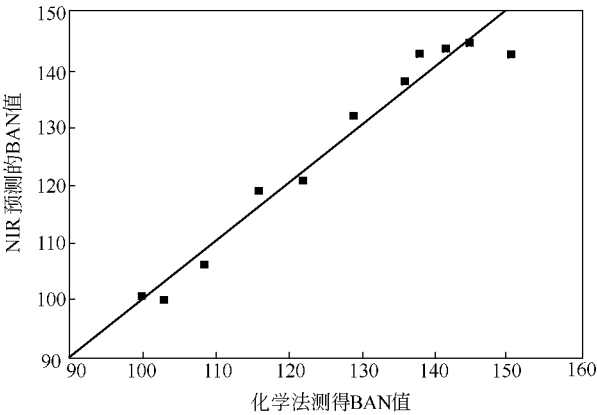


图 20-3 近红外光谱法建立丝光度模型的散点图

20.2.4 棉纤维成熟度的测量

纤维成熟度是指纤维壁的厚度，直接影响产品的质量，所以纺织厂在加工前必须测量棉纤维的成熟度。天然棉的主要成分是纤维，约占总量的 94%，其他成分有石蜡、果胶、糖、有机酸以及蛋白质等。随着棉的成熟，棉中果胶及其他非纤维的成分的相对含量下降，这是定量测定棉纤维成熟度的基础。图 20-4 是成熟和未成熟棉花近红外光谱的二阶导数图，因成熟而引起的变化表现在图中峰值处两者吸光度的差别。

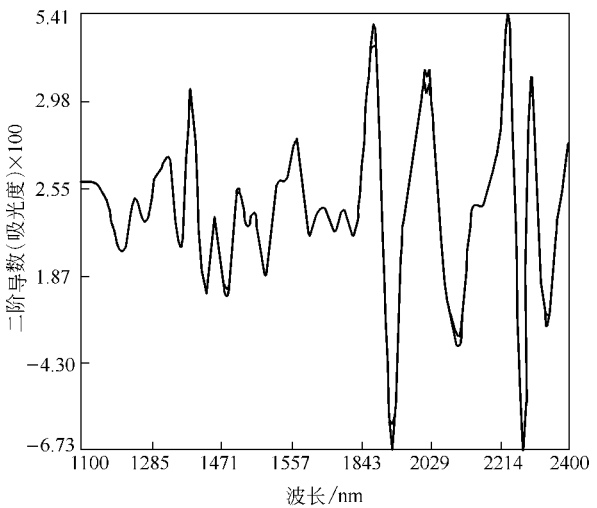


图 20-4 成熟和未成熟棉花近红外光谱的二阶导数图

注：图中横坐标为波长，纵坐标为求二阶导数后扩大 100 倍的吸光度值。

通常有两种方法测量棉纤维的成熟度：Causticaire 法和 Shirley 成熟度测试仪。这些方法既费时又容易将纤维的纯度和成熟度混淆，因此给出的棉纤维成熟度有较大的偏差。

近红外光谱法是测量棉纤维成熟度很好的替代方法，在不同的棉纤维纯度下扫描纤维的光谱，用化学计量学的算法校正棉纤维纯度所带来的影响，从而给出正确的棉纤维成熟度。在挑选样品建立数学模型时一定要选择不同生长地区的样品并且使建模样品的棉纤维纯度的范围尽量宽，这样才能建立稳定、可靠的数学模型。

20.2.5 尼龙织物中湿度的测定

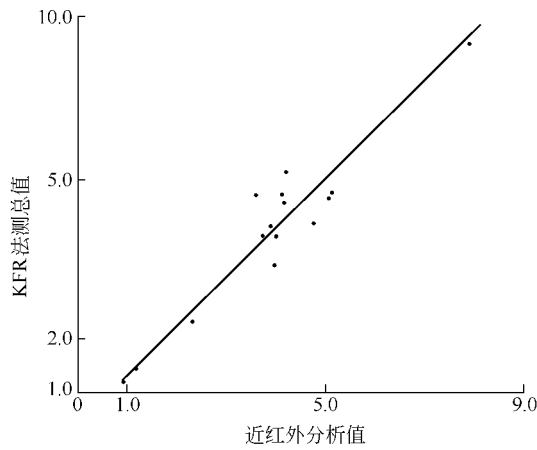


图 20-5 用 3 个波长建立的尼龙 66 湿度模型的散点图

注：图中横坐标是近红外分析值，纵坐标是 KFR 测定值。
散点图见图 20-5。

在人造纤维、纱和织物中湿度是一个重要的参数，对成品的物理性状、加工功率和生产率都有重要影响。对亲水物质如尼龙、纤维的湿度对加工过程中加热温度和加热时间的确定具有参考意义^[25~26]。测量湿度经典方法是 KFR (Karl Fischer Reagent) 滴定法，虽然此方法结果比较准确，但实验过程费时且经常要用到有毒芳香化合物。Rodgers 等人^[27]用 Technicon 400TX InfraAnalyzer 对尼龙 66 进行近红外光谱扫描并建立数学模型，选用 3 个波长 (1445nm、1905nm 和 1940nm) 进行回归，预测标准差达到 0.3%，取得满意结果。模型

20.2.6 热定型温度的测量

人工合成纱的热处理是织物加工过程中的一道重要工序，可以防止织物缩水。为了获得高品质的织物，应该在合适温度、湿度和加热时间下对纱线进行热处理。Rodgers、Ghoh 和 Tincher 等人^[27~29]研究过用近红外光谱法测量尼龙地毯纱的热定型温度。对尼龙 66 地毯纱实际加工过程中的热定型温度和近红外预测值之间建立数学模型，决定系数 R^2 达到 0.99，预测标准差 1.8°F。

我国加入世贸组织后，各行各业都面临着很大的挑战。纺织行业作为我国的一个重要产业，尤其面临着提高产品质量、提高生产工艺的重任，这就需要通过提高生产过程中的检测、测试技术，靠科学的检测手段监控整个生产过程。随着近红外光谱分析方法的完善和近红外光谱仪的改进，近红外光谱分析技术在纺织工业领域的应用也会随着发展。

参 考 文 献

[1] A Himid, W E McClure, W W Weeks. Beitr Tabakforsch, 1978, 9: 267~274
[2] T M Long. Anal Proc, 1983, 20: 69~72
[3] P J Cooper, R J Reynolds. Tobacco Co Unpublished results
[4] J T Diffie, T N Long. NIR Calibration Transfers from the Neotec 6350 to the Technicon InfraAnalyzer

- er 4000, paper presented at the Third International Diffuse Reflectance. Chambersburg, PA, August 1986
- [5] Donald A Burns, Emil W Ciurczak. Handbook of Near-Infrared Analysis. Marcel Dekker, Inc, 1992
- [6] W E McClure, K H Morris, W W Weeks. Beitr Tabakforsch, 1977, 9: 13~18
- [7] R S Pandeya, N Rosa, E H WHITE, et al. Tobacco Science, 1978, 22: 27~31
- [8] A Tonini, G Mellone, S. Commella. Ann Ist Spec Tab, 1981, 8: 89~92
- [9] I Schmeltz, D Hoffman. Chem Rev, 1977, 77 (3): 295~311
- [10] **王文真, 张怀宝. 利用 IA450 近红外分析仪快速测定烟草中的总氮含量. 仪器仪表与分析监测**, 1995 (20): 53~55
- [11] TR Staff Report, Moisture Control On-line, Tobacco Reporter, September 1997: 58~59
- [12] W E McClure, R E Williamson. Status of Near Infrared Technology in the Tobacco Industry, in Advances in the Analytical Methodology of Leaf and Smoke. A symposium presented at the 40th tobacco Chemists' Research Conference, Knoxville, TN, October 14, 1986 (3~53)
- [13] Di Luzio C, Morzilli S, Cardinale E. Rapid Near Infrared Reflectance Analysis of Mainstream Smoke Collected on Cambridge Filter Pades, Beitrage zur Tabakforschung International, Volume 16, No4, Nov. 1995. 171~184
- [14] Williamson R E, Charplin J F, McClure W F. Near-infrared spectrophotometry of tobacco leaf for estimating total particulate matter yield of smoke, CORESTA, 1986, Symposium Taormina, 107
- [15] LO S C. Rapid classification and blend analysis of tobacco mixtures using near IR and artificial neural networks. Proceedings of NIR-95, 1995
- [16] BENSE T, GASTELLU C, LABUONORA D, et al. Computer aided design of tobacco blends. CORESTA, 1994
- [17] **袁行思译. 测量烟草整个制品的技术. 世界烟草动态**, 1996 (4): 41~42
- [18] **赵龙莲, 闵顺耕, 严衍禄等. 傅里叶变换的红外光谱法测定烟草中九种品质参数. 光谱学与光谱分析**, 1998, 18: 4, 89~90
- [19] **张建平, 谢雯燕, 束茹欣等. 烟草化学成分的近红外快速定量分析研究. 烟草科技/烟草化学**, 1999, 3: 37~38
- [20] Warren J Jasper, Eva T Kovacs. Using Neural Networks and NIR Spectrophotometry to Identify Fibers. Textile Research Journal, 1994, 64 (8): 444~448
- [21] Wayne C, Tincher and Anderaluk. NIRS Analysis of Cotton /Polyester Yarns. Spectroscopic Analysis, 1985, 17 (10): 200~204
- [22] **李晓薇, 赵环环, 赵龙莲等. 用近红外光谱定量分析混纺毛织品中羊毛的质量分数. 中国纺织大学学报**, 2000, 26 (3): 72
- [23] V I Nikitin, Vestn Leingr, 1950, 3: 55
- [24] J Gjonnes, N Norman. ACTA Chem Scand, 1960, 14 (3): 683
- [25] J H Saunders. Polyamides, Fibers, in Encyclopedia of Textiles, Fibers, and Nonwoven Fabrics (M. Grayson, ed.). John Wiley and Sons, New York, 1984. 347
- [26] H W Starkweather. Transitions and Relaxations In Nylon. In: Nylon Plastics (M. Kohan, ed.). John Wiley and Sons, New York, 1973. 307
- [27] J Rodgers. Specific Applications of NIRA in the Textile Industry. Proc 7th International NIRA Symposium, Technicon Technical Center, Tarrytown, NY, 1984
- [28] S Ghosh and J Rodgers. Textile Res J, 1985, 55 (9): 556
- [29] W Tincher, et al. Textile Chem Color, 1986, 18 (2): 25

21. 近红外光谱分析技术在生物医学中的应用

21.1 概述

光照射生物体的组织，被其表面反射、内部吸收和散射而衰减。近红外谱区的光对各种物质的吸收系数一般非常小，因而对较厚的生物组织也具有较高的透过性。1977 年 Jöbsis FF 等用近红外光照射猫和人的头部及狗的肝脏，成功地检测到了透过的光，并且发现这些透过光的强度与生物体组织呼吸状态的变化显示相关性，这样利用近红外光可以获得各种生理信息，其中还可获得相关组织氧合状态的信息。自这一开创性的研究报告发表以来，开发用近红外光谱法无创检测生物组织特别是脑组织的血氧浓度的新技术，有望作为普通医疗器械在临床领域推广应用。

人体组织信息的光谱检测法具有可实现无创、连续、实时监测的特点，一直是人们感兴趣的研究课题之一。由于可见光在组织中的衰减过大，光谱检测法首先在从手指尖等较薄部位处无创检测动脉血血氧饱和度上取得了进展。目前，脉搏式动脉血氧计已在临床上得到普遍应用，但是，动脉血氧饱和度只能反映肺部血液的氧合程度，为了把握脑及肌肉组织的氧代谢状况，有必要对组织毛细血管内的血液含氧状态即组织血氧进行检测。作为无创、长时间连续实时监测脑及肌肉组织氧代谢的惟一手段，近红外光谱组织血氧检测法在临床监护、脑机理研究、运动生理研究和康复医学等领域具有广泛的应用前景。

在利用近红外光谱法成功地实现快速、无创检测生物组织血氧、血糖等生化指标的同时，研究者们对近红外光的成像技术或称光断层扫描技术予以特别关注，认为其将成为 21 世纪近红外光谱在生物医学领域广泛应用的重大课题。针对近红外光谱法定量生物组织成分时的缺陷，提出了利用皮秒脉冲光源和快速检测器，使用时间分辨光谱学技术检测光路径长度分布来确定组织吸收或散射的变化，从理论上讲是可以做到被测成分的绝对定量。这样，可将以往独立开发出来的近红外应用技术如光通信技术、短时脉冲技术结合起来，从而产生一个被称为“近红外时间分辨诊断学”的新学科。

21.2 血氧的测定

21.2.1 测定的原理

光在生物组织内同样存在散射和吸收现象，由于近红外光在生物组织中传播受到强散射（散射系数约为 $10^2 \sim 10^3 \text{ cm}^{-1}$ ）和弱吸收（吸收系数约为 $10^{-2} \sim 10^{-1} \text{ cm}^{-1}$ ），在 700~900nm 这段近红外区域内，存在一个“光谱窗”。在这个“光谱窗”内，生物组织对光线的吸收作用大大降低，光线可以进入更深一些的组织。氧合血红蛋白与还原血红蛋白在近红外领域具有不同的吸收光谱。前者使得测量范围能够拓展到离表面较深的部分，而后者则使得组

组织的吸收光谱随血氧含量的变化而变化。同时，由于血红蛋白和细胞色素含氧量不同导致的吸收光谱的差异仍然可以分辨。当波长大于 900nm 时，组织中的水分子对光子的吸收作用十分强烈，光子进入组织几毫米就会被吸收殆尽。而在低于 700nm 的可见光范围内，血红蛋白对光线的吸收作用大大增加，同时，组织的散射作用也十分厉害。所以这两段光谱区均不适合作为光源。如图 21-1 所示是血液中血红蛋白、肌肉中肌红蛋白及所有细胞线粒体中细胞色素氧化酶的近红外吸收光谱，氧合血红蛋白在 930nm 处有吸收峰，但其强度不足可见光区（578nm）的 1/40，还原血红蛋白在 760nm 和 905nm 处都有吸收峰，肌肉中肌红蛋白的近红外吸收光谱与血红蛋白相似，只是整体移了 5nm 左右。细胞色素氧化酶因含有铜，当与氧充分结合时，其近红外吸收光谱在 830 nm 处有一非常明显的吸收峰，当处于还原状态时，吸收峰消失，从而成为生物体本身在近红外谱区血氧浓度的“天然指示物”。由此可见人体组织血管中的血红蛋白对光线的吸收能力与血红蛋白的氧合-还原状态有密切关系，因此通过检测人体组织对光线的吸收情况就可推测出此时血液中的含氧量。

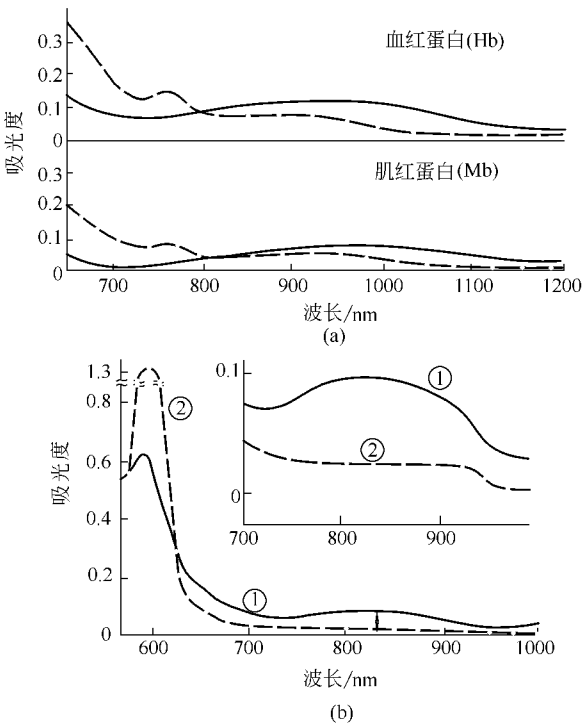


图 21-1 血红蛋白、肌红蛋白和细胞色素氧化酶的吸收光谱

(a) 血红蛋白和肌红蛋白的吸收光谱（实线：氧化型，虚线：还原型）
(b) 细胞色素氧化酶的吸收光谱（实线：氧化型，虚线：还原型）

21.2.2 生物医药领域近红外光谱仪的种类及特点

近红外光谱法可分为连续波光谱法、时间脉冲光谱法和相位调制光谱法。同连续波光谱法相比，时间脉冲光谱法和相位调制光谱法测量光子在组织中的散射路径，它们能够更精确地提取组织的血氧信息。

21. 2. 2. 1 连续波光谱计

连续波光谱法以 Beer-Lambert 定律为理论基础，利用光在组织中传输的衰减来确定含氧血红蛋白和还原血红蛋白的浓度。

因为组织是一个高散射介质，在正常条件下，组织的散射特性随组织的不同而变化，并且即使是同一种组织，其特性并不完全一致，光路径也是变化的，即是一个未知的变量。这样就很难只从测量入射光和散射光的光强度来确定被测物的吸收率，所以连续波光谱法和其他基于光强衰减来定量血氧饱和度的方法一样，一般测量误差较大。为了消除背景的影响，一般采用相对测量的方法，即用某一时间或某一状态的测量值为参比，测量其它时间或状态与参比值的差，以此来确定含氧血红蛋白和还原血红蛋白的浓度变化。由于生物体的组织是多种组分的叠加，在近红外谱区每种成分无法找到具有惟一波长的特征吸收，因此生物体无损近红外检测难以使用单波长方法，通常使用多波长结合化学计量学方法来测定。

21. 2. 2. 2 时间脉冲光谱计

时间脉冲光谱法用大功率脉冲激光二极管作为光源，用计数光电倍增管对出射光进行收集。由于其光源不具有某种特性，所以光检测器件无法把有用光从干扰光中区分出来，因此时间脉冲光谱法受外界光的干扰较大，不便于手术时对病人进行实时监测。

21. 2. 2. 3 相位调制光谱计

相位调制光谱计不同于连续波光谱计，它不再依赖于传输光强度的变化。透射过被测物体（或折射回入射表面）的光，在接收点被光导纤维检测到，并传送到光电倍增管进行信号转换，被检测到的调制光相对于入射光有一个相移，这个相移 θ 就是相位调制光谱计的有用信息。这个相移 θ 正是光子路径（ L ）的量度。当一个高散射的被测物（例如组织）的吸收率增加时，光穿进组织部分的深度将减小，于是被接收光的平均路程缩短，因而相移角 θ 将减小；相反地，当被测物的吸收力减小，光能够穿进被测物的深度加深，因而平均路径变长，相移角 θ 也增加。图 21-2 是标准时域光谱计测量和频域（或称相位调制）测量的比较。与连续波光谱计技术相比，相位调制光谱计能更精确地测量含氧血红蛋白和还原血红蛋白的浓度。

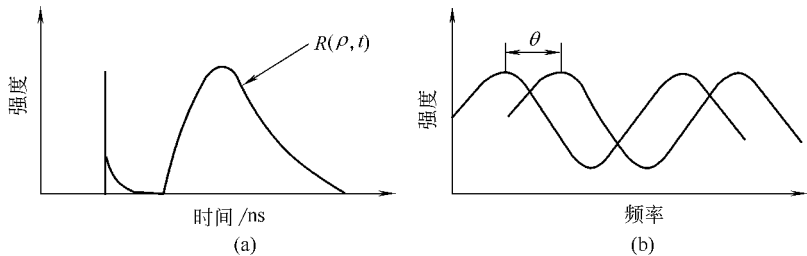


图 21-2 时间脉冲光谱法和相位调制光谱法的比较
(a) 时域光谱计 (b) 频域光谱计

21. 2. 3 近红外光谱组织血氧定量检测的理论基础

目前已有的定量化方法的理论基础主要有 3 种，其一是基本的朗伯-比尔定律；其二是修正的朗伯-比尔定律；三是散射理论。

21.2.3.1 基本朗伯-比尔定律

朗伯-比尔定律描述了光通过均一、无散射介质时入射光强与出射光强之间的关系。即

$$I = I_0 e^{-\epsilon c d}$$

式中 I_0 为入射光强； I 为出射光强； ϵ 为分子吸光系数； c 为介质浓度； d 为介质厚度。将吸光度定义为：

$$A = \lg \frac{I_0}{I} = \epsilon c d$$

在组织光学中通常把易于检测的吸光度的变化作为研究对象，也就是首先取一个参考状态，然后检测其他状态对参考状态的相对变化。即

$$\Delta A = -\epsilon c d + \epsilon c_{\text{ref}} d = \Delta \mu_a d \quad (\Delta \mu_a = \epsilon \Delta c)$$

式中 μ_a 为吸收系数，若假设血液中的吸光物质只有血红蛋白，则吸收系数的变化 $\Delta \mu_a$ 为

$$\Delta \mu_a = \epsilon_{\text{HbO}_2} \Delta [\text{HbO}_2] + \epsilon_{\text{Hb}} \Delta [\text{Hb}]$$

式中 $\Delta [\text{HbO}_2]$ 和 $\Delta [\text{Hb}]$ 分别为氧合血红蛋白和还原血红蛋白浓度的变化。原则上基本的朗伯-比尔定律仅适用于均一、无散射的介质。但实际上，生物组织是一个很强的光散射体，光在组织中经过多次散射才能从光源到达检测器，实际的光路长远大于光源到接收器的距离。因此，1988 年 Cope M 等提出了修正的朗伯-比尔方程。

21.2.3.2 修正的朗伯-比尔方程

Cope M 等的试验表明，光透过高散射介质时，基本的朗伯-比尔方程仍适用，但公式中的光路径已经不是简单的检测器到光源的物理距离，光子从光源到检测器经过多次散射，于是光子实际走过的路径长度比检测器到光源的物理距离大得多。在此基础上 Delpy D T 等提出了微分光路径长度和平均光路长度的概念，并给出了修正的朗伯-比尔方程

$$A = -\lg \left(\frac{I}{I_0} \right) = B \mu_a d_p + G$$

式中 B 是一个光路径长度因子，与吸收系数、散射系数以及散射相位有关； d_p 是光源到探测器之间的距离； $B d_p$ 为光在组织中所走的实际路径长度，定义为平均光路径长度 (L)； G 是由散射引起的光损失，与吸收系数无关。定义微分光路长为 A/μ_a ，则由上式得

$$\frac{\partial A}{\partial \mu_a} = B d_p = (L)$$

于是考虑组织的散射性质，用平均光路长 (L) 代替基本朗伯-比尔方程中物理光路长 d ，得

$$\Delta A = \Delta \mu_a (L)$$

21.2.3.3 散射理论

生物组织是光的强散射体，其散射系数远大于吸收系数，因此可以将 Boltzmann 辐射传输方程进行漫散射近似，以描述组织内的漫散射光与其光学特性参数之间的关系，即散

射介质中的光散射方程，对于半无限、均一介质条件下，超短脉冲的垂直入射光，可由漫散射方程近似求得表面反射光对于时间及空间的解析表达式。

21.2.4 近红外光谱法测血氧实例

21.2.4.1 由动物试验或模型试验直接确定经验公式中的系数

Hazeki O 和 Tamura M. 等直接在老鼠的脑部试验，以确定经验公式中的系数。试验结果表明，组织中肌血球素和细胞色素对近红外光的吸收不大于 10%，于是可以假定组织对光的吸收的变化仅由组织中血红蛋白含量的变化决定，并且用 700nm、730nm、830nm 的近红外光通过老鼠的脑部，当血液中的细胞色素在 15%~50% 时，光通过组织的吸光度的变化与血液中的血红蛋白浓度的变化呈线性关系（见图 21-3）。由此检测出吸光度的变化就可以通过经验公式计算出正常老鼠的血氧含量的变化。

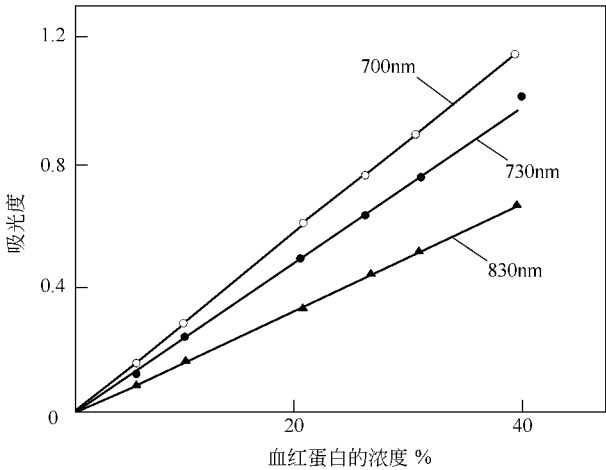


图 21-3 老鼠头部近红外透射光谱与血红蛋白的关系

近年来许多生产厂家研制的近红外监测仪在临床上进行试用，但由于不同的生产厂商使用的波长不一样，如采用三波长或四波长，各自的测定方法（计算公式）不同，并且公式中的系数一般不公开，所测得的结果并不完全一致。如图 21-4 是一种临床用血氧监测仪的示意图。日本岛津公司生产的商用仪器“島津無侵襲酸素モニタ OM100A”型已经上市。

21.2.4.2 由修正的朗伯-比尔方程来确定组织血氧的含量

王峰等利用自行设计的双波长反射式近红外组织氧合变化监护仪，运用修正了的朗伯-比尔定律，提出了在强散射的生物介质中定量检测还原血红蛋白和氧合血红蛋白变化的方法。为进行组织模型氧合程度检测实验，建立了仿真人体组织的液体模型，其结构为：组织缓冲液+光散射体+血液。整个溶液的 pH 为 6.9。为模拟人体组织对光的强散射，在模型中加入俗称脂肪乳的静脉营养液 Intralipid，它对光的散射系数根据浓度不同而变化，浓度 0.5% 的溶液，其光特性（780nm）为吸收系数小于 0.02cm^{-1} ，散射系数约 100cm^{-1} ，与人体组织的光学特性接近。最后通过向模型中加入新鲜动脉血仿真人体组织的供血情况。在实验开始时加入干酵母 2g 以还原血液中的氧，这是仿真人体组织耗氧量增加，过程大约持续 8min，其间观察到还原血红蛋白（Hb）相对浓度上升，这是由于原

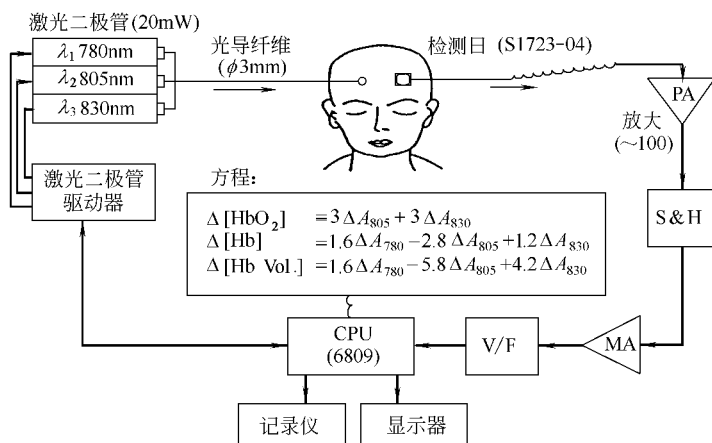


图 21-4 临床用血氧监视器的示意图

血液中的氧合血红蛋白（ HbO_2 ）被还原而生成还原血红蛋白的原因。在时间 600s 时，Hb 浓度不再增加，此时用一个医用供氧器均匀地通入纯度 99.9% 的氧气，随即可观察到还原血红蛋白浓度下降，这解释为 Hb 被氧化又生成氧合血红蛋白的缘故。至 1.1ks 时停止供氧，溶液里的酵母又开始还原 HbO_2 ，Hb 浓度又开始升高，这一过程可又多次重复。

在同样的 1000mL 的模型溶液中递增地每次加入 1mL 新鲜兔动脉血，来检测血容积的变化。可观察到氧合血红蛋白浓度基本上呈线形增加，而还原血红蛋白增加很小，这是由于动脉血的氧饱和度很高的缘故。血容积的增加量 $\Delta C(\text{BV})$ 为氧合血红蛋白浓度和还原血红蛋白浓度增量之和。

完成上述液体模型实验后，在实验室内进行了初步的人体实验。选择肱挠肌作为测试部位，使光源和传感器的轴线平行于该肌肉，利用袖带在后臂进行供血阻断，造成肌肉的缺血缺氧，同时用仪器连续记录了该过程中肌肉的含氧量和血量变化。从 60s 处开始阻断，可以观察到氧合血红蛋白含量降低，这是由于供血减少导致无法得到充足的氧，而氧合血红蛋白又不断代谢生成还原血红蛋白，因此还原血红蛋白的含量升高。至时间为 190s 时，释放气袖，由于供血恢复正常，氧合血红蛋白含量迅速上升，还原血红蛋白被氧化而含量下降。实验证明，近红外组织氧合变化监护仪具有应用价值。

21.2.4.3 由散射理论测定组织血氧含量

由于生物组织是光的强散射体，近几年来，许多学者将散射理论用于生物组织光学的研究。由表面反射光的吸收光谱来定量检测组织的光学特性，从而确定组织的血氧含量。Matcher S J 等用散射理论推导出一组量化计算组织吸收系数的公式，从而确定组织中血氧含量。Suzuki 等用近似相同的算法设计出一种检测组织中各种成分含量的仪器装置。

Matcher S J 等的模型实验和活体实验得出头部和前臂所测得的血氧饱和度比实际值低 5%~10%，这个误差被认为主要是由散射系数的误差引起的，用频率分辨技术可能会改善。近几年基于散射理论的生物组织光学特性的研究取得了很大的进展。但是在这些算法中为了将复杂的散射方程简化，仍需要做很多近似处理而影响了检测精度，这是由于为了确定组织的某一光学特性，需要已知一些其他的特性，由其他的光学特性给检测结果带

来的误差也是不可忽视的，这些都有待深入研究。

自 20 世纪 80 年代后期以来，组织血氧的定量检测已经取得了很大的进展。但是需要指出的是，这些算法都是假定组织为均一的、半无限的单层介质。事实上，人体组织除了待测组织如肌肉组织和脑组织等以外，还包括皮肤、皮下脂肪、颅骨等表层介质，而上述的这些方法都没有考虑表层组织，因此由这些方法算出的组织血氧或其变化量只是待测组织的等效量，并非实际组织的血氧含量。近年来，一些学者在研究组织血氧定量检测技术中考虑了表层组织的影响，并试图寻求克服这种影响的方法，以实现真正意义上的量化检测。

21.3 血糖的测定

糖尿病是一种由遗传和生活方式共同作用而引起的临床综合征，是因胰岛素分泌绝对或相对不足以及靶组织细胞对胰岛素敏感性降低而引起的糖、蛋白、脂肪、水和电解质等一系列代谢紊乱。在欧美等发达国家约有 3% 的人口患有不同程度的糖尿病，在我国随着人民生活水平的不断提高，糖尿病的发病率呈上升趋势。据预测，到 2010 年全世界糖尿病患者将达到 2 亿 2 千万人。目前对糖尿病治疗通常的办法是采用饮食控制和口服降糖药物，对于 I 型糖尿病可能还需要用胰岛素来控制高血糖。因此糖尿病病人需要经常监测体内的血糖浓度，以便在临床治疗和日常生活中及时掌握服药过程中的规律并进行疗效观察和饮食控制，并将血糖控制在适当的范围之内，从而有效延缓糖尿病的病情并减少并发症。病人希望能在家中自我对血糖进行长期监测。目前病人的监测通常到医院或自我进行，监测的方法一般是有创检测，这些方法给患者带来了不可避免的痛苦和不便。在家庭环境中进行血糖监测是病人自我控制的重要手段和方法，在国内外市场上有不少用于自我监测的仪器，如美国强生 Lifescan 的 ONETOUCH、德国的 Accu-chek instant 和罗氏诊断公司的 Glutrend 血糖仪等等，这些仪器在精度和稳定性上存在一定的争议，并且没有改变有创的采样方法。因此，高精度无创血糖监测仪器的开发有着广阔的应用前景。近年来，随着光学仪器设计、制造水平的提高以及计算机技术的发展和化学计量学方法研究的深入，使现代近红外光谱分析技术得到迅猛发展，近红外无创血糖监测仪的开发已成为当今研究的一大热点。

21.3.1 国内外研究进展

血液及其中的糖、蛋白质、尿素等成分的近红外光谱见图 21-5。Heise 等人提出使用去水样品和短光程的方法，来排除由于水的吸收率较大而形成的干扰，运用漫反射技术来建立血糖浓度的校正模型。在无创检测人体血糖浓度方面，采用人体口腔内静脉血作为样品，光谱的采集区间为两个水的吸收峰较高处 1900nm 和 1450nm 之间的区域。这样建立的校正模型误差大约在 2.5mmol/L(45mg/dL) 左右。进行无创监测时，人体部位的选取要考虑许多因素，对于在人体口腔内采集漫反射光谱，人体的组织如皮肤、肌肉等的干扰会小一些，但患者接受起来仍有一定困难。

Kamal Youcef-Toumi 等人认为，用近红外光谱法检测葡萄糖浓度时，近红外区分布着葡萄糖吸收信号的合频和倍频，葡萄糖的吸收峰在 2257nm 处。温度的变化对葡萄糖的

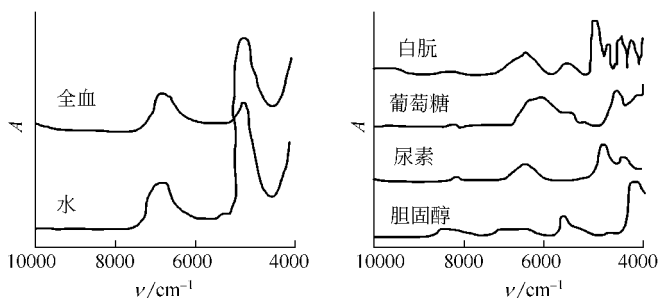


图 21-5 血液中各成分的近红外光谱

吸收也有影响，影响无创血糖检测的关键因素为：信噪比较低，检测的重复性差。在 2100~2320nm 处重复性比较高，因此选用这一波段来对葡萄糖浓度进行回归建模。对于无创监测，他们认为光线需要穿透人体的组织，而组织的变化对光谱的改变远大于血糖浓度变化对光谱的改变，所以测量很可能是不可重复的，需要对每种组织的噪声光谱特点进行研究。影响测量精度的因素很多，特别是水和温度的影响，在建立数学模型时必须采取一定的方法给予矫正。

Michael J. Mcshane 等人，在用近红外光谱对血糖浓度的测量中，认为葡萄糖浓度在 2100~2400nm 的吸收峰比较明显，并且找出四个特征峰分别是：2275nm、2250nm、2290nm 和 2325nm，因此选用这一波段的葡萄糖吸光度的数值采用逐步选择法来确定子集，并使用偏最小二乘法来建立校正模型，取得了比较理想的结果。他们把模型预测浓度和实际浓度进行了比较，两者之间误差很小，相关系数达到了 0.996。

Christopher S. Menulty 和 Ganapati Mauze 等人分别运用小波分析和偏最小二乘法的方法对血糖的浓度进行分析，他们发现在 2200~2400nm 处，水对近红外光吸收相对较小，并且在这段区域血糖有比较明显的吸收峰。因此他们利用这段来进行分析。结果表明用小波分析的方法更适合于进行光谱分析、偏最小二乘法、人工神经网络和小波变换是当前进行近红外光谱分析的常用方法，他们各有特点。还有一些学者在进行无创性测定时，在临床中采用了对一些患者跟踪测量的方法，研究结果表明：建立模型时应该对不同的个体分别建立，而对整体建立就会有较大误差。

Brumeister 等人用舌头作为检测部位，在 1400~2000nm 范围内对葡萄糖的无创性测定进行了研究。最近他们对五个 I 型糖尿病患者进行了 39 天的透射研究，每天测 5 次，他们用这种方法减少了由临时影响带来的问题。对不同的个体都建立了 PLS 定标模型，并用其预测样本的浓度。

StepHen F. Malin 等人探讨了用 1050~2450nm 的漫反射光谱无创测试人体的血糖浓度。他们采用两种方法来建立血糖浓度的预测模型：第一种是采集 7 个糖尿病人在 35 天中任意选取的多个血糖浓度的近红外光谱。第二种是用 3 个健康人采用口服耐糖监测，在数天中，每隔 3.5h 收集他们的近红外光谱，来观察他们血糖浓度的变化。实验结果表明：用糖尿病人血糖的近红外光谱做出的校正模型其标准偏差为 1.41mmol/L(25mg/dL)；利用健康人口服耐糖实验中血糖的近红外光谱做出的校正模型的标准偏差是 1.1mmol/L(20mg/dL)。这说明用波长 1050~2450nm 之间的近红外光进行无创漫反射实

验所建立校正模型是可行的，但模型的长期稳定性和可靠性还待研究。

M. Ries Robinson 等人对三种近红外区常用的光谱仪进行讨论。这三种光谱仪分别是：具有铟化铟探测器的傅里叶透射红外光谱仪；具有硅探测器的光栅单色仪；具有硅探测器和光纤取样装置的光栅单色仪。这三种光谱仪的工作区间均在 $900 \sim 1300\text{nm}$ ，利用化学计量学的方法对其光谱进行分析，用 PLS 和主成分回归法建立校正模型，血糖浓度的标准偏差达到 1.1mmol/L 。

国内的研究与国外相比起步较晚，相关论文和报道也较少。黄岚等利用红外漫反射光谱技术建立了无创检测人体血糖的方法，认为采用个体自我建模方法，在 $9000 \sim 12000\text{cm}^{-1}$ 选择平滑和二阶导数进行预处理，建模结果 $\leq 15\text{mg} \cdot \text{dL}^{-1}$ ，预测结果 $\leq 11\text{mg} \cdot \text{dL}^{-1}$ 。利用信号处理方法进行预处理并进行恰当的谱区选择，采用偏最小二乘 (PLS) 方法建立模型并预测，这种方法可以满足定量测量血糖浓度的要求，但认为建立模型时谱区和预处理方法及样本数目选择很重要，同时受试者组织状态的差异和样本温度也成为影响定量分析的因素。此外，上海中医学院、北京医科大学等也做过相关研究。

21.3.2 NIR 无创监测血糖的技术难点

在用近红外光谱分析技术进行无创血糖监测中存在着以下技术难点：葡萄糖分子在近红外吸收值很小、干扰很大；仪器精度、信噪比要求很高；还需找到更加准确快速的算法；在测量时，受到人体其他成分如骨骼、肌肉、脂肪和皮肤的干扰很大；对外界环境的要求也很高，建立数学模型时不仅需要考虑提高精度，还要考虑实际在仪器开发时的可行性。无创血糖监测是一项先进的血糖监测技术，患者可以在家中随时检测血糖浓度的变化从而控制饮食或及时就医。因此，无创血糖监测技术有着很大的研究价值和实用价值。尽管在研究过程中会有许多难以克服的障碍，相信随着认识的深入一定能够找到解决问题的方法和途径。

21.4 组织及体内外成分分析

21.4.1 组织分析

利用 NIR 光谱法进行组织分析，Dreassi 及其同事发表了一些关于皮肤过敏的文章。首先，他们讨论了 NIR 辐射如何穿透至少 0.20mm 具有复杂结构的细胞间质，研究发现，近红外线可穿透角质层，通过 PCA 分析，他们获得了如水及脂肪等成分的信息。在另一篇文章中，课题小组研究了皮肤与丙烯醇 400 (PEG400)、异戊二烯 (IPM) 及水凝胶间的相互关系。他们分析了皮肤与这些试剂反应后光谱的变化，以及正常、过敏皮肤与水及脂肪等成分反应的差异。

选用 PEG400、IPM 和药剂胶等脂溶性化学试剂以用来促进与心电图的联系。通过使用 PCA 分析，Dreassi 等能容易地从它们与这些试剂接触反应中区别过敏皮肤与正常皮肤。在以后的研究中陆续报告了相似的结果，使用不同分子量及不同黏度的聚乙烯氟可以辨别正常及过敏皮肤的 NIR 光谱。

化学试剂与皮肤之间的相互作用过程包括两阶段：改变皮肤的物理特性，如移走水分并重新分配。化学试剂引起角质层的变化，此变化在正常/过敏皮肤间是不同的，其变化机制也存在差异。

NIR 光谱法一个重要的应用是外伤后组织生存能力的鉴定。一些组织长时间和严重的缺氧可引起组织坏死，Dreassi 研究小组使用 NIR 技术鉴定了低氧供应的组织。试验使用 Mcfarlane 小鼠背部的皮瓣进行试验发现，皮瓣组织的氧气供应应立即下降。众所周知，主血管被切断的严重外伤也可引起皮肤组织坏死。NIR 可作为一种工具来评估损伤皮肤是否成功的重新附着。

Sweden 等也做了相似的工作，他设计了一种监测器来监测高压氧治疗病人组织的氧合过程，组织灌注和氧合是决定是否施行 HBO 治疗和 HBO 治疗次数的客观依据。

Sowa 等研究了另一种组织类型——指甲。他用 MIR（中红外）及 NIR 光谱法来研究人类生长的和修剪后的指甲，并用 MIR 法进行内部剖析及 PAS 法进行无创性的分析，他比较了衰减全反射（NIR-ATR）、NIR 漫反射及光声光谱（PAS）方法，主要用于以 2054nm 为中心的某些氨基酸化合物基础研究。从研究结论看，低能量 NIR-ATR 方法获得了最好的结果。

如图 21-6 所示是肌体脂肪组织的近红外吸收光谱，与水分子在 970nm 处有很强的吸收峰相比，猪脂肪组织在 930nm 左右的特征吸收峰十分明显，由于女性的脂肪多于男性，其吸收峰明显高于男性，男性只有一个很小的肩峰，这样从脂肪吸收光谱的差别中测出脂肪的含量就成为可能。还可利用 NIR 法很容易地测量婴儿的肥胖程度，可用来评价母乳与非母乳喂养新生儿的体胖差异，这比仅仅使用体重这一指标更能反映婴儿肥胖状况。

1997 年 Liu 等发表了胎儿和新生儿测量的另外一个方法。他们提出了用光谱法测量羊水以判定胎儿肺的成熟情况。他们用薄层色谱（TLC）法确定卵磷脂/鞘磷脂（L/N）的比例，并用从孕妇体内抽取约 350mL 羊水以校验 NIR 方程，仪器的扫描范围为 400~2500nm。液体样品与 TLC 样品间相关系数为 0.91，这在快速检测微量和复杂成分样品方面是非常出色的。

1995 年 Barlon 等通过实验发现水在 800~2200nm 间的反射光谱与组织温度相关，获得了一系列组织温度的数据。并得出其建模标准差（ $SEE=0.02\sim0.12^{\circ}\text{C}$ ）和预测标准差（ $SEP=0.04\sim0.12^{\circ}\text{C}$ ）。

由于机体组织为高度光散射介质，需要采取许多方法来减弱这种分散效应。Tsai 等发表的文章很少涉及软组织成分的吸收特性，他们将工作集中于 900~1340nm 的区域，在近红外线区域、较短的波长有较低的分散系数、较低的吸收能力及较深的组织穿透能力。

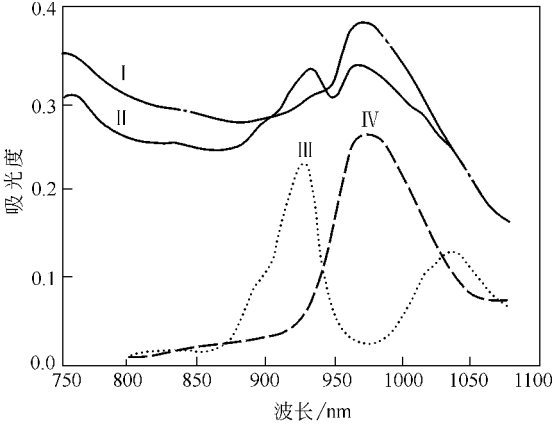


图 21-6 手腕近红外吸收光谱
I — 脂肪少的男性 II — 脂肪多的女性
III — 猪的脂肪 IV — 水

Schmitt 等的文章中讨论了混沌生物组织中 NIR 光谱的作用过程，很多能量被用于消除组织的分散效应，辨别实际的光吸收与扩散引起的光损失需要运用更高级的算法才能实现。

麻省理工学院的 Rava 等通过使用 NIR 拉曼光谱仪产生光谱，使用 Nd: YAG 激光并用足够的能量去照射组织，然后收集合适的光谱。他们还使用 CCD 来采集高噪音光谱信号，来解决组织对光的分散和削弱问题。

Anderson-Engels 及其合作者发表的文章是关于组织的时间-分辨透射光谱，并与肿瘤的发现有关。此文章详细说明了物理学上的皮秒级激光、氩离子/染色激光、蓝宝石激光用于组织方面的时间分辨光谱，这对光学乳房照相术很重要，因为肿瘤附近新血管的生成可引起肿瘤区光吸收的增加，精确预测经验上获得的时间-分辨曲线的模式一直在发展中，这个突破非常有助于乳房照相术的研究。

Anderson-Engels 研究小组在临床上的目的是研究完整的动脉血管壁，并比较正常与动脉粥样硬化患者之间的差异。此技术用来监测与动脉粥样硬化形成密切相关的一些分子包括胶原质、弹性蛋白、胆固醇、胆甾醇酯及钙盐，但其工作的缺点是采集光谱的时间长（需 30min）和功耗较高（500mW）。

评估动脉血管壁的基础性工作是在肯塔基州立大学医疗中心完成的，Robert 及其合作者得到了很好的实验结果，他为此用了十多年时间。在文献中提到了高密度脂蛋白（HDL）、低密度脂蛋白（LDL）及脂蛋白在活体组织中的定位及定量。利用复合集光器（CPC）压缩来自于传递光纤到动脉表面的小点上光线束。

在 1100~2500nm 的近红外光通过集光传递到外部动脉血管壁上方，偏离入射光的硫探测器在 CPC 近末端处探测散射光，各种类型的动脉蚀斑会产生非正常的光谱图，并且能够确定每种蚀斑的数量。

Lodder 研究小组开发出了用于这类研究的功能强大的软件，如名为 CALDATAS 的软件，是以 BEAST 和 BEST 开始的系列软件中的一个，它可用来预测“蚀斑”引起中风和动脉破裂的危险程度。

21.4.2 体内生化分析

马萨诸塞州立大学医疗中心的 Zhang 利用 NIR 光谱法建立了测定组织 pH 的模型，可用来在体内局部缺血时测定心肌的 pH。

NIR 光谱法还可用来测定体内一些重要的生化指标如蛋白酶、脂肪酶、碱性磷酸酶及过氧化物酶等。

21.4.3 癌症判别

NIR 用于癌症研究还不成熟，但由于它的无创性优势，在血液化学研究中正在被广泛使用。约翰霍普金斯大学的研究者在 Chris Brown 的指导下通过使用 NIR 光谱分辨正常或健康者、有非正常细胞患者及有子宫颈癌患者的组织，运用判别分析及主成分分析，发现恶性的和健康的组织在光谱上差异很大，而非正常的组织则兼具两者的特性。

NIR 光谱可用于进行活组织的检查如乳腺癌，国外已建议将它作为 X-射线和物理性有创性活检之外的另一有效选择，可用来消除假阳性，这对早期发现尤为重要。

1994 年, Meurens 等发现在 1200~2370nm 之间的 NIR 光谱能够区分癌组织和非癌组织, 尽管癌细胞的种类繁多, 但癌细胞与非癌细胞的 NIR 光谱具有明显的差异, 这一工作仍在探索之中。

21.4.4 体外化学成分分析

在体外使用 NIR 技术可用于分析化学成分。1998 年 Mcshane 和 Cote 分析细胞培养液, 将 3 日龄纤维培养物所用标准临床技术与用 NIR 技术分析的结果一致。在细胞培养过程中, 葡萄糖、乳酸盐、氨等可用 NIR 技术实现无创检测。

Hall 使用 NIR 去分析人乳的主要成分, 此方法能帮助营养学家很快断定是否需要为母亲加强营养以供其孩子所需。

Shaw 等使用 NIR 技术分析尿样, 他们使用了 127 个样本分别建立了测定蛋白、肌酐和尿素含量的模型, 三个待测指标的预测标准差分别为 0.23mmol/L、0.79mmol/L 和 16.6mmol/L, 结果表明, 蛋白测定用于粗选是可行的, 而其他两种成分的测量可与常规分析方法媲美。

Jackson 等利用 NIR 技术对尿液分析做了进一步的研究, 尿葡萄糖、蛋白、肌酐及尿素浓缩物可运用较简单的算法进行分析。如尿素可简单地由在 2152nm 处的相应吸收进行校准, 分析结果和标准方法呈高度相关; 由于肌酐及蛋白的含量不高, 并且吸收系数也较低, 因此要运用较复杂的算法如 PLS 来分析, 肌酐最好的相关系数达到 0.953, 蛋白则为 0.923。此方法简便快速、不使用试剂, 适于大样本的测定, 尤其在危急时刻, 速度比准确度更为重要, NIR 光谱法可能在此刻成为最重要的工具, 因此在临床上具有广阔的应用前景。

Shaw 等用 NIR 光谱法对从患者膝盖内抽取的滑液进行扫描, 用 PLS 算法建立了分析滑液化学成分含量的数学模型, 并将这些化学成分和关节炎诊断联系起来, 从而可对关节炎患者的病情进行远期预测。

Mark Riley 等利用 NIR 检测技术控制细胞生物反应器的发酵过程, 研究者们运用计算机模拟了在发酵混合物中产生的各种成分的光谱。首先是葡萄糖和谷氨酸盐组成的简单的二元溶液试验。然后他们构造了一个多元混合物溶液模型, 其中含各种浓度的氨、葡萄糖、谷氨酸盐、谷氨酸酯、乳酸盐。他们发现尽管对这些成分的预测误差比标准分析方法的误差大, 但速度快和能实现实时监测是其优势所在。

另外 Hall 等运用 NIR 技术对工业化大肠杆菌发酵肉汤中的醋酸盐、氨、生物数量及甘油的在线分析方法进行了研究。用 PLS 算法建立的数学模型对上述成分进行预测的标准差分别为 0.7g/L、1.4g/L、0.7g/L、7mmol/L, 基本能满足工业生产中质量控制和状态控制在线检测的需要。

21.5 光 CT

利用近红外光谱技术可以获得各种生理信息并可获得相关组织氧化状态的信息。有望在不久的将来, 可开发出对于生物组织尤其是脑内含氧程度的断层扫描技术即光 CT 技术。X 射线 CT 是以直接向组织照射 X 射线为基础, 从不同角度和方向测定其衰减

量, 经过数学处理提供有关组织密度分布的断层图像; 核磁共振成像 (MRI) 是将氢原子核磁共振的信号作为断层图像。光学成像与放射技术相比具有三个优势: 第一, 这种辐射是非电离的, 因此在合适的剂量下可以重复地使用而不会对病人造成伤害; 第二, 光学方法具有区分软组织的优势, 因为软组织对于近红外光具有不同的散射和吸收, 例如胸部肿瘤因其上大量的血管和高线粒体数使得它对光的吸收和散射程度高于周围组织, 因此, 用光 CT 法可以从正常胸组织中诊断癌组织、从肌肉中探测出血块等, 通过光谱分析可以检测出生化指标如葡萄糖、类脂化合物、胆固醇等的含量, 另外, 还可能通过对血红蛋白氧浓度、器官细胞状态及脑血容量和流量的测定进行生物体局部如脑缺氧疾病的早期诊断等; 第三, 自然发色团的特征吸收如血红蛋白使得该方法可以获得生物体组织的功能信息。放射技术只能提供解剖学或结构信息, 很难提供生理学信息, 而光 CT 能将含氧程度等生理学信息作为断层图像展现出来, 人们期望它能够与 X 射线 CT 和 MRI 具有互补性的诊断信息。

MRI 在 X-射线结果可能有问题时被使用, MRI 不具结构图片上的特异性, 只是显示的结果比 X-射线更清晰些, 使用 NIR 同时可得到大量又很好的结构图片。Ntziacgristos 等使用他们自己设计的时间-分辨成像仪, 获得了带有 MRI 图像的 NIR 图像。并证明了 780~830nm 范围内的短波近红外线最合适。

尽管光 CT 在生物医学领域展现出了广阔的应用前景, 但仍有许多问题需要解决, 如细胞级的光散射研究还有待深入, 正常或病理作用下光散射还未弄清; 光 CT 仪器本身还需进一步完善; 由于生物组织对近红外光来说是高散射介质, 在算法上还无法借用现有的 X 射线 CT 技术等等。相信随着研究的深入, 在不久的将来光 CT 技术就会进入临床应用。

参 考 文 献

- [1] 尾崎幸洋, 河田聪稿. 近赤外分光法. 日本分光学会测定法シリーズ32, 学会出版センター, 1995. 170~190
- [2] Emil W Ciurczak. Near-infrared Application In Biotechnology, 2001. 323~343
- [3] 郭海峰, 杨康. 近红外光谱血氧监测仪发展现状. 沈阳工业学院学报, 2002, 21 (4): 73~77
- [4] 苏畅, 丁海曙, 白净. 生物组织光谱学技术. 国外医学生物医学工程分册, 1995, 18 (6): 316~323
- [5] 王峰, 李炜, 林方等. 用近红外光谱技术实现生物组织含氧量的无创检测. 清华大学学报 (自然科学版), 1999, 39 (7): 16~19
- [6] 相韶霞, 林凌, 王艳秋等. 近红外光谱组织血氧检测结果的量化方法. 光学技术, 2001, 27 (5): 451~458
- [7] 黄岚, 丁海曙, 王广志. 用近红外漫反射光谱无创检测血糖的初步研究. 光谱学与光谱分析, 2002, 22 (3): 387~391
- [8] 马海涛, 丁东, 崔继承等. 近红外光谱在无创血糖监测中的应用. 仪器仪表学报, (增刊) 2002, 23 (5): 231~233
- [9] 王宏, 李庆波, 刘则毅等. 遗传算法在近红外无创人体血糖浓度测量基础研究中的应用. 分析化学, 2002, 30 (7): 779~783
- [10] 杨妹清. 光 CT 的现状与未来. 光机电信息, 1998, 15 (7): 1~5

22. 近红外光谱的矿物学应用

矿物是地球物质组成的基本单元，所有近红外光谱的地学应用都是在对矿物的光谱学研究基础上发展起来的。近红外光谱兼有电子光谱和振动光谱的属性。在研究矿物的离子或分子的晶体场跃迁（电子轨道内电子从低能态到高能态跃迁）或价态转变引起的吸收时，近红外光谱被归属于电子光谱（也称为吸收光谱，包括可见-近红外光谱）范围；当研究与矿物分子在较长波区的基频振动、转动能级跃迁相关的倍频、合频吸收时，又归属于红外光谱。

影响矿物近红外光谱吸收光谱的因素有：

① 矿物的化学组成：不同的矿物有不同的近红外光谱，同一种矿物中阳离子取代引起吸收谱带的位置移动；

② 氧阴离子特征：矿物中的 H_2O 、 OH^- 及 O^{2-} ，硅酸盐矿物中的非桥氧 $\text{Si}-\text{O}$ 和桥氧 $\text{Si}-\text{O}-\text{Si}$ ；

③ 阳离子的价态：三价阳离子的晶体场稳定能比二价阳离子的大；

④ 阳离子的配位数：晶体场稳定能按下列配位数顺序变化： $6 > 8 > 12 > 4$ ；

⑤ 氧配位体的对称性：对称性降低引起能级进一步分裂，吸收谱带增多；

⑥ 结晶程度：对同一矿物来说，结晶好、有序度高的有强而窄的吸收峰，结晶差、无序的则有弱而宽的吸收峰；

⑦ 晶体取向：沿晶体的不同方向测量的光谱会有差别，在低级对称晶体中更加明显；

⑧ 阳离子与氧的间距：压力增加使阳离子与氧的间距减小，晶体场稳定能增加，吸收峰移向短波（“蓝移”），升温时移向低能方向（“红移”）。

用近红外光谱进行基础矿物学研究时，通常使用单晶体矿物，如果晶体足够大，可直接按结晶轴方位切磨成平行光面，如果要测量从岩石分离出的粉末矿物或岩石基质中的小晶体，需要在显微镜下使用偏振光测量出 α 、 β 、 γ 三个方向的定位吸收光谱来进行研究。对密封在金刚石砧高压池（DAC）的矿物进行高温、高压测试时，显微吸收光谱是特别重要的技术。其他地学应用则按测量对象包括单晶体测量、或颗粒/粉末样品测量、或块状岩石（矿物集合体）测量，多数情况是测反射谱、漫反射谱，有时需要测透过谱。近红外光谱作为电子光谱，在地学领域的应用已有几十年的历史（Burns, 1970~1993）；作为振动光谱，近十多年来在地学领域的应用发展很快。

22.1 用近红外光谱研究基础矿物学问题

近红外光谱用于研究基础矿物学问题已有几十年历史，可以提供有关矿物结构（包括矿物原子间的键性、键力、配位）、化学组成、物理、化学性质，以及矿物形成时及后期变化的物化环境信息。

22.1.1 阳离子价态及占位研究

过渡金属阳离子的价态决定其在矿物晶格中的位置，阳离子价态变化影响配位体的对称性。同时，矿物中的过渡金属阳离子的价态与其形成的温度、压力及氧化还原环境有关。

Taran and Rossman (2001) 对含钴的合成斜方辉石、橄榄石、绿柱石在室温及液氮温度的可见-近红外光谱研究表明，在 $5000\sim10000\text{cm}^{-1}$ 区间的吸收谱带是八配位 Co^{2+} 的 ${}^4\text{T}_{1g}-{}^4\text{T}_{2g}$ 跃迁，位于 13500cm^{-1} 的谱带是 Co^{2+} 的 ${}^4\text{T}_{1g}-{}^4\text{A}_{2g}$ 跃迁，位于 16000cm^{-1} 附近谱带是 Co^{2+} 的 ${}^4\text{T}_{1g}-{}^4\text{T}_{1g}$ ($4p$) 跃迁。

尖晶石族矿物是通式为 AB_2O_4 的复氧化物。该族矿物属等轴晶系，空间群 $\text{O}_h^7\text{-Fd}3\text{m}$ ， $Z=8$ ，其结构式为 $8[\text{AB}_2\text{O}_4]$ 。单位晶胞中 32 个氧原子作立方紧密堆积，形成 32 个八面体空隙及 64 个四面体空隙。规则四面体位置有 Td 点群对称，畸变的八面体位置有 $\text{D}3\text{d}$ 点群对称。其四面体中 $1/8$ 空隙和八面体的半数空隙分别被三价阳离子 B (Al^{3+} ， Fe^{3+} ， Cr^{3+}) 及二价阳离子 A (Mg^{2+} ， Fe^{2+} ， Fe^{3+} ， Mn^{2+}) 填充。根据晶体场理论计算 (Duniz and Orgal, 1957 年)，在八面体及四面体位置上金属阳离子的分布取决于被八面体或四面体包围的离子的电荷分布的不同而形成的能量差。

通常，铬是以 Cr^{3+} 形式进入矿物晶格的。端员铬尖晶石的八面体位置容许 Cr_2O_3 的最高量是 67.9wt%。值得注意的是 Bunch 及 Olsen (1975 年) 报道了陨石中发现铬异常高 (Cr_2O_3 含量高达 70%~75wt%) 的铬尖晶石，认为这种铬尖晶石中存在四配位的 Cr^{2+} ，推测其形成于地幔环境。合成 Mg-Cr 尖晶石在 6250cm^{-1} 及 6670cm^{-1} 的吸收谱带被认为是四面体 Cr^{2+} 引起的 (Ulmer and White 1966 年; Greskovitch and Stubican 1966 年)。后来对天然铬尖晶石的光谱测量证实存在四面体配位的 Cr^{2+} ，并论证了铬尖晶石中的四配位 Cr^{2+} 在 1400°C 、 $f_{\text{o}_2}=10\sim11.5\text{Pa}$ 条件下平衡 (Mao and Bell, 1975 年)。这种异常高铬含量的陨石铬尖晶石存在于含金刚石金伯利岩中 (Daniels and Gurney, 1986 年)。因此，异常高铬含量的陨石铬尖晶石可以作为寻找金刚石矿的指示矿物。

众所周知，相对于 Mn^{2+} 、 Mn^{4+} 而言， Mn^{3+} 是较不稳定的， $\text{Mn}(\text{III})$ 化合物很容易氧化、还原形成更稳定的 $\text{Mn}(\text{II})$ 、 $\text{Mn}(\text{IV})$ 化合物。天然矿物结构中是否存在含 Mn^{3+} 的矿物呢？

红帘石化学式 $\text{Ca}_2(\text{Al}, \text{Fe}, \text{Mn})\text{Al}_2\text{Si}_3\text{O}_{12}(\text{OH})$ 其 3 个单晶偏振吸收光谱显示不同强度的吸收谱带 (图 22-1)，各个谱带的位置分别在 450nm (22200cm^{-1})、 550nm (18200cm^{-1})、 830nm (12000cm^{-1})。这些谱带的克分子消光系数比其他含 Mn^{3+} 化合物的高一个数量级，证明在红帘石结构中 Mn^{3+} 处

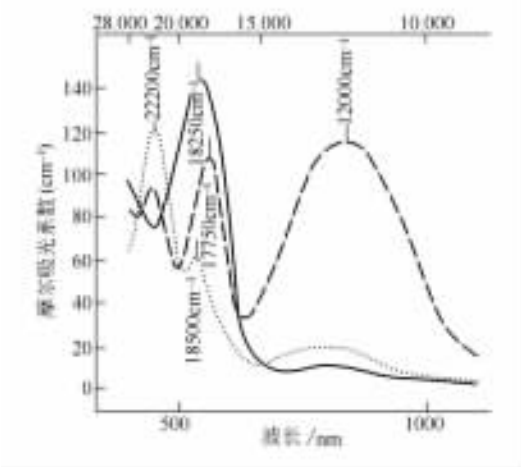


图 22-1 红帘石的单晶偏振吸收光谱
(据 Burns and Strens, 1967)
 α 谱 (点线)， β 谱 (虚线)， γ 谱 (实线)

于无心的坐标位置。后来，一系列锰含量不同的绿帘石的偏振光谱分析发现，随绿帘石中 Mn^{3+} 含量增加而引起相应谱带的消光系数减少，并不符合 Beer-Lambert 定律，证明 Mn^{3+} 不只占一个位置 (Burns and Strens, 1967; Smith et al., 1982; Kersten et al., 1987)。

22.1.2 矿物中铁离子的配位对称性研究

十字石，化学式是 $(\text{Fe}, \text{Mg}, \text{Co}) \text{Al}_9 (\text{SiO}_4)_4 \text{O}_6 (\text{OH})_2$ ，其晶体结构中的 Fe^{2+} 主要占据四面体位置，其偏振吸收光谱在 $1400 \sim 2200\text{nm}$ ($4550 \sim 7000\text{cm}^{-1}$) 间有一个很宽的强谱带。理论计算在 Fe^{2+} 从六配位转为四配位、铁-氧原子间距保持不变时，配位于四面体中 Fe^{2+} 的吸收谱带中心在 2250nm (4444cm^{-1})。实测值比理论计算值更偏向短波。这是由于十字石中 Fe^{2+} 配位的四面体有轻微畸变，其阳离子-氧平均间距为 201pm ，比其他镁铁质硅酸盐矿物中六配位的铁-氧原子间距 (216pm) 短，引起吸收谱带移到高能方向。

对吸收谱带的指派要慎重。Bancroft and Burns (1967) 的研究表明，钙质辉石及斜方辉石分别位于 4420cm^{-1} 、 5400cm^{-1} 的吸收谱带曾经被错误地被认为是四面体 Fe^{2+} ，它们实际上起因于处在辉石中强烈畸变的八面体 (M_2) 位置中的 Fe^{2+} 。又如当初将暗镍蛇纹石的位于 $4100 \sim 4300\text{cm}^{-1}$ 间的谱带指派给四面体 Ni^{2+} ，后来的研究表明是 O—H 的振动光谱 (Faye, 1974)。位于 9000cm^{-1} 、 16000cm^{-1} 、 25000cm^{-1} 的与八面体 Ni^{2+} 有关的吸收谱带只在黏土硅酸盐矿物组合中出现 (Faye, 1974; Manceau and Calas, 1986)。

22.1.3 系列矿物中的阳离子占位研究

自然界中的矿物广泛存在类质同象阳离子替代，形成系列矿物。系列矿物的形成与其所处的物化环境密切相关。系列矿物晶体内的阳离子替代引起吸收谱带频率和强度改变，长期以来引起关注。

镁铁角闪石系列矿物， $(\text{Mg}, \text{Fe})_7 \text{Si}_8 \text{O}_{22} (\text{OH})_2$ ，是双链硅酸盐，有 5 种阳离子位置。 M_I 、 M_II 、 M_III 位于双链中活性氧及 (OH, F) 组成的八面体空隙中，并形成八面体链带。 M_IV 位于八面体链带的两侧。 M_I 、 M_II 、 M_III 、 M_IV 位置主要由 Mg^{2+} 、 Fe^{2+} 、 Fe^{3+} 、 Ca^{2+} 等阳离子填充。在向背的四面体双链间，有连续而较宽大的空隙，由 Na^+ 、 K^+ 、 Ca^{2+} 填充。镁铁角闪石的吸收光谱中位于 $10000 \sim 10400\text{cm}^{-1}$ 、 8500cm^{-1} 、 $4000 \sim 4200\text{cm}^{-1}$ 的谱带与 M_IV 位置的 Fe^{2+} 有关 (图 22-2)，位于 7100cm^{-1} 附近的一系列尖谱带是与 M_I 、 M_III 位置的阳离子键合的 OH^- 伸缩的一级倍频谱 (Mao and Seifert, 1974; Hawthorne, 1981)。

又如顽火辉石 ($\text{Mg}_2 \text{Si}_2 \text{O}_6$)—铁辉石 ($\text{Fe}_2 \text{Si}_2 \text{O}_6$) 系列矿物。该系列矿物的吸收光谱中位于 $10500 \sim 11000\text{cm}^{-1}$ 及 $4900 \sim 5400\text{cm}^{-1}$ 两个谱带的强度和 Fe^{2+} 含量有很好的线性相关，可以用 Beer-Lambert 定律来定量测量顽火辉石-铁辉石系列矿物中 Mg^{2+} 、 Fe^{2+} 的占位状态 (Goldman and Rossman, 1979)。

属斜方晶系的镁橄榄石 ($\text{Mg}_2 \text{SiO}_4$)—铁橄榄石 ($\text{Fe}_2^{2+} \text{SiO}_4$) 系列矿物的近红外吸收光谱有 3 个吸收谱带，其吸收最大值分别位于 11000cm^{-1} 、 8000cm^{-1} (α 及 β 偏振光谱) 和 9300cm^{-1} 附近 (γ 偏振光谱)。所有吸收最大值显示随 Fe^{2+} 含量减少向短波方向移，与系列矿物中的 Mg-Fe 成分变化有良好的线性关系 (图 22-3, Burns et al. 1972)，反映了随

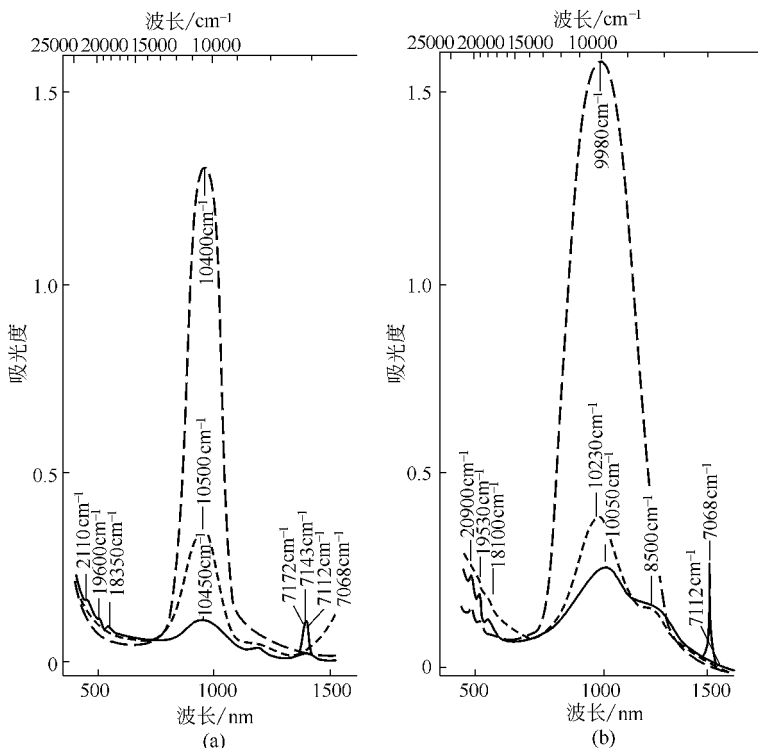


图 22-2 Mg-Fe 闪石的单晶偏振吸收光谱 (据 Burns 1993)

α 谱 (点线), β 谱 (虚线), γ 谱 (实线)

(a) 含 35.8% 的 $\text{Fe}_7\text{Si}_8\text{O}_{22}(\text{OH})_2$ 克分子

(b) 含 95.3% 的 $\text{Fe}_7\text{Si}_8\text{O}_{22}(\text{OH})_2$ 克分子

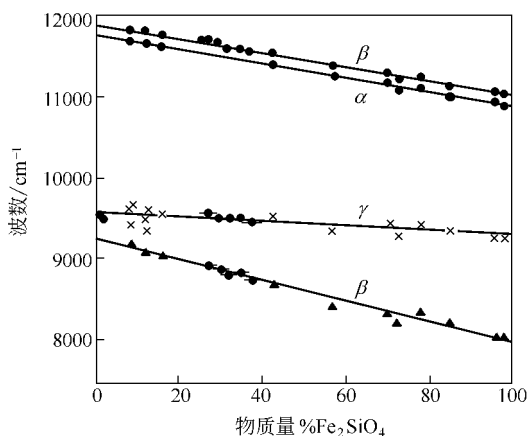


图 22-3 Mg-Fe 橄榄石偏振光谱的吸收最大值与 Mg, Fe 含量关系 (据 Burns et al. 1972)

着 Mg 含量增加, M_1 、 M_2 位置上阳离子-氧间距逐渐减少。根据铁橄榄石的吸收光谱推导出在 M_1 、 M_2 位置上 Fe^{2+} 的能级图。 11000cm^{-1} 及 8000cm^{-1} 附近的吸收谱带指派为 M_1 位置 Fe^{2+} 的 e_g 轨道的分裂; 位于 9300cm^{-1} 附近的谱带代表 M_2 位置 Fe^{2+} 的 e_g 轨道的跃迁 (Gonschorek 1986)。

对镁尖晶石 (MgAl_2O_4) - 铁尖晶石 (FeAl_2O_4) 固溶体系列矿物的最新室温 FT-IR 光谱研究表明, 铁尖晶石的克分子量引起在 $2500 \sim 7000\text{cm}^{-1}$ 区间的光谱变化与 $^{\text{IV}}\text{Fe}^{2+}$ 的含量线性相关, 符合在 $^{\text{IV}}\text{Fe}^{2+}$ 中的自旋允许的 $d-d$ 电子跃迁, Jahn-Teller 分裂

(Skogby and Halenius, 2003)。

22.1.4 不同温度、压力下矿物的近红外光谱研究

矿物受温度、压力作用引起内部结构改变, 其物化性质随之改变。这对认识地表及深部的地球化学及动力学规律非常重要。目前通过钻探可以直接采到 12km 深的岩石样品。

对于地球的更深部物质组成及结构，一方面要靠研究由火山岩浆带到地表的很少的深部物质和地球物理测量来推测；另一方面是通过实验室模拟深部环境进行的矿物学岩石学实验。近红外光谱是高温高压实验的有力观测手段，早在 20 世纪 70 年代前后开展的矿物高温高压实验研究，获得了大量近红外光谱数据。

根据晶体场模型，晶体结构中配位体的晶体静电场影响配位体中心过渡金属离子的 3d 轨道能级。温度变化影响原子在结晶学位置的振幅，引起内部原子间距的微小改变。随着温度增高，晶体场裂变参量 Δ 降低。在震荡体系中，晶体场谱中吸收谱带的宽度与摆动的金属-配位体间距相关。热膨胀导致吸收谱带变宽，谱带中心略微移向长波。这种吸收谱向低能方向的移动叫“红移”。

另一方面，增加压力使晶体结构压缩，氧化物和硅酸盐的克分子体积随压力加大而减小，配位体中原子间的距离缩短。金属-氧间距的收缩引起巨大的 Δ 增加，结果是加大了 3d 轨道能量分裂，吸收谱带移向短波。吸收谱带的这种向高能方向的移动叫“蓝移”。高压和增温还使共价性及电子振动的耦合增加，引起吸收谱带增强。

如铁铝榴石， $\text{Fe}_3^{2+}\text{Al}_2(\text{SiO}_4)_3$ ，其位于 8130cm^{-1} 、 5880cm^{-1} ，属于畸变八面体位置上的 8 配位 Fe^{2+} 的自旋允许谱带在 5GPa、10GPa、17.3GPa 压力下显示 $50 \sim 60\text{cm}^{-1}/\text{GPa}$ 的蓝移（Shankland, Duba and Moronow 1974, Balchan and Drickamer 1959）。在 12GPa 压力下，位于 7800cm^{-1} 、 6100cm^{-1} 的 8 配位 Fe^{2+} 的自旋允许谱带分别有 $60\text{cm}^{-1}/\text{GPa}$ 、 $90\text{cm}^{-1}/\text{GPa}$ 的蓝移（Smith and Langer 1983）。加热到 400°C ，所有 Fe^{2+} 的自旋允许谱带显示红移，位于 7690cm^{-1} 、 5850cm^{-1} 、 4175cm^{-1} 的吸收谱带分别移到 7620cm^{-1} 、 5695cm^{-1} 、 4150cm^{-1} （Parkin and Burns, 1980）。

又如硅铁钡矿（ $\text{BaFe}^{2+}\text{Si}_4\text{O}_{10}$ ）是由硅氧四面体组成的特殊双层结构的层状硅酸盐。两层硅氧四面体四元环共角顶形成双层，硅氧四面体间共用三个角顶，余一个活性氧，Fe 原子位于四个相邻活性氧组成的四配位平面上。一个大气压下硅铁钡矿的吸收光谱在 20000cm^{-1} 、 8000cm^{-1} 附近的吸收谱带与四配位平面上的 Fe^{2+} 有关。2.6GPa 压力作用引起硅酸盐层褶皱，四配位位置从平面对称转变为四面体对称。相应的吸收谱带移到 17200cm^{-1} 和 7150cm^{-1} （Abu-Eid et al. 1973）。

22.1.5 矿物的近红外振动光谱特征

前面主要涉及与矿物电子光谱有关的近红外吸收光谱。近红外区的振动光谱是由矿物分子伸缩、弯曲振动引起的合频及倍频谱带，包括 H_2O 、 OH^- 、 Al-OH 、 Fe-OH 、 Mg-OH 、 CO_3^{2-} 、 NH_4^+ 等。这些分子常见于含羟基/水矿物和碳酸盐矿物中。这些分子水和阳离子-羟基（M-OH）的伸缩振动的一级倍频谱带出现在 1400nm （ 7143cm^{-1} ）附近，水的伸缩-弯曲振动的合频谱带位于 1900nm （ 5263cm^{-1} ），羟基（M-OH）的伸缩-弯曲振动的合频谱带位于 2200nm （ 4545cm^{-1} ）附近的长波区（表 22-1）。

表 22-1 一些常见矿物的倍频、合频谱带振动谱带位置

基团/分子	谱带位置/nm	有关矿物
H_2O	$\sim 1400, \sim 1900$	水、绿柱石
Al-OH	$\sim 1400, 2170 \sim 2221$	叶蜡石、高岭石、蒙脱石、绢云母

续表

基团/分子	谱带位置/nm	有关矿物
Si-OH	~1400, ~2250	蛋白石、石英
Fe-OH	~1400, 2210~2300	黄钾铁矾、囊脱石、皂石、锂皂石
Mg-OH	~1400, 2300~2380	绿泥石、滑石、绿帘石、角闪石
CO ₃ ²⁻	1850~2200, 2300~2350	方解石、文石、白云石、菱镁矿、菱铁矿
NH ₄ ⁺	~2120	水铍长石

对含水矿物的近红外反射谱研究表明，所有含 OH 基团的矿物在此区间都有明显的吸收峰，吸收谱带的精细结构可以用来确定主要矿物成分及区分矿物的种类（Wetzel, 1983；Clark et al. 1990）。

矿物结构中的水来自地球内部流体，流体在地壳及地幔地球化学过程中的重要作用越来越受到地质学者重视。水是最重要的流体组分。水组分在熔体形成、演化及侵蚀过程中起着极重要的作用。水组分在地幔中的含量、保存状态及分布至今仍是难题。被广泛接受的假说认为地幔中水含量约为 0.1wt%，分布不均。在 0~70km 深度，主要保存在角闪石中。在 150~400km 深度，推测水以 OH 形式存在于橄榄石、辉石、石榴石中，极少量活动流体位于颗粒间。而过去认为这些矿物是不含水的，被称为无水干相。近十多年来，人们陆续发现上述幔源矿物含有微量 OH，Smith（1987）从理论分析认为，作为转换带的主要物相 β -橄榄石晶体结构适于容纳 OH 组分。McMillan 等（1989）证实了 β -橄榄石结构中存在 OH，含量约 0.06wt%。上述新进展是依靠分子光谱学实验研究（主要是红外光谱学研究）取得的。这些矿物中水的含量非常低，通常仅几个到几千个 mg/kg，需要使用单晶体进行测量。由于矿物在 $>2000\text{cm}^{-1}$ 的长波段有强吸收，影响了对水的存在形式的判定。使用近红外光谱可以帮助鉴别是 H₂O 还是 OH。如用近红外光谱就可能确定钾长石（KAlSi₃O₈）、斜长石（NaAl Si₃O₈-CaAl₂Si₂O₈ 系列矿物）中含有 H₂O 还是 OH（Johnson and Rossman, 2003）。

楣石（CaTiSiO₅）在 4100~4500cm⁻¹ 区间的吸收谱带是 OH 伸缩与 CaTiSiO₅ 骨架振动的合频。4100cm⁻¹ 附近的是 OH 伸缩与 Ti-O 弯曲振动的合频。在 4400cm⁻¹ 和 4500cm⁻¹ 谱带的归属还不确切，后者可能是 OH 伸缩与 Ti-OH 弯曲的合频。5000~7000cm⁻¹ 区间除了在 6968cm⁻¹ 有一个 OH 伸缩的倍频外，谱线是平的（Zang et al. 2001）。

含水矿物的结晶度、化学组成（特别是与 OH 键合的阳离子）和阳离子占位直接影响水的倍频、合频谱带的频率、强度及形态。

通常，特征吸收谱带形态的变化反映了结晶度的变化。特征吸收谱带的尖锐程度和强度随结晶度增加而增加，如同用 XRD 来测定结晶指数一样。通过近红外光谱测量矿物结晶度变化可以帮助从风化的蚀变矿物中区别出未蚀变的风化矿物（Fraser S. J. et al, 1990；Kruse and Hauff, 1991b）。

含水矿物结构中的阳离子取代强烈的引起吸收谱带的频率位置移动，因为不同的阳离子半径改变键长和晶体的几何结构。在组合谱带中，伸缩谱带和弯曲谱带位置的小改变是加和的，使得它们更容易区别（Post & Noble, 1993）。

含水的系列矿物中阳离子（如绿泥石、黑云母、金云母中的 Al、Fe、Mg；白云母、伊利石、钠云母、明矾石中 Al、Na、K）取代，会引起近红外光谱的吸收峰的位置、强

度、宽度和对称性的改变。通过测定相关吸收谱带的波长位置，可以半定量地计算出绿泥石、黑云母、金云母中的 Fe 和 Mg 的比例（McLeod et al，1987）。

绿泥石族矿物，其化学式为 $(\text{Mg}, \text{Fe})_5\text{Al}(\text{Si}_3\text{Al})\text{O}_{10}(\text{OH})_8$ ，随着晶格中 Fe^{2+} 含量增加，位于 2340nm、2260nm 和 1400nm 附近的倍频、合频谱带向长波方向移。

蒙皂石是一族有膨胀性、结构层间能吸收水或有机分子和具有进行阳离子交换能力的黏土矿物。蒙皂石族中蒙脱石、贝得石、绿脱石和皂石属二八面体结构；锂皂石、锌皂石属三八面体结构。这两族矿物的结构和化学组成分别与叶蜡石及滑石有关。有可能从这些蒙皂石的近红外光谱的谱带位置来鉴别该族矿物的种属和估计其 Al_2O_3 含量（Post and Noblr，1993）。

伊利石是结构上与云母相关的黏土矿物，伊利石因含有较多的硅、较少的钾以及黏土尺寸的粒径而有别于白云母。但伊利石族矿物因其基本层间阳离子是钾而彼此最相似。大多数伊利石同白云母一样是二八面体结构的。但有少数像黑云母一样是三八面体结构的。

伊利石多型变化是由于其结构中成分层的堆垛顺序不同，高温形成有序的 1M 型变种和低温无序的 2M 变种。XRD 是最常用而有效的鉴别伊利石多型的分析技术。

Kruse and Hauff（1991）总结出伊利石多型可供鉴别的吸收光谱特征（表 22-2）。接近 1400nm、1900nm 和 2200nm 以及 2300~2500nm 间的吸收谱带分别是层间水、结构水（ OH^- ）及一些八面体层的特征。2M 伊利石多型的反射光谱显示出和白云母的相同的特征。

表 22-2 伊利石多型的近红外光谱特征（据 Kruse and Hauff，1991）

光谱位置/nm	相关基团	谱带形态(2M)	谱带形态(1M)
1400	$\text{H}_2\text{O}+\text{OH}$	尖	宽,不对称
1900	H_2O	尖	宽,不对称
2200	OH	尖	宽,不对称
2300~2500	八面体层阳离子	较清晰	

Taranik（1990）对赤铁矿-针铁矿混合物的近红外光谱测量发现，纯赤铁矿在 850nm 的吸收谱带随着针铁矿的含量增加平稳地移向针铁矿的 925nm 位置（表 22-3）。

表 22-3 不同比例赤铁矿、针铁矿混合物的吸收谱带位置

矿物混合比例	吸收谱带的近似位置/nm	矿物混合比例	吸收谱带的近似位置/nm
100%赤铁矿	850	25%赤铁矿 75%针铁矿	900
75%赤铁矿 25%针铁矿	863	100%针铁矿	925
50%赤铁矿 50%针铁矿	875		

这种混合物的光谱可在有 8m 像素的机载 GERIS 数据中频繁见到。

碳酸盐类矿物的近红外光谱在 2300~2400nm 处有与羟基无关的特征吸收谱带（表 22-4）。方解石、文石、白云石在 1600~2550nm 区间都有至少 7 个起因于碳酸根离子振动的吸收谱带。这些谱带的位置及宽度具有鉴定特征，可用于鉴定这 3 种常见矿物，即便当 1900nm 处的吸收谱带被存在于流体包体中的少量水覆盖，也能鉴别。在古老岩石中，

最富集的矿物是方解石和白云石（Pettijohn, 1977）。位于 2000~2500nm 区间的 4 个谱带相对不受流体包体含量的影响。碳酸盐在 1200nm 附近的宽双峰不仅证明 Fe^{2+} 的存在，还有助于矿物的鉴定，其吸收谱带强度与 FeO 的百分含量有关。

表 22-4 一些碳酸盐矿物的 CO_3^{2-} 合频谱带位置

方解石 CaCO_3	白云石 $(\text{Mg}, \text{Ca})\text{CO}_3$	菱铁矿 FeCO_3	菱镁矿 MgCO_3
2335~2340nm	2316nm>2330nm(含 Fe^{2+} 的)	2330nm	2310nm

矿物颗粒的大小对碳酸盐光谱参数的影响仅是样品的总体亮度和吸收谱带的绝对强度。谱带数目、位置、宽度及相对强度是不受影响的。0.01wt% 的有机质通常会引起样品亮度及吸收强度的明显降低，这意味着有机质是呈十分细的浸染状。需要用此光谱与粒状蒙脱石和炭黑的混合物进行比较。在许多碳酸盐岩石中，直到炭黑加到 2%，此混合物不显示激烈的光谱弱化（Hunt and Salisbury, 1971）。

天然的水铵长石， $(\text{NH}_4)\text{AlSi}_3\text{O}_8 \cdot 1/2\text{H}_2\text{O}$ ，是 1964 年在北 California 的一个热泉沉积物中首次发现并描述的含水的铵铝硅酸盐矿物。离子半径为 14.3nm 的 NH_4^+ 取代长石结构中半径为 13.3nm 的 K^+ 或其他碱金属阳离子。水铵长石以在 2020nm 的大肩状吸收和在 2120nm 附近的 NH_4^+ 离子的宽吸收谱带特征有别于其他黏土矿物（Krohn, 1986）。

22.2 近红外光谱的地学应用

22.2.1 近红外光谱在遥感测量中的应用

近红外波段在遥感测量中占有十分重要的位置。与地球科学有关的近红外光谱分析包括地球和行星表面的物质组成、结构及成因、矿产资源调查等。对天体的测量主要是通过地面或航天器上的望远镜实现。少量陨石及通过登月火箭或宇航员从月球带回地球的样品是在实验室测量的。对地球资源的遥感调查是将光谱仪放在飞机或卫星上测量的。所有遥感测量获得的是表面反射光谱，对这些光谱数据的判读和解释，是以前述的矿物近红外光谱的实验室研究及野外样品实测数据为基础的。但是，遥感测量数据一般不能与实验室实测数据直接对比，如在地球、月球、火星、水星、金星的赤道位置白天最高温度分别是 315K、400K、280K、650K、740K。还要考虑到大气圈中 H_2O 及 CO_2 分别在 1330nm、1440nm、1600nm、1950nm、2080nm 及 820nm、950nm、1400nm、1900nm、2850nm、3100nm 吸收光谱的干扰，进行必要的校正。

对来自月球的矿物的吸收光谱研究（Adams, 1974；Burns, 1989a）表明，其钙质斜长石结构中含有少量 Ca^{2+} ，它取代了位于大配位体位置中的 Fe^{2+} ，引起在 1250nm 的 Fe^{2+} 的弱晶体场吸收。橄榄石在 M_1 位置的 Fe^{2+} 产生 1050nm 谱带，在 M_2 位置的 Fe^{2+} 则产生 850nm、1250nm 谱带。辉石中在 M_2 位置上的 Fe^{2+} 在 2000nm 区产生晶体场吸收，该谱带比位于 1000nm 的更适合来辨别辉石类型。但是，氧化物结构中或硅酸盐玻璃中的四配位 Fe^{2+} 在 1800~1900nm 区间的吸收（Bell et al 1975；Nolet et al 1979）会使辉石晶

体化学和含玄武质玻璃的混合样品的遥感反射光谱的解释变得困难 (Farr et al, 1980; Dyar and Burns, 1981)。

根据从地面望远镜得到的月球上 Copernicus 火山口地区的表面反射光谱 (图 22-4), 在火山口壁 (图 22-4a) 及底部 (图 22-4b) 的反射光谱有属于斜方辉石的、中心在 920nm 附近的弱谱带, 推测是具有苏长岩成分的典型高原土壤。其他底部地区 (图 22-4c) 是含辉石及玻璃的冲击熔体。火山口中央山峰地区 (图 22-4d) 有中心在 1050nm 附近、呈现宽的复合谱带, 是以橄榄石为主的镁铁质硅酸盐矿物, 推测为橄长岩。

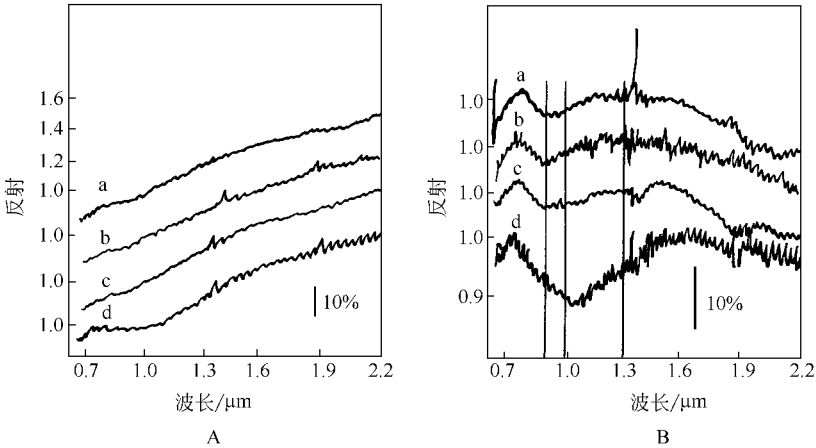


图 22-4 地面望远镜观测月球表面反射光谱 (据 Pieters et al. 1985)

A—反射谱 B—为连续剥离后的吸收谱 a—火山口壁 b—底部地区含斜方辉石推测为苏长岩成分 c—含辉石及玻璃的冲击熔体底部地区 d—含橄榄石的中央山峰地区推测为橄长岩

水星离太阳最近, 很难用望远镜测量它的表面反射光谱。只用 1000nm 谱带来鉴定水星上的辉石并推测它们的组成只会得到模糊的结论。早期研究参照在月球温度测量的斜方辉石光谱的最低限, 将在 890nm 处分辨出来的一个弱宽谱带归因为斜方辉石中 Fe^{2+} 的高温晶体场谱带, 推测在水星表面的低铁玄武岩中有顽火辉石存在。后来的研究提出 890nm 波长的光谱可能是含三价铁辉石的 Fe^{3+} 的晶体场跃迁。

火星大气圈中的 CO_2 对近红外光谱有干扰。火星表面按反射率的高低分为亮区和暗区。通过地面望远镜观测到火星亮区光谱除了大气中 CO_2 的光谱外, 占优势谱带位于 870nm 和 620nm, 归因于 Fe^{3+} 的晶体场跃迁。推测火星表面有含三价铁的物相的氧化的风化产物, 包括结晶度差的氧化物、黏土矿物、磷酸盐。暗区的光谱在 1000nm 区显示与亚钙辉石有关的

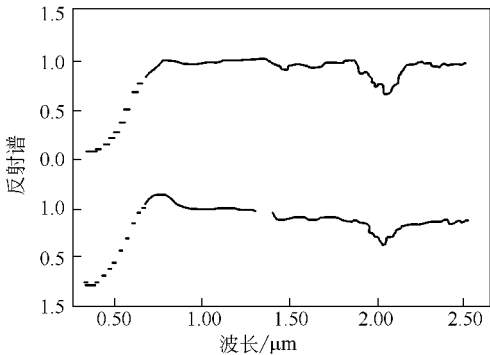


图 22-5 地面望远镜观测到火星亮区及暗区反射光谱 (Singer 1985)
亮区光谱 (上), 暗区光谱 (下)

Fe^{3+} 的谱带。但是, 由于大气 CO_2 的干扰, 不能根据 2000nm 的谱带来鉴定辉石的种属和组成。通过暗区望远镜光谱与被认为来自火星表面的辉玻无球粒陨石的反射光谱(含易变辉石-普通辉石)进行对比, 证明火星表面有玄武质岩石(Singer and McSween, 1992)。

近红外遥感技术研究地球表面的地学问题至少已有 30 多年。如利用遥感数据研究小比例尺地质填图、资源、土地利用及环境污染状况等。由于近红外辐射对 M-OH 键振动的高灵敏性, 近红外光谱技术是研究含水矿物的有力手段(Farmer, 1974; Hunt, 1977; Hunt and Ashley, 1979; Hunt and Salisbury, 1970)。利用近红外遥感技术进行地球资源、环境调查等方面的应用, 就是以含水矿物结构中的 H_2O 、OH 的倍频、合频谱带变化为基础, 来确定它们与矿化作用、环境变化的关系。

很多金属矿床是通过热液交代作用形成的。热液交代作用还引起矿体周围的岩石发生蚀变, 形成以特定含水矿物组合的蚀变带。含水矿物在矿产资源勘察中的意义是:

- ① 提供矿化环境的特征, 如交代类型和交代带等;
- ② 鉴别原岩类型, 如: 鉴定出高岭石, 表明其原岩是长英质岩石, 发现蒙脱石表明原岩是镁铁质岩石;
- ③ 指示矿化关系, 富 Mg 绿泥石接近矿化中心; 和钠云母比起来, 富 K 的白云母更和矿化作用有关;
- ④ 指示风化范围和过程, 如三水铝石表示晚期的铝土质环境;
- ⑤ 指示矿化作用的化学过程(如 K/Na 交代)及温度(叶蜡石、黄玉、地开石等矿物是高温类矿物)。

在内华达的许多浸染型金矿中, 伊利石是与金矿化伴生的黏土矿物。由于它反映金矿体周围的水平、垂直带, 是这种矿床类型中重要的矿物种。伊利石随温度变化改变其晶体结构, 可能作为地质温度计测定成矿, 及深部和地表流体的温度。对 Carlin 金矿中的 2M 伊利石多型高品级金矿带伴生。

在一些内华达矿床中(如 Preble), 相对高温形成的 2M 型伊利石也产在矿带中。与蒙脱石族黏土交生的无序的低温 1M 型伊利石位于主矿带周围。蒙脱石族和高岭石族黏土, 通常在未交代的岩石中, 很少或不在这些带中。在 Gold Bar, 伊利石的近红外吸收光谱的强度和对称性与结晶度呈正相关。2M 多型在 2200nm 的尖谱带和 1M 多型在 2200nm 的明显不对称谱带间的差别是很清晰的, 机载和卫星载光谱仪有应用潜力。

Kruse 和 Hauff 1991 年着手为在 2002 年发射的澳大利亚资源信息和环境卫星 (ARIES-1) 的商业合作计划进行了前期遥感研究。根据在美国内华达州 Comstock 采矿区收集到的卫星多光谱反射数据鉴定出与深部岩浆热液事件有关的相对高温成因的地开石、叶蜡石, 成因不同的低温高岭石、明矾石、绿泥石, 并按伊利石的 Al-OH 的合频谱带位置将其分为: Al-OH 吸收谱带位于 2194nm 的短波伊利石(与高岭类黏土、明矾石共存, 代表前期的泥质组合)、Al-OH 吸收谱带位于 2212nm 的长波伊利石(构成青盘岩质或绢英质组合的一部分)和 Al-OH 吸收谱带位于 2204nm 的中间伊利石 3 种。遥感光谱数据进行单矿物填图, 在此基础上形成蚀变图(Yang et. al. 1999)。使用便携近红外光谱仪进行了野外现场测量, 根据这些蚀变矿物组合划分出浅色带、青盘岩化带及云英岩化带, 分别

形成不同时代，不同成因类型的 Au-Ag 矿化。从遥感数据分析得到的蚀变矿物类型和分布同野外直接测量的数据一致（图 22-6）。证明了使用 ARI-ES-1 卫星在选定地区提供高精度矿物填图的多光谱数据的可行性，以及在矿产勘察方面的潜在应用前景（Yang et. al. 1999）。

根据蚀变矿物的可见-近红外光谱学特征，利用从东天山大东沟至赤湖地区的遥感数据确定出了蚀变异常。经异常点实地查证，实测坐标与蚀变异常中心坐标吻合率达 85.9%（张玉君等，2002）。

以地面实测岩石矿物的反射光谱特征测量为基础进行的用高光谱遥感技术识别岩石类型的试验研究表明：依据不同岩石的反射光谱特征，通过对高光谱遥感图像与光谱信息处理，可以实现对各种岩石类型的定性分析到定量识别（王青华等，2000）。

利用近红外遥感技术进行环境监测也很有成效。如澳大利亚最大的物资运输港 Hedland 港，年吞吐量超过 7000 万 t，相当部分是铁矿石。铁矿石装卸过程扬起的矿石粉尘对沿港湾周围环境造成严重污染。野外试验表明，使用 500 ~ 2500nm 波长的遥感光谱对沿港湾生长的红树林测



图 22-6 据遥感光谱数据绘出的 Comstock 矿区水热蚀变图（据 Yang et. al. 1999）

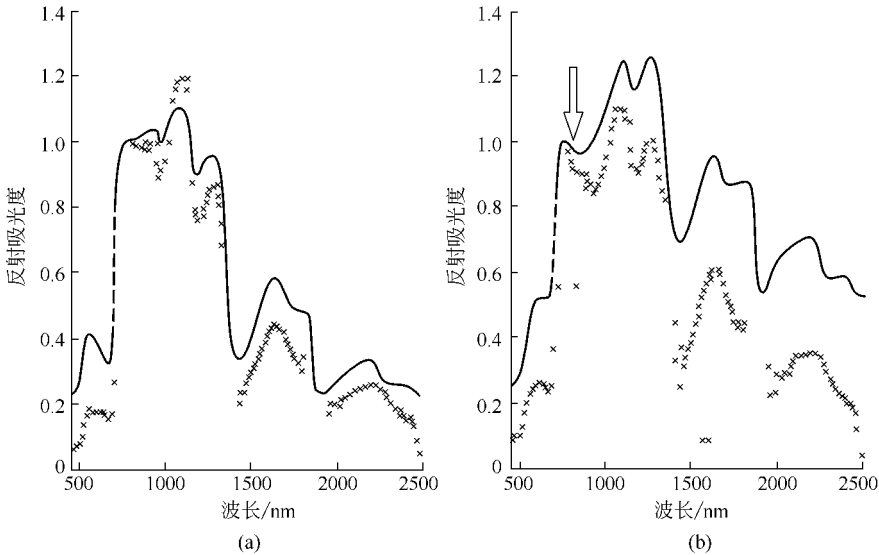


图 22-7 地面（实线）及航测（叉线）红树叶反射光谱
(a) 清洁的 (b) 有粉尘的箭头指示氧化铁的吸收区

量证实（图 22-7），红树叶表面尘土附着物的反射光谱位于 860nm 的吸收正是富含赤铁矿的铁氧化物粉尘（Ong et. al. 1999）。

煤层自燃不仅损失大量的煤炭资源，还引起一系列经济、环境及生态问题。通过对煤矿煤层自燃形成的烧变岩的反射光谱特征分析，来实现应用遥感技术完成火区位置、范围的快速圈定和动态监测（祝善友等，2003）。

22.2.2 近红外光谱在地质调查及矿产资源勘察中的应用

相对地面地质调查、资源调查而言，遥感近红外光谱提供范围广、小比例尺、区域性的地球表面信息，是普查性的工作。遥感光谱数据的解释及应用，是以对应的地面测量工作为基础的。地面近红外光谱测量直接面对样品，可以深入到局部细微变化进行测量，可以测量钻机提取的深部样品，是勘察、评价工作。

在实验室，有很多分析测量手段可供选择。近红外光谱技术的特点是，除了对矿物结构的基础矿物学研究需要单晶体定向制备外，大多数测量不用进行前处理，不污染或损伤样品。模型化的定性/定量分析技术，分析者不用去理会光谱的理论解释。操作简便，测量效率高。近二十年来，由于便携近红外光谱仪的出现，实现了在野外现场、原位测量。通过包括黏土在内的层状硅酸盐矿物的光谱特征来作出热液蚀变带，使勘察周期缩短，成本降低（Pontual and Merry, 1985；Yang et. al. 1998）。图 22-8 是一些常见矿物的近红外光谱。

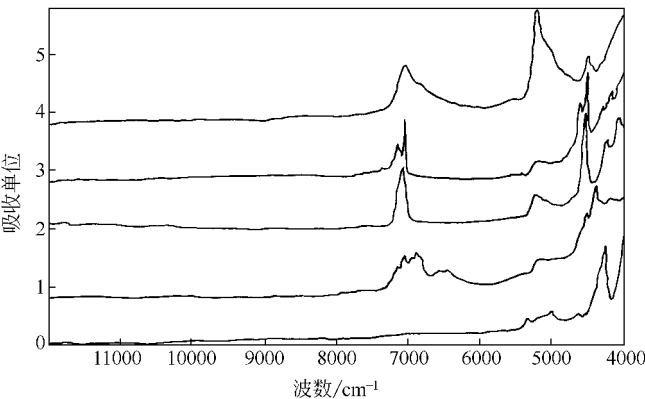


图 22-8 岩心剖面的近红外光谱（据 Yang et. al. 1998 年）
从上到下：蒙脱石、高岭石、伊利石、方解石

在对日本的 Hishikari 低温热液 Au-Ag 矿床钻孔岩心的近红外光谱测量中，根据 Al-OH、Mg (Fe)-OH 吸收谱带的存在与否，将剖面粗分成两带。位于 2200~2225nm 的吸收谱带是白色云母的 Al-OH 合频，位于 2250~2260nm 的吸收谱带是绿泥石的 Mg (Fe)-OH 合频，位于 2160~2180nm 的吸收谱带是高岭石的 Al-OH 合频，位于 2200~2208nm 的吸收谱带是蒙脱石的 Al-OH 合频。从上而下，71.5~79.5m 段是蒙脱石（2200~2208nm）±高岭石（2160~2180nm）组合，81.7~99.6m 段是高岭石±蒙脱石组合，189.6~229.2m 段是绿泥石±蒙脱石组合，248.2~256.2m 段是白色云母+绿泥石组合。根据绿泥石的 Mg (Fe)-OH 合频谱带细分为位置在 2318~2330nm 的富 Mg 绿泥石，在 2340~2348nm 的富 Fe 绿泥石。含金石英脉位于下部含 Mg 绿泥石的火山岩中（图 22-9 Yang et. al. 1998）。

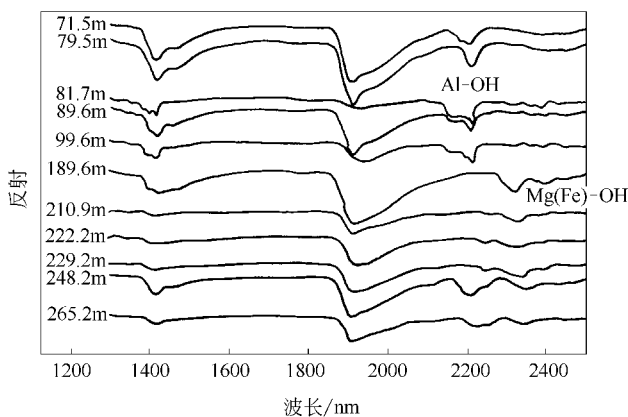


图 22-9 一些矿物的近红外光谱

从上到下：蒙脱石、高岭石、伊利石、三水铝石、方解石

通过蚀变矿物的近红外光谱测量分析，还可以对地热资源进行评价。如对新西兰的 Broadlands-Ohaaki 地热场的 8 个钻井岩心进行近红外光谱测量，测量出伊利石含铵伊利石、蒙脱石、绿泥石等和热液蚀变有关的含水层状硅酸盐矿物。根据蚀变矿物的光谱特征及组合特征作出上部低温蚀变组合，中部蚀变组合和下部蚀变组合绘出的蚀变带剖面图和现在地热场的热状态基本一致。有的蚀变带分界线切过等温线，是矿物组合与当前温度间由于地热系统演化所引起的局部不平衡 (Yang and Browne, 1999)。

金刚石是在 140km 以下的地幔形成的，由金伯利岩带到地表。金伯利岩是在蛇纹石、云母、钙钛矿、钛铁矿、尖晶石基质中有橄榄石斑晶和捕虏晶的斑状岩石。基质中的碳酸盐可能是原生的或次生的。橄榄石和蛇纹石是最主要的矿物。来自南非金伯利岩样品的典型光谱在 1600nm 附近有接近 10% 的低反射，在 1400nm, 1900nm, 2320nm 附近有窄而明显的谱带，在 2380nm 附近有一非常弱的谱带。后两个谱带主要归因于蛇纹石中 Mg-OH 的弯曲及伸缩振动模式。在 1400nm 和 1900nm 的组合谱带证明水的存在。在 1400nm 的是蛇纹石和金云母中的羟基。许多金伯利岩样品在可见光区间完全消失。两个切出新鲜表面的和风化表面的南非样品光谱显示相同的光谱特征，风化表面的整个光谱更清晰，吸收谱带更强 (Kingston, 1986)。

岩石、土壤水系沉积物是地球化学调查的研究对象。早在 20 世纪 60 年代就开始了土壤样品的近红外反射光谱测量研究，发现水分含量、有机质、颗粒大小都对光谱有影响 (Bowers and Hanks, 1965)。

湿样品中的水会降低样品光谱的强度，降低吸收谱带的对比度，甚至导致完全模糊的光谱。所以，样品应该是“干”的，潮湿的样品可用微波来干燥。水与样品有没有化学键，其近红外光谱是不同的，水的主要吸收谱带位于 1450nm (6896nm)、1920nm (5208nm)、2700nm，位于 1900nm 的吸收谱带显示水的含量。

通常，颗粒越小（粉末），在近红外区产生越明晰的光谱。相对于块状岩石，粉末样品由于其较强的表面散射及较弱的体散射，产生较弱的吸收。因此，块状岩石的近红外光谱要比它们的粉末光谱好。有些包裹在表面的细粒风化物质会影响岩石的光谱感应度。被研磨过的土壤的近红外光谱通常比原岩的弱。

有机质在碳酸盐及石墨页岩中很常见。岩石表面的地衣难于被发现，有机纤维的宽吸

收位于 1650nm、2100nm、2300nm 附近。它们抑制岩石的吸收强度（对比度），伴随噪音增加，可能需要对光谱做平滑处理。有机质也影响土壤光谱的对比度。

应用近红外光谱学方法快速测量土壤中化合水、有机碳总量是可行的（龙梅及裴世桥，2003 年）。对我国不同地区土壤剖面样品矿物组成（包括高岭石、水云母、石英、长石、碳酸盐矿物等）、总水、总有机碳、高岭石有序度的初步近红外定量分析测试表明，近红外反射光谱作为快速可靠的技术，在地质调查中是很有潜力的。

22. 2. 3 近红外光谱研究宝石学问题

早期开展的围绕对矿物的基础晶体化学问题的可见-近红外光谱学研究是用单晶体来测量的，它们当中有相当大部分就是宝石级矿物。所以，很多涉及宝石的物化性质（包括宝石的颜色、多色性等）有关的基础矿物学研究，在 20 世纪 70 年代前后都已详细研究过，三十多年来不断有新的进展，特别是在宝石的质量鉴定方面。

例如对经过漂白并经聚合物充填的 B 货翡翠的鉴定（图 22-10）。B 货翡翠以它的近红外光谱在 5982cm⁻¹、5916cm⁻¹、5882cm⁻¹，5781cm⁻¹、5673cm⁻¹ 的吸收谱带是树脂类聚合物的饱和、不饱和烃 CH₂、CH₃ 的一级伸缩倍频振动，位于 4679cm⁻¹、4620cm⁻¹、4560cm⁻¹、4331cm⁻¹、4258cm⁻¹ 的吸收谱带以 CH₂、CH₃ 的伸缩-弯曲振动的合频为特征。通常测量透过谱效果最好。镶嵌好的宝石有时无法测透过谱，可以使用显微镜、漫反射方法或近红外光学纤维探针来测量反射光谱（Quek and Tan, 1997；Gao and Zhang, 1998）。

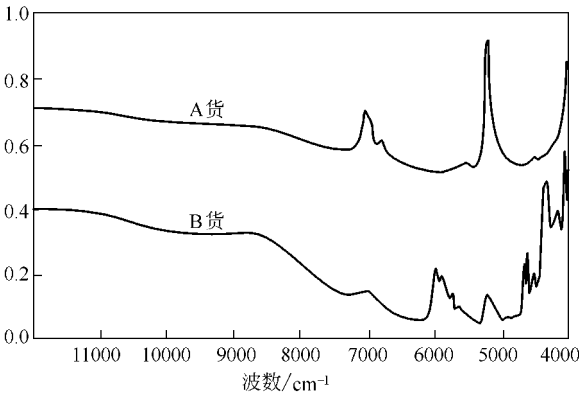


图 22-10 翡翠的近红外漫反射光谱

欧泊是有变彩效应的蛋白石（SiO₂ · nH₂O），天然欧泊的近红外光谱位于 5270nm 的尖峰及 5150nm 的肩分别是孤立水分子和液态水伸缩振动 + 歪曲振动的合频，位于 4500nm 的谱带是 Si-OH 的合频振动（Langer and Florke, 1974；Graetsch, 1994）。聚合物优化处理欧泊及塑料仿欧泊以在 5700~6100nm 的 C-H 倍频谱带有别于天然欧泊。含少量水的合成欧泊在近红外区的吸收谱带弱，也有别于天然欧泊（Fritsch and Stockton, 1987）。

23. 近红外遥感在农业上的应用

23.1 近红外遥感的数据获取

23.1.1 航天、航空平台

23.1.1.1 含近红外波段的传感器简介

近红外遥感技术在农业上应用的主要目标是为实施精准农业提供信息源。没有高密度、低成本、快速准确的农田信息获取技术，精准农业也就难以得到大面积应用。作物信息，如叶面积指数、植株形态信息、生化组分信息及干物质积累等，目前也还依赖于破坏性采样。卫星遥感和航空遥感正成为农田信息获取的主要来源，用遥感获取土壤和植物参数已经得到广泛的应用，随着卫星遥感精度的提高，卫星遥感技术在农业上的应用前景更为广阔。

目前在轨运行的主要商用遥感卫星有美国的资源卫星 Landsat-7（1999 年）、IKONOS（1999 年）、EOS-AM1（1999 年）、Quick Bird（1999 年）、Orbview3；法国的地球观测卫星 SPOT-4（1998 年）、SPOT-5（2002 年），印度的遥感卫星 IRS-1；中国的资源卫星 CBERS-1（1999 年）；加拿大的雷达卫星 RadarSAT-1（1995 年）等。遥感器的光谱范围涵盖了可见光-近红外-短波红外-热红外和微波；空间分辨率从数百米提高到了几米甚至小于 1 米；用小卫星组网可将重复观测时相由原来的数十天提高到几天，甚至几个小时，极大地拓展了遥感应用范围。表 23-1（a）列出部分运行和将要发射运行的光学卫星。

此外，也研制了许多基于航空平台的机载传感器，大都包含近红外波段，如表 23-1（b）所示，为近红外遥感技术在农业上的应用奠定了基础。

表 23-1（a） 含有近红外波段的部分卫星及传感器一览

高空间分辨率卫星				
卫星	传感器	分辨率	波段设置	幅宽
<u>ORBView-3</u>	ORBView-3	1m Pan 4m MSS	MSS 4 bands Pan 1 band ±45° off nadir	8 km
<u>QuickBird 1 & 2</u>	QuickBird	1m Pan 3. 5m MSS	Pan 450-900 nm MSS - VNIR	22km×22km 40km×40km 22km×220km 704km Accessable
<u>Ikonos 1&2</u>	Ikonos	1m Pan 4m MSS	MSS 4 bands Pan 1 band	11km
<u>ALOS</u>	AVNIR-2	2. 5m MSS	ch1:0. 42~0. 50μm ch2:0. 52~0. 60μm ch3:0. 61~0. 69μm ch4:0. 76~0. 89μm ±40° (cross track)	70km

高空间分辨率卫星				
卫星	传感器	分辨率	波段设置	幅宽
<u>SPOT 5</u>	SPOT 5 Vegetation	2.5 & 5 Pan 10m MSS 20m SWIR band 1km Vegetation	B1(0.50~0.59) μ m B2(0.61~0.68) μ m B3(0.79~0.89) μ m SWIR 1.58~1.75 μ m Pan 0.51~0.73 μ m	60km
高光谱分辨率卫星				
卫星	传感器	分辨率	波段设置	幅宽
<u>OrbView-4</u>	<u>WarFighter</u>	1m Pan 4m MSS 8m HSI	200 bands 0.45~5.0 μ m @11.4nm 80 band 3.0~5.0 μ m @25nm MSS 4 bands Pan 1 band	8km $\times$ 8km Pan/MSS 5km $\times$ 20km HSI
SIMSA	HSI	25m	220 bands 430~2400nm @6~20nm $\pm$ 30 $^{\circ}$ cross-track pointing	25km
<u>NEMO</u>	COIS & PIC	5m Pan 30m HSI	210 bands 400~2500 nm @10 nm PAN 0.5~0.7 μ m ORASIS-real-time HSI processing in space	30 km
<u>ARIES-1</u>	ARIES-1	30m nadir 10m Pan	32 bands 400~1100nm @22nm 32 bands 2000~2500nm @16nm 32bands 1000~2000nm @31nm Pointable $\pm$ 30 $^{\circ}$ Panchromatic	15km nadir
PRISM		50m	158 bands 450~2350nm@12nm 1 bands 3800nm@600nm 3 bands 8000~12300@1000nm	50km
<u>EO-1</u>	<u>Hyperion</u>	30m	220 bands 0.4 to 2.5 μ m @10nm Grating Imaging Spec	7.5 km by 100 km
<u>EO-1</u>	LEISA AC	250m	309 Bands 0.850~1.60 μ m @2.4nm Wedge Imaging Spec	7.5 km by 100 km
<u>Envisat-1</u>	<u>MERIS</u>	300m @nadir and 1200m global	15 band programmable 390nm to 1040nm @2.5nm	1150 km
ADEOS-2	GLI	250/1000	19 bands 375~865nm@8~20nm 4 bands 460~825nm@50~110nm 6 bands 1050~2210nm@20~220nm 7 bands 3715~12000nm@330~1000nm	
多光谱卫星				
卫星	传感器	分辨率	波段设置	幅宽
<u>Landsat 7</u>	ETM+	15m Pan 30mMSS 60mTIR	7 Bands MSS VNIR-TIR	185km $\times$ 170km

多光谱卫星				
卫星	传感器	分辨率	波段设置	幅宽
<u>EO-1</u>	<u>ALI</u>	10m Pan 30m MSS	PAN 0.480~0.680 μ m 9 band MSS 0.433~2.35 μ m	185 km
<u>Envisat-1</u>	AATSR	1km	VNIR radiometer	500km
<u>Terra (EOS AM-1)</u>	<u>ASTER</u>	15m VNIR 30m SWIR 90m TIR	3 bands VNIR, 7 bands SWIR 5 bands TIR Stereo 0.76~0.86 μ m	60 km
<u>Terra (EOS AM-1)</u>	<u>MODIS</u>	250m, 500,1000m	36 Bands VNIR-TIR	2330km
<u>Terra (EOS AM-1)</u>	<u>MISR</u>	275m, 550m,1100m	4 Bands MSS VNIR 443nm,555nm,670nm,and865nm Multi Angle+Fore/Aft/Nadir	360 km
<u>CBERS-1</u>	CCD Camera	20m	CCD(high-res. ,5bands)	113 km

表 23-1 (b) 含有近红外波段的部分机载成像光谱仪一览

遥感器	近红外波段/nm	总波段数	波段宽/nm	研制国家
PLI-PMI	403~805	288	2.5	加拿大
CASI	430~870	288	2.9	加拿大
SFSI	1200~2400	122	10.0	加拿大
AIS-1	900~2100	128	9.3	美国
AIS-2	800~1600	128	10.6	美国
AVIRIS	400~2450	224	9.4~16.0	美国
ASAS	455~873	28	15.0	美国
改进 ASAS	400~1060	62	11.5	美国
GERIS	1000~2000	63	120.0	美国
ROSIS	450~850	128	<5.0	德国
IMS	800~1700	128	12.5	法国
AMSS	500~980	46	20.0~71.0	澳大利亚
DAIS-7915	1000~1800	79	100.0	GER 为 ESA 研制
MIVIS	433~833	102	20.0	美国 Daedalus
HYDICE	400~2500	256	4.7/11.7	NRL
MAIS	460~1100	71	—	中国
OMIS	460~1100	128	10	中国
PHI	400~850	244	<5	中国

注：通常按遥感器的波段划分为可见光（400~750nm）、近红外（750~1300nm）和短波红外（1300~2500nm），为与其他章节统一，本章内容中除特别指出之外，将 750~2500nm 范围统称为近红外。

23. 1. 1. 2 PHI 成像光谱仪

目前国内机载研究上使用的面阵推扫型成像光谱仪 PHI（Pushbroom hyperspectral imager），其主要光学参数如表 23-2 所示，波段设置如表 23-3 所示。

表 23-2 PHI 仪器主要参数

光学参数	数据采集参数	光学参数	数据采集参数
总视场:21° 瞬时视场:1mrad 光谱范围:400~800nm 光谱分辨率:<5nm 光谱采样:244	输入数据速率:14MB/s 帧存容量:128KB (120 波段) PC 插槽:ISA 总线 内存地址:(Hex)D000;0-D000;FFFF	像元数:360P/L 最大扫描率:60s/s 数据量化:12bit 电源:27V DC	端口:(Hex)300 中断:IRQ5

表 23-3 PHI 波段设置

波 段	波 长	波 段	波 长	波 段	波 长	波 段	波 长
1	411.9	21	569.3	41	647	61	723.3
2	419.7	22	573.2	42	650.8	62	727
3	427.5	23	577.1	43	654.6	63	730.7
4	435.4	24	581	44	658.5	64	734.4
5	443.2	25	585	45	662.3	65	738.2
6	451.1	26	588.9	46	666.1	66	741.9
7	458.9	27	592.8	47	669.9	67	745.6
8	466.7	28	596.7	48	673.8	68	749.2
9	474.5	29	600.6	49	677.6	69	752.9
10	482.3	30	604.6	50	681.5	70	760.2
11	490.1	31	608.5	51	685.3	71	767.4
12	498	32	612.4	52	689.2	72	774.7
13	506.1	33	616.3	53	693	73	782
14	514.1	34	620.2	54	696.9	74	789.2
15	522.1	35	624.2	55	700.7	75	796.4
16	530.2	36	628.1	56	704.6	76	803.6
17	538.1	37	632	57	708.4	77	810.9
18	545.9	38	635.8	58	712.1	78	818.2
19	553.7	39	639.5	59	715.8	79	825.5
20	561.5	40	643.3	60	719.6	80	832.8

23. 1. 1. 3 OMIS 成像光谱仪

机载实用型模块化成像光谱仪 OMIS（Operative Modular Imaging Spectrometer）采用光机扫描方式成像，扫描部件为一个 45°旋转反射镜。OMIS 成像子系统采用特别的双光学通道设计，即使用成像主镜的中心遮光区，形成两个独立的光学通道——主通道和辅助通道。由辅助通道承担短波红外 I 光谱区（1.2~1.9μm），而主通道承担其他 4 个光谱区，即使困难的光谱分离得以实现，又提高了光学效率。工作波段覆盖了 0.46~12.5μm 可见近红外至热红外的全部大气窗口，设计了两种工作模块 OMIS-Ⅰ型和 OMIS-Ⅱ型。

仪器包括机载和地面两大部分。机载部分主要有：光机头部，数据采集电子学系统，高速大容量数据记录系统，机上辐射定标系统，实时图像监视系统，高精度陀螺稳定平

台，GPS 定位系统等；地面部分主要有：地面数据快速回放及预处理系统，地面检测及光谱与辐射定标系统。成像光谱仪系统的性能评价包含了对仪器技术参数的测试和定标，以及各分系统功能检测。

OMIS 主要技术参数见表 23-4，其光学设计示意图和仪器外观图如图 23-1 所示。

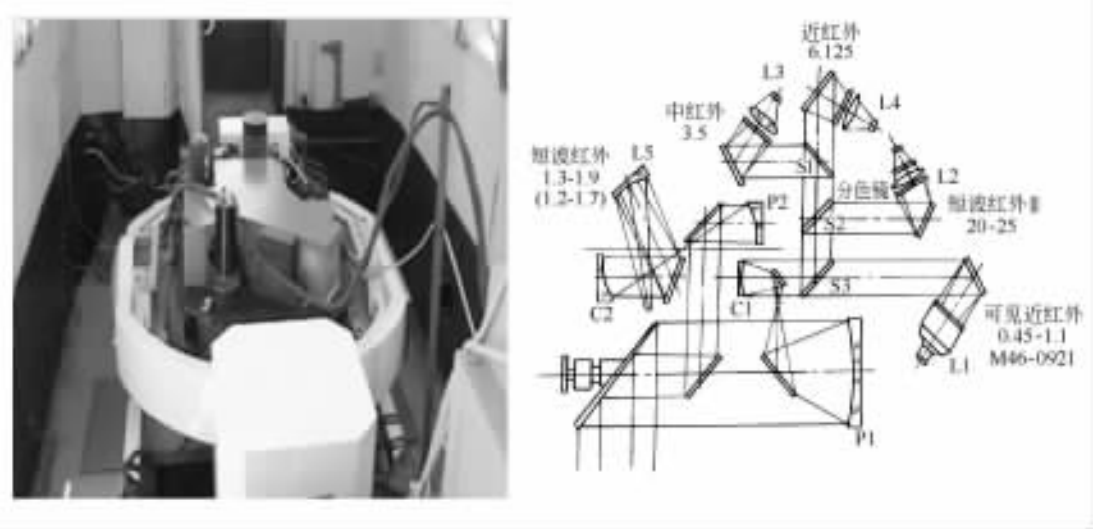


图 23-1 OMIS 光学设计图及安装在飞机上的仪器及平台的外观图

表 23-4 OMIS 成像光谱仪主要技术参数

OMIS- I				OMIS- II			
总波段数			128	总波段数			68
光谱仪	光谱范围/ μm	光谱分辨率/ nm	波段数	光谱仪	光谱范围/ μm	光谱分辨率/ nm	波段数
	0.46~1.1	10	64		0.4~1.1	10	64
	1.06~1.70	40	16		1.55~1.75		1
	2.0~2.5	15	32		2.08~2.35		1
	3.0~5.0	250	8		3.0~5.0		1
	8.0~12.5	500	8		8.0~12.5		1
瞬时视场/ mrad		3		1.5/3 可选			
总视场($^{\circ}$)		>70					
扫描率/(s/s)		5、10、15、20 可选					
行像元数		512		1024/512			
数据编码/bit		12					
最大数据率(Mbps)		21.05					
探测器		Si、InGaAs、InSb、MCT 线列		Si 线列、InGaAs 单元、InSb/MCT 双色			

23.1.2 地面平台

大田冠层光谱测量常用的仪器为 ASD Fieldspec FR2500 光谱仪（图 23-2），该光谱仪

的主要参数如下：

光谱范围：350～2500nm；采样间隔：1.4nm（350～1000nm 区间）和 2nm（在 1000～2500nm 区间）；光谱分辨率：3nm@700 nm、10nm @400nm&2100nm；波长精度：1nm；测定速度：固定扫描时间为 0.1s，光谱平均最多可达 31800 次；输出波段数：2150（重采样间隔 1nm）；测定记录方式：3 分段探测，自动优化设定增益和积分时间，自动消除暗电流；记录参数：DN（raw）值、相对反射率（ref）或辐射值（rad）可选；观测通道：单通道，光纤传输，非同步参考板测定；等效辐射噪声（NE Δ L）：UV/VNIR1.4×10⁻⁹ W/cm²/nm/sr @ 700nm，NIR2.4×10⁻⁹ W/cm²/nm/sr @ 1400nm，NIR8.8×10⁻⁹ W/cm²/nm/sr@2100nm。



图 23-2 美国 ASD 公司的 ProFR 型地物光谱仪



图 23-3 美国 Trimble Ag-132 型亚米级 GPS 系统

具体方法是在测量目标反射太阳辐射的前后，准同步地测量参考板反射的太阳辐射光谱，参考板的光谱反射率是经过严格的实验室标定的，通过比值方法，计算目标的光谱反射率。公式如下：

$$R_T^n = \frac{DN_T^n}{DN_R^n} \times R_R^n \tag{23-1}$$

式中 R_T^n 和 R_R^n 分别为目标和参考板在第 n 个波段的光谱反射率， DN_T^n 和 DN_R^n 为目标和参考板在第 n 个波段反射的太阳光强度。

星-机-地同步测量时通常需采用 GPS 系统定位，图 23-3 为美国 Trimble 生产的 Ag-132 型亚米级 GPS 系统。

23.2 近红外在作物水分监测上的应用

23.2.1 植被在可见近红外波段的光学属性

植被地物的生物、化学和物理参数决定了其光学属性，如光谱反射率、透过率和吸收。对高光谱精准农业遥感而言，理解电磁波辐射与植被叶片及冠层的相互作用是开展作物生化、物理参量遥感反演的基础。

由于植被地物共有特征是叶片丰富的含水量和服务于光合作用的叶绿素等色素，所以

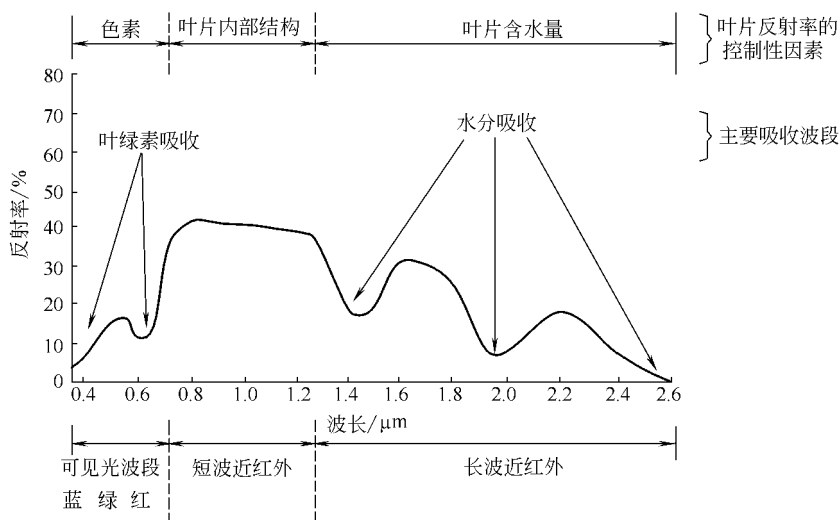


图 23-4 叶片的光谱反射率特性

植被地物在可见至近红外波段呈现特有的光谱反射率特性，如图 23-4 所示。主要表现在：

0.4~0.70 μm 可见光波段：由于色素吸收，特别是叶绿素 a、b 在 450nm 和 660nm 附近的强吸收，在可见光波段植被有其独特的光谱特征，即“蓝边”、“绿峰”、“黄边”、“红谷”等区别于土壤、岩石、水体的独特光谱特征。

0.7~0.78 μm 波段：这是叶绿素在红波段的强吸收到近红外波段多次散射形成的高反射平台的过渡波段，又称为植被反射率红边。红边是植被营养、长势、水分、叶面积等的指示性特征，并得到了广泛应用与证实。当植被生物量、色素含量高，生长力旺盛时，红边会向长波方向移动，而当病虫害、污染、叶片老化等因素，红边便会向短波方向移动（刘良云，2002 年）。

0.78~1.35 μm 波段：叶片内部结构能够解释其光谱反射率特性。由于光线在叶片内部的多次散射形成，且色素和纤维素在该波段来说是近似透明的（多次散射最多 10% 被吸收），即便是叶片含水量也只在 0.97 μm 、1.2 μm 附近有两个微弱的吸收特征，所以多次散射的结果便是近 50% 的光线被反射，近 50% 被透射。该波段反射率平台（又称为反射率红肩）的光谱反射率强度取决于叶片内部结构，特别是叶肉与细胞间空隙的相对厚度。但叶片内部结构影响叶片光谱反射率的机理比较复杂，已有研究表明，当细胞层越多，光谱反射率越高；细胞形状、成分的各向异性及差异越明显，光谱反射率也越高。部分研究表明，该波段的光谱反射率或光谱指数与叶片含水量有显著的相关特性（刘良云，2002 年）。

1.35~2.5 μm 波段：叶片水分吸收主导了该波段的光谱反射率特性。由于 1.45 μm 、1.94 μm 、2.7 μm 的强吸收特征，这些吸收光谱位置中间，形成 2 个主要反射峰，位于 1.65 μm 和 2.2 μm 附近。部分作者（王纪华等，2001 年；Tian 等，2001 年）在室内条件下利用该波段的吸收特征反演叶片含水量，但由于叶片水分的吸收波段受到大气中水汽的强烈干扰，而将大气水汽和植被水分对光谱反射率的贡献相分离的难度很大，虽取得了部分进展，但仍满足不了植被含水量的定量遥感需求。

23. 2. 2 室内条件下叶片含水量的光谱反演

离体条件下利用叶片光谱特征构建水分遥感模型是最基础的工作。Gregory 等（1991 年）认为水分含量对光谱特性的初级影响是由水分含量对辐射的直接吸收造成的，次级影响是因为水分含量造成了叶片内部结构的变化引起的。Penuelas 等（1993 年）提出用近红外波段反射率可以反演作物水分含量。Inoue 等（1993 年）得出了用 950~970nm 处的反射率来预估叶片相对含水量的结论，后来又有学者陆续提出了 1450nm、1940nm、2200nm 附近为水分的敏感光谱波段。王纪华等（2001 年）等利用归一化定量描述技术研究了小麦离体叶片的反射光谱及其用于叶片水分状况诊断的可行性，并结合叶绿素荧光动力学参数测定、水分胁迫阈值分析等探讨了光学检测方法在小麦水分诊断及受胁迫机理研究中的应用。

水分对叶片光谱反射率的影响比其他生化组分大得多。在叶片的反射光谱区内，存在多个水特征吸收峰。其中 1.45μm 附近的光谱反射率强吸收特征可敏感地反映叶片的含水状态，但该波段在大气窗口之外，为大气水吸收波段，因此利用高空遥感技术获取该波段的高质量数据比较困难。1.65~1.85μm 可用来反映植被水分状态。而且该波段在大气窗口之内，受大气层水的干扰较小，适于作为航空或卫星遥感探测指标应用。

图 23-5（a）所示在室内光照、离体测定条件下，1.65~1.85μm 处的光谱特征参量随叶片含水量呈负向波动的变化。为了对这种变化的幅度进行量化描述，引入吸收深度（Depth）和吸收面积（Area）两个参量。深度指特征吸收峰的极小值点相对 100% 线的距离；面积指特征吸收峰深度一半处宽度与深度的乘积，以相对于最大量的百分比表示，如图 23-5（b）所示。

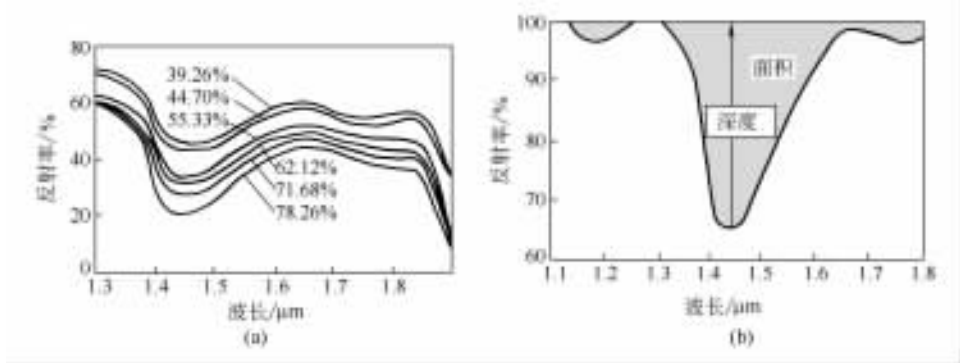


图 23-5 小麦叶片的光谱吸收特征

(a) 不同含水量小麦叶片的光谱吸收特征 (b) 吸收深度与吸收面积

根据从大田不同灌溉量节水试验田取样的小麦旗叶样品，1.65~1.85μm 水特征峰与含水量之间具有很强的相关性。小麦叶片含水量与 1.65~1.85μm 处特征峰的吸收深度和吸收面积均呈现为显著的线性负相关关系（图 23-6、图 23-7）。即随着叶片含水量的下降，其在 1.65~1.85μm 处特征峰的吸收深度和吸收面积增加。其回归表达式为：

$$y=88.31-5.6728x_1 \quad (r=-0.8823); \quad y=82.34-0.6207x_2 \quad (r=-0.8669) \quad (23-2)$$

式中 y 为叶片含水量（%）； x_1 为吸收深度（%）； x_2 为吸收面积（%）。经查 r 显著检验，吸收深度和吸收面积与叶片含水量的负相关均达到极显著水平。

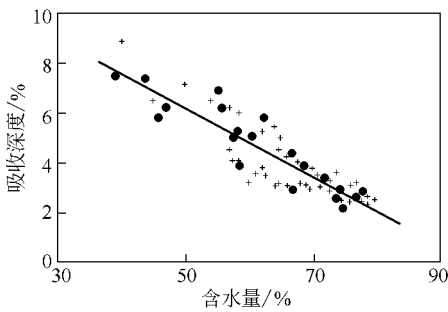


图 23-6 含水量与 1.65~1.85 μm 处吸收深度的相关

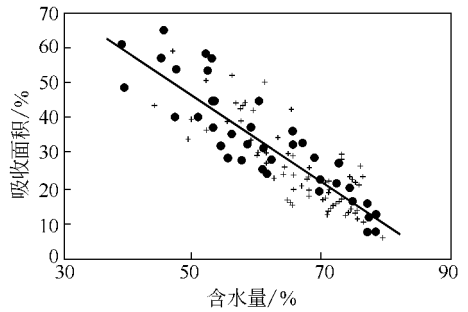


图 23-7 含水量与 1.65~1.85 μm 处吸收面积的相关

23.2.3 田间冠层含水量的光谱反演

田间条件下利用冠层光谱特征反演植株水分含量更接近应用，但存在很多困难，主要受到两个方面的干扰：一方面是土壤背景的干扰，土壤裸露程度及土壤的水分含量不同会影响测定结果；另一方面是来自大气中水汽的干扰。为实现冠层水分的光谱反演，前人进行了广泛的探索。

红边是植被地物的最重要光谱特征，它表现为在红波段 670nm 左右的光谱反射率很低，但此后光谱反射率迅速上升，到近红外波段 780nm 左右光谱反射率才趋于稳定。植被红边光谱反射率及光谱位置主要由植被叶绿素含量和叶片形态结构共同决定，而植株含水量是叶片形态结构的最关键因素，所以利用红边位置和形状信息是有可能反演植株含水量的。

红边位置定义为光谱反射率在红边光谱范围上升速率最快的波段，数学上定义为植被红边上升沿的拐点或一阶导数最大值。但直接采用导数光谱最大值来定义红边位置是十分不利的，一方面，由于红边导数光谱存在双峰特性，并且部分导数光谱有较宽平的峰，红边光谱位置容易受到噪声干扰；另一方面，计算的红边光谱位置的精度由光谱采样间隔决定，而对于同类作物而言，其红边位置差异较小，所以该方法计算的红边位置实用性较低。

为了定量分析红边特性，采用倒高斯模型模拟小麦红边特性。图 23-8 为高斯模型模拟植被红边处的光谱反射率的示意图。倒高斯模型能够很好地模拟植被地物在 670~780nm 的反射率光谱，其定义如下，

$$R(\lambda) = R_s - (R_s - R_0) \exp\left(\frac{-(\lambda_0 - \lambda)^2}{2\sigma^2}\right) \quad (23-3)$$

式中， R_s 为红肩处的光谱反射率； R_0 为叶绿素吸收谷的光谱反射率； λ_0 为叶绿素吸收红谷的光谱位置； λ_p 为红边光谱位置； σ 为倒高斯模型的方差项，也是植被地物光谱红边光谱位置与红谷光谱位置之差，定义为红边宽度， $\sigma = \lambda_p - \lambda_0$ 。

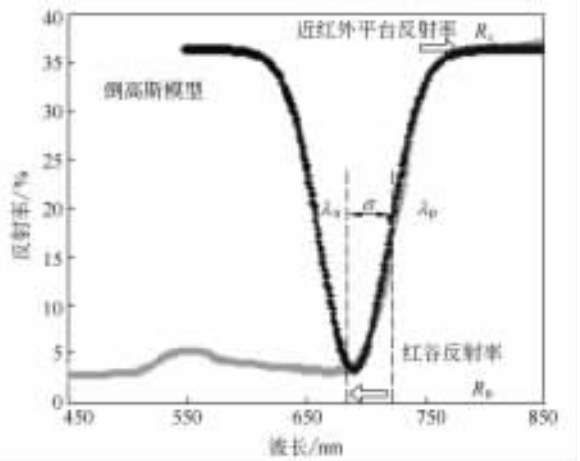


图 23-8 倒高斯模型

红边位置的物理意义为红边处最大上升沿的光谱位置，而红边宽度为植被红边光谱反射率的变化快慢程度，集中体现了红边的形状信息。

在编程处理时，定义 670~680nm 范围光谱反射率的平均值为 R_0 ，780~795nm 光谱范围的反射率均值为 R_s 。挑选 685~780nm 光谱范围的反射率红边作为倒高斯模型的模拟对象。

利用 R_0 和 R_s 两个参数，对小麦作物红边处的反射率光谱进行对数变换，如下所示，

$$B(\lambda) = \left[-\ln\left(\frac{R_s - R(\lambda)}{R_s - R_0}\right) \right]^{\frac{1}{2}}$$

(23-4)

$B(\lambda)$ 为小麦红边反射率光谱的对数变换后的取值，对 $B(\lambda)$ 和 λ 进行线性拟合，可以得到斜率 a_1 和截距 a_0 ，红谷光谱位置和吸收谷宽度 λ_0 ， σ 分别为

$$\lambda_0 = -\frac{a_0}{a_1}; \sigma = \frac{1}{\sqrt{2}a_1}$$

因此，对于红边的倒高斯模型，一共可以得到 3 个拟合参量，即红边位置、红谷位置和红边宽度。通过对倒高斯模型计算的红边位置信息，如红边位置、红谷位置和红边宽度与各个生育期的植株含水量进行统计相关分析，发现利用红边宽度能够很好地反演冬小麦在各个生育期的植株含水量，建立的冬小麦在各个生育期植株含水量与红边宽度的统计模型如表 23-5 所示。

表 23-5

冬小麦在各个生育期的植株含水量遥感模型 ($n=48$)

生育时期	统计模型	复相关系数 R^2	均方差
起身前期(4/2)	$y = -1.9367x + 139.6$	$R^2 = 0.5106$	0.9234
起身后期(4/10)	$y = -1.9741x + 141.03$	$R^2 = 0.3859$	0.9622
拔节期(4/18)	$y = -3.8784x + 195.21$	$R^2 = 0.456$	1.2254
灌浆始期(5/17)	$y = -3.3945x + 175.34$	$R^2 = 0.4589$	2.3253
灌浆盛期(5/24)	$y = -2.8596x + 157.77$	$R^2 = 0.5165$	3.3484
乳熟期(5/31)	$y = -1.015x + 96.183$	$R^2 = 0.4543$	3.2705

注：y 为植株含水量（%），x 为红边宽度（nm）；95%置信度的显著复相关系数 R^2 为 0.0827，99%置信度 R^2 为 0.1145，99.9 置信度 R^2 为 0.2160。

上述研究表明，植株含水量处理对冠层光谱的影响除了水分光谱吸收外，还包括植株含水量对叶片内部结构的影响，从而影响冠层光学特性，即植株水分是叶片内部细胞大小和结构的重要决定因素，从而影响近红外波段的光谱反射率，即植株水分决定叶片内部细胞的空气间隙，影响多次散射，而多次散射是形成近红外波段的光谱特征最重要的因素。

并且，利用红边宽度反演植株含水量，它克服了以往常用遥感参量只能适用于作物特定生育期的叶片含水量的局限性，另一方面，该红边宽度光谱参量不仅能够很好地消除土壤背景、大气干扰等不利因素，而且该模型是基于室外作物冠层光谱建立的，与基于室内叶片光谱模型相比较，能够很好地推广到大面积遥感应用。

23.3 近红外在作物营养监测上的应用

近红外光谱含有丰富的植物生化组分信息，利用光谱仪测得的冠层光谱反射率特征变

化对作物的生长、营养状况进行监测，其在现代农业生产中将发挥越来越重要的作用。20 世纪 80 年代末，利用遥感估测干叶片蛋白质、淀粉、木质素等组分的含量同利用试验室方法获得的结果已十分相近。使用高光谱分辨率遥感数据估计作物农学参数多采用以下两种方法：一是通过多元回归方法建立光谱数据或由此衍生的植被指数与作物农学参量间的关系；二是通过作物的红边参数，尤其红边位置，来描述作物的物候变化及其农学参数，建立包括生化组分含量的叶片散射和吸收模型。这些方法为遥感技术在农业上的应用提供了理论基础。

23.3.1 叶片氮素的光谱反演

作物氮素是对作物生长、产量和品质影响最为显著的营养元素。叶片含氮量一般用单位干重叶片中的氮含量来表示，不受叶片含水量的影响，且叶片含氮量在植株中含量较高。通过遥感的光谱数据来研究植被光谱反射率与叶绿素浓度的关系，可为简便、快速监测作物长势的方法提供理论依据。

植物在可见光波段的反射率主要受叶绿素的影响。为了探索植物叶片氮素遥感诊断的可能性，20 世纪 80 年代以来国内外就水稻氮素营养水平与光谱特性进行了深入系统的研究 (Shibayama, et al. 1986; 王人潮等, 1993; 周启发等, 1993; 王珂等, 1996), 指出诊断水稻叶片氮素营养的敏感波段为近红外 760~900nm; 可见光 630~660nm 和 530~560nm。Shihayama 等 (1986) 研究了氮素营养对水稻叶片光谱特性的影响, 认为缺氮时的水稻叶片和正常营养的水稻叶片的光谱特征显著不同, 并且认为叶绿素是导致光谱特征差异的主要内在因素。王珂等 (1998) 研究指出, 氮素营养对从可见光至近红外较大范围内水稻反射光谱特性有较大影响。基于氮素与叶绿素间的相关关系, 以及叶绿素在红光波段的弱反射和近红外波段的强反射光谱原理, 美能达公司研制出了便携式叶绿素计, 建立了根据叶片叶绿素相对含量 (SPAD 值) 诊断田间的作物氮素状况并指导施肥的方法。

通过光谱测定及其变量的运算可以增加信息量, 更准确地区分不同氮素营养水平。Shibayama 等 (1986) 在水稻上的研究发现, 单位土地面积上的叶片氮含量与 R620 和 R760 的线性组合以及与 R400、R620 和 R880 的线性组合均有较好的回归关系, 且不受品种类型的影响。Lee 等 (2000) 研究了棉花叶片氮浓度与光谱比值指数的关系, 认为用红边位置与短波近红外波段的比值预测的精确度和准确度都比较高。牛铮等 (2000) 的研究也表明, 用 2120nm 和 1120nm 处反射率一阶导数的线性回归方程可以较好地预测小麦叶片氮素含量。

23.3.2 冠层氮素的光谱反演

小麦冠层叶片反射率与全氮含量, 在近红外 1141~1300nm 和 820~1140nm 两波段的正相关分别达到显著和极显著水平 (表 23-6)。经过适当修正小麦能够反映冠层叶片全氮含量的敏感波段为 820~1100nm 和 1150~1300nm。表 23-7 为小麦叶片含氮量与光谱反射率间的关系方程, 不同生育时期叶片含氮量与叶片光谱反射率之间的方程不同: 返青期和灌浆中期以光谱反射率的对数形式建立的回归方程最为理想 (复相关系数 R^2 分别为 0.9657 和 0.9325); 拔节期以光谱反射率的倒数形式建立的回归方程具有较高的复相关系

数 ($R^2=0.6897$)，因而比较可靠；成熟期以光谱反射率本身建立的回归方程最好，复相关系数 R^2 为 0.9325。因此不同生育时期应选用适当的光谱反射率表现形式来建立与叶片含氮量之间的关系。

表 23-6 小麦近红外波段反射率与冠层叶片含氮量的相关系数

波长范围/nm	671~819	820~1140	1141~1300
相关系数/ r	-0.03	0.8325	0.5138

注： $r_{(0.05,17)}=0.456$ ； $r_{(0.01,17)}=0.575$ 。

表 23-7 不同生育时期小麦叶片全氮与光谱反射率间的回归方程

生育时期	波长范围 (820~1140nm)	因变量/%	回归方程	复相关系数 (R^2)
返青期	叶片含氮量($x, \%$)	ρ	$=5.8412x+1.9519$	0.9132
		$1/\rho$	$=-0.0049x+0.0566$	0.7814
		$\ln\rho$	$=0.1688x+2.5958$	0.9657
拔节期	叶片含氮量($x, \%$)	ρ	$=-1.6081x+51.022$	0.6154
		$1/\rho$	$=0.0011x+0.0181$	0.6879
		$\ln/\rho$	$=-0.0415x+3.966$	0.6647
灌浆中期	叶片含氮量($x, \%$)	ρ	$=-3.9018x+54.915$	0.8012
		$1/\rho$	$=0.0038x+0.0091$	0.7568
		$\ln/\rho$	$=-0.1218x+4.1777$	0.9325
成熟期	叶片含氮量($x, \%$)	ρ	$=6.1315x+19.793$	0.9789
		$1/\rho$	$=-0.0057x+0.0427$	0.9127
		$\ln\rho$	$=0.1877x+3.0916$	0.9345

由以上结果看出，单叶光谱与冠层光谱之间存在不一致性，表明光谱响应不仅收到 NH、OH 等化学键的支配，同时收到叶片内部结构、田间群体大小、结构及土壤背景等因素的干扰和影响。

23.3.3 碳组分及糖氮比的光谱反演

碳组分是作物产量形成的基础。碳氮代谢平衡的变化，会引起叶片颜色、厚度、形态结构、水分含量等发生一系列的变化，从而引起植被冠层光谱反射特性的变化；另一方面，冠层叶片生化组分的化学键受到太阳光的照射，会发生振动，从而形成特定的吸收光谱特征，所以利用冠层光谱反射率及波形特征来反演冠层植株体内碳氮组分是可能的。根据对冬小麦在返青期至乳熟期的可溶性糖和全氮参量与作物冠层光谱反射率或反射光谱特征进行了统计相关分析，发现在不同生育阶段，叶片和茎秆的糖、氮组分和糖氮比的敏感波段差异较大，但利用冬小麦冠层光谱在近红外 1690nm 处反射峰特征，能够很好地反演各个生育期糖氮比，并建立了叶片和茎秆在起身期至乳熟期的遥感反演模型。

23.3.3.1 光谱反射特征

定义了 6 个光谱反射特征，每个光谱特征定义了 3 个特征参量。如图 23-9 所示，即反射峰深度、反射峰特征面积及归一化反射峰深度。共 18 个特征。

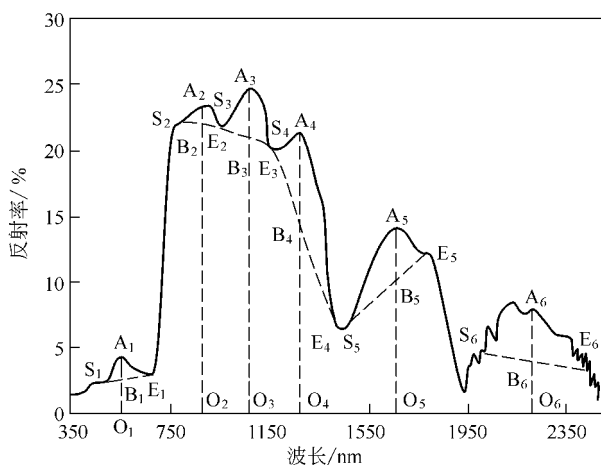


图 23-9 冬小麦光谱在可见~近红外的 6 个反射峰光谱特征

其中，反射峰深度的定义表达式为：

$$P_Depth_i = \frac{A_i B_i}{A_i O_i} = 1 - \frac{(R_{Si} + (R_{Ei} - R_{Si}) / (\lambda_{Ei} - \lambda_{Si})) (\lambda_{Gi} - \lambda_{Si})}{R_{Gi}} \quad (23-5)$$

式中 R_{Gi} , R_{Si} , R_{Ei} 分别为反射峰特征中心点、起点和结束点处的光谱反射率； λ_{Gi} , λ_{Si} , λ_{Ei} 分别为反射峰特征中心点、起点和结束点处的波长。

反射峰面积为包络线与光谱反射率之间的面积，其表达式如下：

$$\begin{aligned} P_Area_i &= \int_{\lambda_{Si}}^{\lambda_{Ei}} (R(\lambda) - R_{line}(\lambda)) d\lambda \\ &= \int_{\lambda_{Si}}^{\lambda_{Ei}} \left(R(\lambda) - \left(R_{Si} + \frac{R_{Ei} - R_{Si}}{\lambda_{Ei} - \lambda_{Si}} (\lambda - \lambda_{Si}) \right) \right) d\lambda \end{aligned} \quad (23-6)$$

式中 λ 和 $R(\lambda)$ 为反射峰特征范围的波长和光谱反射率，其他符号同前式。

归一化的反射峰深度为反射峰深度与反射峰特征面积的比值，即

$$P_ND_i = \frac{P_Depth_i}{P_Area_i} \quad (23-7)$$

定义的 6 个反射峰光谱特征的光谱位置与范围如表 23-8 所示。

表 23-8 六个光谱反射特征的光谱位置及范围

编号	中心波长/nm	光谱范围/nm	编号	中心波长/nm	光谱范围/nm
1	560	500~670	4	1280	1200~1480
2	920	780~980	5	1690	1480~1780
3	1100	980~1200	6	2230	2000~2400

23.3.3.2 糖氮比的特征光谱参量

利用上述 6 个反射峰共 18 个光谱特征，与冬小麦叶片和茎秆可溶性糖与全氮比值进行统计相关分析。从返青期至乳熟期（每期 48 个样本），冬小麦叶片与茎秆糖氮比与吸收或反射光谱特征的统计相关系数如表 23-9 所示。

表 23-9

冬小麦糖氮比与光谱特征参量的统计相关系数汇总

	冠层叶片糖氮比与光谱参量相关系数								冠层茎秆糖氮比与光谱参量相关系数						
日期	0325	0402	0410	0418	0505	0517	0524	0531	0402	0410	0418	0505	0517	0524	0531
P_Depth ₁	0.17	0.09	0.06	0.24	−0.11	0.30	0.57	0.27	0.66	0.34	0.37	0.12	0.61	0.64	0.13
P_ND ₁	−0.28	−0.15	−0.14	−0.37	−0.06	−0.42	−0.55	−0.59	−0.59	−0.10	−0.27	−0.25	−0.65	−0.48	−0.27
P_Area ₁	−0.29	−0.15	−0.13	−0.35	−0.04	−0.41	−0.56	−0.60	−0.61	−0.13	−0.28	−0.25	−0.65	−0.49	−0.27
P_Depth ₂	−0.06	−0.14	−0.31	−0.22	−0.21	−0.44	−0.47	−0.56	−0.53	−0.43	−0.09	−0.35	−0.56	−0.55	−0.28
P_ND ₂	−0.10	0.10	0.12	−0.15	−0.21	−0.19	−0.38	−0.29	−0.04	0.17	−0.13	−0.42	−0.43	−0.22	−0.02
P_Area ₂	−0.08	0.10	0.15	−0.11	−0.19	−0.14	−0.32	−0.11	0.01	0.22	−0.10	−0.39	−0.37	−0.14	0.06
P_Depth ₃	−0.31	−0.20	−0.09	−0.16	−0.14	−0.34	0.04	−0.32	−0.41	−0.15	−0.19	−0.15	−0.42	0.02	−0.29
P_ND ₃	−0.05	0.06	0.00	−0.42	−0.24	−0.37	−0.50	−0.60	−0.22	0.11	−0.13	−0.44	−0.64	−0.45	−0.21
P_Area ₃	−0.03	0.06	0.01	−0.42	−0.24	−0.37	−0.50	−0.50	−0.21	0.12	−0.12	−0.45	−0.64	−0.45	−0.12
P_Depth ₄	−0.16	−0.04	−0.07	−0.01	−0.01	0.14	0.19	−0.19	0.10	−0.18	0.21	−0.07	0.12	0.13	−0.16
P_ND ₄	−0.09	−0.07	−0.13	−0.59	−0.21	−0.44	−0.52	−0.44	−0.40	0.08	−0.32	−0.44	−0.71	−0.47	−0.28
P_Area ₄	0.00	0.01	0.15	0.41	0.16	0.34	0.55	0.02	0.37	0.13	0.40	0.47	0.59	0.43	−0.04
P_Depth ₅	0.18	0.42	0.40	0.38	0.21	0.50	0.16	0.34	0.60	−0.05	−0.05	0.16	0.55	0.17	0.25
P_ND ₅	−0.12	−0.10	−0.25	−0.57	−0.33	−0.41	−0.51	−0.60	−0.48	−0.12	−0.42	−0.49	−0.66	−0.45	−0.28
P_Area ₅	−0.14	−0.12	−0.27	−0.58	−0.33	−0.42	−0.50	−0.57	−0.49	−0.11	−0.40	−0.48	−0.67	−0.45	−0.29
P_Depth ₆	0.01	0.09	0.08	0.13	−0.41	0.20	0.45	0.44	0.05	0.07	−0.05	−0.28	0.17	0.37	0.18
P_ND ₆	0.00	−0.11	−0.12	−0.34	−0.44	−0.28	−0.34	−0.25	−0.51	0.10	−0.27	−0.38	−0.50	−0.24	0.03
P_Area ₆	−0.02	−0.13	−0.14	−0.38	−0.44	−0.31	−0.40	−0.31	−0.54	0.09	−0.28	−0.39	−0.54	−0.29	0.00

23.3.3.3 糖氮比的遥感统计模型

从表 23-9 可以看出，冬小麦冠层光谱在 1690nm 处的反射峰的特征，能够较好地反演冬小麦在各个生育期的叶片和茎秆的糖氮比。

所以，利用式（23-5）和式（23-6）定义的反射峰深度和面积，建立的冬小麦在各个生育期的叶片及茎秆糖氮比的遥感模型如表 23-10 所示。

表 23-10

利用 1690nm 处的光谱反射峰特征反演叶片和茎秆糖氮比（S/N）

日 期	叶 片	茎 秆
起身期	$S/N=8.62E-5\times P_Depth+0.0042$	$S/N=4.04E-5\times P_Depth+0.0042$
拔节前期	$S/N=3.19E-5\times P_Depth+0.0042$	相关未达到显著标准
拔节后期	$S/N=-7.1546\times P_Area+80.586$	$S/N=-0.8141\times P_Area+76.158$
抽穗期	$S/N=-5.3162\times P_Area+82.312$	$S/N=-0.696\times P_Area+80.097$
灌浆初期	$S/N=-12.618\times P_Area+77.305$	$S/N=-1.266\times P_Area+76.916$
灌浆盛期	$S/N=-15.456\times P_Area+84.674$	$S/N=-1.3\times P_Area+72.917$
乳熟期	$S/N=-6.57\times P_Area+66.121$	$S/N=-0.8133\times P_Area+56.725$

23.3.4 其他生化组分的光谱反演

除了氮素、碳素和可溶性总糖之外，对于叶绿素、淀粉、纤维素等生化组分也有大量研究，表 23-11 是归纳了经大量研究确定的冬小麦主要生化组分的光谱特征波长或波段。

表 23-11 小麦主要生化组分的光谱特征波长或波段

编号	生化组分	光谱特征或波长/nm
1	叶绿素 a	430、660
2	叶绿素 b	460、640
3	叶绿素总量	699、红边参数、NDVI
4	全氮	515、699、756、1402~1531、红边参数
5	可溶性总糖	577、843、1490、1580、2080
6	淀粉	970、1530、1900
7	纤维素	1200、1540、1820、1940、2270
8	水分	733、970、1450、1650~1850、2040

需要指出的是，迄今利用遥感手段获取的植被信息侧重于水平方向的“面状”信息，而对于植株营养状况尤其是氮素随冠层高度的垂直分布状况重视不够。由于氮素的易运转特性，缺氮时老叶中的氮素会向新叶中转移，作物缺氮的显著特征是植株下部叶片首先褪绿黄化，然后逐渐向上部叶片扩展。在缺肥初期即轻度氮素胁迫时，下层叶片因缺氮而引起早衰；在持续缺肥即中度胁迫时，中、下层叶片均明显衰老，田间条件下肉眼可见中、下层叶片由深绿变浅绿或变黄，但此时上层叶片很少发生变化。因此，生产上迫切需要拔节期对中、下层叶片氮素或叶绿素状况的监测来指导适时和适量施肥。为此，需要通过进一步深入的分层光谱分析，以及采用多角度研究方法等途径，以提高对作物冠层中、下部生育状况的诊断精度。与可见光和短波红外波段相比，作物叶片对近红外波段的光谱透射率相对较高，可能对于了解植被冠层中下部的信息有所帮助，需要进一步深入探讨。

23.4 近红外在作物病虫草害监测上的应用

23.4.1 小麦条锈病的遥感监测

23.4.1.1 小麦条锈病的冠层光谱特征

当绿色植物叶绿素含量高、生长旺盛时，“红边”会向红外方向偏移（称红移，即往长波方向移动）（Vane &Goetz, 1998）；当植物由于感染病虫害或因污染或物候变化而“失绿”时，则“红边”会向蓝光方向移动（称蓝移）。F. Boochs 等指出，由高光谱分辨率光谱曲线所得红边，可提供足够的信息以探测出植物形态和生化状况的微小差异。通过研究分析可以看出，对每个生育期的不同梯度的光谱反射率进行一阶微分处理，发现每个生育期正常叶片的“红边”（即一阶导数在 680~750nm 之间反射率最大值所对应的波长）

先向长波方向移动，即发生“红移”，后向短波方向移动，即“蓝移”，后又发生“红移”。而病害处理有不明显“红移”，表现为“蓝移”趋势。

比较拔节、孕穗、灌浆和成熟期病害与对照区光谱特征，在可见光波段处，特别在红光波段处，对照区的反射率，要比发病区低，而在近红外波段处则相反，综合看条锈病病害条件下的光谱有着如下的特征：

① 在任何时相下，黄边波段（550~660nm）处，条锈病病害冠层反射率都高于正常生长冬小麦当冠层光谱反射率。并且在近红外波段 1350nm 处，发病区冠层光谱反射率均位于对照处理区光谱上方；

② 任何时相，近红外波段 760nm 处，拔节期后的条锈病病害冠层反射率都要大大低于对照区冬小麦当冠层光谱反射率；

③ 整个生育期，近红外波段 760~1300nm 反映叶片细胞结构变化的波段范围内，条锈病病害的冠层反射率低于正常对照冬小麦的冠层光谱反射率；

④ 在近红外波段 1450nm、1950nm 水分敏感波段处，条锈病病害的反射率要低于正常生长的冬小麦当冠层光谱反射率。

23.4.1.2 小麦条锈病的单叶光谱特征

采用美国 LI-Cor1800-12 外置积分球，与 ASD Field Spec Pro FR 2500（350~2500nm）型光谱仪耦合进行冬小麦条锈病单叶片的反射率测定。积分球的优点在于它是密闭环境下的测定，操作稳定，测量科学，在一定程度上减少了误差，提高了研究的可靠性和精确性，并且积分球适合于作物单叶片的研究。光谱仪采样间隔在 350~1000nm 范围内为 1.4nm；在 1000~2500nm 范围内为 2nm。光谱分辨率在 700nm 为 3nm，1400nm 与 2100nm 为 10nm。每张单叶片按 SPAD 叶绿素测量步骤，分叶尖、中部和叶基部 3 个部位，各部位测 3 次，共测量 9 次，每次测量前后均用积分球的内置标准参考白板进行校正。

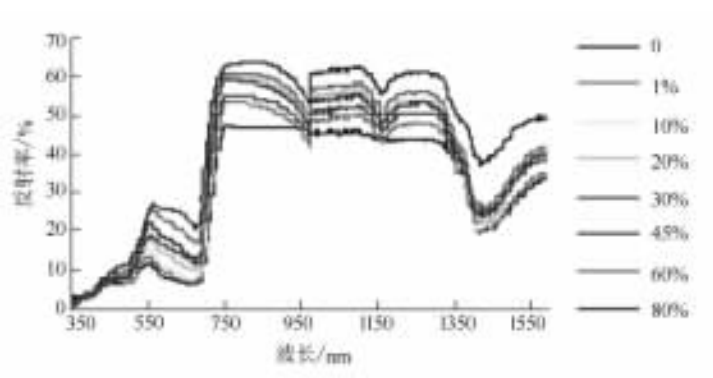


图 23-10 条锈病不同严重度单叶光谱特征

由图 23-10 的冬小麦条锈病不同严重度的单叶光谱特征可以看出，条锈病单叶片光谱的反射率在整体上都随叶片严重度的增加而增加，严重度为 0 的单叶片光谱反射率最低，严重度为 80% 的单叶片光谱反射率最高。

单叶光谱与前期的冠层水平光谱研究表明，在可见光处两者的反射率变化相当一致，而近红外平台反射趋势不同。对这种单叶光谱变化的初步机理解释是：条锈病菌孢子侵染叶片后，破坏叶表皮和叶肉细胞，造成叶片水分含量和叶绿素含量减少，且随着条锈鲜黄

色孢子堆突破表皮而使叶片增厚，也改变了叶片结构，由于冠层水平和单叶水平下，叶绿素与水分所影响的波段区域一致，所以条锈病胁迫时可见光 550nm 附近黄绿光反射率与 1450nm 水分敏感波段区的反射率随严重度均呈上升的趋势对两者来说是一致的。但在近红外 750~1350nm 波段，植物冠层反射特性受叶面积指数、水分和土壤背景影响较大，所以在条锈病胁迫增加时，其冠层叶面积指数变小，植株含水量下降，所以近红外平台反射率下降，但单叶片条件下，不受群体与土壤的影响，其光谱特征主要受叶绿素 a 和 b、叶片单位面积含水量、干物质含量和叶片内部结构四个主要因子影响，所以在此波段两种条件的反射特性不一致。

对于病虫害胁迫，作物的病理、生理及形态变化复杂，但仍有规律可循，需要进一步结合条锈病胁迫下的生理生化参数及光谱信息探讨条锈病害的遥感监测机理。

23.4.1.3 小麦条锈病的遥感监测模型

红边参数、植被指数等常被用于病情监测。其中归一化差值植被指数 (NDVI) 即近红外波段反射率值与可见光红波段反射率值之差和这两个波段反射率值之和的比值。用光谱数据建立的监测病情指数的几种模型可用于病害程度监测和损失评估：

模型一 自拔节期后，光谱的红边位置 (x) 与病情指数 (y) 存在着显著的正相关，所以在一定时期内可用红边位置来反演病情指数。

$$Y = -3.031x_{\text{红边位置}} + 2262.7 \quad r = 0.918^{**}, r_{(0.01, 18)} = 0.561$$

模型二 $NDVI = \frac{\rho_{976nm} - \rho_{680nm}}{\rho_{976nm} + \rho_{680nm}}$ (归一化植被指数 (x))、 RVI [$\text{比值植被指数 } (x) = \frac{\rho_{976nm}}{\rho_{680nm}}$]、 TVI [变换植被指数 (x) = $\sqrt{NDVI + 0.5}$ (Rouse 等, 1974)] 与病情指数 (y) 也存在着显著的正相关。

$$Y = -143.11x_{NDVI} + 165.41 \quad r = -0.871^{**}, r_{(0.01, 13)} = 0.641$$

$$Y = -7.288x_{RVI} + 111.74 \quad r = -0.900^{**}, r_{(0.01, 13)} = 0.641$$

$$Y = -304.16x_{TVI} + 398.17 \quad r = -0.866^{**}, r_{(0.01, 13)} = 0.641$$

模型三 通过数据的相关系数分析，发现一阶微分后的 1536nm 处的反射率值 ($d\lambda_{1536}$) (x) 与病情指数 (y) 之间也存在着明显的正相关。

$$Y = 135.43e^{-15.115x} \quad r = -0.880^{**}, r_{(0.01, 13)} = 0.641$$

以上条锈病光谱的特征也可以作为遥感在监测病害时的多时相、多波段组合或综合诊断模式方法。此外，由于微分 (导数) 光谱会使信噪比大幅度下降，应用时需谨慎。

23.4.1.4 利用高光谱遥感影像监测小麦条锈病

发病处理叶片由于条锈病夏孢子堆突破表皮，破坏了大量的叶绿素，从而在各个生育期的叶绿素含量均比正常处理要低。这与条锈病发生后，冬小麦叶片退绿、发黄的状况吻合。这种病害植株绿度值变化在遥感影像或光谱数据曲线上均有所反映，从而为利用高光谱遥感影像对条锈病害的监测提供了一种可视信息。孕穗期以后，与正常处理叶片相比较，病害处理叶片由于病害胁迫作用，叶绿素下降的速率加快，下降的速率与发病的程度的不同而有所区别。经显著性差异检验，正常处理叶片与重度发病叶片叶绿素含量在拔节

* * 极显著相关。

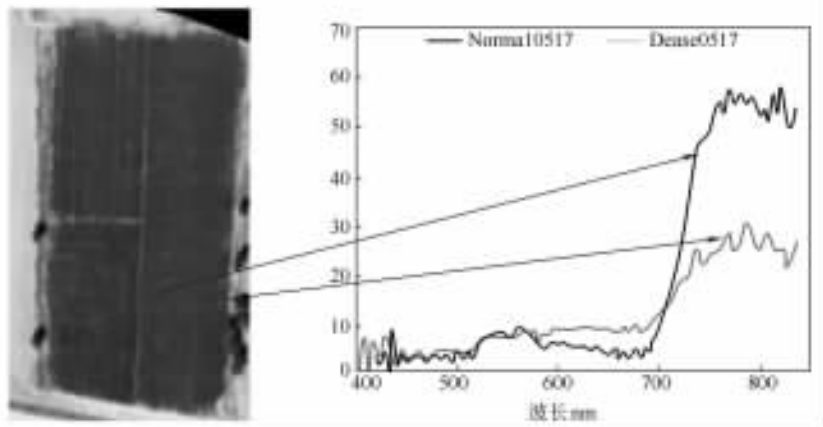


图 23-11 条锈病严重时光谱曲线与正常光谱曲线对比（5 月 17 日，灌浆期）

期在 5%水平上呈显著性差异；在孕穗期和灌浆期，正常处理与各病害处理叶绿素含量在 1%水平上均呈显著性差异。因而在一定程度上，可以把拔节期作为病害发生的判别点，把孕穗期作为区分病害程度的关键点。

根据获取的经辐射、几何校正和反射率转换的 PHI 图像数据可见，病害点与正常生长的对照点的光谱差异十分明显（图 23-11），可以直接从 CIR 显示 PHI 图像上发现病害发作较严重区域，从而监测病害严重程度和评估损失。从对照点与发病处理点的 PHI 图像光谱的比较图上可以发现，在近红外 730~830nm 波段范围内，对照点光谱反射率远比发病点高，说明对照点的叶面积指数、干物质含量远比发病区大，并且病害严重破坏了叶片正常的组织结构，造成近红外平台的光谱反射率的下降。

23.4.2 小麦蚜虫的遥感监测

23.4.2.1 小麦蚜虫发生状况及其光谱特征

冬小麦在发生病虫害的情况下，首先在叶子的表面反映出来。因此通过探测其生长条件下的光谱特征也可以做出判别。图 23-12 为 2001 年 5 月 26 日在小汤山国家精准农业示范基地水胁迫试验区获取的冬小麦叶片附着不同大小和稀密程度蚜虫情况的光谱数据。实测结果显示，随着叶片上附着的蚜虫个体和密度不断增大，光谱反射率随之明显降低。在近红外反射平台的 1100nm 处，反射率相对值随蚜虫发生程度密集而下降，降幅达 33%，表明高光谱遥感可以作为诊断冬小麦蚜虫发生程度的辅助手段。

23.4.2.2 小麦蚜虫发生程度的遥感监测

在当前的卫星遥感中，技术较成熟应用较广泛的卫星数据是美国陆地卫星 TM、法国 SPOT 和 NOAA 卫星 AVHRR 图像。法国 SPOT 卫星图像成本高，美国陆地卫星相邻两次过境时间间隔为 16 天，间隔过长且图像质量受天气的影响很大而蚜虫发生和繁殖较快。NOAA 卫星的 AVHRR 图像虽然空间分辨率较低，但可以利用卫星覆盖面大的优点，配合地面的测量结果，对麦蚜灾害进行大范围宏观监测分析，以确定最佳防治时间。

卫星遥感数据的植被指数（RVI，近红外波段反射率值与可见光红波段反射率值的比值）与虫情监测的关系研究表明，RVI 变化与蚜虫量变化具有相关性。以北京地区

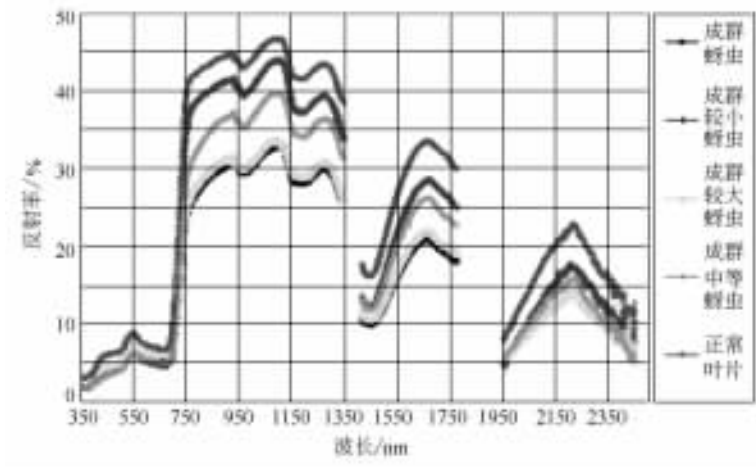


图 23-12 小麦蚜虫密集程度及其光谱特征

为例，5 月份小麦的主要灾害为蚜虫危害，在此阶段若不考虑自然因素、管理因素的影响，RVI 值的变化可认为是蚜虫危害的光谱表现，且 RVI 值降低越大，蚜虫危害越严重。

23.4.3 田间杂草的遥感监测

杂草的防治在很大程度上依赖于除草剂的喷洒，然而杂草在地块中的生长不是均匀分布的而是呈片状分布的，因此以最少的投入测量这些呈片分布的除草剂是减少除草剂喷洒量的方法之一。由于作物的叶片和杂草的叶片具有不同的光谱特征，通过光谱分析即可以对田间的杂草情况有一个粗略的估计。图 23-13 为 2001 年 5 月 23 日在小汤山国家精准农业示范基地对冬小麦与杂草共生情况所测光谱结果。图中两种有杂草的情况，近红外反射平台的反射光谱数值都明显高于无杂草的情况。这是因为有杂草的区域行间绿色覆盖度较高，土壤背景的影响进一步减小所致。图 23-13 中杂草很多田块在近红外反射平台高于杂草较少的田块。

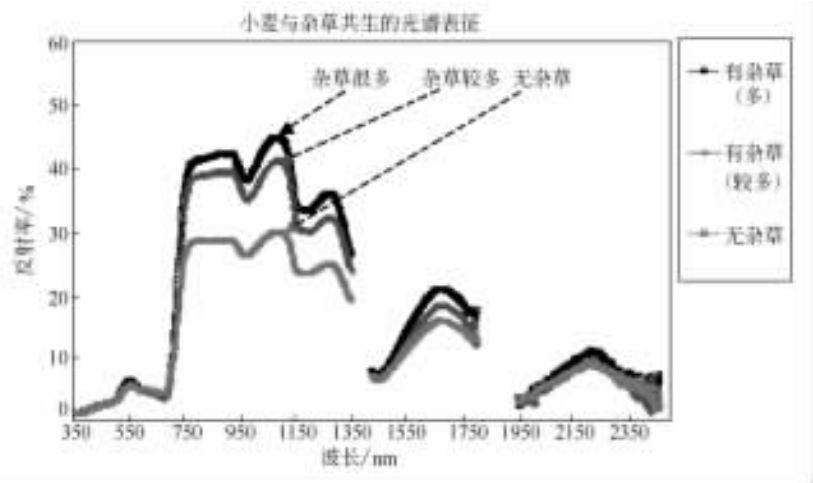


图 23-13 小麦与杂草共生的光谱表征

23.5 近红外在作物长势监测上的应用

通常利用长势绿度-时间方法监测作物长势，即首先依据植被指数（即绿度）表征地表植被的数量状况，再通过历年指数比较法，判断同一地区作物在不同时间段长势的好坏。常用的植被指数有：归一化植被指数（*NDVI*）、差值植被指数（*DVI*）、比值植被指数（*RVI*）等。

23.5.1 叶面积指数的遥感反演

在小麦主要生育阶段利用地物光谱仪测得的数据表明，拔节期在近红外 720～810nm 波段的反射率与叶面积指数（*LAI*）显著正相关。表 23-12 为 PHI 图像光谱特征与 *LAI* 的统计相关系数，从中可以看出，利用 PHI 图像的 *NDVI* 能够在拔节期、灌浆期和乳熟期反演 *LAI*，反演模型如表 23-13 所示。

表 23-12 PHI 图像提取 *NDVI* 与 *LAI* 的统计相关系数

	拔节期 <i>LAI</i>	灌浆期 <i>LAI</i>	乳熟期 <i>LAI</i>
<i>NDVI</i>	0.40	0.55	0.69

表 23-13 *LAI* 与 PHI 图像光谱特征反演模型

日期	模型	日期	模型
拔节期	$LAI=7.12\times NDVI-3.09$	乳熟期	$LAI=3.9178\times NDVI-1.8743$
灌浆期	$LAI=6.4613\times NDVI-3.2719$		

23.5.2 利用生化参量填图结果进行长势评估

利用红边位置（ λ_p ）、红谷位置（ λ_0 ）及近红外与红光波段组合的 *NDVI* 反演的作物

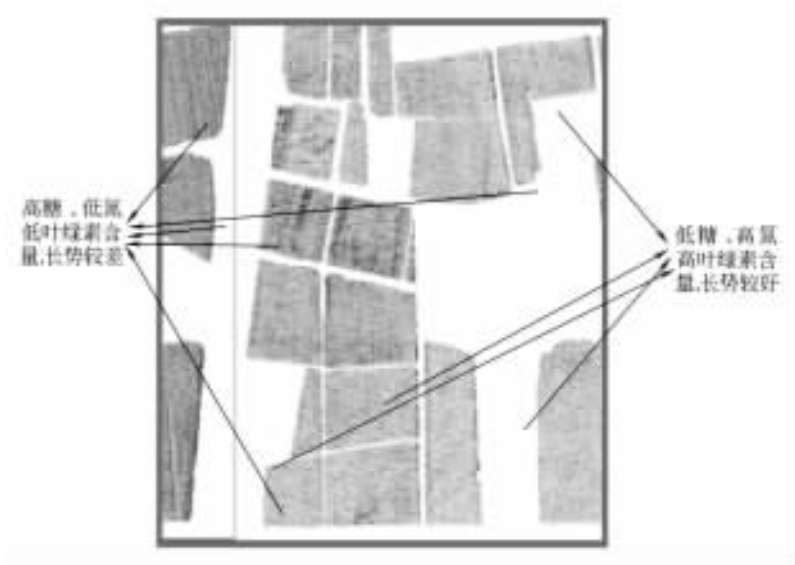


图 23-14 利用叶绿素、全氮和可溶性糖合成的假彩色作物长势图

生化组分和 *LAI* 等参数，辅以地理位置（空间分布）信息，可以对叶绿素、全氮、可溶性糖、叶片水分等生化参量填图进行假彩色合成，并得到作物长势、营养分布图。图 23-14 为叶绿素含量（R）、全氮含量（G）和可溶性糖（B）3 个生化参数“图像”的彩色合成结果。3 个参数后括号中的 R、G、B 分别代表参加彩色合成时 3 个参数被赋予的三原色：红、绿、蓝。

对于拔节后期的冬小麦作物，高糖、低氮、低叶绿素的作物地块呈蓝色显示，这些地块作物缺肥，过早进入生殖生长阶段，而高氮、高叶绿素、低糖地块呈黄色显示，表示这些地块营养充足，仍处于营养积累阶段。所以，利用色彩差异，可以决定冬小麦长势，并辅助决定农业管理。

23.6 近红外在作物估产上的应用

23.6.1 作物遥感估产原理

遥感估产是根据生物学原理，在收集分析作物光谱特征的基础上，通过卫星传感器记录地表信息，辨别作物类型，提取作物种植面积，监测长势，并在作物收获前预测产量。它包括两项重要内容：作物种类识别和播种面积提取、依据长势分析预报产量。

23.6.1.1 作物种类识别及播种面积提取方法

① 成数抽样方法。在了解某种作物种植面积占总面积的比例（成数）基础上，将区域分成许多格网，从格网中随机选取样点，将调查结果进行分类统计，核实在全部的调查点中小麦所占成数，根据测区总面积的大小，计算出小麦的种植面积。以前用这种方法时，都采用实地调查方法，现在用遥感调查代替实地调查。

② 通过对遥感数据进行分类，根据分类结果，按每类的像元数计算作物的种植面积。这种方法要求遥感数据获得的时间符合要求。如冬小麦最好是 3 月份，或冬天 11 月份获得的遥感数据。遥感直接分类方法的精度不高，一般在 70%~80% 之间。为了提高精度，在分类时需要使用其他数据，如数字地面高程，并采用地理信息系统技术。

23.6.1.2 小麦遥感估产模型

以近红外与红光波段组合的 *NDVI* 和两个近红外波段 [890nm, 1200nm] 定义的归一化差值光谱指数为例，建立的冬小麦产量的分生育期遥感模型如表 23-14 所示。

表 23-14 基于 *NDVI* 和 [890nm, 1200nm] 归一化差值光谱指数的产量预报模型

生育期	基于 <i>NDVI</i> 的统计模型		[890nm,1200nm]光谱指数的统计模型	
返青期	未达到显著相关		$y=1602.8x+289.61$	$R^2=0.2876$
起身前期	$y=376.78x-32.096$	$R^2=0.2801$	$y=1068.9x+187.28$	$R^2=0.3698$
起身后期	未达到显著相关		$y=1297.1x+164.35$	$R^2=0.2979$
拔节期	$y=861.63x-486.08$	$R^2=0.234$	$y=1770.3x+76.912$	$R^2=0.4501$
抽穗期	$y=1065x-696.17$	$R^2=0.577$	$y=979.85x+68.167$	$R^2=0.5248$
灌浆始期	$y=505.3x-150.41$	$R^2=0.5614$	$y=837.01x+173.97$	$R^2=0.5666$
灌浆盛期	$y=455.55x-91.673$	$R^2=0.6595$	$y=1084.1x+172.69$	$R^2=0.6908$
乳熟期	$y=370.04x+25.512$	$R^2=0.7011$	$y=1134.3x+191.31$	$R^2=0.6324$

小麦产量形成的过程较为复杂，并且产量的物质积累与转化也主要发生在拔节期、抽穗期、灌浆期、乳熟期，尽管越早的遥感产量预报对指导农业生产与国家政策的意义也越大，但早期的产量预报也需要后期数据的验证，特别是当后期遭遇到气候、病虫害等不利因素的干扰时，后期验证就更显得必要。对于正常生长与管理的作物，依据作物生长发育的继承性进行早期的产量预报也是可行的，尤其在生产力水平较低的大面积中低产地区，小麦产量主要受到春季田间肥水管理水平的制约，返青后一起身—拔节前的群体数量大小与产量密切相关，通过遥感手段掌握冬小麦生长早期田间群体动态，可以用于估算后期产量。但对于生产力水平较高的高产地区，群体通常已经较大，群体过大引起倒伏进而影响产量的潜在因素显现，这时群体质量上升为主要因子，早期预报产量难度增加，对此，还需要深入研究。

23.7 近红外在作物品质监测预报上的应用

23.7.1 影响作物品质的主要因素及遥感监测模式

随着近年来航空航天技术的发展和传感器性能的提高，使利用遥感信息反演作物体内生化组分含量进而监测品质成为可能。小西教夫等（2000 年）报道利用卫星影像数据监测稻谷氮素、直链淀粉、支链淀粉等品质指标取得重要进展，其中日本北海道中央农业试验场以蛋白质含量为主要监测指标，利用卫星遥感成像技术指导区域施肥，有效地提高了稻谷品质，经济效益显著，该项研究成果在当地获奖。中村宪知等（2001 年）报道了利用 TM 卫星影像数据反演茶叶的氮素/纤维素指标，进而大面积监测茶叶品质的方法。王纪华等（2002 年）的研究表明，后期的小麦叶片全氮与成熟期籽粒品质组分之间具有很强的相关性，用开花期的叶片全氮能很好地预测籽粒蛋白质和面筋含量，这些为利用遥感技术监测小麦品质和调优栽培提供了依据。

大多数情况下，决定作物品质的因素是复杂的。一方面是影响作物品质的生化组分的多样性；另一方面是决定品质形成的遗传与环境作用的复杂性。例如小麦加工面粉的商品品质主要由籽粒中粗蛋白、面筋、淀粉等生化组分的种类及含量决定。而最终品质的形成，除了品种遗传因素外，栽培过程中环境气象条件、氮素肥料和水分供给、土壤质地以及病害发生与否等均影响到品质的形成。华北地区小麦灌浆过程中的高温逼熟、水分过大和倒伏等都会引起面粉商品质量下降。

随着近年来航空航天技术的发展和传感器性能的提高，使利用遥感信息尤其是近红外光谱信息反演作物体内生化组分含量进而监测品质的技术逐渐成熟。利用遥感手段对麦田冠层的叶绿素、氮素等生化组分以及温度、水分、病害及倒伏状况进行监测，并通过遥感与农学模型的链接，可初步实现对小麦调优栽培的指导和对籽粒品质的预测预报（图 23-15）。

23.7.2 遥感监测预报作物品质应用案例

在栽培上，多时相卫星数据或航空影像数据可实现地表水分、温度、叶绿素、*LAI* 等的反演，根据结果指导调优栽培中的肥水管理；在收购加工方面，利用遥感数据将监测

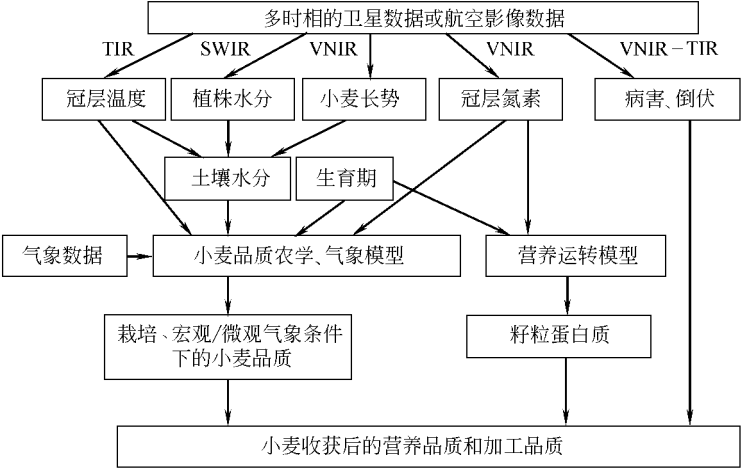


图 23-15 小麦品质遥感监测综合模式

注：TIR 为热红外波段、SWIR 为短波红外、VNIR 为可见近红外波段。

区域内某种作物的品质初步划分出若干个等级，优等的可以免去实验室检测程序；中等的可以进入实验室抽样检测；劣等的不予收购或降级降价收购。在此基础上进行分级加工，就可以从总体上提高商品等级，并大幅度降低品质分析检测成本。

日本北海道纲走地区东藻琴村及十胜地区士幌町、音更町，2000 年以来利用 LANDSAT、SPOT 和 IKONOS 等卫星影像数据，以近红外与红光波段组合的 NDVI 参数应用为主，通过监测氮素吸收量、贪青晚熟发生情况以及确定适宜收获期等调控和预测预报小麦籽粒品质取得了较好的效果。

国家农业信息化工程技术研究中心于 2000 年率先在国内开展了利用遥感手段大面积监测麦田氮素、温度、水分及病害倒伏等主要指标并预报面粉品质的研究，建立了利用不同生育时期遥感数据反演主要品质相关因子的模型。如表征小麦籽粒和面粉品质的主要指标有：蛋白质含量、面筋含量（本文用干面筋含量）、沉降值、形成时间、稳定时间等。表 23-15 为开花期小麦叶片含氮量与以上指标间的相关系数。由表 23-15 可知，叶片含氮量与蛋白质含量和干面筋含量之间相关达极显著水平，而与其他指标相关不显著，尤其是负相关均很弱，表明干扰很小。

表 23-15 开花期小麦叶片含氮量与籽粒品质指标间的相关系数							
籽粒品质指标	蛋白质	干面筋	沉降值	形成时间	稳定时间	耐揉指数	断裂时间
相关系数	0.5265	0.4940	0.2343	-0.0114	-0.1540	-0.0351	-0.1253

注： $r_{(0.05,26)}=0.374$ ； $r_{(0.01,26)}=0.478$

根据以上关系我们建立了开花期小麦叶片含氮量（ x ）与籽粒部分品质指标（ y ）间的回归方程。小麦叶片含氮量与籽粒蛋白质含量间的方程为：

$$y=6.3761x-7.8417 \quad (4\leq x\leq 5, R^2=0.9012)$$

式中， x 为小麦叶片含氮量（%）； y 为小麦籽粒蛋白质含量（%）

小麦叶片含氮量与籽粒干面筋含量间的方程为：

$y=6.8815x-24.546 \ (5\leq x\leq 6, R^2=0.8579)$

式中， x 为小麦叶片含氮量（%）； y 为小麦籽粒干面筋含量（%）

根据多年结果建立的模型，利用 LANDSAT 卫星的 TM 遥感影像数据，分析提取田间地块尺度的冠层水分状况、温度、植株叶绿素、氮素和叶面积指数（LAI）信息，及时
进行调优栽培决策和品质预报，并通过地理信息系统（GIS）进行可视化表达后提供给用户。图 23-16 的画面为精确到田块的小麦籽粒蛋白质含量分布，在计算机上移动鼠标即可
显示每 30m×30m 的蛋白质含量预报结果。

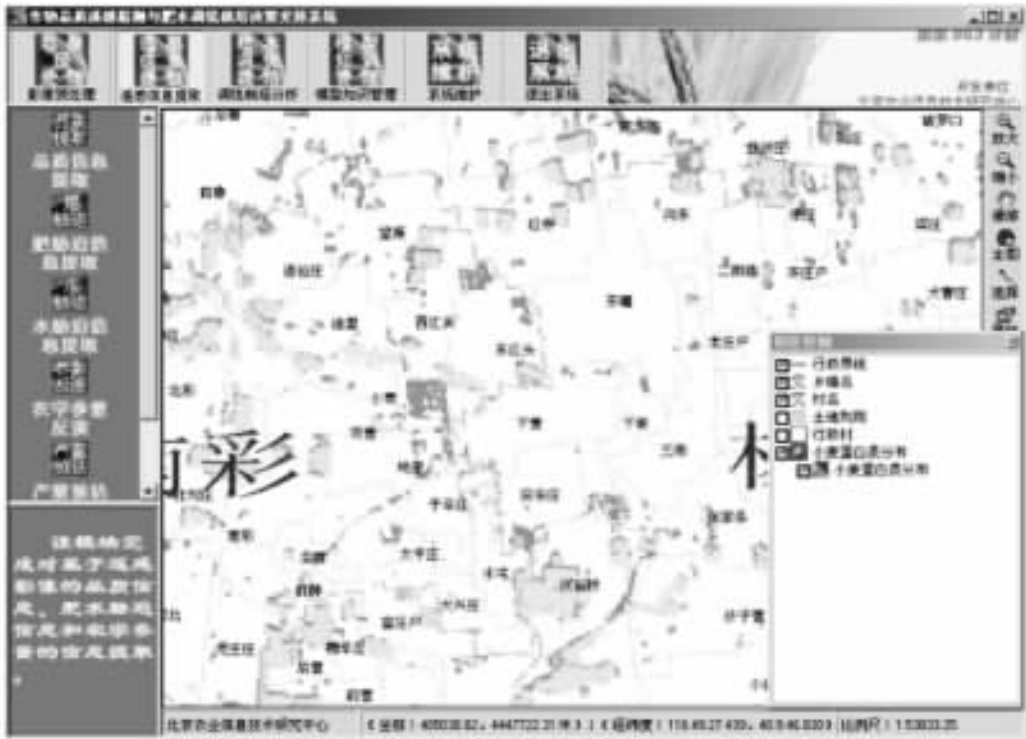


图 23-16 利用 TM 遥感影像预报的田块小麦籽粒蛋白质含量（北京顺义区南彩镇）

参 考 文 献

[1] Holben B N, Schutt J B, Mcmurtrey J. Leaf Water Stress Detection Utilizing Thematic Mapper Bands 3, 4 and 5 in Soybean Plants. International Journal of Remote Sensing, 1983, 4: 289~297

[2] Shibayama M, Akiyama T. A Spectroradiometer for Field Use. VII. Radiometric Estimation of Nitrogen Levels in Field Rice Canopies. Japanese journal of Crop Science, 1986, 55 (4): 439~445

[3] Gregory A, Carter. Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves [J]. American Journal of Botany, 1991, 78 (7): 916~924

[4] Danson F M, Steven M D, Malthus T J, et al. High-spectral Resolution Data for Determining Leaf Water Concentration [J]. International Journal of Remote Sensing, 1992, 13: 461~470

[5] 周启发, 王人潮. 水稻氮素营养水平与光谱特征的相关研究初报. 遥感应用, 1993, (5): 9~16

- [6] 王人潮, 陈铭臻, 蒋亨显. 水稻遥感估产的农学机理研究. I. 不同氮素水平的水稻光谱特征及其敏感波段的选择. 浙江农业大学学报, 1993, 19 (增刊): 7~14
- [7] Penuelas J, Filella I, Biel C, et al. The Reflectance at the 950~970 nm Region as an Indicator of Plant Water Status. International Journal of Remote Sensing, 1993, 14: 1887~1905
- [8] Inoue Y, Morinaga S, Shibayama M. Non-destructive Estimation of Water Status of Intact Crop Leaves Based on Spectral Reflectance Measurements. Japanese Journal of Crop Science, 1993, 62: 462~469
- [9] Shibayama M, Takahashi W, Morinaga S, et al. Canopy Water Deficit Detection in Paddy Rice Using a High-resolution Field Spectro-radiometer. Remote Sensing of Environment, 1993, 45: 117~126
- [10] Fernandez S, Vidal D, Simon E, et al. Radiometric Characteristics of Triticum aestivum cv. Astral under Water and Nitrogen Stress International Journal of Remote Sensing, 1994, 15 (9): 1867~1884
- [11] 王珂, 沈掌泉, 王人潮. 利用冠层及叶片反射光谱分析速测水稻氮素的研究. 生命科学研究与应用, 杭州: 浙江大学出版社, 1996: 523~526
- [12] 秋山侃, 福原道一, 斎藤元也ら. 農業リモートセンシング. 東京, 養賢堂, 1996, 32~34
- [13] Stone M L, Solie J B, Raun W R, et al. Use of Spectral Radiance for Correcting In-season Fertilizer Nitrogen Deficiencies in Winter Wheat [J]. Transaction of the ASAE, 1996, 39 (5): 1623~1631
- [14] Penuelas J, Pinol J, Ogaya R, et al. Estimation of Plant Water Concentration by the Reflectance Water Index WI (R900/R970), International Journal of Remote Sensing, 1997, 18: 2869~2872
- [15] 蒲瑞良, 宫鹏. 森林生物化学与 CASI 高光谱分辨率遥感数据的相关分析. 遥感学报, 1997, 1 (2): 115~123
- [16] Kokaly R F, Clark R N. Spectroscopic determination of leaf biochemistry using Band-Depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing of Environment, 1999, 67: 267~287
- [17] 王纪华, 赵春江, 郭晓维等. 利用遥感方法诊断小麦叶片含水量的研究. 华北农学报, 2000, 15 (4): 68~71
- [18] 王纪华, 赵春江, 郭晓维等. 光学检测方法在小麦水分诊断及受胁机理研究中的应用. 两高一优农业与农业创新. 北京: 中国农业科技出版社, 2000, 306~310
- [19] 牛铮, 陈永华, 隋洪智等. 叶片化学组分成像光谱遥感探测机理分析. 遥感学报, 2000, 4: 125~129
- [20] 蒲瑞良, 宫鹏. 高光谱遥感及其应用, 北京: 高等教育出版社, 2000, 86~96
- [21] Lee Tarpley, Raja K, Reddy F, et al. Reflectance Indices with Precision and Accuracy in Predicting Cotton Leaf Nitrogen Concentration. Crop Science, 2000, 40: 1814~1819
- [22] 牛铮, 陈永华, 隋洪智等. 叶片化学组分成像光谱遥感探测机理分析. 遥感学报, 2000, 4: 125~129
- [23] 王纪华, 赵春江, 郭晓维等. 用光谱反射率诊断小麦叶片水分状况的研究. 中国农业科学, 2001, 34 (1): 104~107
- [24] 小西教夫, 志賀弘行, 安積大治ら. 美味しいお米を作るリモートセンシング, 測量, 2000, (7): 14~20
- [25] 田庆久, 宫鹏, 赵春江等. 利用光谱反射率诊断小麦水分状况的可行性分析. 科学通报, 2001, 45 (24): 666~669
- [26] 王秀珍, 王人潮, 李云梅等. 不同氮素营养水平的水稻冠层光谱红边参数及其应用研究. 浙江大学学报 (农业与生命科学版), 2001, 27 (3): 301~306

24. 近红外技术在过程分析中的应用

在现代工业生产中，质量保证和质量控制问题越来越受到人们的重视。在企业内部建立完善的质量保证和质量控制体系，不但可以降低生产成本、保证产品质量，而且可以节约资源、减少环境污染、保护环境、实现企业的可持续发展。近红外分析技术因其可运用于复杂背景下样品的多组分、瞬态分析和结合现代光纤技术进行可实现现场遥测而不破坏样品和干扰生产过程等特点，使其在工业现场控制中得到了广泛的应用。

24.1 过程分析技术

在传统的工业生产质量保证体系中，对中间产品或成品的检验工作由分立的化验室完成。从生产线上采集样品送到实验室进行化学分析，最终提交分析报告给生产部门，整个过程至少需要几个小时，有的甚至需要几天。一旦生产过程出现问题，需要滞后几个小时或几天的时间才能发现和采取相应的措施，也就是说这段时间生产线上还在继续生产不合格产品，从而造成巨大的经济损失和资源浪费。在生产过程控制中，常规的监控方法是采集一些温度、压力和流量等物理参数作为生产过程的控制变量，而实际影响生产过程的参数远远不止这些，只是过去由于受到检测技术和方法的限制不容易测定。有效的质量控制方法必须是在整个生产过程中确定一些直接影响最终的产品性能的关键质量控制点，并对这些控制点进行实时检测，获得这些测量点样品的各种组分信息，将这些过程信息反馈给过程控制系统，实现对生产过程的优化控制。过程分析化学对过程分析技术做了系统的研究。图 24-1 显示了传统分析方法和过程分析方法在生产过程中的工作流程。

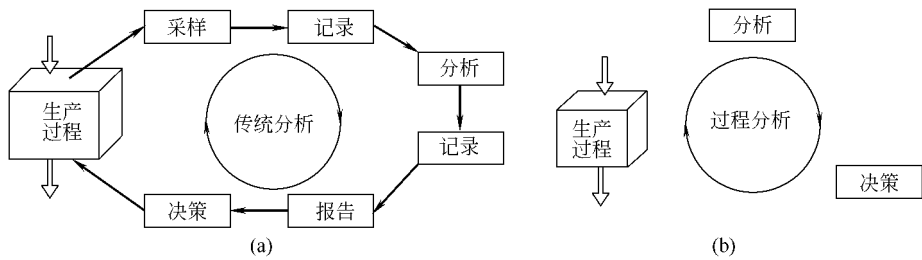


图 24-1 (a) 传统分析流程 (b) 过程分析流程

24.1.1 过程分析化学

在分析化学基础上快速发展起来的过程分析化学（PAC），对过程处理中的样品组分信息的实时分析技术做了系统的理论研究和实用技术的研究和开发，极大地促进了过程分析技术的发展。其主要研究领域包括化学计量学和过程控制算法、过程分析传感器和仪器、采样系统、过程控制系统等。至今已有三大比较有影响力的研究机构着力于过程分析

化学的研究，并在理论和应用研究领域做出了很大的贡献。它们是美国华盛顿大学的 CPAC、美国田纳西大学的 MCEC 和英国考文垂大学的 CTAC。PAC 的专业国际会议有一年一度的过程分析化学国际论坛 IFPAC 年会，FACSS、Pittsburg 和 ACS 会议也增加了过程分析化学的讨论内容。Jerome Workman 最近提出，使用过程分析技术 (PAT) 比过程分析化学这个术语更为贴切，因为过程测量参数不仅包括物质的化学特性，还应包括其他物理参数，如在制药业中需要测定混合度、药片的溶解度、干燥度等。

24.1.2 过程分析技术和方法

常用的过程分析技术包括色谱技术、近红外光谱技术、拉曼光谱技术、中红外光谱技术、X 光谱技术、核磁共振谱技术等。这些分析技术在测量参数、测量条件、分析精度、分析时间等方面各有自己的特点，本章将重点讨论近红外的过程分析技术，在此不详细讨论其他分析技术。

过程分析方法根据采样方式的不同可分为临线（线边）分析（At-Line）、侧线在线分析（On-Line）和原位在线分析（In-Line）。不接触样品分析（Noninvasive）也是一种原位在线分析。人们可以根据具体的应用需求如生产环境、样品性能、过程分析的实时性要求选择合适的分析方式。

临线分析（At-Line）方法将分析仪器如光谱仪安放在生产线附近，由人工操作仪器，通常需从生产线物质流中手动取样进行分析。分析仪器要求在生产现场正常工作，必须具有较好的抗干扰、防尘、耐高温、耐腐蚀等性能。临线分析方法可对生产线上的半成品和成品实现快速分析，但不能做到连续分析。

侧线在线分析（On-Line）方法需要在生产线物流管道上安装一个旁路取样系统，通过取样系统将生产线上的样品进行部分分流，并对样品做适当的前处理，使之适合分析仪器对样品状态的要求，再引入分析仪器进行分析。这种分析方式省去了人工取样的环节，减少了因取样引起的分析时间的延迟，提高了数据分析的反馈速度。可以避免操作人员暴露在恶劣的、有一定危险性的高温、高压、有腐蚀或有毒环境，提高了安全性能。通过样品前处理可对测量条件如温度、压力、流速等进行适当控制，并在样品被引入仪器前做一些必要的处理如过滤杂质、消除气泡等。

原位在线分析（In-Line）是直接将探头安装在生产线上或插入特定的测样部位，通常用于光谱分析特别是近红外光谱分析，将光纤探头固定安装在生产线上采集当前物流上的光谱，而连接光纤的光谱仪可以放置在生产现场之外，对生产线上的产品进行远距离的、连续、实时的监测。许多近红外光谱仪厂商可以提供耐高温、高压、腐蚀的光纤探头，使得实时过程分析更加容易实现。

不接触样品分析（Noninvasive）是指分析仪器探头不与样品进行直接接触的分析方式。如在生产线管道测量点部位安装上视窗，光谱仪探头放置在视窗外侧采集样品的透射或反射光谱进行光谱分析，这样探头与样品之间不产生直接接触。又如用光谱漫反射测量方式测量生产线传送带上的样品，探头与生产线之间保持一个固定的距离。显而易见，这种测量方式对生产线上的产品不会产生任何干扰，仪器与探头的安装对生产线的生产也不会带来太大的影响。

24.1.3 过程分析技术的优化选择

生产厂家在引入过程分析系统之前需要考虑许多问题。首先，需要了解过程分析能为工厂带来哪些收益，如生产量的提高和产品质量的控制，生产成本和能源消耗的减少，生产条件和安全性能的改善等。其次，需要估算引入过程分析系统的投资和运行成本，如仪器与采样系统的购置、安装费用，安装调试分析系统可能带来的停产，过程分析系统的使用和维护成本，操作人员培训费用等。最后，生产厂家需要从众多的分析技术做出选择，并确定将分析仪器安装在生产线上的哪些部位。

过程分析系统的任务是实时获得尽可能多的过程信息，以满足优化过程控制的实际需求。影响过程分析技术性能的主要因素包括过程变量参数以及测量位置的选择、仪器的测量精度、分析时间（取决于采样频率、采样时间和仪器的响应时间等因素）。每一种分析方法都有一定的测量精度和分析时间的限制，完全连续和百分之百精确的测量是不现实的。色谱分析方法有很高的分析精度但需要较长的分析时间，每分析一个样品需要 5~10 分钟。光谱分析方法的精度比色谱方法低但分析时间短，只需要一秒钟甚至几百毫秒就能分析一个样品。侧线在线分析方法的延迟时间比临线长，但又不如原位在线分析方法可以做到即时分析。在具体应用中需要通过分析过程信号变化的特点，综合考虑分析技术所能检测的信号强度、测量精度和分析时间，确定最佳的分析方法和安装部位。

Age K. Smilde 等在 van der Grinten 测量理论的基础上，发展了过程分析技术的性能评价理论和方法，提出了过程分析可测量度（Measurability） M 的定义，用于评价分析系统采集的离散信号重建真实连续过程信号的逼真程度。可测量度计算公式如下：

$$M = \frac{s_{\text{true}}^2 - s_{\text{error}}^2}{s_{\text{true}}^2}$$

式中 s_{true}^2 为真实信号在一定时间 T 内的离散度， s_{error}^2 为实际时间 T 内测量信号与真实信号的均方差。当可测量度 M 的值在 0.5~1.0 之间时，认为这种分析方法可以接受。Age K. Smilde 等使用测量度分析对长度为 1m 的反应罐内苯乙烯聚合反应过程的过程分析方法和检测位置进行评价和选择，分析方法包括色谱临线分析（At-line XSEC）、近红外临线分析（At-Line NIR）、短波近红外侧线在线分析（On-Line SW-NIR）、短波近红外原位分析（In-Line SW-NIR），如图 24-2 所示。这些分析方法的测量度 M 的值分别为

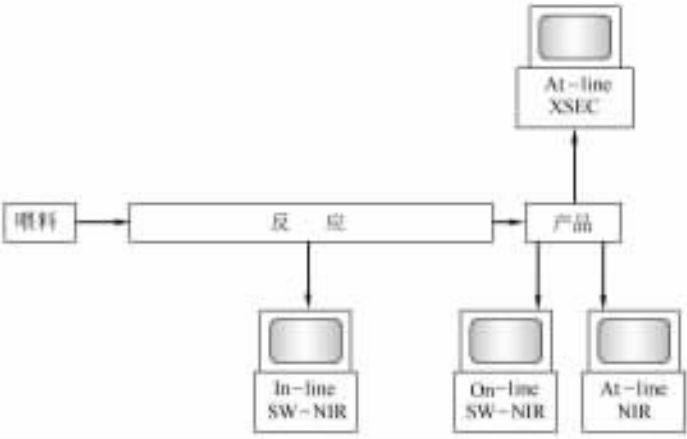


图 24-2 4 种被试验的过程分析方法在化学反应流程中测样点的分布

0.4、0.72、0.95、0.98，图 24-3 显示了色谱临线（At-Line XSEC）分析、近红外临线分析（At-Line NIR）和原位在线近红外分析（In-Line SW-NIR）这 3 种过程分析方法测定信号进行重建与真实信号的比较。显然，原位在线近红外分析方法为这些过程分析方法中的最佳选择，它基本能达到过程实时监控的需求。为了选择测量位置，将原位在线近红外分析方法分别在反应器的不同位置进行试验，图 24-4 显示了原位在线近红外分析方法在反应罐不同位置的可测量度的变化，说明最佳测量位置应是反应罐的出口处附近。

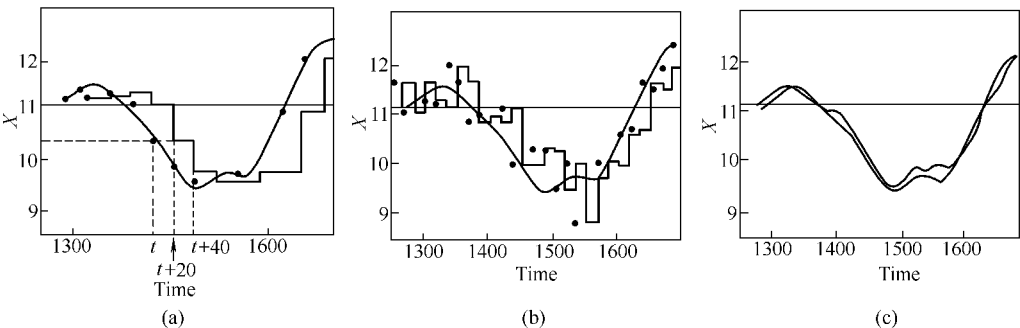


图 24-3 信号重建度比较

(a) At-line XSEC (b) At-line NIR (c) In-line NIR

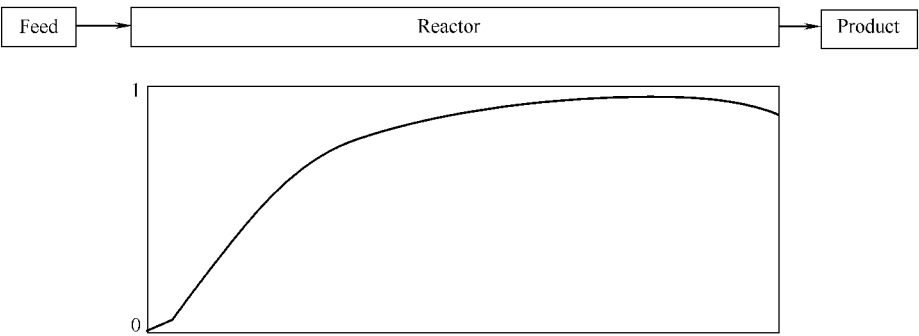


图 24-4 原位在线近红外分析方法的可测量度在反应罐不同位置的变化曲线

作为用户的一般生产厂家在过程分析仪器方面的经验和专业知识是非常有限的，他们也不可能花费很多的精力去做广泛深入的调查和研究。较好的情况是在同行业中有些厂家率先成功地使用了一些过程分析系统，为其他厂家提供借鉴经验。有一些厂家已经将一些分析方法用于实验室的常规分析，则比较容易进一步将该分析技术推广到过程分析中。作为应用厂家，当他们引进过程分析技术时，所需要的不仅仅是分析仪器，他们更需要一套完善的过程分析系统。

24.1.4 近红外过程分析技术

近红外技术在过程分析中一直处于主导地位，这是由近红外技术的特点所决定的。近红外分析技术不需要做样品制备而可以直接在复杂背景下进行无破坏、快速、多组分同时分析，它既可以分析物质的化学特性，也可以分析其物理性状。这些特点使得近红外分析

对生产线上的复杂背景下的样品不做预处理而直接进行原位瞬态分析，不干扰其生产过程。多组分同时分析的特点使得一台近红外光谱仪可以替代多种多台分析仪器，从而节省大量设备费用和维护费用，也节省大量人力和物力，因此近红外分析技术具有“分析巨人”的美誉。近红外分析因没有化学试剂的消耗和污染等副作用，被称为绿色分析技术。在人们生态环境保护意识和环保立法日益加强的现代社会中，近红外分析技术所带来的社会环境效益也是该技术得以被广泛重视的重要原因之一。

近红外光谱分析技术的迅速发展为其在过程分析中的广泛应用提供了可靠的技术保证。自 1980 年以来已经有很多仪器厂商推出了各种类型的近红外在线分析仪器，这些仪器可以将分析得到的过程参数及时传输到用于过程控制的分散控制系统（DCS）或现场总线（FIELD BUS）系统，也可以通过以太网与信息管理系统集成。生产操作优化系统可根据这些参数调整装置控制条件，实现生产过程的平稳运行，获得可靠的质量保证。近十年来光纤探头技术的快速发展和应用，更加进一步促进了近红外在线分析技术的发展。使用光纤探头可以使得光谱仪从几百米外对生产线上的产品进行远距离测量，而对生产线上的物质流不产生任何干扰，因此可直接用于易燃、易爆、有毒环境下的快速在线检测，是一种非常安全的检测方法。总之，近红外分析方法是一种理想的过程分析技术，它为生产质量保证和质量控制提供了可靠的技术支持。从生产原材料的鉴定，到生产过程中各种参数的测定和中间产品的组分分析，直至最后的成品品质分析，近红外分析技术几乎可用于其中的每一个质量控制环节，实现对生产过程的连续、实时的过程监控，确保原材料和生产过程的长期稳定和一致，获得稳定的高质量的产品。

在近红外过程分析中，比较常用的是在线分析方法，如侧线在线分析、原位在线分析包括非接触式分析。本章将重点讨论这些在线近红外分析技术的系统开发和应用。

24.2 在线近红外分析系统

在线近红外过程分析系统在硬件配置上除了作为主体部件的光谱仪外，还需要配置自动采样系统用于对生产线上的样品采集光谱，以及各种标准的通讯接口实现与过程控制系统和企业信息管理系统的连接。在软件配置上需要安装用于过程分析的校正模型，提供系统自动诊断和维护软件，提供分析报告文档自动生成软件等。

近红外在线分析系统包括采样系统、光谱仪器、分析模型、分析和维护软件、数据通讯等多个子系统，在实际应用中，需要针对具体的测试对象、分析要求和测量环境进行合理的系统配置。

24.2.1 近红外在线采样系统

在线过程分析系统中，采样系统方面的花费通常将近占用整个分析系统总成本的 40%~50% 和光谱仪器的价格相当。它是过程分析系统中的关键部件，在整个测量系统中的维护率是最高的，也是在整个分析系统中最有可能导致分析项目失败的部件。如果采样系统不能正常工作，整个分析系统将毫无价值。自从 20 世纪 80 年代中期引入在线分析技术以来，开始几年的应用并不顺利，除了仪器硬件成本高、建立校正模型费时等因素以外，还有一个非常重要的原因是受到了采样系统的可靠性和稳定性的限制，特别是对于恶

劣的测量环境，采样系统的安装容易失败。近几年来，在线分析技术发展非常迅猛，其中很重要的一个原因是采样技术的发展。

在设计过程采样系统时，必须慎重考虑以下几个问题：①采集的样品是否能代表过程处理中的样品；②采样系统是否能以实时方式为分析仪提供样品；③样品状态能否满足分析仪进行准确分析的要求；④采样系统是否有较好的成本效率；⑤采样系统是否可靠；⑥采样系统是否易于维护；⑦采样系统能否在任何条件下安全运行。

在线近红外过程分析中，采样方式大体上可分为旁路分流采样、原位采样和非接触式采样三大类。样品的物理状态和所处环境是影响采样方式的重要因素，下面分别就液体类、固体类样品和悬浮液与乳状液样品的在线采样方法做分类介绍。

24.2.1.1 液体类样品

液体类样品的光谱采集一般采用透射或透反射方式。采样方式可选用分流采样、原位采样和非接触式采样等多种方式。

旁路分流采样是液体在线过程分析中最为经典的采样方式。样品需要经过一系列处理：通过连接在生产线输流管道的旁路，从过程流中分流，传送给样品调节系统做适当的前处理，使得样品能满足分析仪器对温度、压力等物理状态的要求，然后传送给光谱仪进行测定，最后将分析完后的样品送入废料流中进行适当的处理。

常见的液体分流采样系统如图 24-5 所示，它通常包括两个子系统：快速循环子系统和样品调节子系统。快速循环子系统（Fast Loop）的液流速度必须足够大，以保证经此系统流入分析仪（Analyzer）的液体样品能代表当前处理过程中物流或反应器内的物体状态。样品调节系统（Sample Conditioning System）必须维持流入分析仪的样品状态尽量和处理过程中的物流的真实状态保持一致，还要去除液体中容易干扰分析的气泡和杂质。典型的分流采样系统一般包括：采样器、阀门、过滤器和分离器、液流测速控制仪、液压测定和调整仪、液体泵等。

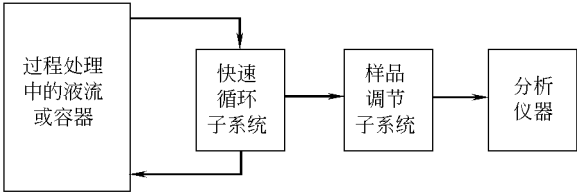


图 24-5 液体分流采样系统结构

采用分流系统的最大优点是可以根据分析仪器对被测样品的要求对样品做适当的前处理，还可以通过安装过滤网和气泡消除装置去掉液体样品中的杂质和气泡，提高测量精度。另外，用于模型建立和校正的标准样品和实验室分析样品的也可以方便地在分流取样系统中采集。

对于大多数的处理过程如石油调和过程，样品的前处理比较简单。但是，有一些处理过程样品变化非常快，温度和压力很高，化学物质有较强的腐蚀性，环境非常恶劣，有一些物质离开其反应条件，特性将发生变化。这些情况下样品的前处理较为困难，有的甚至不可能，需要考虑采用原位采样。

原位采样是把分析仪的传感器（探头）直接安装在过程处理中装载样品的容器（如管

道、反应罐) 内进行直接采样的方式。光纤探头是现代光谱过程分析中最常用的原位采样探头。常用的插入液体内的光纤探头有两种类型。一种是透反射型光纤探头, 这类探头的一头是一个全反射镜子, 另一头连接光纤束, 其中心部分光纤用于接收透射光, 外围的光纤传送入射光, 测量有效光程是探头液体槽宽度的 2 倍, 图 24-6 显示了透反射光纤探头的结构。透反射测量方式可用于测量透明或有轻微散射的液体测量。另一种是透射型光纤探头, 这类探头的一头连接入射光纤, 另一头连接出射光纤, 如图 24-7 所示。透射型光纤探头可用于测量透明和不透明的液体。

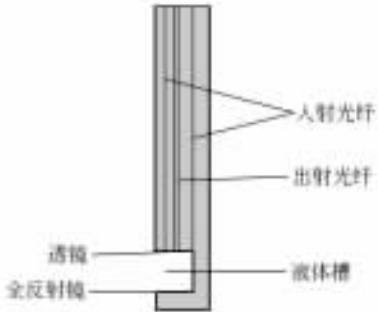


图 24-6 透反射液体光纤结构

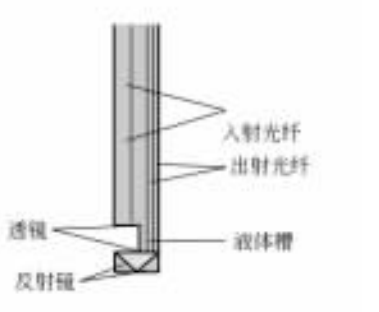


图 24-7 透射液体光纤结构

有一些专用于液体过程分析的在线近红外分析系统, 直接把液体流动池安装在生产线的物流管道上, 进行原位在线过程分析, 分析仪器与生产线集成在一起。图 24-8 显示了将液体流动池安装在牛奶高温消毒生产线上, 进行牛奶成分的近红外在线分析的过程。

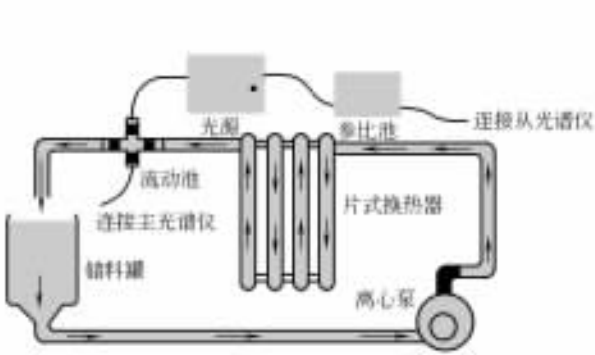


图 24-8 液体流动池用于牛奶原位在线分析

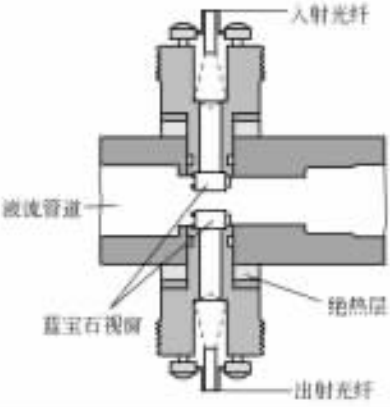


图 24-9 液流的非接触测量

将一对互成 180° 的透射光纤探头安装在液流管道两侧, 可以实现液体的非接触测量。这种测量方式需要在管道的测量部位安装两个相对的视窗, 管道的粗细程度也要适合测量光程的需求。这种测量方式可用来测量透明和不透明的液体, 也可用于分析浆状、糊状物质。图 24-9 显示了这类采样方式的结构示意图。

在选择液体采样方案时, 需要重点考虑测量光程问题。选择较长的光程有利于黏滞型液体的流动, 也可以减弱气泡、杂质的干扰影响, 还便于探头和样品池的清洗。为提高测量精度, 光程的选择要考虑到样品的吸光度变化范围应在仪器的线性响应区间内。测量波段的选择也影响光程的选择, 由于长波区 ($1500 \sim 2500\text{nm}$) 的散射和吸收较强, 光程的

可选择范围在 0.5~5mm 之间，短波区（700~1500nm）测量的光程可能在 5~30mm 之间。波段的选择与被测物质的成分以及要求的精度有关。长波区可以提供更为精细的结构和测量灵敏度。

选择和设计液体采样方案时还要注意测量样品的温度、压力、流速等物理状态以及液体中可能出现的杂质和气泡等因素的影响。温度的变化将会引起光谱的基线漂移和官能团的吸收峰产生位移，压力变化会使得液体内部产生气泡，液体中的杂质和气泡进入光路后将对光产生散射从而降低的信号强度，流速过高将会产生湍流引起光谱的基线扭曲。

采用分流采样系统可以适当降低压力和流速，也可以在液体流动池前面安装恒温装置，通过安装水浴恒温装置，对于在 30~80℃ 之间变化可以把温度控制在 $\pm 0.5^{\circ}\text{C}$ 的变化范围。对于超过 100℃ 的情况和不适合旁路引流的黏性液体，很难实现恒温控制，需要对模型做温度校正。在样品前处理系统中安装过滤网和气泡消除器以减弱气泡和杂质对测量结果的影响。适当增加流动池的宽度或光纤数目可以提高信号强度，可以适当补偿气泡与杂质所导致的信噪比的降低。

用于原位在线测量的光纤探头和液体流动池，需要有较好的耐高温、高压和耐腐蚀等性能。许多专业的厂商提供了专用的工业在线近红外探头，有些探头可在 1000℃ 高温和 6kPa 高压下正常工作。工业用探头一般采用不锈钢或哈司特镍合金铸造，所有的接头使用电子束焊接技术焊接，接口封装处不使用橡胶等弹性材料或铜等熔点低的金属材料，用蓝宝石材料作视窗。图 24-10 显示了 Axiom 的透射和反射光纤探头。图 24-11 显示了 Desmond Brennan 等设计的液体流动池，池体使用 316L 不锈钢材料制造，采用耐酸碱、耐高温的蓝宝石做视窗材料。

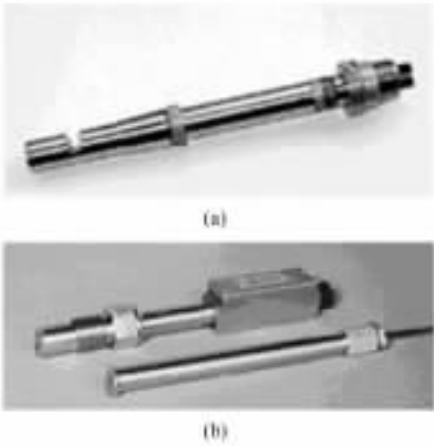


图 24-10 耐高温、高压的光纤探头
(a) 透射光纤探头 (b) 反射光纤探头



图 24-11 耐酸碱、耐高温的液体流动池

24.2.1.2 固体类样品

固体样品的在线测量一般采用漫反射方式，波长范围在 1100~2500nm 之间。在漫反射方式下采集光谱，样品的物理特性（如颗粒度、表面粗糙度、形状、方向、密度、厚度等）对光谱特征的影响很大，样品必须满足以下几个条件以保证样品光谱的重现性：①样品的厚度为无限厚，即在此基础上再增加样品的厚度不会增加样品的反射吸光度；②样品的颗粒度均匀；③样品的松紧度一致。在生产过程中，固体样品具有不同的形态，需要针

对不同类型的样品形状设计相应的采样系统。

测量粉末或颗粒状样品，可以将反射光纤探头直接安装在生产线上，插入容器内。在选择安装位置时，注意保证样品状态在测量前后的一致性。图 24-12 显示了混合器中安装 NIR 探头的示意图 [S. Sonja Sekulic]，这是一种典型的固体样品的原位测量方式，漫反射光纤直接插入被测样品进行光谱采集。

对于不均匀的粉末或较粗的颗粒可以采用非接触采样方式，将探头安装在料斗或传输管道上并与内壁平面齐平，通过玻璃把探头与样品隔开。图 24-13 显示了 Foss 在线近红外分析仪 NIR 探头与装料斗内壁齐平的非接触方式测量。

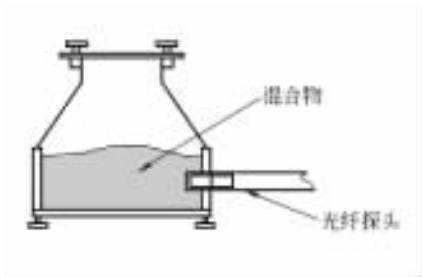


图 24-12 光纤探头插入混合器内

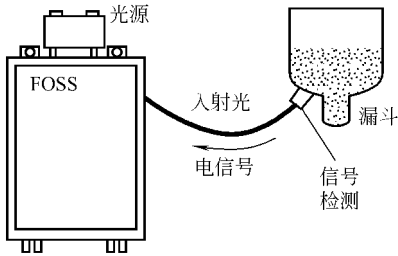


图 24-13 NIR 探头与装料斗内壁齐平

另一种开放式的非接触的测量方案是 NIR 探头在与固体物流保持一定距离，这个距离可根据被测样品的反射率、仪器的探测效率以及期望的光斑面积等因素来确定。在实际测量时有可能出现样品中有空隙或传送带上无物质等情况，在模型中可考虑使用模式识别方法检测这些特异情况。图 24-14 [G. Tügersen99] 显示一种了开放式的非接触式采样方式，NIR 探头安装在绞肉机的出口上方，对肉末进行在线测量。

对于网状或片状样品，一般采用非接触方式测量。这类样品均匀性较好，尺寸较大，在生产线上传送速度较快，可能达到每秒钟数米的移动速度。因此，这类样品测量的结果通常是平均值，如成分的平均含量、样品的平均厚度等。测量这类样品时，可以考虑把测样点选择在滚筒附近以防止样品飘动。图 24-15 所示为 Bruker 的 Matrix-E 光谱仪在滚筒附近对纸做在线分析。

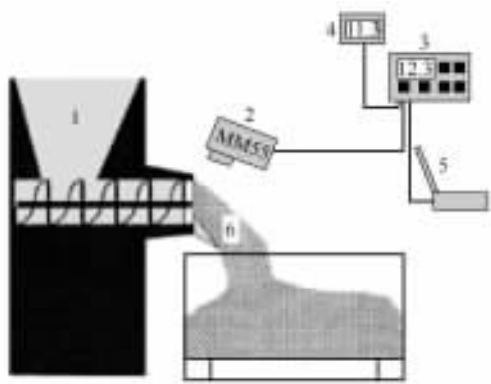


图 24-14 NIR 探头安装在出料口上方

1—绞肉机 2—MM55 NIR 探头 3—控制单元
4—远程显示 5—微型电脑 6—肉末流

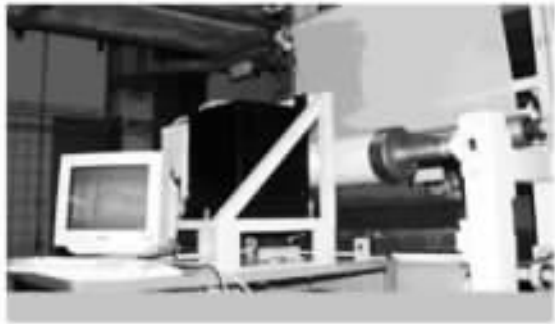


图 24-15 在滚筒附近对纸做在线分析

薄膜或纤维状样品的测量难度较大，因样品厚度达不到无穷厚度，用通常的漫反射方法测量时信号较弱，光谱重现性差，影响检测精度。用漫反射方法测量时，在被测物底层放置一层漫反射物，可以加强光的散射程度，增加有效光程。这种方法还可以改善因样品不均匀所造成的误差。

24.2.1.3 悬浮液与乳状液

这类物体的测定比液体和固体样品的难度更大，其测量方式可以根据物质的光谱特性与分析要求选择反射或透射。如果需要测量的是液体相部分物质，推荐使用反射方式测量，这是由于颗粒的后向散射将使大部分的光集中在液体内，使反射光谱携带丰富的液体相物质的信息；相反，透射方式则适用于测量固体相物质。在生产过程中可能在前期为较透明的液体，逐渐的液体变得混浊，需要设计既可用于反射方式也可用于透射方式的专用探头。

在过程分析的实际应用中，人们已经越来越认识到采样系统在整个过程分析中的重要地位，更加强了采样系统的研究。采样系统正朝着模块化、微型化、智能化的方向发展。

24.2.2 近红外在线测量仪器

20 世纪 80 年代以来，在线 NIR 技术应用发展很快，对性能的要求也在不断提高。下面列出了用于控制塑料薄膜和涂层质量的 NIR 在线分析仪的性能需求的变化：

	精度	响应时间
1975 年	$\pm 1\mu\text{m}$	1s
1985 年	$\pm 0.1\mu\text{m}$	200ms
1995 年	$\pm 0.05\mu\text{m}$	10ms~100ms

除了仪器本身的测量性能以外，随着应用领域的拓展，近红外在线分析仪器从可靠性、采样方式、通讯接口、智能化程度等方面提出了更多更高的要求。

NIR 在线测量仪器被用于控制循环需要一天 24 小时不间断工作。这就要求光谱仪必须长期稳定可靠、坚固耐用，以保障仪器使用的低故障率和尽可能少的维护。仪器的自检功能对于在线分析也很重要，它有助于避免仪器性能的变化对分析结果的影响。现场使用和近距离测量方式的仪器或探头必须能在非常恶劣的环境下正常工作，如抵抗温度、湿度的变化和粉尘、振动、光电供应不稳定等因素的影响，仪器还应符合 EMC 规范。在易燃易爆环境下使用的仪器还需要考虑防爆设计，以保障其使用的安全性。

NIRS 在线测量光谱仪还应配置连接探头的接口以便于在线采样，配置的光纤应具有耐弯曲和抗挤压性能，光纤的最大长度决定了远程测量的距离，配置多路转换器，则可以实现一台仪器对多个流的循环分析。

NIR 在线测量最终目的是为过程控制服务，这就需要 NIR 仪器能把所测得的样品组分信息按照控制系统能够识别的方式传送给控制器。对于常用的分布式控制系统 DCS，需要把这些组分信息与温度、压力等参数一样转换为 4~20mA 之间的模拟量送至 DCS。过程逻辑控制器 PLC 则可以接受更丰富的数字信息用于过程控制，NIR 仪器一般提供了数字通讯接口，常用的数字通讯协议有 ModBUS、FieldBUS、OPC 等。有些仪器配置了以太网接口，可以实现与计算机网络的连接。

NIRS 在线光谱仪和实验室光谱仪一样可以分为离散波长和连续波长两大类型，离散

波长仪器主要有滤光片型、LED 型两种。连续波长仪器有光栅扫描、FT 变换、CCD 阵列和 AOTFS 等结构类型。仪器的选择需要考虑过程控制对仪器的选择性、灵敏度、可靠性等性能的要求。在选择仪器时，很难绝对地说哪一种仪器最好，主要看仪器的分析能力和稳定性能是否能达到过程控制的实际需求。

24. 2. 2. 1 离散型在线光谱仪

离散型仪器价格较低，一般适用于分析含量较高（百分之几的水平），吸收峰比较独立的组分。滤光片型仪器具有较高的信噪比而且坚固耐用，可安装在工业现场进行近距离测量。这类仪器较早用于在线测量，分析结果可靠。典型的滤光片型在线分析仪器如 MM55（Infrared Engineering, Inc., Irwindale, CA, USA）型光谱仪。MM55 近红外在



图 24-16 MM55

线分析仪在食品加工工业中应用较为广泛，该仪器采用干涉滤光片作为分光器件。可以根据不同的应用需要和经验知识选择滤光片组，并且建立了通用模型。该仪器提供了多种通讯接口，可与 PLCs, PCs 和 CSADA 控制系统集成。图 24-16 为将 MM55plus 安装在炸土豆片生产线上，监测从油锅和烤箱出来的土豆片的含油量和含水量，用于优化控制系统调节处理时间和温度。

另一类离散型仪器是由几个不同波长的 LED 管作为离散光源的 LED 型仪器，其最大优点是仪器内部没有任何移动部件，较滤光片仪器更为坚固耐用。现有的 LED 仪器工作波长一般在 700~1100nm 之间，可用以测量固体以及其他散射较强的物质，配上光纤探头既可用于近距离测量，也可用于远距离测量。值得注意的一个问题是，LED 管的替换会影响波长的重现性，从而影响测量性能。近些年推出的以 LED 管作为光源的滤光片仪器，结合了传统的滤光片仪器与 LED 型仪器的优点。Foss 公司开发的一款 LED 型仪器用光栅作为分光器，光谱分辨率为 13nm。图 24-17 为该仪器的外形，已被成功用于农产品和石油化工生产的过程分析。

24. 2. 2. 2 连续型在线光谱仪

连续型光谱仪因其能提供更多的光谱信息而在工业控制中具有更广泛的应用。连续型光谱仪提供连续的光谱信息，结合化学计量学多变量分析方法和光谱预处理方法，可获得较高的测量精度，选择性好。使用连续光谱结合模式识别技术还可以判别异常样品。虽然连续光谱仪有以上诸多优点，但仪器价格比离散型仪器高。

传统的光栅型光谱仪采用机械传动方式转动光栅，扫描速度慢（扫一条光谱至少需要 1min），抗振性能差，只能用于实验室测量。现代的光栅光谱仪将光栅驱动改进为光控方式，提高了扫描速度（一秒钟可扫描多次）而且抗振性好，配置光纤探头可进行近距离或远距离在线测量。FOSS 的 NirSystems Process Analyzer 为典型的现代光栅型在线近红外光谱仪，如图 24-18。该仪器可连接一对互成 180 度的透射光纤探头进行透射方式测量，配置透反射光纤探头可实现透反射方式测量，配置反射光纤可进行反射方式的测量。光纤



图 24-17 Foss LED 型仪器

探头可工作在 35MPa 高压和 300℃ 高温环境下。透射测量方式的光程可在 1~20mm 间连续可调。透反射光纤的液体槽在 1~10mm 范围内可调,有效光程 2~20mm。

Foss 的 MODEL SY-4402-X, SY-4467-X 光栅型在线光谱仪是为非接触测量方式设计的,采用反射测量方式,适合用于环境光较弱的场合。测量距离可在 1mm~30cm 范围内调整。可用于分析传送带上运动的物质,也可以安装在管道或漏斗内壁。仪器还提供 NEMA 4 和 NEMA 4X 防护罩。主要用于塑料、造纸和纺织生产中的过程控制。

二极管阵列光谱仪,具有更高的扫描速度,其工作波长在短波区,没有可移动部件使得仪器更为可靠,在现场控制有较好的应用前景。

图 24-19 所示的 OMA-517 为一种典型的二极管阵列在线分析光谱仪。该仪器被放置在 NEMA4 保护箱中以抵御恶劣的环境,仪器提供 4 路 4~20mA 的模拟输出量和 RS232、RS485 通讯接口,还提供以太网接口。被应用于石油化工、制药等行业的在线分析。

AIT 的 PIONIR 1024 (图 24-20) 也是一种二极管阵列型的在线近红外光谱仪,响应时间为 1min。可在 -40~50℃ 的环境温度下正常工作,仪器放置在一个防爆设计的防护箱内,也能防雨,可放置在户外工作。提供控制系统接口、远程调制解调器,光纤长度可达 1km,可用于石油炼化厂的过程控制。



图 24-18 NirSystems Process Analyzer



图 24-19 OMA-517



图 24-20 PIONIR 1024

PetroMetriX Beacon 2000-II 二极管阵列型在线分析系统,通过放置在一个中央仪器室的主分析仪器能够实现全厂范围内的多路近红外过程分析。可连接 15 个测量分路,各个分路的分析时间 10~30s,测量点到主分析仪器的最大距离可达 3km。提供 Modbus 接口,可选择 8~16 路 4~20mA 的模拟量输出和 16 路 0~5VDC。该系统被成功应用于石化生产工艺的过程控制。

傅里叶变换光谱仪的最大优点是波长精确度高、扫描速度快,重复性好。现代傅里叶

光谱仪解决了以往抗振性能差的问题，已被广泛用于测量精度、选择性和灵敏度要求高的在线测量。Bruker 的 MATRIX™-E 和 MATRIX™-F（图 24-21）是典型的傅里叶型在线近红外光谱仪。MATRIX™-E 仪器设计为漫反射非接触测量方式，最佳测量距离 17cm，最大测量距离可达 50cm。体积小，可直接安装在传送带上。仪器外面罩上了坚固的外壳，振动和热浪冲击不影响仪器的性能。仪器还具有自诊断和周期性自动测试功能。可选择使用工业标准通讯协议 Modbus 或 4~20mA 的模拟量输出与 DCS 过程控制系统集成。该仪器已成功用于造纸、纺织、化工以及制药生产的过程分析。MATRIX™-F 型仪器提供光纤连接和多路转换测量功能，一台仪器可以连接 48 个测量点。可放置在标准 19 英寸的工业机柜内。可用于测定液体、乳状液、粉末和颗粒状样品。主要应用于石油、化工、造纸、制药生产的过程控制。

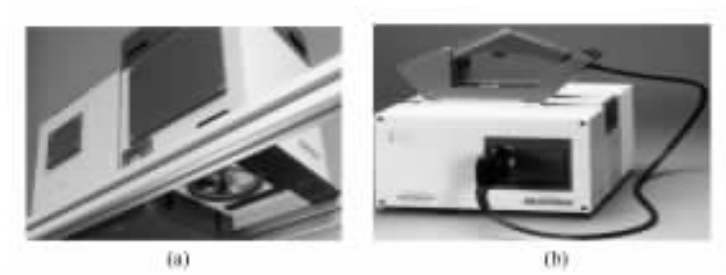


图 24-21

(a) MATRIX™-E (b) MATRIX™-F

AOTF 型仪器扫描速度很快，100ms 可以扫描一条光谱，因此适合于对实时性要求很高的控制过程中，可通过多次扫描提高 S/N 比，提高测量精度。工业现场的电磁场和振动可能会干扰 AOTF 的测量，宜采用远距离测量方式。



图 24-22 AOTF-NIR Analyzer

ROSEMOUNT Analytical AOTF-NIR Analyzer（图 24-22）为典型的 AOTF 型近红外在线分析仪器。该仪器波长稳定性好，采样系统使用单芯光纤，可使光纤长度达到几百米。分离的模块化仪器结构，非常适合于远程测量，仪器提供自诊断功能。典型的应用分析如过氧化氢、乙基氯，有机

物低 PPM 的水含量，无机物的水分百分含量分析和腐蚀物洗涤控制。

24.2.3 在线分析模型

对于近红外分析来说，模型是近红外分析技术的核心。分析模型的建立和使用，需要遵循一般的近红外分析模型的步骤，首先是收集标准样品并建立校正集与检验集，然后是选择光谱预处理方法和定量分析方法标定模型和检验模型直至获得最优化模型，最后是在实际应用中使用时维护模型。但是，由于在线分析模型用于实时测量生产线上流动的样品，其测量状态和测量背景不像实验室分析那样易于控制，这就需要在在线分析模型的开发与应用中更多考虑生产现场的实际问题。稳定可靠的在线过程分析的模型，其测量结果不应受温度的变化、振动、气泡和杂质等影响，并且能够适应过程的变化。

一个稳定可靠的在线分析模型的建立，是一个极其复杂和费时的过程。这就需要在标

准样品的收集和贮存、化学方法标准值的测定、标准样品的光谱采集、建模方法的选择、模型的检验和评价等每一个过程进行规范操作。ASTM 为近红外分析模型的建立、检验和维护制定了具体的标准化操作规范。建立分析模型可参照 ASTM E-1655 标准, ASTM D-2885/3764 则提供了模型自动检验标准, ASTM D-6122 为自动检验特异样品和判定测量值漂移标准。石化行业的近红外过程分析模型的建立、检验和维护操作已经使用了 ASTM 标准, 模型的在线分析的精度和稳定性有了可靠的保证。

24.2.3.1 样品的收集

开发一个稳定可靠的 NIR 定量分析模型, 必须收集一批具有代表性的样品作为校正样品集。这个样品集不仅需要覆盖被测成分所有可能测试的浓度范围, 而且还要包含其他影响光谱的因素, 如基体组分的变化、温度的变化等。在实验室很难模拟出在线测量时的各种变化, 校正集样品需要从生产过程中采集, 才能使得建立的模型适用于在线分析。但问题是生产过程在短时间内不会有很大的变化, 也就意味着不可能在短时间内建立一个完全代表在线测试样品的校正集。在实际建模过程中, 可以先建立一个初始基础模型, 模型的校正集包含 60%~80% 的样品变化范围。然后在较长的一段时间内, 使用基础模型对生产过程进行在线分析, 并进一步检测光谱与样品的变化, 把这些变化的样品不断补充到校正集, 逐步扩展校正集直至该校正集包含了生产过程中被检测样品所有可能的变化。建立一个初始样品集, 对于连续生产过程至少需要收集两个循环, 特别要注意收集过程处理的起始点和终止点的样品, 它们包含更多的样品变化。对于成批处理的过程需要至少收集 5~6 个不同批次。除了考虑过程处理的变化, 还要注意不同季节环境温度和湿度的变化。有一些复杂的过程在线分析稳定模型的建立需要长达两年时间。

采样过程和样品的贮存、运送操作必须严格按照操作规范, 参考化学方法的标准值测定也必须严格按照相应的标准规范操作, 以确保样品和标准值稳定的重复性。

24.2.3.2 建模方法的选择

NIR 建模方法分为多波长分析和全波长分析两类, 它们各有自己的优缺点。对于在线分析, 模型的稳定可靠性(即适应预测分析过程中可能产生的不同变化的分析能力)非常重要, 甚至超过对分析精度的要求。因此在线分析模型建立的准则是, 只要能满足过程控制的精度要求, 建立模型尽量使用简单的方法, 多波长法比全波长法简单, 采用多波长方法如 MLR 则要求选择较少的波长点, 使用全波长法如 PLS、PCR 方法则尽可能选用较少的主成分。选择光谱区间时要考虑吸光度变化范围应在仪器的线性响应区间, 同时尽量避免包含水的吸收峰区间, 因其随着环境和气候的变化较大而容易导致误判。

对于离散型光谱测量, 一般采用 MLR 等多波长建模方法。波长的选择主要考虑与被测成分相关的特征波长, 还需要选择一些波长用于校正基线或重叠吸收峰的干扰以及样品的散射。建立的模型对于没有包含在校正集中的光谱变化不敏感, 只要收集几十个样品校正集就能建立稳定的模型, 它更适合于较为简单的应用。

对于连续型光谱测量, 一般可以对光谱做预处理如求导、附加散射校正等。可以采用多波长方法建模, 校正集需要几十个样品, 模型稳定性较好, 样品的背景变化对测量结果的影响较小。但多波长方法对于样品背景极其复杂, 特征吸收峰重叠很严重的应用极为有限, 也不适于测定物理参数。更为复杂的 PLS、PCR 等全波长建模方法可以充分应用连续光谱的丰富而又精细的信息, 能够分析复杂样品背景中基体组分和过程控制条件变化所

引起的样品光谱的细微变化特征，获得较高的分析精度，也可用于测量物理参数如温度的变化。全波长建模方法需要的校正集较大，一般需要上百个样品，并且对于未加入校正集的各种变化较为敏感。

建立在线使用模型时需要特别考虑一年四季温度的变化。温度的变化对 NIR 光谱的影响主要在以下两个方面：一是使液体样品光谱产生基线漂移，它是由温度的变化而产生液体的密度的变化而引起的，一般通过基线校正如求解光谱的一阶导数和二阶导数等方法修正。二是使得极性官能团的吸收峰产生位移，这是由于温度影响着极性官能团与氢键的结合程度，从而改变其特征吸收峰的位置。图 24-23 (a) 显示了 Desmond Brennan 等通过实验测得的温度与水峰位置之间的线性关系，温度改变 1°C 将引起水峰 0.2nm 的位移。图 24-23 (b) 显示了 G. Toersen 等测定的冻牛肉在不同温度下的光谱。在 NIR 在线过程测量中，温度所引起的峰位移动将严重影响测量结果，这就要求在建立模型时需要考虑对温度的修正。通常有两种校正方法：一种是建立一个不考虑温度影响的模型，使用它分析样品光谱时，用其他传感器测定的温度作线性校正。这种线性补偿的校正方法适合用于温度变化范围较小，而且产生的峰位漂移也较小的情况。当温度变化较大时，这种方法就不再适用，需要用另一种更为复杂的温度校正模型方法。温度校正模型把温度作为影响光谱变化的一个重要因素，这种校正方法需要采集在线过程中所有可能温度的样品和光谱构造校正集，采用 PLS、PCR 等全波段建模方法分析光谱。

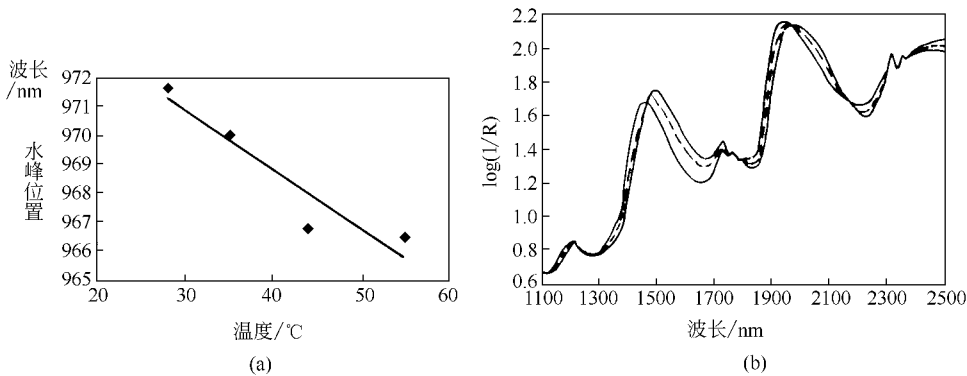


图 24-23 温度对光谱的影响
(a) 水峰位置随温度的变化 (b) 不同温度下的牛肉光谱

生产过程中的液体不可能是完全纯净的，其中会夹杂一些杂质，也会产生气泡。液体中的杂质和气泡进入光路后将对光产生散射从而影响测量结果。使用过滤网可以减少杂质，防止杂质液体聚集在流动池或光纤探头内形成污垢，但它并不能过滤掉所有的杂质，而液体在流动中也会产生新的气泡。气泡和杂质对光的散射将降低光谱信号的信噪比，改变有效光程，产生基线漂移。对光谱做一阶导数和二阶导数等基线校正，可以适当减弱杂质与气泡产生散射效应。

24.2.3.3 模型的检验

模型的优劣主要看它的预测精度，在实际应用中可以预先建立多个模型，通过检验集对模型的预测能力检验来确定最佳模型。

对于在线测量模型的检验需要直接从生产线上采集检验样品。为了使得这些检验样品

能够完全反应将来被分析的样品，要求在较长的一段时间内收集样品，对于连续过程一般需要持续多个生产周期，而成批处理过程则需要多个批次。如果预测标准差比检验标准差大得多，可能是模型的建立产生了过度拟合，需要减少模型中的波长点或者主成分个数；也可能是校正集不能代表检验集的样品，需要扩展校正集。对于特异样品，在排除样品采集或标准值的问题以后，确定校正集中缺少此类代表样品，加入校正集。

24.2.3.4 模型的维护

当在线测量模型投入分析应用后，NIR 测量就成为过程控制中的一种常规测量，其测量结果需要连续不断地送给控制系统。模型的维护工作应该由普通的操作员而不是一名近红外应用专家来担任，这就需要系统提供自动化程度较高的维护方案。模型预测性能受到两大基本因素影响：一是仪器的系统漂移，二是样品的变化。

为确保仪器的可靠性，常规的仪器诊断数据如波长准确度、噪声水平、带宽以及参考标准样品的光谱响应等应该做自动记录。还需要经常性地抽取一些控制样本进行 NIR 测量和参考方法测量的对比以检验 NIR 方法的性能，一般每隔 4~8 小时需要做一次验证工作，并记录检验结果。把这些记录结果绘制成一个控制图表可以有效地监控仪器和测量模型的性能。

连续型光谱仪提供的全光谱还可用来直接识别样品的变化是否在校正集覆盖的范围。用模式识别或异常检测技术可以检测出被测样品光谱与校正集和检验集光谱的差异，虽然它不能确认差异是否来源于仪器或被测样品，但还是可以作为一种辅助的诊断方法。

对于异常样品的处理，在实验室测量中，对于通常需要判断样品是否未包含校正集中的新样品，扩展加到校正集并重新建模。这种处理方法不适用于在线测量，在线模型用于控制循环中以后，不允许进行频繁的模式重建工作。因此，在线测量模型必须在确定建立完善后才能投入使用，特异样品的出现则被认为是过程控制出现异常，应发出报警信息。

24.3 NIR 在线过程分析应用

进入 20 世纪 90 年代以来，NIR 作为一门现代测试技术已基本成熟，其应用领域迅速扩大，从最初的食品行业，迅速渗透到石油、化工、环境和生化等行业，在工艺控制、产品质量分析检测方面发挥了突出的作用，并创造了巨大的经济效益。在前面的章节中已经详细介绍了这些领域的近红外应用，这里将重点介绍近红外在过程分析方面的应用。

24.3.1 农业和食品加工业应用

近红外技术最早应用研究工作始于农产品的快速品质分析，这些应用研究作为 NIRS 在线分析技术在农业和食品加工业的应用打下了良好的基础，使其在农产品和饲料加工、食品和饮料生产中得到了极其广泛的应用。在 20 世纪 80 年代中期，专用于农产品和食品加工在线过程分析的近红外分析系统被开发并得到了成功应用。以后，专用于肉类、乳品类在线分析的近红外分析系统逐渐被成功用于过程控制。

近红外分析技术在饲料加工业中被成功用于分析混合饲料中的蛋白质、纤维素、油脂、氨基酸、糖分、灰分等含量。FKRA 饲料厂在 1995 年底引入了近红外在线分析仪，用于在原料入口处进行原料自动分析和 7 条生产线上的生产成品的实时分析。试运行两个

月共分析了 25000 份样品。他们使用在线近红外分析系统实时检测原材料是否合格，并根据原材料的组分含量优化饲料配方，优化原材料的利用。使用近红外分析方法对饲料产品的蛋白质含量的分析，实现了饲料蛋白质含量的控制，既获得了稳定质量的饲料，又大大降低了生产成本。近红外在线分析系统的使用，还为 FKRA 生产过程提供全天 24h 在线控制。他们原来的分析实验室开放时间是 8:00am~16:00pm，也就是说在一天 24h 的生产过程中，有 16h 的生产无法对蛋白质、脂肪、水分等含量进行控制。使用了近红外在线

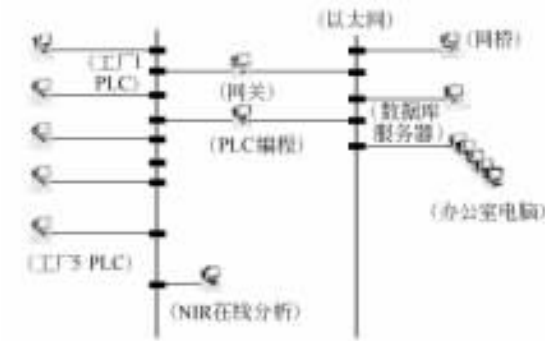


图 24-24 NIR 与控制系统和以太网集成

分析，全天 24h 的连续分析获得的组分信息被实时传送给饲料配方优化系统。FKRA 在 1994 年获得了 ISO 9002 认证，NIR 在线分析提供了质量保证的技术支持，并实现了给予用户以质论价的承诺。FKRA 将近红外在线分析系统与 PLC 连接，实现了饲料生产过程的控制，并通过以太网与信息管理系统集成，为系统决策及时提供可靠的测量数据。图 24-24 为近红外在线分析系统与 PLC 控制系统和信息管理系统集成的结构示意图。

在农产品加工过程中，NIR 探头被直接安装在传送谷物、面粉的传送带上，测定农产品的水分、蛋白质等含量。图 24-25 为在线近红外分析方法测定小麦面粉的蛋白质含量，测量的结果送给反馈控制系统来调节面筋的速度，控制蛋白质的水平。图 24-26 显示了传送带上柑橘的近红外在线分析。

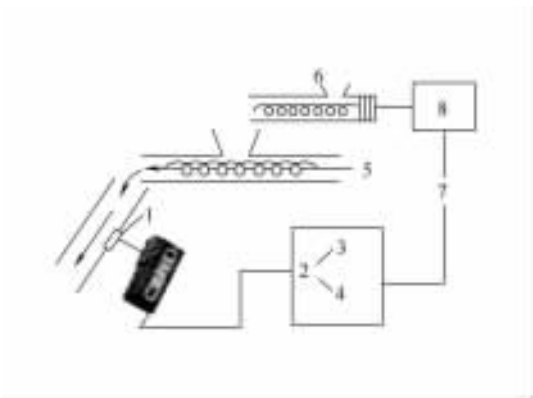


图 24-25 近红外在线分析用于面粉加工

1—蓝宝石视窗 2—电路控制单元 3—图表 4—PID 控制器 5—面粉 6—面筋 7—控制信号（4~20mA） 8—马达速度控制



图 24-26 柑橘的近红外在线分析

在线近红外分析仪被用于分析食品的水分、蛋白质和脂肪含量，可用于快餐、巧克力、咖啡、茶、马铃薯粉、面粉、淀粉、坚果、大豆等食品加工过程的质量控制。

麦芽生产厂商和其他相关行业的生产厂商需要每年加工处理上千吨的谷物和麦芽。使用在线近红外分析对麦芽生产从原料的购入到生产最终产品的所有过程进行实时监控，可

以保证稳定的高标准质量的产品。对麦芽的成分监测有蛋白质、可溶氮、水分、榨出物、水分、黏性度等。

在乳制品加工行业，在线近红外分析仪被用于分析牛奶原料和成品奶的脂肪、蛋白质、乳糖和固形物，与 PLCs 等控制系统集成，实现对牛奶加工处理过程如混合过程的自动标准化控制。图 24-27 显示了乳酪加工过程中的近红外在线过程分析系统。

在肉类加工中，近红外在线分析仪可以在搅拌机或传送带上分析肉末或肉粉的脂肪、蛋白质、水分和胶原汁组含量。结合反馈控制系统调整传送给搅拌机的配料比例，实现肉类肥瘦比例标准化控制。

食用油碘值的测量非常重要，它决定了食用油的饱和度。近红外在线分析系统用于榨油厂对 27 种食用油的碘值进行在线分析，控制食用油的氢化过程。

在制糖业中，通过测定蔗汁蔗糖分、转光度、锤度和还原糖分等指标进行原料蔗压榨汁的品质监控，还用于在蔗糖煮炼过程中在线检测糖膏转光度和锤度等数据，提高糖品质和产量。

在制浆造纸业中用于检测纸页涂料中的水分含量和纸页的水分，测定纸浆中木素含量或卡伯值。

下面介绍两个具体的近红外在线过程分析应用系统开发实例。

实例 1：肉末的脂肪、蛋白质、水分含量的工业级在线检测

该项目的 NIR 分析目标是通过在线测定研磨后肉类的脂肪、蛋白质、水分含量，实现对肉类质量的实时监控。使用的 NIR 光谱仪是滤光片型的 MM55 NIR 仪，在滤光片轮上安装有 5 个干涉滤光片，转动频率为 20Hz。滤光片的中心波长分别为 1441nm，

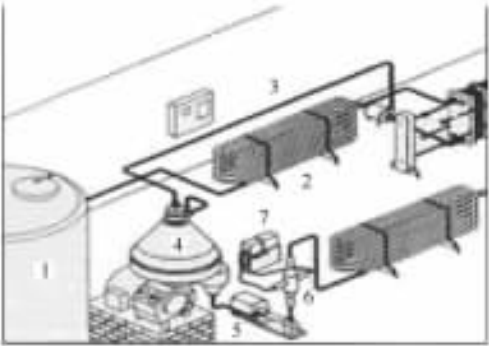


图 24-27 干酪生产过程中 NIR 在线分析
1—奶罐 2—加热器 3—输送管 4—分离器
5—泵 6—NIR 探头 7—NIR 光谱仪

1510nm，1655nm，1728nm 和 1810nm，带宽均为 25nm。滤光片波长是根据牛肉的波长特征选择的。图 24-28 是均匀牛肉的反射光谱，① 1441nm、② 1510nm、③ 1655nm、④ 1728nm、⑤ 1810nm 为选择的 5 个滤光片的波长。

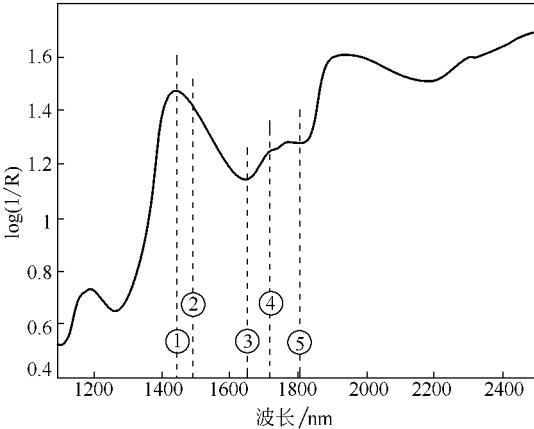


图 24-28 牛肉的反射光谱与特征波长

采用开放的漫反射采样方式。将 NIR 检测仪固定在研磨机出口的侧上方，在绞肉机的出口安装一个微微下斜的不锈钢喷管，使得出口处的肉末形成一个稳定的肉末流，肉末流与探头的距离大约 25cm，表面的光斑直径 40mm。图 24-14 显示了

NIR 探头的安装示意图。采用连续不间断方式采集光谱，每 5s 计算一次平均谱，并对采集的光谱做如下处理：1441nm，1510nm，1655nm，1728nm 波长点的 lg 值减去 1810nm 处 lg 值。采用多元回归方法建立定量分析模型。在建模过程中选用 112 个鲜肉（牛肉 61 个和猪肉 51 个）作校正集，每个样品重量在 200~800kg 之间。42 个（牛肉和猪肉各 21）用来作检验集。化学组分分布如下：脂肪 7%±26%，水分 58%±75%，蛋白质 15%±21%。所有的鲜肉用大型研磨机研磨，研磨盘孔径为 13mm。表 24-1、表 24-2 分别列出了校正集和检验集的预测标准差和相关系数。

表 24-1 校正集内部交叉证实结果（RMSECV-标准差，r-相关系数，I-样品数）

	校正集								
	全集(牛肉+猪肉)			牛肉子集			猪肉子集		
脂肪	1.43	0.91	106	1.40	0.87	57	1.48	0.93	49
水	1.25	0.88	106	1.23	0.84	57	1.09	0.93	49
蛋白质	0.71	0.72	106	0.56	0.62	57	0.56	0.38	49

表 24-2 检验集预测结果（SEP-预测标准差，r-相关系数，n-特异样品数）

	检 验 集								
	全集(牛肉+猪肉)			牛肉子集			猪肉子集		
奇异样品阈值:3x 标准差									
脂肪	1.49	0.95	0	1.30	0.96	0	1.35	0.94	0
水	1.33	0.92	0	1.21	0.95	0	1.18	0.92	0
蛋白质	0.30	0.85	0	0.45	0.39	0	0.57	0.68	0
奇异样品阈值:2x 标准差									
脂肪	1.16	0.97	4	1.16	0.96	1	0.82	0.98	3
水	1.11	0.94	3	0.94	0.96	2	1.03	0.93	1
蛋白质	0.57	0.90	2	0.45	0.39	0	0.35	0.87	1

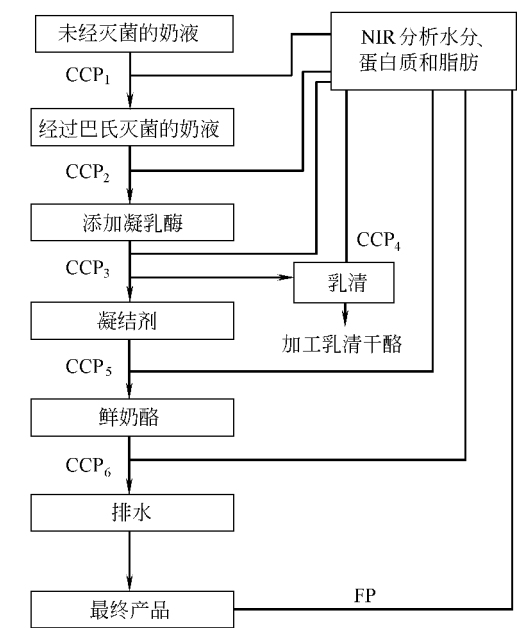


图 24-29 羊乳酪生产流程与关键控制点

实例 2：乳酪生产中的质量控制

该项目的 NIR 在线分析目标是通过乳酪生产过程的 6 个关键质量控制点的水分、蛋白质和脂肪成分的实时分析，实现乳酪生产过程的质量控制。图 24-29 显示了典型的乳酪生产流程以及关键控制点。CCP₁~CCP₆ 这 6 个点分别代表 6 个质量控制点，FP 为获得最终产品的终止点。

光谱仪选用了 Instalab 600 滤光片型近红外分析仪，并根据水分、蛋白质、脂肪的特征吸收波长选择了 1.94μm，2.18μm，2.31μm 和 2.3μm 为中心波长的滤光片。在建立定量分析模型时，每个控制点和最终产品点均分别采集 60 个样品作为校正集。每个控制点采集 15 个样品作为检验集，分别对最终产品 5 个种类的每个种类各取 12 个样品作为终止点的检验集。在建模过程中，为各个

控制点中的每个被测成分建立独立的多元回归分析模型。图 24-30 (a)、(b)、(c) 分别显示了检验集样品 NIR 测定值与化学值之间的相关图。

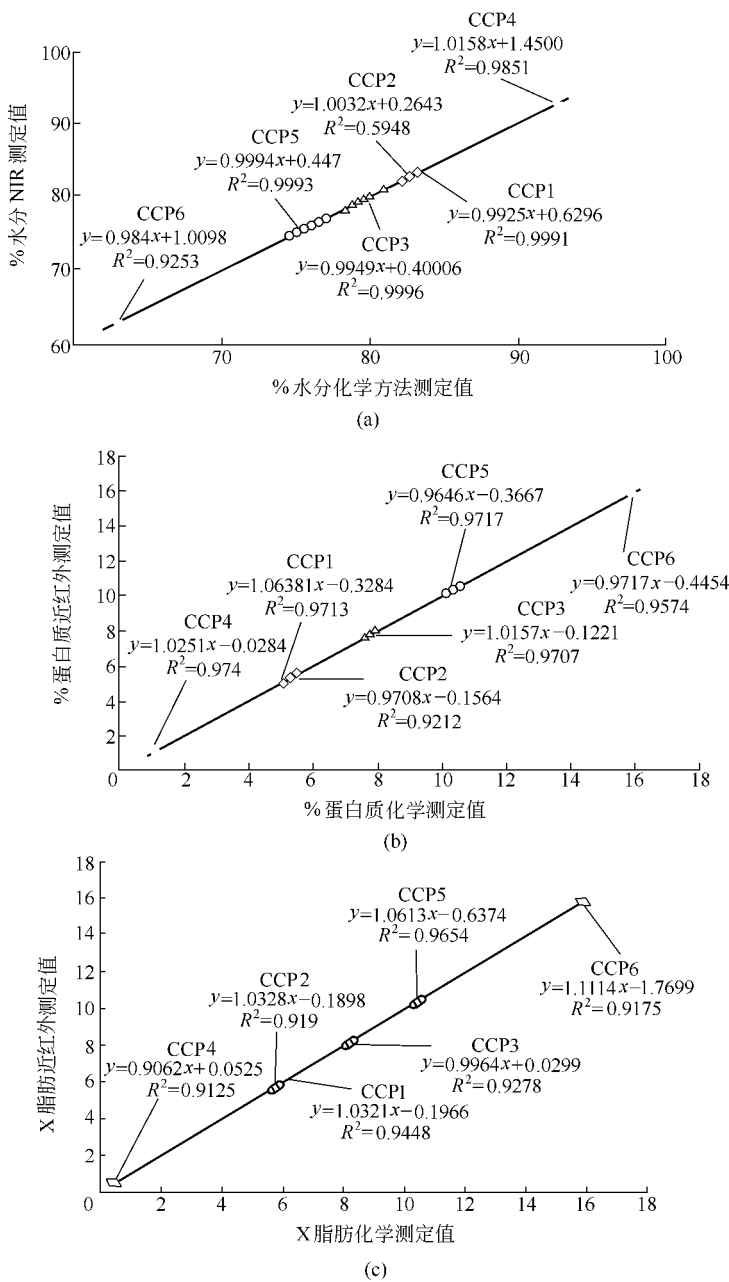


图 24-30 水分、蛋白质和脂肪模型的检验结果

(a) 各个控制点检验集的水分 NIR 测定值 (b) 各个控制点检验集的蛋白质 NIR 测定值 (c) 各个控制点检验集的脂肪 NIR 测定值

24.3.2 制药工业应用

随着药品工业现代化进程的加快及药物加工过程中生产效率的提高，药物的制备过程

要求引入越来越多的自动化分析技术。在药物混合、加工、成型及包装等一系列过程的每一个环节均需进行质量监控。这就要求在每一步加工过程中都必须对所有组分进行全面的检测，各环节中任何不慎的操作都可能会对后续环节产生严重的影响。利用近红外光谱分析的技术特点，在制药过程中可以实现药物制剂的无损分析、生产过程的实时分析和在线分析，从而达到控制药物制剂品质的目的。目前 NIR 已成为 AOAC (Association of Official Analytical Chemists) 一种标准分析方法应用于药品检测中。仪器生产商和药物分析专家的合作开发已使 FDA、欧洲和加拿大药物管理局正式研究用近红外光谱分析技术取代繁琐费时的常规分析方法的可行性，部分测试项目已被 FDA 批准为标准方法。USP (United States Pharmacopoeia 第 25 版) 最近已在附录中增补近红外分析方法。欧美的制药行业根据制药过程中每个生产环节控制质量的要求，设计了分别用于原料分析、混合物分析、临床用药处理分析、全程扫描分析、稳定性检验分析和药物包装、标签、贮存样品分析的 NIR 分析仪器，把 NIR 光谱分析仪器用作一种首选的过程控制仪器。

NV Organon 在 1955 年利用 NIRSystems Rapid Content Analyzer (RCA) 对测量生产过程中的药片和药粒进行非接触的测量，仅水分含量的过程分析每年可以节省 1 万美元的费用。

NIR 在药物生产过程控制中主要应用包括生产原料的质量评价、固体混合物的分析

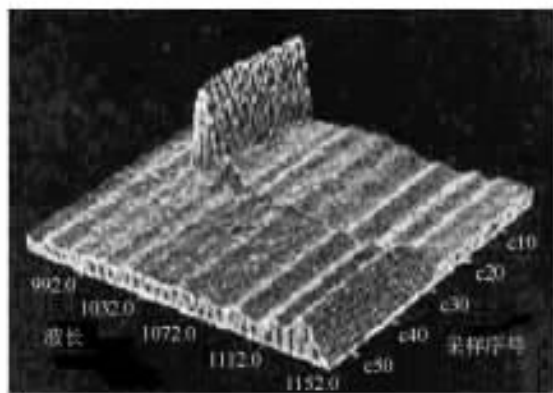


图 24-31 结晶反应过程中采集的光谱

测定、流化床干燥程度测定、发酵过程分析、片剂和胶囊的有效成分在线分析、包衣剂的质控分析、包衣厚度、释放速率、失活时间分析等多个方面。NIR 在生产线上的测量速度可以达到每秒 30 个药片，可以分析线上的每一个药片或胶囊，实现对产品真正的 100% 检测，而使用传统的检测方法只能达到 1% 的检测率。将光纤探头插入药物反应器中，通过采集反应物的光谱可以精确测定出最佳反应终止点，既缩短了生产周期又能获得最高产量。图 24-31 显示了药物结晶反应不同阶段中采集的光谱，大约在

第 28 次采样得到的光谱有明显的变化，表示结晶反应完成。在溶剂的回收纯化过程中，通过蒸馏器蒸馏，使溶剂纯度由 50% 纯化到 95%~98%。使用光纤多路转换器可以对多个蒸馏器监控，获得多组分信息，对回流率、进样速度、温度、压力等参数进行实时的优化和控制，既节约了能源又稳定了纯化质量。

下面介绍两个制药过程控制的应用开发实例。

实例 1：使用在线近红外光谱仪定量监测药物粉末的混合过程

该项目的目标是在线分析粉末混合物含量，监控粉末的混合过程。近红外光谱仪使用 Bomem 的 MB-160 光谱仪，仪器配置光纤接口和电子制冷的 InGaAs 检测器。使用不锈钢双向漫反射光纤探头采集样品光谱，探头外径为 13mm，光斑面积 0.12cm²。光纤探头固定在搅拌机侧壁，探头穿过外壁，向内腔突出内壁面 0.5~1mm，图 24-32 为实验室用搅

拌机光纤安装示意图，在搅拌机上下部各安装了一个探头，生产规模的搅拌机只在下部安装一个探头。光谱扫描的分辨率为 32cm^{-1} ，扫描区间 $1082\sim 2025\text{nm}$ ，每条光谱包含 280 个数据点，扫描频率为 4.78次/s 。用粗粒度粉末 ($d>300\mu\text{m}$) 和细粒度粉末 ($d<200\mu\text{m}$) 共 10g ，配置出含量 $0\text{m/m}\sim 100\%\text{m/m}$ 之间的样品 11 个，每个样品各分成 100 个子样用于构建校正集。对采集的光谱先做 MSC 校正，再求多次测量光谱的平均谱 (图 24-33)，采用非线性 PLS 方法建立分析模型。图 24-34 显示了校正集中样品的粗粒度粉末含量的 NIR 测定值与真实值的相关图。

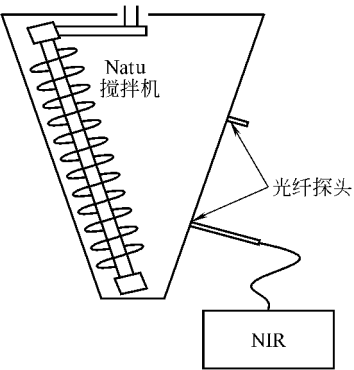


图 24-32 光纤探头插入搅拌机

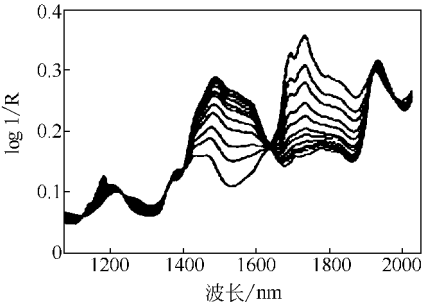


图 24-33 预处理后的光谱图

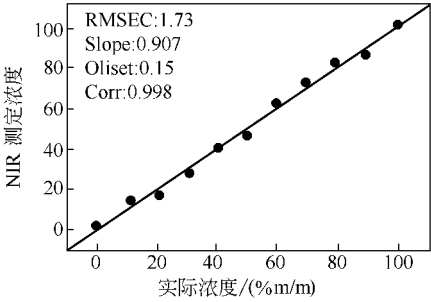


图 24-34 NIR 预测值与实际浓度的相关图

使用校正模型对实验室规模的搅拌过程采集的光谱进行分析，计算混合粉末中粗粒度粉末的含量。图 24-35 (a) 显示了粗粒度粉末含量随混合时间的变化曲线，整个混合过程分为 100 个单位处理时间。图 24-35 (b) 显示了每隔一定的时间间隔求得的粗粒度含量的

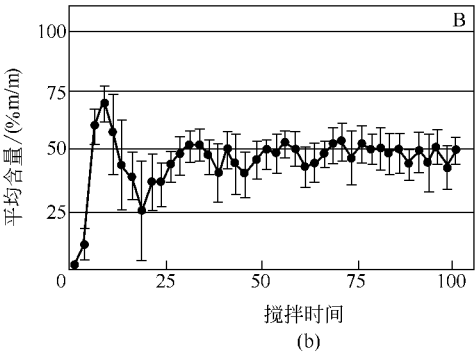
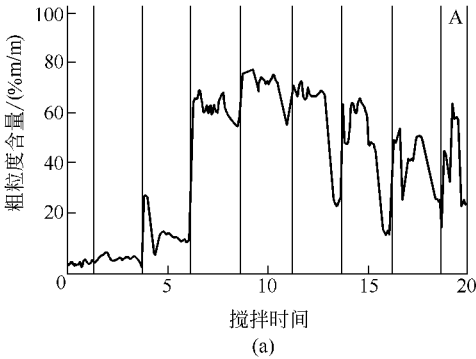


图 24-35

(a) 浓度随搅拌时间的变化 (b) 均值和标准差随搅拌时间的变化

平均值和标准差。将标准差作为衡量混合均匀度的指标，分别对实验室规模与生产规模的搅拌进行监测，如图 24-36 所示。实验室规模搅拌在 30 个单位时间后比较稳定，即粉末混合较均匀。生产规模的搅拌则需要在 50 个单位时间后均匀。

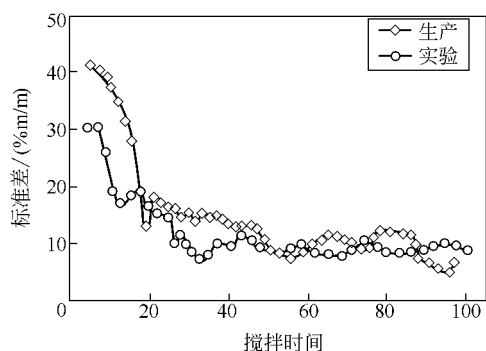


图 24-36 监测标准差随搅拌时间的变化

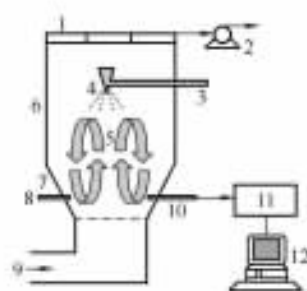


图 24-37 流化床 NIR 检测

1—过滤器 2—风扇 3—喷洒杆 4—喷嘴 5—转动方向
6—膨胀腔 7—装样池 8—采样器 9—加热气流入口 10—NIR 探头 11—光谱仪 12—电脑

实例 2：流化床颗粒化形成过程的监控

该项目的在线分析目标是通过监测喷洒和干燥过程中的湿度和颗粒度，精确地监控影响颗粒的湿/干平衡的过程参数。近红外光谱仪使用 NIR Systems 65000，配置光纤探头和 1m 长的光纤束包括 210 条入射光纤和 210 条出射光纤。采样采用光纤探头直接插入方式，如图 24-37 所示，直接将光纤探头插入流化床一侧内壁，与样品池内的样品直接接触。采集光谱的扫描波长范围为 1100~2500nm，间隔 2.5min 扫描一条光谱。对原始光谱做二阶求导预处理后，用 MLR 方法进行定量分析。图 24-38 显示了流化床内样品的原始光谱。图 24-39 显示了模型校正集样品含水量的 NIR 测定值与标准方法测定值之间的相关图。图 24-40 (a)，(b) 分别显示了造粒过程中的颗粒增长曲线和波长为 2282nm 的吸光度值的变化曲线。显然，这两条曲线的变化趋势基本一致，说明 2282nm 吸光度值可用于监测造粒过程重颗粒的增长量。

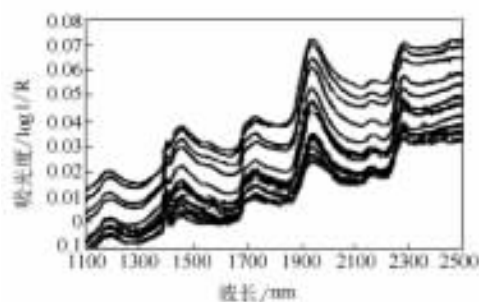


图 24-38 样品原始光谱

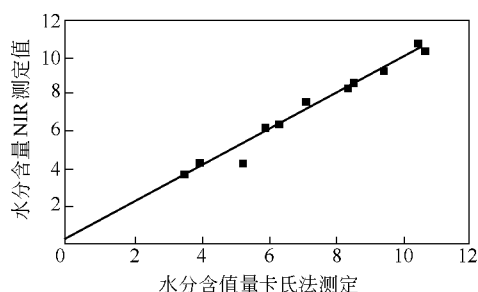


图 24-39 水含量的 NIR 测定

目前，中药产品质量的稳定性和一致性较低，极大地阻碍了中药现代化与国际化进程，其主要原因之一是缺乏中药制药过程分析方法，诸多过程参数尚无法在线检测，更谈不上实施有效控制。因此，如何有效地控制中药制药过程以保证产品质量、提高生产效率、节约生产成本是中药现代化进程中急需解决的重大课题。

浙江大学杨南林等研究了使用近红外光谱分析方法，在线测定大孔吸附树脂洗脱物中

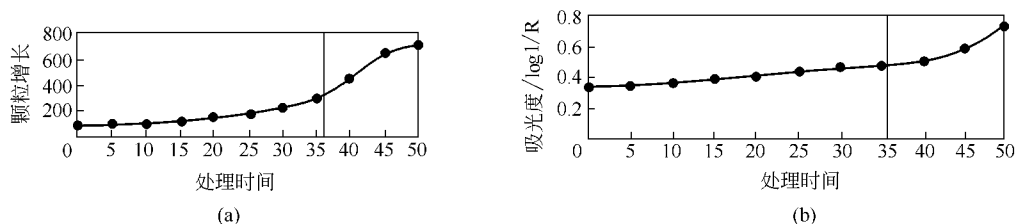


图 24-40 造粒过程中的颗粒增长和吸光度变化

(a) 造粒过程中的颗粒增长 (b) 造粒过程中 2282nm 处的吸光度变化

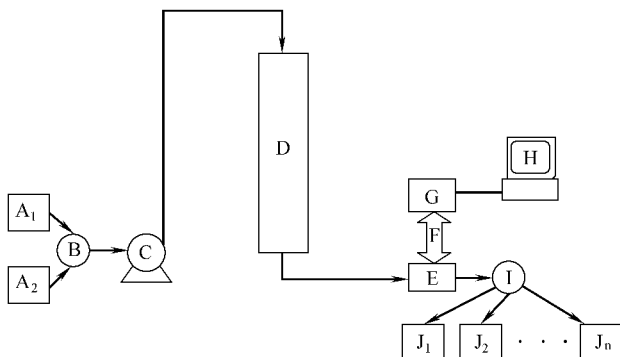


图 24-41 中药纯化过程的 NIR 在线检测实验装置示意图

A_1 , A_2 —溶剂池 B—混合器 C—蠕动泵 D—纯化柱

E—检测池 F—光纤 G—NIR 光谱仪 H—计

算机 I—切换阀 J_1 , J_2 ... J_n —组分收集器

盐酸小檗碱、盐酸巴马亭、盐酸药根碱和黄连总生物碱的浓度，以实时监控中药纯化过程。使用的光谱仪是德国布鲁克光学仪器公司的 IFS28/N 傅里叶变换近红外光谱仪，配置有光程 2mm、长度 2m 的石英光导纤维透射式探头，半导体冷却的 InGaAs 检测器，OPUS 光谱采集和处理软件。在线检测实验装置如图 24-41 所示，其采样方式是将光纤直接插入检测池采集光谱。光谱扫描范围为 $4000 \sim 11000 \text{ cm}^{-1}$ ，扫描次数 32 次，分辨率为 8 cm^{-1} 。对光谱做 13 点平滑处理后，用 PLS 方法建立校正模型。图 24-42 显示了黄连生物碱洗脱物的 NIR 原始光谱，表 24-3 显示了黄连洗脱液样品中生物碱组分含量的建模结果，图 24-43 显示了 NIR 在线分析得到的黄连生物碱洗脱曲线和乙醇洗脱剂梯度变化曲线。

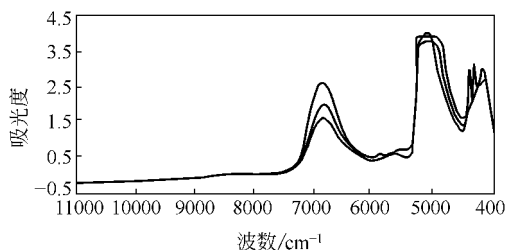


图 24-42 黄连生物碱洗脱物的 NIR 光谱图

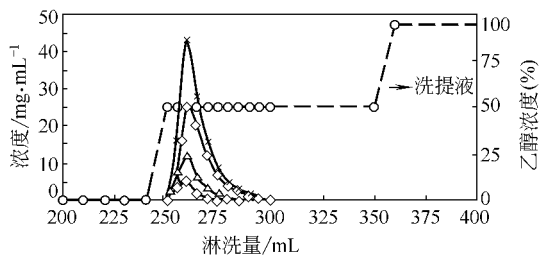


图 24-43 黄连生物碱洗脱曲线 (实线) 和乙醇洗脱剂梯度变化曲线 (虚线)

○—药根碱 △—巴马亭 ◇—小檗碱 ×—总生物碱

表 24-3

黄连洗脱液样品中生物碱组分含量的建模结果

项目	校正集			总生物碱	预测集			总生物碱
	药根碱	巴马亭	小檗碱		药根碱	巴马亭	小檗碱	
样品数	29				11			
平均值	1.052	2.732	7.452	11.236	1.056	2.595	7.403	11.054
标准差	1.383	3.536	8.409	13.016	1.280	3.060	7.312	11.501
最大值	4.988	12.658	28.651	45.647	3.739	9.102	22.637	35.478
最小值	0.003	0.031	0.162	0.206	0.085	0.130	0.729	0.966
分布范围	4.985	12.627	28.489	45.442	3.654	8.972	21.908	34.512

中国农业大学和世农绿方公司联合研究的近红外在线监测系统成功地用于葛根素生产工艺。通过在线测定葛根异黄酮、葛根素和酒精浓度等含量，实时监控大孔树脂吸附解吸过程和聚酰胺树脂层析分离纯化过程。采样系统采用原位在线采样方式，将液体流动池直接安装在生产线的测量部位。图 24-44 为吸附和解吸过程的近红外在线监控示意图。图 24-45 显示了其中 1 号近红外在线分析仪器黄酮分析模型的性能，NIR 测量值与标准方法测定值的相关系数为 0.966。图 24-46 显示了 1 号仪器在线分析得到的黄酮和葛根素浓度变化曲线和洗脱过程中乙醇浓度的变化曲线。

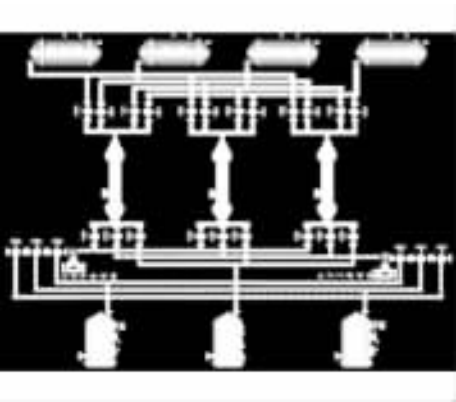


图 24-44 近红外在线监测吸附解吸过程

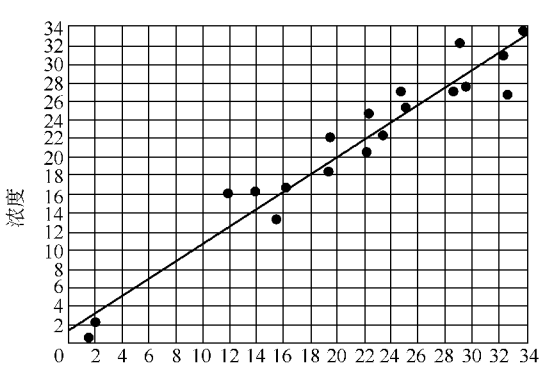


图 24-45 1 号仪器黄酮模型结果

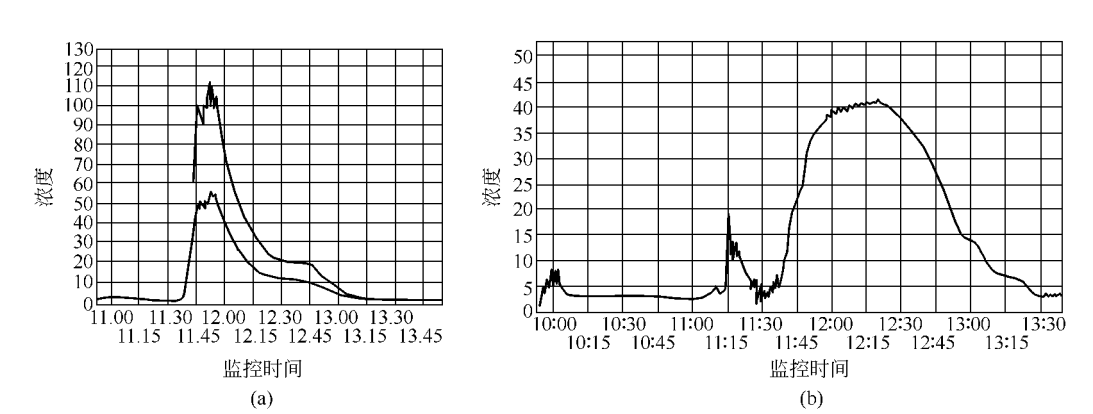


图 24-46 监控洗脱过程

(a) 洗脱过程中黄酮和葛根素浓度的变化 (b) 洗脱过程中乙醇浓度的变化

近红外在线分析技术在中药制药过程中的应用，将有效地改进传统的中药生产工艺，促进中药生产的现代化进程。

24.3.3 生物工程技术应用

20 世纪 90 年代以来，近红外分析在生物技术中的应用越来越受到重视，许多生物过程尝试使用近红外过程监控。尤其在发酵过程中，近红外过程分析技术被用于监控和优化整个发酵过程。近红外过程分析方法在生物催化反应过程中的应用尝试，为改进催化反应工业工艺过程开辟了新的途径。

24.3.3.1 微生物发酵过程的监控

在微生物发酵过程中，随着生物量的不断生长，营养水平将稳定下降，为优化发酵过程，需要监控发酵液的营养状态，使之维持在一个最佳的水平，同时生物量也需要被实时监控，以避免过度生长和抑制生长。图 24-47 显示了放线菌（*Streptomyces Fradiae*）发酵过程中发酵液所含的各种营养源组分和生物量的动态变化。

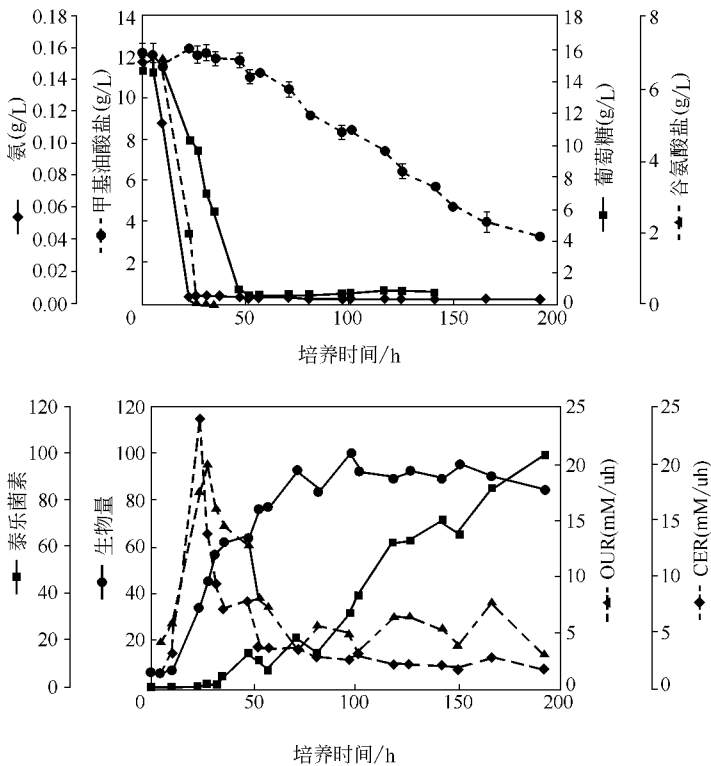


图 24-47 发酵液组分在发酵过程中的动态变化

近红外过程分析方法可以快速地测定发酵液的营养成分浓度和生物量累计程度，已被成功应用于各类发酵过程中。如酵母发酵过程中酒精和细胞密度的测定，乳酸发酵过程中生物量、葡萄糖和乳酸的过程分析，大肠埃希氏菌工业发酵过程中实时分析发酵液中的醋酸盐、铵、甘油的含量和生物量，在酶生产发酵过程中测定生产的碱性蛋白酶的活性，放线菌（*Streptomyces Fradiae*）发酵过程中实时分析营养物质糖、谷氨酸盐、甲基硫酸盐的

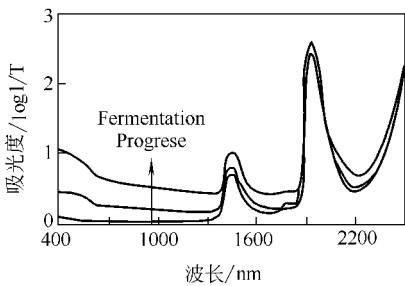


图 24-48 大肠埃希氏菌发酵液光谱

浓度等。

在发酵过程中，随着生物细胞的生长，发酵液将由透明状态逐渐变成高散射度的不透明状态。光谱采集方式可以采用透射（用于初始的透明状态）和反射方式（用于后期的不透明状态）。波段选择一般低于 2000nm，这是由于水在 1940nm 的特征吸收已经饱和。图 24-48 显示了工业级大肠埃希氏菌发酵过程中发酵液的漫反射吸光度光谱（400~2500nm）。从光谱图可以看到，水的 OH 键一级合频 1450nm 和组合频（1940nm）的特征吸收均已达到饱和。在总体趋势上，反射吸光度光谱的基线随发酵时间和生物量的增长而提高，但也受到发酵液其他组分的影响。对原始光谱做二阶导数处理，可以消除基线漂移。

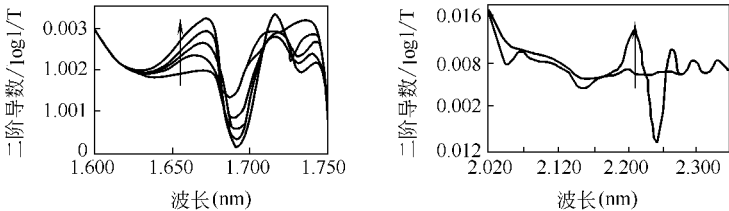


图 24-49 大肠埃希氏菌发酵液二阶导数光谱

图 24-49 显示了二阶导数光谱。定标方法一般采用 MLR 和 PLS。使用 MLR 法建立定标模型时，通常首先扫描各被测组分纯净样品的的光谱，找出它们的特征吸收波长，进行多元回归。使用 PLS 方法时，需要连续收集发酵过程的发酵液标样，在线采集光谱时要注意搅拌速度和气流的影响。表 24-4 显示为复杂的放线菌发酵过程分析，分别使用 MLR 和 PLS 方法建立定量分析模型的结果。

表 24-4 放线菌发酵液 NIR 定量分析模型

材 料	RE ^a (g/L)	校正集			检验集	
		回归模型	r^b	SEC ^c (g/L)	r^b	SEP ^d (g/L)
铵	0. 008	PLS ^e : 3 个主成分 2100-2300nm	0. 960	0. 018	0. 941	0. 021
葡萄糖	0. 37	PLS ^e : 6 个主成分 2080-2340nm	0. 977	0. 45	0. 994	1. 12
谷酰胺	0. 16	PLS ^e : 3 个主成分 2222-2342nm	0. 987	0. 53	0. 996	0. 58
油酸甲酯	0. 50	MLR ^f : 1718nm	0. 983	0. 52	0. 990	0. 56
油酸甲酯	0. 50	PLS ^e : 5 个主成分 2200-2400nm	0. 989	0. 44	0. 983	0. 93
油酸甲酯	0. 50	PLS ^e : 4 个主成分 1600-1800nm	0. 986	0. 50	0. 991	0. 76

RE^a 为参考方法的误差

r^b 为相关系数

SEC^c 为校正标准差

SEP^d 为预测标准差

PLS^e 为偏最小二乘分析

MLR^f 为多元线性最小二乘分析

24.3.3.2 生物催化过程监控

使用生物催化方法生产活性物质和化学中间物的主要优点是选择性高、副产品少。然

而在工业生产过程中，一般不使用 Baeyer-Villiger 单氧酶生物催化方法，其主要原因是当反应物和产物达到一定的浓度时，生物酶的活性将受到抑制，而不能实现完全反应，使产出率大大降低。图 24-50 显示了使用环己酮单氧酶（CHMO）作催化剂，将反应物酮转化为内酯的催化反应过程中，酮和内酯的浓度对酶活性的影响。随着反应产物内酯浓度的增长，酶的活性将迅速降低。为维持稳定的催化反应过程，当生成物达到一定浓度时，需要及时将反应产物从反应器中提取出来，同时也需要适量的补给反应物原料。

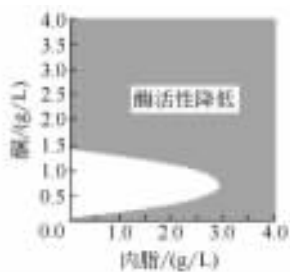


图 24-50 酮和内酯浓度与酶活性关系图

显然，在整个处理过程中，关键是需要实时测定反应物（酮）和产出物（内酯）浓度，才能控制好反应产物内酯提取和反应物酮的添加过程，将酮和内酯的浓度稳定控制在一个最佳浓度范围（酮为 0.2~0.4 g/L，内酯<4.5 g/L），使得催化酶的活性保持在较高的水平。

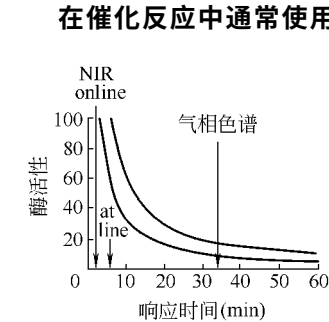


图 24-51 酶活性随响应时间的变化

在催化反应中通常使用传统的气相色谱方法分析反应物和产物的浓度，这种分析方法虽然精度较高，但分析速度不能满足催化反应过程控制的需求。从取样到气相色谱分析获得结果，整个过程一般需要持续 35 分钟。分析时间的滞后，使得对反应物和产出物的抑制作用不能得到有效的控制，从而限制了这种催化反应过程的工业化应用。近红外分析方法由于其快速和易于在线分析的特点，被尝试用于对反应过程中的反应物和生成产物浓度的实时分析。图 24-51 显示了酮的浓度为 0.5 g/L，内酯浓度为 0.75 g/L 内时酶的活性降低 1% 时，使用气相色谱分析方法和近红外分析方法进行行程测量时，其响应时间对催化反应过程的影响。酶的活性降低 1% 时，气相色谱在 35 分钟后才能获得分析结果，而到那时候酶的活性已降到了 10%，近红外在线分析方法则可给予几乎实时的响应。

图 24-52 显示了 Paul A. Bird 等使用在线近红外方法对反应器内的酮和内酯的浓度进行过程监控，实时调整反应物的添加速率和产出物的提取。实验结果证明，使用近红外过程分析在催化反应过程的实时监控，可以有效地控制生物的抑制作用，提高 30% 的生产率。图 24-53 显示了催化反应过程中反应物和产出物浓度的变化，图（a）代表进料的流速根据 NIR 方法测定的浓度变化进行控制的情况（虚线表示流速），图（b）代表未使用近红外分析方法，而按一个恒定的流速（1.5 g/L/h）连续供给原料的反应情况。显然采用近红外分析方法的反应过程持续时间长，产出物最终浓度高。近红外过程分析技术的应用，可能使 Baeyer-Villiger 单氧酶生物催化方法作为一种高效的反应过程用于工业生产。

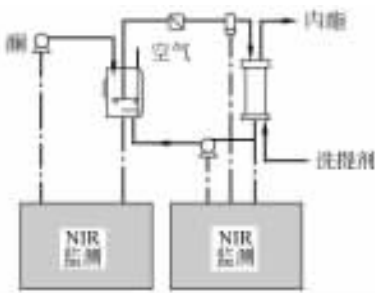


图 24-52 NIR 监控催化反应过程

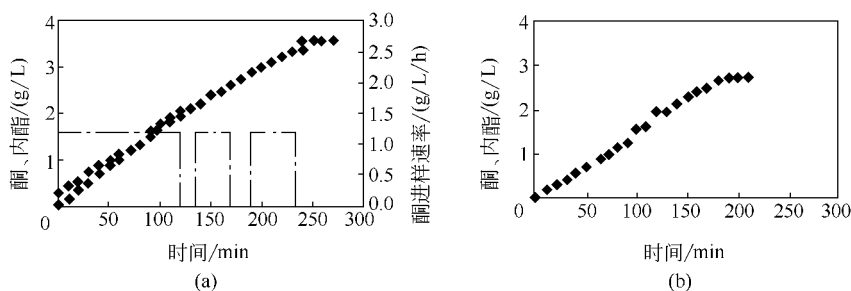


图 24-53 催化反应过程中酮和内酯浓度的变化

(a) 根据近红外在线分析结果控制进料速度 (b) 以恒定的速度进料

近红外分析方法除了在以上生物反应过程的应用以外，还被用于其他的生物过程分析。如有研究采用近红外分析方法分析替代动物生物学实验检验药物的生物活性。还有研究采用近红外分析方法对酵母菌株和细菌种类进行快速鉴别。有一些厂商也已致力于专用生物过程分析的近红外过程分析系统的开发，将加快近红外过程分析在生物工程中的应用步伐。

24.3.4 石油、化工、纺织工业的应用

近十几年来，NIR 在石油、化工、纺织方面的应用发展很快，尤其是在石油炼制和高分子合成过程中，NIR 在线过程分析技术在优化生产工艺、提高产品质量，降低能源消耗、减少环境污染方面发挥了极大的优势，得到了迅速的发展。NIR 分析技术被广泛应用于炼油厂燃油调和、原油蒸馏、催化裂化、催化重整、乙烯裂解、加氢裂化等工艺控制，为过程控制器 APC 提供原料、各个馏出口产品及中间产物性质等信息，便于实时调整各个控制单元的操作条件。NIR 在线分析也被成功地用于监测纺织生产系统中的许多关键的物理化学特性，如：含湿量、整理剂的施加量、热定型温度，涤/棉混合比、棉纤维的成熟度和聚合物的类型等。在分子合成和加工过程中，NIR 在线分析技术被成功地用于控制纤维的合成过程，测定聚合物的塑化度、结晶度、定向性、含水量，确定聚合反应最佳终止点，在挤塑过程中测定聚合物的熔融指数。袁洪福等在第 18 章已经对石油、化工业的 NIR 在线分析做了详细介绍，在此不再重复。

24.4 小结

近红外过程分析技术正在进一步走向专用化、小型化、模块化、智能化和网络化。许多仪器厂商与应用行业密切合作开发出一些专用的在线分析系统，如乳品、油品、食品在线分析系统等。仪器的小型化和尺寸标准化，更便于在生产线上的安装而不需要改造生产线。如一台 FT 型仪器现在可以做到只有 19 英寸，直接安装在生产线上。采样系统和主机系统的模块化分离结构，使得在线分析系统更便于安装和调试，也使得分析系统更加安全可靠。智能化近红外仪器不但提供分析模型，而且可以根据样品特征自动选择模型，仪器还能够自我诊断和校正，以避免过多的人工干预。仪器不但能提供与控制系统的标准接口，还能提供以太网连接，实现与其他信息系统和控制系统的集成，还能实现仪器的

远程诊断和监控。

任何技术都会有自己的弱点,近红外技术也不例外。近红外技术并不是万能的分析方法。在有些过程分析应用中,需要把近红外技术与其他分析技术结合起来使用。如在食品工业中,近红外光谱仪与折光仪、导向微波谱仪(GMS)结合起来使用,实现对食品生产线的连续实时控制。为实现对生产线上的药片进行100%的检验,高分辨率的摄像机与NIR仪结合使用,已达到对被测药片的精确定位。在化学工业中将近红外与拉曼光谱技术联合使用于化学反应过程控制。在汽油的在线分析中将NIR技术与激光诱导荧光光谱集成在一起,辛烷值的预测标准差比单独使用NIR分析方法有显著的提高。

近红外过程分析技术能得以迅速发展,除了由于近红外技术本身的特点,更是因为有一大批非常活跃的近红外研究和应用专家,为该技术的进一步发展和应用推广做了大量的工作。他们举办了专业的近红外杂志、国际会议和时事通讯,进行广泛的学术交流。东部分析专题会议和Pittsburgh会议每年都为在该领域做出特殊贡献的研究者颁发终生成就奖。

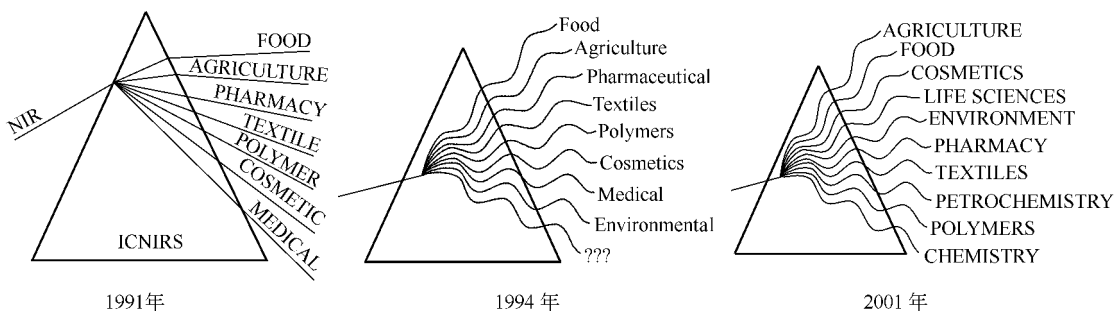
近红外过程分析技术在国际上获得迅速发展的同时,国内的研究专家和企业也非常重视NIR过程分析技术的应用与开发研究。在石油、化工、饲料、造酒、造纸、制药等生产过程中,已经逐步引入了近红外过程分析技术。中国石化研究院陆婉贞等研制了用于石化在线过程分析的在线分析系统,北京理工大学张化朋等研制了光纤传感的近红外漫反射测量系统用于遥测生产线上呈颗粒状或纤维状的样品中的水分和挥发分,中国农业大学严衍禄等研究了用于中药生产过程分析的在线近红外分析。国内大量的近红外分析技术还是在实验室分析水平,近红外过程分析技术的研究将进一步推动该技术在现实生产中的广泛应用。

参 考 文 献

- [1] Workman J, Mel Koch Jr, et al. Process Analytical Chemistry. Anal Chem, 2002, 75: 2859~2876
- [2] Workman J, Veltkamp D J, et al. Process Analytical Chemistry. Anal Chem 1999, 71: 21R~180R
- [3] Blaster W W, Bredeweg R A, et al. Process Analytical Chemistry. Anal Chem 1995, 67: 47R~70R
- [4] P A Hailey, P Doherty, et al. Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy. Part I. System development and control. Journal of Pharmaceutical and Biomedical Analysis, 1996 (14): 581~599
- [5] Smilde A K, Frans W J, van den Berg, et al. How to choose the right process analyzer. Analytical Chemistry, 2002. 368A~373A
- [6] Van den Berg, F W J, Hoefsloot, H C J, et al. Selection of optimal process analyzers for plant-wide monitoring. Anal Chem, 2002, 74: 3105~3111
- [7] Alexander Horch, Alf J Isaksson. Assessment of the sampling rate in control systems. Control Engineering Practice, 2001 (9): 533~544
- [8] JEROME J WORKMAN, JR. Review of Process and Non-invasive Near-Infrared and Infrared Spectroscopy; 1993-1999. APPLIED SPECTROSCOPY REVIEWS, 1999, 34 (1, 2): 1~89
- [9] Brimmer, P J, DeThomas F A, Hall J W. Method Development and Implementation of Near-Infrared Spectroscopy in Industrial Manufacturing Processes. Near-Infrared Technology in the Agricultural and Food Industries, P. Williams and K. Norris, Eds. (American Association of Cereal Chemists, St. Paul, MN, 2001). 199~214

“应用篇” 结束语

近红外光谱由于信息量丰富，近红外技术可以进行无损的多组分检测，因此使用范围极为广泛，而且还在不断的发展之中。从近红外光谱国际会议（International Conference on Near-Infrared Spectroscopy）会标的变化可以看到这种发展的趋势。下图是 1991 年、1994 年与 2001 年该会议的会标。



近红外光谱国际会议会标的变化

1991 年的会标是近红外技术“折射”出食品、农业、医学、纺织、高聚物、化妆品和制药等 7 个领域的应用；1994 年的会标中增加了近红外在环境科学中的应用，又加上了“???”，表示应用领域有待发展；2001 年的会标又增加了生命科学与石油化学等应用领域。会标的变化说明了该项技术是正在发展中的技术，应用领域在不断扩展。

本书的“应用篇”着重介绍了近红外光谱分析技术在农业、粮油加工业、饲料、食品、石化、制药、烟草与纺织等 8 个行业以及在生物医学、矿物学两个学科中的应用，还介绍了近红外技术在过程控制与农业遥感中的应用。有些近红外技术的应用领域，如环境科学、化妆品、化学、生命科学等本书未做介绍，有待再版时结合该技术的发展做进一步介绍。

附录一 近红外光谱分析常用名词术语解释

近红外光谱分析技术是一门涉及光谱分析、化学计量学（数理统计学）及所分析对象的专业等诸多领域知识的交叉应用科学，所涉及的知识面广，而且新技术特别是化学计量学部分的知识更新也比较快。因此，在各类近红外光谱分析仪的使用手册、说明书及科技论文报道中，常出现较多的名词术语，而且大都是英文。另一方面，近红外光谱仪器的实际使用者，往往熟悉所分析对象的专业知识，对其他知识比较陌生。作者根据自己的理解，将近红外光谱分析技术中常见的名词术语，做最简单的解释，希望对仪器的使用和操作者有所帮助。

1. 光学仪器部分

① 波长（Wavelength）

近红外光是电磁波的一部分，在光波的运动中，2个相邻波的波峰（或波谷）之间的直线距离称为波长，一般用 λ 表示，如图1所示。在近红外谱区波长的单位用纳米（nm）或微米（ μm ）表示。

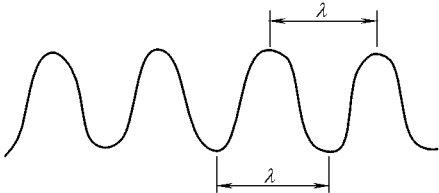


图1 波长示意图

② 波数（Wave number）

表示光的频率的单位，指1cm中所含光波的数目（ ν ），其单位是波数（ cm^{-1} ）。它与以微米（ μm ）为单位的波长具有下列关系：

$$\nu(\text{cm}^{-1}) = 10000 / \lambda(\mu\text{m})$$

③ 透光度（Transmittancy）

透光度，又称透过率。透射光强度与入射光强度之比，也称透光比、透光率。符号为T，通常以百分比表示： $T = (I/I_0) \times 100\%$

④ 反射率（比）（Reflectancy）

反射光强度与入射光强度之比，计算方法同透光度。

⑤ 吸光度（Absorbance）

吸光度又称光密度（OD）、消光度。在数值上等于入射光强度与反射光或透射光强度之比的对数，或透光度的负对数，用A表示： $A = \text{Log}(I_0/I) = \text{Log}(1/T)$

⑥ 透射（Transmission）

光穿透某一媒质而不改变其频率。可分为单向透射、漫透射和混合透射。近红外透射光谱属于漫透射。

⑦ 漫透射（Diffuse Transmission）

光在透过媒质时分散在各个方向，即不呈现折射的规律。如果在各方向的光亮度完全相等，则称为均匀漫反射。

⑧ 反射（Reflection）

光进行到两种媒质的分界面（可以是平面也可以是曲面）将有一部分返回原媒质。如

果面是光滑的，即其不平整度比波长小，则称为单向反射，否则称为漫反射。如果既有单向反射，又有漫反射，则称为混合反射。

⑨ 漫反射 (Diffuse Reflection)

与单向反射不同，漫反射是将以一定方向入射的光投到各个方向去。实际的物体是或多或少的漫射体，只有特别洁净的反射面，例如没有灰尘或杂质的液体表面，才是单向反射。近红外反射光谱分析的对象大多不是这样的液体，因此属于漫反射。

⑩ 光程 (Path Length)

样品槽内两个平行的面之间的垂直距离，即光线垂直射到样品槽面上时，光线通过样品层的厚度。

⑪ 伸缩振动 (Stretch Vibration)

原子沿连接它们的键做周期的伸长及缩小的振动叫伸缩振动，符号为 V 。伸缩振动的力常数比变角振动的力常数要大，因而同一基团的伸缩振动常在高频端出现吸收。周围环境的改变对频率的变化影响较小，由于振动耦合作用，原子数 $n>3$ 的基团还可分为对称伸缩振动和不对称伸缩振动。

⑫ 弯曲振动 (Bent Vibration)

也叫变角振动，是基团键角发生周期变化的振动，一般用 δ 表示。同一基团的变角振动都在共伸缩振动的低频端出现。它对环境变化较敏感，所以一般不把它作为基团频率处理。根据其振动特点，又可分为对称弯曲振动和不对称弯曲振动。

⑬ 差示光谱 (Difference Spectra)

它是补偿光谱的一种。待测的吸收谱带用参比光路中相近的吸收谱带做补偿。这种方法可改变单位浓度变化所引起的记录笔偏转，因而可提高定量测定的精确度。测定时须预先精确地配制出一系列浓度差别很小的标准溶液，用其中的一种溶液放置于参比光路中，其他的溶液与之做差示光谱，然后样品溶液与同一溶液做差示光谱，便得到样品光谱记录笔的偏转，在标准溶液的笔偏转之间进行内插，即可得出高精度的分析结果。

⑭ 导数光谱 (Derivative Spectra)

以吸光度随波长改变的速度即导数 dA/dv 为纵坐标，波数为横坐标所记录的光谱图称作导数光谱。纵坐标也可以是二次以上的高阶导数。吸收谱带的导数光谱如图 2 所示。

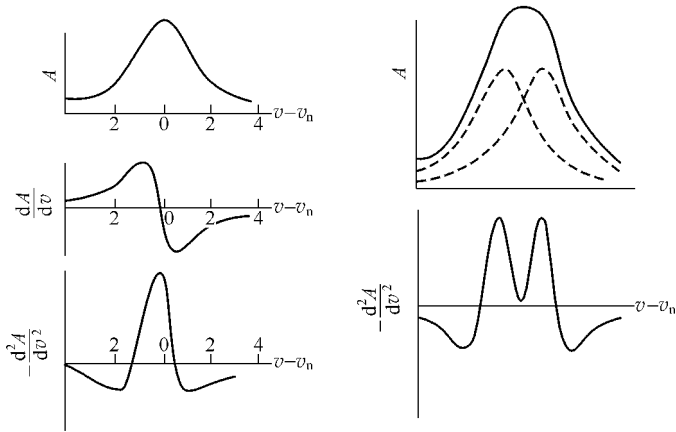


图 2○导数光谱示意图

一阶导数作图使原吸收曲线形状大为不同，原峰值相应导数为零。而二阶导数同原吸收曲线相似但谱带变窄。导数光谱不改变峰值位置，但减弱背景对吸收谱带强度的影响。对于经典的光谱计量学来讲，导数光谱主要用来分开叠加在主吸收带上的小吸收带，适于小组分的分析。另外，两个能级相近的振动产生的宽谱带，用导数光谱可以将它们分开。

⑮ 波长重复性 (Repeatability)

用一个或多个已知样品吸收峰作标准（对于近红外光谱该标准物质通常为聚苯乙烯），在相同条件下多次重复读取峰位，计算每次观察的波长对平均值的偏差，这些偏差的平均值就是此分光光度计的波长重复性。

⑮ 谱带半宽度 (Half Band Width)

谱带半宽度又称有效带宽。这里是指离开单色器出射狭缝的辐射光谱峰高的一半处的谱带宽度，如图 3 所示。检查谱带的半宽度必须要用锐线光源。

⑮ 杂散光 (Stray Radiation)

杂散光是指由检测器接收到的任何由仪器单色器分离的光谱范围以外的辐射。杂散光可以分成两种形式，一种杂散光波长与测定波长相同，它是由于测定波长因种种原因偏离正常光路，在不通过样品的情况下直接射到检测器上造成的。另一种杂散光波长与测定波长不相同。

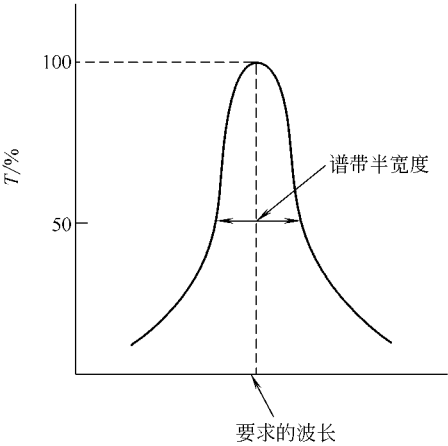


图 3○谱带半宽度示意图

⑮ 分辨率 (Resolution)

分辨率是指仪器对于紧密相邻的峰可分辨的最小波长间隔。对于经典光谱分析，分辨率是衡量仪器的重要指标之一。但对于近红外光谱分析则不然，因为大多数近红外光谱图都进行了多点平滑，不希望将紧密相邻的峰分开。单色器输出的单色光的光谱纯度、强度以及检测器的光谱灵敏度等是影响仪器分辨率的主要因素。

⑮ 波长范围 (Wavelength Range)

这是指仪器上、下限波长之间的工作范围，它是由光源、单色器及检测器的光谱响应特性决定的。有的仪器在波长范围两端缺乏足够的能量，不能正常工作，表现为 100% T 或 OA 设定困难，基线两端不平直等。

⑮ 波长准确度 (Wavelength Accuracy)

波长准确度是指仪器波长指示器上所指示的波长值与仪器出射的实际波长值之间的符合程度，可用两者之差（即波长误差）来衡量其准确性。如某分光光度计的波长准确度为 1.0nm，那么测定某汞灯辉线时的最大能量应落在此辉线的 ±1.0nm 之间。

⑮ 噪声 (Noise)

此词来源于声学，指一种不希望接收的能量对于探测器的干扰。在近红外光谱仪器中，它是指检测器上接收的非所需的样品信号。噪声根据来源可分为仪器噪声和样品或操作噪声，仪器噪声又可分为热噪声（噪声）、散射效应噪声、环境噪声和闪变效应噪声。

仪器的随机噪声常可用 20 到 30 次空白测量的方法求得，这一组所得数值的标准差就

是噪声，即： $\text{Noise} = \sqrt{\frac{1}{n} \sum_1^n (x_i - \bar{x})^2}$ ， $\bar{x}$ 为 n 次测量值的平均。

⑫ 信噪比 (Signal to Noise Ratio (SNR 或 S/N)

测量的信号和噪声有关的电流、电压或功率的比率，也可以用分贝表示：

$\text{SNR} = 20 \text{Log} (\text{信号电流、电压} / \text{噪声电流、电压})$

⑫ 基线 (Baseline)

光谱仪的基线是指没有置入样品时，仪器所记录的谱线，基线主要决定于检测器的暗电流与噪声。

⑫ 近红外漫反射光谱 (Near Infrared Diffuse Reflectance Spectrum)

在近红外区域，以波长或波数为横坐标，以反射光密度为纵坐标作图，即获得反射光谱。由于所研究对象通常没有平滑的光学面，所产生的反射为漫反射，因此，其光谱就是漫反射光谱。

⑫ 近红外透射光谱 (Near Infrared Diffuse Transmission Spectrum)

在近红外区域，以波长或波数为横坐标，以吸光度为纵坐标作图，即获得透射光谱。由于所研究对象通常没有平滑的光学面，所产生的透射为漫透射，因此，其光谱就是漫透射光谱。

⑫ 滤光片型近红外光谱分析仪 (Filter Near Infrared Spectrometer)

根据产生单色光部件的不同，通常将近红外光谱仪分为滤光片型、光栅型和傅里叶型三种。一个滤光片产生一个波长的单色光，由于滤光片数量的限制，滤光片型仪器产生的光谱通常是波长不连续的谱图。

⑫ 光栅型近红外光谱分析仪 (Grating Near Infrared Spectrometer)

它是以衍射光栅作为色散元件的近红外光谱分析仪，它可获得波长连续的谱图。由于激光制作光栅、旋转光栅等技术的应用，提高了该类仪器的扫描速度和信噪比。

⑫ 傅里叶型近红外光谱分析仪 (Fourier Transform Near Infrared Spectrometer)

用干涉仪将近红外光分为两束后，再以不同的光程差重新组合，而形成光信号的干涉图。此图中光信号的强度是光程差，即时间的函数，将此函数做傅里叶变换，可得光强为波长或频率的函数曲线，即普通光谱图。因此，利用干涉方法，并经过傅里叶变换而获得近红外光谱的仪器称为傅里叶型近红外光谱分析仪。

2. 光谱分析中的化学计量学部分

① 因子 (Factor)

这是一个有多种含义的数学概念，在多元统计分析中，是指对可观察变量有影响的不可控或不可观测的随机变量。

② 得分 (Score)

多元样本观察值对均值离差的主成分。

③ 因子载荷 (Loading)

也称因子权值。因子模型中公共因子的系数。

④ 标准正态化处理 (Standard Normal Variate)

这是对近红外光谱散射校正的技术之一，以 SNV 表示，即每个波长点处的光密度减去整个波长区光密度的平均值，再除以整个波长区光密度的标准差。光谱进行 SNV 处理

后，减少了非线性的弯曲作用（Bend effect）产生的偏差。

$$SNV_{(1\sim 100)} = [Y_{(1\sim 100)} - Y_1] / \{[(\sum Y_{(1\sim 100)} - Y)^2 / (n-1)]^{1/2}\}$$

式中 Y —— 某样品某个波长点吸收值的负对数（ $\lg 1/R$ ）；

Y_1 —— 该样品在所有 100 个波长点上吸收值的平均值；

$SNV_{(1\sim 100)}$ —— 某样品某个波长点经标准化后的吸收值。

⑤ 附加散射校正 [Multiplicative Scatter Correction (MSC)]

这是对近红外光谱散射矫正的另一个技术，以 MSC 表示。每个光谱回归到这一批样品的平均光谱，其中的截距和斜率被用于矫正，即每个光谱的每个波长点处的光密度减去截距，再除以斜率。

⑥ 去偏差 (De-Trend)

去偏差是对近红外光谱基线矫正的技术之一，它是用一个人工模拟的模型（一般为多项式）来描述基线，然后将整个波长区内每一点光密度减去该点的基线。

⑦ 奇异点 (Outlier)

那些与全群平均值差异很大的个体称为奇样品，也称奇异点，它们处于全群分布的边缘或之外。分布在 NIT 或 NIR 分析中，有 4 种类型的奇异点：

X 奇异点：一个样品系列中某样品的总体 H 值（Globe H statistics GH）大于 5，则为该样品系列的 X 奇异点。

Y 奇异点：一个样品系列中某样品的某营养指标化学分析值超过 $3 \cdot SD \pm X$ 的范围时，则为该样品系列该营养指标的 Y 奇异点。

T 奇异点：在建立光谱与样品中某成分含量回归关系的校正过程中，如果标准预测误差大于 2.5，则为该样品系列的 T 奇异点。

H 奇异点：在建立光谱与样品中某成分含量回归关系的校正过程中，如果某样品的“邻居” H 值（Neighbourhood H statistics NH）大于 5.0，则为该样品系列的 H 奇异点。

⑧ 广义距离 (Generalized Distance)

也称马氏距离。两个具有相同方差矩阵和不同均值向量的多元总体之差异的一种度量。设两个多元总体（多元随机变量） X 和 Y 有相同的方差矩阵 Σ ，其均值向量应为 μ 和 ν ，则广义距离定义为

$$D^2 = (\mu - \nu)' \Sigma^{-1} (\mu - \nu)$$

⑨ H 统计值

H 统计值（简称 H 值）是多元回归统计中观察值（变量）的力臂（Leverage）的概念（或称 Mahalanobis 距离）（Barnes 1982, Shenk 1990），也就是某变量的独立成分相对与其他变量的位置。其数学定义是： $H = 1/I + 1/n \sum H_i$, $H_i = X(X^T X)$ ，

式中 H_i —— 帽子矩阵；

X —— n 个样品 100 个波长点的光谱吸收值矩阵；

X^T —— X 矩阵的逆矩阵。

⑩ 最小二乘法 (Least Squares Method)

也称“最小平方法”。一种统计估计方法，主要用于数据的分析修匀和估计回归函数中的未知参数，其基本原则是使观测值与修匀值（或回归值）的偏差平方和最小。设随机

变量 Y 对 (随机或非随机) 变量 $X_1, \dots, X_m (m \geq 1)$ 有随机相依关系; $Y_j, X_{1j}, \dots, X_{mj} (j=1, \dots, n)$ 是对 $Y, X, \dots, X_m$ 的 n 次独立重复联合观测结果; $f(X_1, \dots, X_m)$ 是可供选择的修匀函数, $Y_j = f(X_{1j}, \dots, X_{mj})$ 是 $Y_j (j=1, \dots, n)$ 的修匀值. 最小二乘法在一切修匀函数中选择, 使残差平方和最小者. 实际应用中, 假设可供选择的修匀函数具有已知数学形式, 如 $f^*(X_1, \dots, X_m) = b_0 + b_{1x_1} + \dots + b_{mx_m}$. 这时选择最优修匀函数的问题, 归结为求 $b_i (i = 0, 1, \dots, m)$ 的估计值, 使残差平方和 $\sum_{j=1}^n (Y_j - \hat{Y}_j)^2$ 达到最小值的问题.

称 $f^*(X_1, \dots, X_m)$ 为最小二乘修匀函数, 称 $\hat{b}_i$ 为 b_i 的最小二乘估计.

④ 回归模型 (Regression Model)

回归分析的数学模型. 一个随机变量 Y 对另一个或一组 (随机或非随机) 变量 $X_1, \dots, X_m$ 的随机相依关系的数学表达式及有关假设条件的总称. 常用回归模型有如下一般形式:

$$Y = \mu(X_1, \dots, X_m) + e$$

其中 $\mu(X_1, \dots, X_m)$ 是 Y 对 $X_1, \dots, X_m$ 的回归函数, e 是数学期望为零的随机变量, 称作随机误差, 也称模型误差. 方程 $Y = \mu(X_1, \dots, X_m)$ 称作 (理论) 回归方程. 自变量 $X_1, \dots, X_m$ 称作 “回归变量 (或因子)”、“说明变量”、“解释变量” 或 “预报因子”; 因变量 Y 称作 “响应变量”、“被说明变量”、“被解释变量” 或 “预测量”. 只有一个回归变量的回归模型称作一元的, 否则称做多元的. 应用回归分析中, 假定回归函数的数学形式已知, 但含未知参数, 最简单、最重要且最便于处理的情形是 $\mu(X_1, \dots, X_m)$ 关于未知参数 $b_0, b_1, \dots, b_m$ 线性情形:

$$\mu(X_1, \dots, X_m) = \sum_{i=0}^k b_i g_i(x_1, \dots, x_m), \text{ 其中 } g_i(X_1, \dots, X_m) (i = 0, 1, \dots, k) \text{ 是不含未知参}$$

数的已知函数, $b_0, b_1, \dots, b_m$ 称作 “回归系数”. 这样的回归模型称作线性的, 否则称作非线性的, 许多非线性回归模型, 经变换可以化为线性的, 直接确定回归函数 $\mu(X_1, \dots, X_m)$ 的形式往往是困难的, 常根据对 $Y, X_1, \dots, X_m$ 的观测结果 $Y_j, X_{1j}, \dots, X_{mj} (j = 1, \dots, n)$ 选择适当函数 (修匀函数) $f(X_1, \dots, X_m)$, 使 Y_j 对 $\hat{Y}_j = f(X_1, \dots, X_m) (j = 1, \dots, n)$ 偏差在一定意义下最小. 这种对观测值的拟合, 是回归模型的直观背景, 在应用回归分析中被广泛采用.

④ 回归系数 (Regression Coefficient)

线性回归函数所含未知参数 (见 “回归模型”).

④ 回归方程 (Regression Equation)

回归方程是理论回归方程和经验回归方程的总称. 设 $\mu(x) = E(Y | X=x)$ 是随机变量 Y 对 X 的一元回归函数, $\mu(X_1, \dots, X_m) = E(Y | X_1=x_1, \dots, X_m=x_m)$ 是随机变量 Y 对 $X_1, \dots, X_m$ 的 m 元回归函数, 则曲线方程 $y = \mu(x)$ 及曲面方程 $y = \mu(X_1, \dots, X_m)$ 都称作理论回归方程. 根据观测值 (X_j, Y_j) 或 $(Y_j, X_{1j}, \dots, X_{mj})$ 建立的回归方程, 称作经验回归方程, 它是理论回归方程的估计.

④ 一元回归模型 (Univariate Regression Model)

仅含有一个回归变量的回归模型 (见 “回归模型”).

多元回归模型 (Multivariate Regression Model)

含有两个或两个以上回归变量的回归模型 (见“回归模型”)。

⑮ **线性回归模型** (Linear Regression Model)

回归函数关于未知参数是线性的回归模型。线性回归模型的一般形式为：

$$Y = \sum_{i=0}^k b_i g_i(x_1, \dots, x_m) + e, \text{ 其中 } k \geq 1, b_i (i = 0, 1, \dots, k) \text{ 为回归系数 } g_i(X_1, \dots,$$

$X_m) (i = 0, 1, \dots, k) \text{ 是不含未知参数的已知函数, } e \text{ 是数学期望为零的随机变量。}$

⑮ **数据修匀** (Data Smoothing)

也称“数据平滑”。旨在消除随机波动以揭示变量间随机相依关系的数据处理方法。常用的数据修匀方法有：图形修匀法、机械修匀法和分析修匀法。

⑮ **线性逐步回归** (Stepwise Multiple Linear Regression, SMLR)

选择回归变量，产生最优回归方程的一种常用数学方法。其基本思想是，将回归变量一个一个地选入，选入的条件是其偏回归平方和显著；每选入一个新变量后，对已入选的各变量逐个进行显著性检验，并剔除不显著变量。如此往复选入、检验、剔除，直至无法剔除且其余变量也无法选入为止。

⑮ **主成分分析** (Principal Components Analysis, PCA)

也称“主分量分析”、“分量分析”，是将多个指标的问题简化为少数指标问题的一种多元统计分析方法。其目的在于：a) 简化统计数据 (降维)；b) 揭示变量间的关系。许多情形下，由于 m 个变量的大部分变异性，实际上可以由它们的 $k (k \leq m)$ 个主成分—— m 个变量的线性组合的变异性来概括，从而使数据简化。主成分分析还用于揭示用其他方法难以揭示的变量间的关系。主成分分析法，以所研究变量的方差矩阵为基础，一般不要求满足更多的条件，一般也不假定变量有多元正态分布。

⑮ **聚类分析** (Cluster Analysis)

研究事物或变量分类的数学方法，多元统计分析的重要分支，旨在由统计数据出发找出事物或变量的自然类别。聚类包括对个体 (事物、项目、样品等) 的聚类和对变量 (数量标志及可数量化的属性标志) 的聚类。聚类分析，在事先不知类型的个数和结构的情形下，要根据 n 个数据点，a) 将 n 个个体分类-对个体的聚类；b) 将 m 个统计标志分类-对变量的聚类。聚类的基本原则，是将在一定意义下“相似”或“接近”的对象 (个体或变量) 聚为一类，作为“接近”或“相似”程度的度量，对于个体的聚类，使用各种距离型度量 (如欧几里得距离、马哈拉诺比斯距离等)；对于变量的聚类，多使用各种相似系数 (如两数据点的夹角余弦、样本相关系数等)。聚类分析，尽管其理论尚不完善，其方法还显得相当粗糙，但已被广泛用于解决许多实际问题，与回归分析和判别分析齐名，被誉为多元统计的三种重要方法。

⑮ **聚类统计** (Cluster Statistic)

聚类分析使用的统计量，反映个体或变量“接近”程度 (相似性或相异性) 的数值度量，常称作“相似性度量”。对于个体 (事物、项目或样品) 的聚义，多使用数据点间的“距离”做聚类统计量；对于变量的聚类，多采用表征不同变量之间相依性或关联性的样本特征做聚类统计量。聚类统计量的选择，要根据研究对象、有关专业知识以及具体问题的要求而定，还要考虑所考察变量的类型。

㉑ 主成分回归法 (Principal Components Regression, PCR)

如果在回归中应用所有的 100 个 (NIT) 或 700 个 (NIR) 波长点光谱的信息, 这样在建立回归模型时, 至少需要 101 个或 701 个样品建立 101 个或 701 个线性方程组。主成分分析可用于压缩所需要的样品数量, 同时又应用光谱所有的信息。它将高度相关的波长点归于一个独立变量 (Factor) 中, 进而就以为数不多的独立变量建立回归方程, 独立变量内高度相关的波长点可用主成分得分 (Score) 将其联系起来, 并用内部检验 (Cross Validation) 防止过模型现象。

㉑ 部分最小偏差回归法 (Partial Least square Regression, PLSR)

部分最小偏差回归法是 20 世纪 80 年代末应用到近红外光谱分析上的。该法与 PCR 很相似, 仅是在确定独立变量时, 不仅考虑光谱的信息 (X 变量) 还考虑化学分析值 (Y 变量)。该法是目前近红外光谱分析上应用最多的回归方法。

㉑ 多元线性回归 (Multivariate Linear Regression, MLR)

也称多重线性回归, 回归函数关于未知参数和回归变量都是线性的多元回归, 其变量关系为: $Y = b_0 + b_1x_1 + \cdots + b_nx_n + e$, 其中 b_i ($i=0, 1, \cdots, n$) 是回归系数, e 为随机误差。

㉑ 残差 (Residue)

近红外方法测定值与真实值 (经典方法测定值) 的差值, 以 d 表示: $d = \hat{y} - \bar{y}$

㉑ 偏差 (Bias)

残差的平均值, 以 Bias 表示, $\text{Bias} = \frac{1}{n} \sum_{i=1}^n d_i$

㉑ 残差的平方 (Mean Square Error)

残差的平方是反应定标模型预测误差的重要指标之一, 以 MSE 表示,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n d_i^2$$

㉑ 残差平方根 (Root Mean Square Error)

以 RMSE 表示, $\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n d_i^2}$, 要注意的是, 在不同的文献、仪器使用手册中, RMSE 有多种表达: 如 RMSEP, 即为预测样品的残差平方根; RMSECV, 即为交叉检验中的残差平方根; RMSELC, 即为杠杆矫正中的残差平方根; RMSEE, 含义与 RMSE 相同。

㉑ 标准分析误差 (Standard Error of Estimation)

残差平方根是测定中的总体误差, 包含了偏差。如果将其中的偏差去除, 即为标准分

析误差, 以 SEE 表示: $\text{SEE} = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$, 同样 SEE 也有多种表达形式, SEC 为定标样品中的标准分析误差; SECV 为交叉检验样品中的标准分析误差; SEP 为预测样品中的标准分析误差; SEP, 有文献将其定义为预测样品中的标准分析误差, 也有文献将其定义为分析中的标准分析误差。

㉑ 相对分析误差 (RPD)

SEC 或 SEP 可用于比较不同回归模型对同一样品系列某一成分预测的优劣，以便于寻找最佳的模型。但是它们还受到所分析样品系列的真值的分布和变异的影响。不能用于不同成分间，或不同样品系列间的比较。而相对分析误差（Ratio of Performance to Standard Deviate）则可克服这个缺点。 $RPD=SD/SEC$ 或 SEP 。

③⑩ 相关系数（Correlation）

用 $Y=a+b(X)$ 的简单直线模型描述 NIR（NIT）预测值与真值的相关性。其两者的相关系数以 r 表示。有时为区别起见，校正系的相关系数以 R 表示，检验系的相关系数以 r 表示。 a 为截距， b 为斜率，理想的情况是 $a=0$ ， $b=1$ 。

③⑩ 斜率（Slope）

参见相关系数条。

③⑩ 截距（Offset，Intercept）

参见相关系数条。

附录二 近红外光谱仪器的日常维护

近红外光谱分析仪是综合光学、精密机械和电子学的精密仪器。为了使测定准确、稳定、可靠以及延长仪器的寿命，应注意正确的安装、使用，定期检查各项性能指标，并进行仔细的维护和保养。

1. 开箱和清点

开箱时，应注意避免强烈振动。保留包装材料，以备用于仪器的贮存或再运输。检查各包装件及机壳外表有无损伤，如有损坏，随即与厂商联系。

先按合同单和装箱单清点各部件，然后按说明书清点各细小部件，如有漏缺，及时与厂商联系。开箱清点后，仔细阅读说明书，按照说明书的要求安放仪器，连接电缆，检查电源额定电压、频率。并做好一切清点、安装的工作记录。

2. 选择合适的安放地点

① 温度要求

由于环境温度变化对仪器，特别是单色器精度有明显的影响，所以安装大型精密分光光度计的实验室最好是冬夏室温变化不大、不受阳光直接曝晒的房间。

近红外光谱分析仪中另一个易受环境温度变化影响的部件是检测器。目前，许多近红外光谱分析仪使用半导体的硫化铅（1100~2500nm）、氧化硅（400~1100nm）或铟镓砷作检测器。温度的变化也会影响电子的激发，检测器电流的变化给仪器带来噪声。尽管这类检测器都配有温控装置（一般为3级温控），但要求工作室环境温度变化不能过大，一般是控制在最适温度（以仪器说明书为准） $\pm(3\sim5)^{\circ}\text{C}$ 。当然以有空调房间最为理想。

此外，仪器附近不应有高温的暖气管道。仪器后面留有一定空间，以便于散热和维修。

② 湿度要求

在潮湿的环境中，除了仪器会生锈外，许多精密光学镜面反射率会降低，甚至引起反射膜层斑驳、脱落。所以仪器室的相对湿度最好保持在40%~75%。

③ 防振、防尘、防腐和防电磁干扰

振动会造成单色器的波长误差，所以仪器应放在稳固的实验桌上，并在底座下垫适当的软垫。

尘土会使光学元件的透光度或反射率降低，也是杂散光的主要来源之一。对精密机械部分来说，灰尘也是零件磨损、生锈、腐蚀的原因，因此仪器室应有一定的防尘条件。此外，仪器室不应有腐蚀性气体或溶剂蒸气。

安装近红外仪器时最好用专用线连接电源，并通过专用的电子稳压器供电。电源电压波动不能过大，附近不要有产生强烈电磁辐射的电子系统。

④ 其他措施

(a) 配线时，要把输出强电部分的引线 with 装置内部的信号线分开，不能扎在一起。交流线和直流线最好分开走，否则强电引线和信号引线之间就像接了一个小电容。当强电部分的电流通断时，或电网中有强干扰时，就会在信号线上感应出干扰信号，影响仪器的正

常工作。

(b) 插接件的质量对仪器的可靠性影响很大，为了克服插接件的接触不良现象，接线板插头座全部采用双接点并联传输信号。另外，接线板之间信号的传输采用低电平传输，并且在接受信号的输入端，通过电阻接高电位（电源接正）。这样可以避免由于接触不良而造成的悬空电压。

(c) 加强仪器的抗干扰能力

I. 对于地线上的干扰，必须保持良好的接地措施才能克服。地线干扰是由于接地不良而造成地线之间存在有电阻，接地电流流过这个电阻时产生了尖峰电压造成的。

II. 仪器的接线板上铜箔导线应尽可能宽些，围绕着接线板形成一个完整的回路，从板两边经各自的引线到系统的地线，这样接地才能良好。

III. 对直流电源线要加适当地去耦电容，对 TTL 集成电路来说，一个较好的方法是：每个被同步驱动的门，应加 $0.01\mu\text{F}$ 的电容，每 20 个门，不管是否同步驱动，应至少加 $0.1\mu\text{F}$ 的电容。这些电容可以集中，但如果分布在印刷版上就更加有效。对于驱动长线的门，共直流电源的去耦电容要接在该器件的直流供电一端并且把电容器、器件和传输线的地线引到一个公共点上。

3. 定期检验

尽管 NIRS 的定量分析是根据光谱的定标模型来测定某成分的含量，但仪器的表现也是 NIRS 分析的关键因素，因为仪器的表现影响着 NIRS 分析的检测限、精度和方法的线性。20 世纪 80 年代，由 Karl Norris 教授领导的 ASTM 专业委员会，制定了一个评价近红外分析仪器表现的测试规程，它包括以下 7 个测试：

①光度计噪声；②反射元件；③线性；④波长精度；⑤波长准确度；⑥光谱带宽；⑦杂散光。

(1) 光度计噪声

光度计噪声（随机噪声）为多次扫描参比（标准物质）的标准差。对于透射型仪器，参比即为空气；对于漫反射型仪器，参比为反射标准物质。光度计噪声是以多次扫描光密度的标准差 RMS 来表示。一般来说，在整个近红外工作波长范围内，32 次扫描的 RMS 应小于 20×10^{-6} 光密度单位。

仪器光度计噪声应表现为随机性，如果光度计噪声出现峰值或整个噪声水平较高，则说明仪器出现问题。因此，仪器光度计噪声测试对评价仪器的工作状态是非常有效的，仪器的误操作或异常，往往可通过光度计噪声测试发现。下面是光度计噪声测试可推断仪器的异常源：

- A. 光源发光强度低：在 2300~2500nm 处出现过高的噪声峰；
- B. 光栅运动机械部件磨损：在 1500nm 处出现过高的噪声峰；
- C. 检测器温度控制不稳：RMS 的系统偏差过高；
- D. 检测器损坏：在整个近红外工作波长范围内噪声峰较高。

(2) 反射元件测试

仪器反射元件测试是测定炭黑的光谱，如图 1 所示。因为炭黑被认为可完全吸收弥散的散射光，因此，对于炭黑物质，检测器检测到的光就来自于反射元件。

(3) 仪器信号的线性

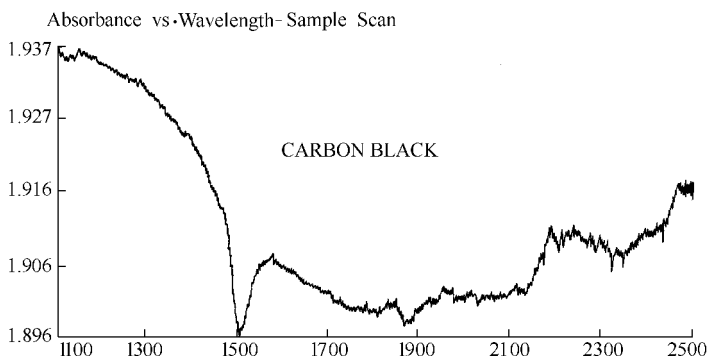


图 1 炭黑的近红外吸收光谱

仪器信号的线性测试需要 3 个标准物质：纯 HALON 和 2 个由 HALON 和炭黑组成的混合物。纯 HALON 可 100% 被反射，吸收为 0；两个由 HALON 和炭黑组成的混合物的吸收分别为 0.5 和 1.0 个吸收单位。

仪器信号的线性测试通常可检查仪器电信号测量是否正常。一般仪器信号的线性的测试是在仪器安装调试时进行。

(4) 波长精度、准确度、光谱带宽的测试

对仪器的波长精度、准确度、光谱带宽的测试可同时进行，使用的标准物质为聚苯乙烯和镨钕稀土，前者为长波长区的标准物质，后者为短波长区的标准物质，它们的 NIRS 光谱见图 2。测试时，每个标准物质扫描 20 次，计算机对这 20 次扫描结果的分析，可获得仪器的波长精度、准确度、光谱带宽的测试结果。

聚苯乙烯的 4 个吸收峰的位置分别为 1143nm、1681nm、2164nm 和 2303nm，而镨钕稀土的 3 个吸收峰的位置分别为 683nm、805nm 和 878nm。20 次扫描标准物质吸收峰位置的变化反映了仪器的波长精度、准确度；每次扫描与平均值的差表示为波长准确度，20 次的变异系数表示为波长精度。光谱的带宽由 2164nm 和 2175nm 处，吸收峰和吸收谷比来表示。

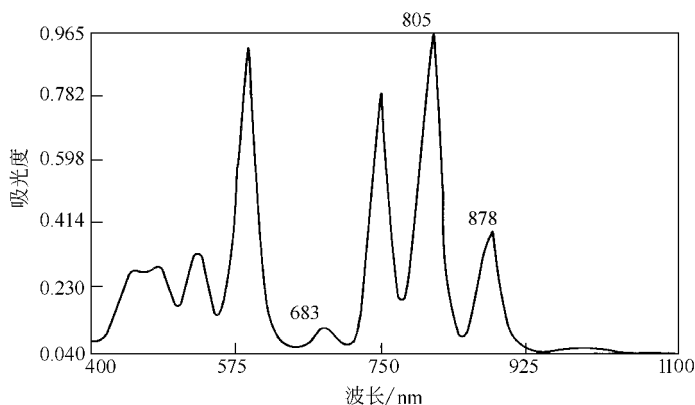


图 2A 镨钕稀土的 3 个吸收峰

(5) 杂散光的测试

仪器杂散光的测试也是用聚苯乙烯作为标准物质，扫描其在 1、2、3 倍厚度时，NIR

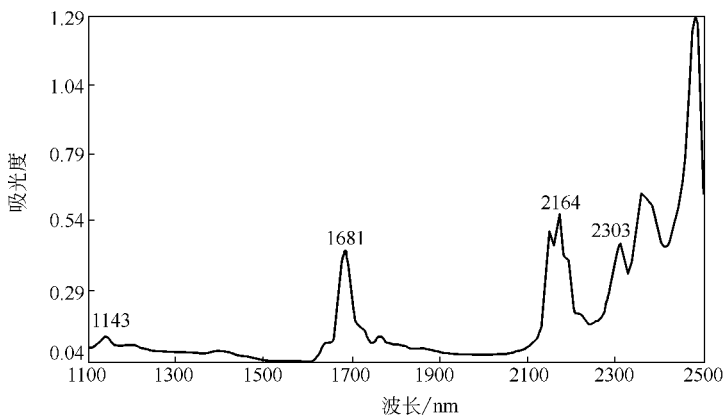


图 2B 聚苯乙烯的 4 个吸收峰

的透射光谱。如果没有杂散光，则在 2304nm 处，吸光度呈线性增加。如果存在杂散光，则线性就会出现离散。此外，要说明的是，对于色散型仪器的杂散光是光栅性能的函数，光栅确定后，仪器的杂散光就确定了，而不随仪器的使用操作变化。因此，仅在仪器出厂前做其杂散光的测试，用户不再进行测试。

4. 仪器故障发生几率分析

近红外光谱仪同其他分析仪器一样，故障出现的几率在安装运行早期较高。随着时间的增长，由于电子器件衰老、机械磨损等原因，达到耗损期时，出现的故障几率又逐渐增高。在运行早期和耗损期之间，有一段时间是比较稳定的，即发生故障几率比较低。对此，王化正等编著的《现代分析仪器维修手册》中有详细的论述。如果以时间 t 为横坐标，以出现故障几率 y 为纵坐标，即可绘制出一曲线，若经过规范化之后，便同可靠性的特性“澡盆曲线”完全吻合，如图 3 所示。

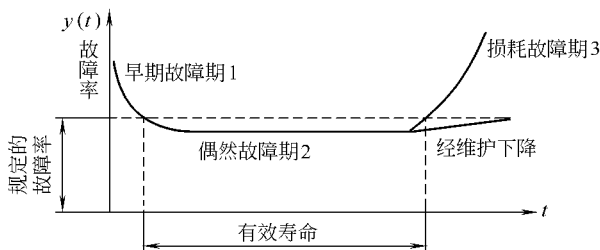


图 3 仪器故障变迁曲线

若以数学观点描述，在早期和磨损期出现故障几率是按高斯分布—正态分布的曲线给出的。数学表达式为：

$$y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

式中 x ——横坐标，为时间；
 y ——纵坐标，为故障几率；

μ —— 故障几率的平均值；

σ —— 标准偏差；

e —— 自然对数的底；

i —— 代表任意一次。

经验上，近红外光谱仪早期出现的故障，大多数发生在电子元器件上。而耗损期，出现的故障多数发生于机械部件或者光学零件。造成早期故障一般是由于设计上的问题，比如元件的选用，逻辑电路的设计等。生产工艺也是一个不可忽视的原因，新元件的老化筛选、印刷版的加工、焊接质量的保证都可减少早期故障的出现。但是，制造仪器的公司或厂家，一般都是整机厂，大多数单位并不生产电子元器件，都需要外购，其筛选和老化处理这些电子元器件的时间和手段也是有限的，如果再加上人为的操作失误等因素，仪器在使用初期发生故障的几率就高了。

为什么在使用中间一段时期发生故障的几率又低呢？这是因为在初期发生故障后，经过修理维护，仪器元器件都已得到老化处理和考验，而机械、光学、电子部件又未损耗或衰老，性能趋于稳定，操作人员也逐渐熟练，故这段时间发生故障几率一般都低，故障多数是因为随机因素造成。使用后期，仪器故障几率增高，显然是由于损耗原因，因此仪器使用到一定时期后必须更新换代。

5. 仪器的常见故障类型及排除

仪器的各部分都是按逻辑关系有条不紊地进行工作，以保证实现仪器所规定的各种功能。因此，不论组成仪器的任一部分或其中某一元器件损坏，或其性能参数发生变化，都可导致仪器发生工作紊乱、工作不正常或根本不能工作。王化正等编著的《现代分析仪器维修手册》中将分析仪器的常见故障分为 6 类，笔者认为近红外仪器也是如此，因此，将这方面的论述介绍给读者。

① 外因造成的故障

尽管现代近红外仪器对工作环境的要求不如早期仪器那么严格，但仍要求具备一定的安装条件，如对于市电、杂散磁场、温度、湿度、振动等都有一定的要求。由于这些因素造成的故障称为外因故障。外因故障轻，可以造成仪器工作不正常，严重时将会损坏仪器。

(a) 市电因素造成的故障

我国市电电压允许波动范围一般规定为 $\pm 10\%$ ，如果市电电压波动超过规定值或伴随其他干扰信号，仪器就可能发生故障。在这种情况下，应该使用 220V 交流稳压器。但是，选择稳压器时一定要注意稳压器的输出功率是否能满足仪器的需要。一般稳压器的功率应大于仪器所需功率的 2 倍，以免过载损坏稳压器。

市电中的干扰信号；我国市电规定为 50Hz，220V，波形为正弦波。如果附近有大型电气设备以及能产生干扰的电气设备时，就会把干扰信号传入市电网络中。例如大电流的可控硅整流器工作时，可使市电的交流正弦波上叠加尖脉冲。该尖脉冲严重时，可以引起高倍图像边缘不整齐。此外，有些高频干扰信号也会传入市电网络。但它的影响较小。

(b) 环境造成的故障

环境是指仪器周围的温度、湿度、杂散磁场、振动等外界因素，这些因素对仪器有不同程度的影响。

部分近红外仪器要求安装在有空调的房间内，温度保持在 20°C 左右，湿度在 60% 以下。这是因为温度变化太大和湿度过高都会影响仪器的正常运转。振动的影响：一般的安装条件要求为：振动不能超过 $3\mu\text{m}$ 。尤其振动在 50Hz 以下时，振动如果超过 $3\mu\text{m}$ 就会造成图像边缘有毛刺，数字式仪器读数不稳。但如果仪器设计中有较强的防振措施，则对安装的振动限制要求不高。杂散磁场的强度大小也影响仪器的工作。一般要求周围杂散磁场不能超过 $3\times 10^{-4}\text{T}$ ，尤其是 50Hz 以下的杂散磁场，对图像的影响较大，严重时图像周围也会产生毛刺。

② 仪器早期故障

仪器早期故障一般是指仪器安装时或开始运行期间内发生的故障。这是由于仪器不合理或装配工艺不佳造成的故障，特别是元器件的质量不高给仪器带来的故障。

(a) 元器件的质量造成的故障

由于近红外仪器所用元器件较多，因此，这类故障在实际操作发生的故障中占有很大的比重。例如，由于硅钢片质量差造成变压器发热，电容器断路或短路，电阻烧焦，电位器动态噪声过大，各种开关接触不良等，都会给仪器带来轻重不同的故障。尤其是二极管、三极管以及运算放大器、线性放大器、计数器、译码器、数码管、各种“与非”门等。这些器件都是电路中的重要部件，往往因为质量差，在使用过程中，各项参数常有变化。如：耐压不够而被击穿，有的温漂过大，改变了放大器的工作点和频率响应等等。元件质量不稳定，是电路中的最大隐患，有时由于一个元件失灵，会造成整台仪器瘫痪。所以，提高元器件的质量，进行筛选老化是一件非常重要的工作。

(b) 装配工艺造成的故障

一台仪器的质量，除了与元器件质量有关以外，装配工艺也是一项很重要的因素。在现代分析仪器的电路系统中，经常会发现由于配装工艺问题造成的故障。例如，电阻、电容、二极管、三极管各种组件以及各种电位器，开关等，常有接触不良、虚焊和开路等现象，同轴电缆的插头装配工艺处理不当往往会造成屏蔽层和芯线碰到一起，使得输入和输出信号短路。电位器、波段开关装配时，因没有定位，紧固不牢造成随着旋钮的转动而移动的现象，结果把焊接在上面的导线扭断。类似这种装配工艺所造成的故障、事例很多。

③ 自发性和磨损性故障

自发性故障也就是由于自然规律造成的故障，元器件有一定的使用寿命，随着时间的推移，元器件衰老到一定程度后，仪器性能逐渐发生变化或失效而引起故障。因此，当仪器使用到一定时间后，对某些易损元器件应进行及时更换，这样就可避免仪器发生故障。

磨损性故障一般指仪器的机械部位，由于运行疲劳或振动造成结晶构造变化或产生化学物理变化而引起的种种故障。对于机械部件，要特别注意维护，定期清洗上油，以延缓磨损件的寿命是机械维护方法的一种。

④ 附属故障

由于仪器中某一个元器件的故障而引起其他故障称为附属故障。而且，在修理过程中，由于处理不当而扩大的故障是时有发生。因此，要求修理人员进行修理工作时，不能轻举妄动，随便拆卸，随便变动电路中的各种可调元件，更不能毫无目的地东寻西找，企图侥幸找到故障所在，这样的做法很可能将仪器越弄越坏，以致无法收拾。

掌握修理技术的要领，在于对仪器的工作原理、逻辑关系、各单元电路的作用、某些

重要部件在线路中的作用以至工艺技术都要有一定的了解。只有在掌握了这些知识之后，才能比较准确地判断出故障所在。因此，在修理之前，应该首先仔细阅读仪器使用说明书，研究电路图、整机结构、仪器各部分的作用，在弄明白了这一切之后，才能着手进行修理。

修理的工作过程是：首先观察故障的现象。为此应按仪器的使用程序，转动面板上具有各种功能的旋钮，以观察仪器的各种反应。在必要时还应该加被测信号来进行观察，并将所观察到的故障现象记录下来，然后根据仪器的工作原理，逻辑关系，顺理推断以初步确定产生故障的大致部位，例如是转换部分呢，还是计数器或逻辑控制部分，并通过一定的测试手段以证实初步判断的正确性。在大部位确定之后，再进一步判断故障的单元电路，乃至最后找到产生故障的元件。当然，在这一过程也是离不开必要的测试工作的。故障找到后还需研究如何排除故障，更换已经损坏的元件，或重新调整电路的工作状态，所有这些步骤都应小心谨慎，切不可粗心大意操之过急。

有时产生某种故障的原因，从理论上分析可能是多方面的。遇到这种情况，尤其要谨慎。要从可能产生这种故障的多种原因中分出主次，从主要到次要，有条不紊地通过检测最终找到发生故障的部位。

修理工作的关键就是寻找故障，要想比较迅速而准确地找到故障部位，就要求修理者对仪器能有较为深入地了解。这是掌握修理技术的重要一环，也是防止故障扩大化的基本条件之一。

⑤ 人为的故障

所谓人为的故障，就是因违反操作规程和使用规则而造成的故障。

(a) 操作不当引起的故障。这种故障是由于操作人员的粗心大意和不熟练造成的，属于非损坏性的不正常现象。这样的问题往往误认为仪器有问题，实际上仪器是正常的，只是操作不当的问题，这类问题的解决办法，关键是提高操作人员的技术水平。

(b) 误操作引起的仪器损坏。仪器若处在自动控制状态下使用，一般都有保护装置，即便出现误操作，也不致使仪器造成故障。误操作形成的大部分故障是在手动状态下出现的。

(c) 电压相序变化造成的故障。这类故障常发生在仪器供电线路维修时，无意地改变了三相电源的相序。结果，开机后装有机泵的设备导致机械泵倒转，造成返油现象污染真空管道。类似这种故障完全是人为造成的。

⑥ 预防性维护的内容和方法

预防性维护包括日常维护和定期检查。根据仪器的特点、结构、出现故障的性质和经验等制订维护内容，包括各种光源、单色器、电源、声响、探测器、气味有无异常，示值和记录曲线精度情况，磨损件的运行状态等。潜在故障检查，包括可动件和静止件的检查。短寿命部件如电解电容、继电器等，由于极间短路或断路，引起的突发性故障和由于电流泄漏增大，静电容量减少引起的老化性故障；各种灯由于振动、气密性下降，部分氧化供电不稳等原因，可定期进行更换；保险丝因为慢性氧化、成分蠕变、应力集中、振动等原因造成报废，可定期更换。

仪器的精度应根据有关规定定期检定。每个操作者都应仔细阅读仪器使用说明书，并写出适当的工作手册和操作程序，仪器应有专人保管，并建立定期维护制度。

保证分析仪器工作正常，增长其使用寿命，最有效的方法是统计方法。像机械传动零部件定期清洗油垢、加油是增长平均无故障工作时间和减小磨损的好方法。对于光学、电子部件，保障适当的环境条件也是延长使用寿命和减少故障行之有效的方法。

采取适当措施，在故障出现前加以排除，对潜伏性故障能及时发现和处理，对有故障的零部件处理好以防再发生故障。预防性维护应贯穿仪器的整个运行期间。