

近红外光谱技术在食品品质检测方法中的研究

刘翠玲 吴静珠 孙晓荣 著



机械工业出版社
CHINA MACHINE PRESS

近红外光谱技术在食品品质检测方法中的研究

刘翠玲 吴静珠 孙晓荣 著



机械工业出版社

本书从社会实际需求出发,作者多年从事测控技术及信息处理,并与食品相关专业等研究人员合作,融入许多为了解决实际问题而进行的研究和实践经验,系统介绍了本课题组基于近红外光谱分析技术在蔬菜农药残留量的检测、食用植物油、小麦粉、淀粉的品质检测中的应用研究成果。本书主要内容:较系统地综述了近红外光谱法的分析特点及发展应用过程;总结阐述了近红外光谱分析的基本理论和流程;阐述了基于近红外光谱分析技术的蔬菜中农药残留量检测方法及模型建立;采用近红外光谱技术结合模式识别方法建立了食用油种类、花生油和橄榄油掺伪定性识别模型;采用近红外光谱技术结合化学计量学方法建立了食用油脂肪酸含量、过氧化值和酸价的定量分析模型;提出了基于近红外光谱检测小麦粉品质;运用近红外光谱技术可实现快速、无损地鉴别淀粉品质等。

本书适合从事食品、卫生、检测技术、自动化、信息处理等专业内容方面的研究生、本科生及相关工程技术人员、科研人员阅读。

图书在版编目(CIP)数据

近红外光谱技术在食品品质检测方法中的研究/刘翠玲,吴静珠,孙晓荣著. —北京:机械工业出版社,2015.11

ISBN 978-7-111-52026-9

I. ①近… II. ①刘… ②吴… ③孙… III. ①红外光谱-应用-食品检验-检验方法-研究 IV. ①TS207

中国版本图书馆CIP数据核字(2015)第260199号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑:顾谦 责任编辑:顾谦

责任校对:刘志文 封面设计:陈沛

责任印制:李洋

北京机工印刷厂印刷(三河市南杨庄国丰装订厂装订)

2016年1月第1版第1次印刷

184mm×260mm·13.25印张·326千字

0 001—2 500册

标准书号:ISBN 978-7-111-52026-9

定价:39.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线:010-88361066 机工官网:www.cmpbook.com

读者购书热线:010-68326294 机工官博:weibo.com/cmp1952

010-88379203 金书网:www.golden-book.com

封面无防伪标均为盗版

教育服务网:www.cmpedu.com



目前,食品安全问题备受关注,食品品质的快速、无损检测技术尚不健全,有碍于食品安全事件的快速监控,影响人民生活质量。本著作是研究团队从实际的社会需求出发,关注当前的社会热点,与多年从事测控仪器及信息处理、食品相关专业等研究人员合作,融入为解决实际问题所进行的科学研究和实践经验而著的。

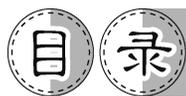
本著作主要是课题组在完成了北京市自然科学基金“基于光谱技术的农药残留量快速智能检测方法研究(4073031)”、北京市优秀人才基金“基于近红外技术的食用植物油品质检测方法研究(20081D0500300130)”、“基于多光谱信息技术的面粉品质快速检测方法研究(2012D005003000007)”等课题研究基础上,针对近红外光谱技术在测量食品品质时的信息提取与建模问题,以蔬菜农药残留、食用油、面粉为研究对象,采用近红外光谱技术与模式识别及智能信息处理相结合的方法,研究了近红外光谱技术在食品安全检测中的透射、漫反射的信息提取及测量模型的建立,确立实用的定性、定量模型,揭示了近红外光谱在食品测量中的信息重叠问题,研究了基于光谱技术的食品安全无损快速检测方法,建立了一种基于光谱技术的共享、稳定、自适应校正模型。

本书介绍了本课题组基于近红外光谱分析技术在蔬菜类农药残留量的检测、食用植物油品质、小麦粉、淀粉的品质检测中的应用研究成果。本书主要内容:较系统地综述了近红外光谱法的分析特点及发展应用过程;总结阐述了近红外光谱分析的基本理论和流程;阐述了基于近红外光谱分析技术的蔬菜中农药残留量检测方法及模型建立;采用近红外光谱技术结合模式识别方法建立了食用油种类、花生油和橄榄油掺伪定性识别模型;采用近红外光谱技术结合化学计量学方法建立了食用油脂肪酸含量、过氧化值和酸价的定量分析模型;提出了基于近红外光谱检测小麦粉品质;运用近红外光谱技术可实现快速、无损地鉴别淀粉品质等。本书适合从事食品、卫生、检测技术、自动化、信息处理等专业内容方面的研究生、本科生及相关工程技术人员、科研人员阅读。

本书由北京工商大学刘翠玲、吴静珠和孙晓荣著,共分9章内容,刘翠玲著作了第2~4章、第9章;吴静珠著作了第5章和第6章;孙晓荣著作了第7章和第8章;三位作者共同完成了第1章。在本书撰写过程中,得到了北京工商大学研究生隋淑霞、郑光、索少增、李慧、吴胜男、苗雨晴、董秀丽、胡玉君、赵琦等在校期间的协助。同时本书在著作过程中,参考了国内外一些优秀的相关研究内容(见各章参考文献),在此一并表示诚挚的感谢。

由于作者研究尚浅,有些学术观点乃一家之言,仅供参考,错误和不妥之处在所难免,欢迎读者批评指正,共同探讨。

作者



前言

第1章 绪论	1
1.1 近红外光谱技术的分析特点及在各领域中的应用现状	1
1.1.1 近红外光谱技术的分析特点	1
1.1.2 近红外光谱技术的发展及应用	2
1.2 近红外光谱技术在食品安全检测领域的研究概述	3
1.2.1 近红外光谱技术在农残检测领域的研究进展	3
1.2.2 近红外光谱技术在食用油品质检测领域的研究进展	6
1.2.3 近红外光谱技术在面粉品质检测领域的研究进展	11
1.2.4 近红外光谱技术在淀粉品质检测领域的应用	13
1.3 基于近红外光谱的数据处理分析方法概述	13
1.3.1 光谱数据预处理方法概述	14
1.3.2 校正模型方法概述	15
1.4 完成的相关科学研究项目概况	16
1.5 本书的主要内容	16
参考文献	19
第2章 近红外光谱分析技术基础	24
2.1 近红外光谱分析的理论基础	24
2.2 近红外光谱分析流程	26
2.3 近红外光谱分析预处理方法	28
2.3.1 主成分分析法	29
2.3.2 马氏距离法	30
2.3.3 杠杆值法	31
2.3.4 半数重采样法	31
2.3.5 蒙特卡洛交叉验证法	31
2.3.6 数据规范化处理	32
2.3.7 光程校正	32
2.3.8 基线校正	33
2.3.9 导数法	34
2.3.10 平滑处理	34
2.3.11 多元散射校正	35
2.3.12 傅里叶变换	35
2.3.13 正交信号校正	35
2.3.14 标准正态变量变换	36
2.3.15 基于分段窗口的特征谱区筛选方法	36
2.4 近红外光谱模型多元校正分析方法	37

2.4.1 多元线性回归 (MLR) 法	37
2.4.2 主成分回归 (PCR) 法	38
2.4.3 偏最小二乘 (PLS) 法	39
2.4.4 人工神经网络 (ANN) BP 法	40
2.4.5 聚类分析法	41
2.4.6 支持向量机 (SVM) 法	45
2.5 近红外光谱定标模型的评价指标	47
2.6 近红外光谱仪器介绍	47
2.6.1 光谱仪发展介绍	47
2.6.2 本研究应用光谱仪平台介绍	50
2.7 小结	53
参考文献	54
第 3 章 基于近红外光谱的溶液中农药残留检测方法研究	57
3.1 简介	57
3.2 基于近红外光谱的溶液中毒死蜥 PLS 定量分析方法研究	59
3.2.1 样品制备	59
3.2.2 光谱采集	59
3.2.3 结果与分析	60
3.2.4 重现性实验结果与分析	63
3.3 基于近红外光谱的溶液中毒死蜥 BP 网络定量分析方法研究	64
3.3.1 BP 算法的实验设计	64
3.3.2 样品制备	64
3.3.3 结果与分析	64
3.4 溶液中毒死蜥的 PLS 和 BP 定量分析模型性能的比较	71
3.4.1 PLS 法的样品选择	71
3.4.2 BP 法的样品选择	71
3.4.3 比较分析结果	71
3.5 基于近红外光谱的溶液中毒死蜥 PLS 定性分析方法研究	72
3.5.1 样品制备	72
3.5.2 光谱采集	72
3.5.3 结果与分析	72
3.5.4 重现性实验结果与分析	72
3.6 光谱预处理对毒死蜥和炔螨特近红外光谱定量模型性能影响分析	75
3.6.1 样品制备	75
3.6.2 光谱采集	76
3.6.3 不同预处理方法对毒死蜥 PLS 模型性能影响分析	76
3.6.4 不同预处理方法对炔螨特 PLS 模型性能影响分析	78
3.6.5 毒死蜥和炔螨特的近红外光谱 ANN BP 检测试验研究	80
3.7 小结	83
附录	85
附录 A 附表	85
附录 B 标准物质相关说明	87
参考文献	88

近红外光谱技术在食品品质检测方法中的研究

第4章 基于近红外光谱的蔬菜中农药残留检测方法研究	89
4.1 简介	89
4.2 蔬菜中常见的有机磷农药残留量检测方法	91
4.2.1 色谱法	93
4.2.2 生物化学测定法(胆碱酯酶抑制法)	96
4.2.3 发光菌检测法	100
4.2.4 生物传感器法	101
4.2.5 免疫分析法	101
4.2.6 光谱法	102
4.2.7 实验室机器人	110
4.3 基于近红外光谱的大白菜中毒死蜱残留量检测方法研究	110
4.3.1 样品制备	110
4.3.2 光谱采集	111
4.3.3 光谱预处理对模型性能影响分析	112
4.3.4 近红外光谱波长选择对模型性能影响分析	113
4.3.5 模型性能验证	113
4.4 基于近红外光谱的菠菜中毒死蜱残留量检测方法研究	114
4.4.1 样品制备	114
4.4.2 光谱采集	115
4.4.3 光谱预处理对模型性能影响分析	115
4.4.4 近红外光谱波长选择对模型性能影响分析	116
4.4.5 模型性能验证	116
4.5 基于近红外光谱的萝卜中毒死蜱残留量检测方法研究	117
4.5.1 样品制备	118
4.5.2 光谱采集	118
4.5.3 萝卜中毒死蜱残留量的PLS模型建立与分析	118
4.5.4 萝卜中毒死蜱残留量的BP模型建立与分析	119
4.6 小结	120
参考文献	120
第5章 近红外光谱技术在食用油品质定性分析中的应用研究	122
5.1 简介	122
5.2 模式识别方法	122
5.2.1 系统聚类分析法	122
5.2.2 支持向量机分类器	122
5.3 基于近红外光谱的食用油种类定性识别分析	127
5.3.1 样品制备	127
5.3.2 光谱采集	127
5.3.3 结果与分析	128
5.4 基于近红外光谱的纯花生油掺假鉴别分析	130
5.4.1 样品制备	130
5.4.2 光谱采集	130
5.4.3 结果与分析	131
5.5 基于近红外光谱的纯橄榄油掺假鉴别分析	134

5.5.1 样品制备	134
5.5.2 光谱采集	134
5.5.3 结果与分析	135
5.6 小结	136
参考文献	136
第6章 近红外光谱技术在食用油品质定量分析中的应用研究	138
6.1 简介	138
6.2 光谱分辨率的选取对食用油近红外模型性能影响分析	138
6.2.1 样品制备	139
6.2.2 光谱采集	139
6.2.3 结果与分析	139
6.3 基于近红外光谱的食用油脂肪酸定量分析	141
6.3.1 常见谱区优化算法	142
6.3.2 CARS 波长变量选择法	145
6.3.3 基于 CARS - PLS 的食用油脂肪酸定量模型的优化	146
6.4 基于近红外光谱的食用油过氧化值定量分析	148
6.4.1 样品制备	149
6.4.2 光谱采集	149
6.4.3 结果与分析	149
6.5 基于近红外光谱的食用油酸价定量分析	152
6.5.1 样品制备	152
6.5.2 光谱采集	152
6.5.3 结果与分析	152
6.6 小结	155
参考文献	155
第7章 基于近红外光谱的小麦粉品质检测方法研究	157
7.1 简介	157
7.2 基于近红外光谱的小麦粉 PLS 模型的建立与分析	158
7.2.1 样品制备	158
7.2.2 光谱采集	158
7.2.3 结果与分析	159
7.3 基于近红外光谱的小麦粉异常样品剔除方法研究	160
7.3.1 基于马氏距离法的异常样品判别分析	161
7.3.2 基于蒙特卡洛交叉验证法的异常样品判别分析	161
7.3.3 剔除异常样品后的模型评价	162
7.4 基于近红外光谱的小麦粉特征谱区筛选方法研究	162
7.4.1 小麦粉样品集划分方法比较	163
7.4.2 基于常规区间偏最小二乘法的小麦粉近红外光谱模型优化	164
7.4.3 基于联合区间最小二乘法的小麦粉近红外光谱模型优化	165
7.4.4 基于向后区间最小二乘法的小麦粉近红外光谱模型优化	166
7.4.5 基于 CARS 谱区筛选法的小麦粉近红外光谱模型优化	168
7.4.6 特征谱区筛选法的比较	170
7.5 SVM 和 ANN BP 在小麦粉品质定量分析中的应用研究	170

近红外光谱技术在食品品质检测方法中的研究

7.5.1 小麦粉近红外光谱的主成分分析	171
7.5.2 基于近红外光谱的小麦粉 ANN BP 模型	171
7.5.3 基于近红外光谱的小麦粉支持向量回归模型	172
7.5.4 小麦粉模型比较分析	174
7.6 小结	175
参考文献	175
第8章 基于近红外光谱的淀粉品质检测方法研究	176
8.1 简介	176
8.2 基于聚类分析的淀粉种类定性识别方法研究	176
8.2.1 样品制备	176
8.2.2 光谱采集	176
8.2.3 结果与分析	177
8.3 基于 SVM 的淀粉种类定性识别方法研究	181
8.3.1 样品制备	181
8.3.2 光谱采集	181
8.3.3 结果与分析	181
8.4 基于近红外光谱的淀粉含水量快速分析方法研究	187
8.4.1 样品制备	187
8.4.2 光谱采集	187
8.4.3 淀粉含水量 PLS 模型的建立与分析	187
8.5 马铃薯淀粉掺杂玉米淀粉的近红外光谱检测方法研究	192
8.5.1 样品制备	194
8.5.2 光谱采集	194
8.5.3 定量分析模型的建立	194
8.5.4 定量分析模型的验证	194
8.6 淀粉掺杂滑石粉的近红外光谱检测方法研究	196
8.6.1 样品制备	197
8.6.2 光谱采集	197
8.6.3 实验方法设计	198
8.6.4 定量分析模型的建立	198
8.6.5 定量分析模型的验证	198
8.7 小结	199
参考文献	200
第9章 总结与展望	201
9.1 本书的主要内容总结	201
9.2 展望	203
参考文献	203

1.1 近红外光谱技术的分析特点及在各领域中的应用现状

1.1.1 近红外光谱技术的分析特点

近红外光谱 (Near Infrared spectroscopy, NIR) 是介于可见光与中红外光之间的区域, 波长范围为 780 ~ 2526nm, 波数范围为 12500 ~ 4000 cm^{-1} 。近红外光区域是人们最早发现的非可见光区域, 距今已有 200 多年的历史。自 1800 年被天文学家 William Herschel 发现后, 经研究者的不断努力, 近红外光谱技术日益完善。由于近红外光谱兼有红外光区光谱数据信息量丰富和可见光区光谱数据容易获取的优点, 因此被广泛应用于食品及农产品作物加工检验、石油化工行业、生态及社会环境污染检验、生物医学等各学科的分析研究领域。

近红外光谱属于分子振动光谱, 产生于共价化学键的非谐能振动, 是分子振动的倍频和组合频, 源于 X-H 键 (X 为 C、O、N、S 等) 的吸收。由于不同的含氢基团所产生的光谱在吸收峰强度和位置上有所不同, 根据朗伯-比尔吸收定律, 如果样品的成分含量变化, 其近红外光谱特征也会随之发生变化。近红外光谱分析技术正是根据不同的光谱特征信息, 从而进行定性和定量分析的一种技术。

近红外光谱分析技术近年来得到了越来越广泛的关注, 成为了一种高效、快速、便捷、同时适合于定量和定性分析的工具, 正是由其技术特点所决定的:

1. 样品无需预处理

近红外光具有很强的散射效应和穿透能力, 在进行检测时, 所测样品无需进行任何预处理, 近红外光可以穿透玻璃、塑料包装及石英等进行直接检测, 也无需使用任何化学试剂, 因此不会破坏样品, 可进行无损检测。

2. 多组分同时测定

近红外光谱图能够提供大量的样品信息, 并且可以对多种样品成分同时进行分析, 能够在很大程度上简化测定操作。样品的不同组分对测定结果会有一定的影响, 是由于在测定过程中, 其他的组分对近红外光也有不同程度的吸收, 可以利用化学定量计算软件采用数学的方法来消除这种影响。这也是近红外光谱技术能够得到大力推广的主要原因。

3. 分析速度快

近红外光仪器的测定时间短, 利用傅里叶变换近红外光谱仪扫描一次样品一般只需 1s 左右; 传统的光栅型近红外光谱仪扫描一次样品也至多只需要几分钟; 而目前广泛使用的固定光路多通道检测型近红外光谱仪的扫描时间只需要几十毫秒。

4. 无污染、成本低

测量过程中不需使用化学试剂或特殊测试条件, 检测后不会产生污染, 也不会对人体造

近红外光谱技术在食品品质检测方法中的研究

成伤害，既环保又能够节省大量化学试剂所需费用。

5. 重现性好、效率高

近红外光谱技术具有近红外光谱仪、应用模型和化学计量学软件的三位一体的成套性特点。近红外光谱测量技术可以与化学计量方法相结合来建立分析模型，有很好的预测效果。

6. 对测试人员的要求不高，易培训、推广

当然，近红外光谱分析技术目前在实际应用中也存在一定的局限性，需要进一步探索和改善：

- 1) 对痕量分析和分散性样品分析有很大困难，其检测限常在 0.1%。
- 2) 模型建立需要投入一定的人力、财力和时间，而且建立后的模型还需要不断修正。
- 3) 模型的适用性受一定的地域或者环境影响，不容易建立通用模型。
- 4) 需要大量有代表性且化学值已知的样品建立模型。这样，对小批量样品的分析用近红外光就显得不实际了。
- 5) 模型需要不断更新，由于仪器状态改变或标准样品发生变化，模型也要随之变化了。
- 6) 模型不通用，每台仪器的模型都不相同，增加了使用的局限性。
- 7) 建模成本高，测试用度大。

1.1.2 近红外光谱技术的发展及应用

到目前为止，近红外光谱分析技术研究大体上经历了五个发展阶段：

第一个阶段是从 1800 年发现近红外光之后的一个半世纪，由于受当时的技术水平和实验条件限制，无法将近红外光谱信息充分提取出来，不能进行深一步的研究，使近红外光谱技术沉睡了近 150 年的时间。

第二个阶段是 20 世纪 50 年代，随着近红外光谱技术和仪器的发展，Karl Norris 率先将近红外光谱技术应用于农副产品分析，开启了近红外光谱技术在农副产品应用的时期。

第三个阶段是 20 世纪 60 年代中后期，随着红外光谱技术的发展，加上当时近红外光谱分析无法突破灵敏度低、抗干扰性差的弱点，自此近红外光谱分析技术又进入徘徊不前的状态，被人们称之为光谱技术中的沉睡者。

第四个阶段是进入 20 世纪 80 年代后期，随着计算机技术的发展，加上中红外光谱技术积累的经验以及化学计量学方法的应用，解决了近红外光谱谱峰重叠、提取光谱信息困难及背景干扰等问题，使近红外光谱分析技术在各个领域得到了迅速推广，成为一门独立的分析技术。

第五个阶段是 20 世纪 90 年代以后，近红外光谱分析技术进入一个快速发展时期，在工业领域的应用全面展开。同时计算机软硬件技术突飞猛进，数字化的光谱仪器与化学计量学方法相结合，形成了现代近红外光谱分析技术，有关近红外光谱的研究及应用文献呈几何级增长。

与国外的研究相比，由于缺乏先进的光学系统、优良的电子线路、精密的机械结构和高质量的数据处理技术，我国对近红外光谱技术的研究从 20 世纪 90 年代中后期才逐步开始。近几年随着人们关注度的不断提高，已经在基础研究、仪器和软件开发等方面取得了巨大进步。

近红外光谱分析作为一种快速、高效、无损的检测技术，可以对气、液、固等各种样品进行定量或定性分析。目前，近红外光谱分析方法已经在农业、食品工业、石油化工、制药、医学等领域广泛应用。

1. 农产品中的应用

近红外光谱技术可测定谷物中的水分、蛋白质、淀粉、油脂、糖分、脂肪及食用纤维、矿物质含量等；水果和蔬菜中的水分、甜度、酸度、成熟度、农药残留分析；饲料中纤维、蛋白质、木质数、活体和离体消化度、灰分及消化摄入量等；此外还在烟草作物和木材等产品中应用。

2. 食品工业中的应用

鱼类肉类、奶制品、食用油和液体饮料。测定包括水分、糖分、脂肪、蛋白质、氨基酸、纤维、灰分等营养成分含量。

3. 石油化工行业的应用

汽油炼制中辛烷值、芳香烃含量、苯含量、乙醇、蒸馏值、挥发值、添加剂、黏度、闪点、相对密度等的测定；柴油、润滑油的组成及性质分析；高分子合成及加工中单体纯度、残余单体量、聚合度、相对分子质量、交联度、密度等性质指标。

4. 制药工业及临床医学的应用

工业中原料和活性组分的测定，固体药剂的湿度、含量均一性、颗粒大小分布、片剂膜衣厚度、结晶度及硬度的定量表征；临床医学中用于估测组织中氧合血红蛋白和无氧血红蛋白的含量、血液的体积、血液流速、组织耗氧量、血糖含量等，还可离体测定血浆和尿中蛋白质、葡萄糖、胆固醇、甘油三酸酯及其他类脂化合物的含量。

5. 其他应用

近红外光谱分析技术还广泛应用于纺织、造纸、地质学、生物化工、天体学及地理等领域。

1.2 近红外光谱技术在食品安全检测领域的研究概述

1.2.1 近红外光谱技术在农残检测领域的研究进展

从20世纪50年代起到最近几年，经过多年的研究和发展，农药残留量检测分析技术日趋多样化、便捷化、低成本化和快速化。目前，农药残留量检测方法有：色谱法、毛细管电泳法、酶抑制检测法、生物传感器检测法、酶联免疫分析检测法、蛋白芯片技术检测法、酶分析技术检测法、活生物体分析技术检测方法、光谱检测法（荧光光谱法、紫外-可见光谱法、近红外光谱法、红外光谱法、高光谱成像检测技术）等。何永红在基于生物学原理的农药残留量检测技术的进展综述中提到，生物学检测农药残留量的主要方法之一：免疫分析法，其优点表现为独特并且敏感，然而，获取因各类农药而产生的抗体很难。生物学检测农药残留的主要方法之二：生物传感器方法，此方法的特点是迅捷和同步性，但这种方法的敏感性、适用强健性尚且有待于进一步的研究。生物学检测农药残留的主要方法之三：酶分析技术和活生物体分析技术，此类方法相对于免疫分析法和生物传感器法，且更加廉价，但其与化学检测方法都需要对样品进行相对繁杂的处理。

近红外光谱技术在食品品质检测方法中的研究

在生物、化学方法的农药检测中，为了克服样品组成成分复杂性，便于检测和得到准确、可靠的检测结果，需采用必要的预处理方法，常用的样品制作提取方法包括提取和净化，提取和净化主要采用的方法包括浸渍-振荡法、索氏抽提法、超声波提取法、净化技术、固相萃取、固相微萃取（SPME）、基体分散固相萃取（MSPD）、加速溶剂萃取（ASE）、超临界萃取（SFE）、凝胶色谱（GPC）净化、固相萃取（SPE）净化等。样品的预处理技术虽然为微量农药残留量检测技术提供了便利和作出了巨大的贡献，但是其操作过程却是繁冗复杂的。而光谱检测法作为近几年发展起来的新检测技术，由于可弥补上述生物、化学方法的缺点，引起了广大研究者的兴趣。

使用光谱技术测定样品的基本内容如下：首先，采用现有的标准分析技术测出待测样品的性质和数据；然后，使用光谱装置和仪器采集样品的光谱数据，并用各种分析方法对光谱数据进行分析；最后，建立光谱数据的数学模型并将光谱法的结果与标准法的测量结果进行关联，即得到以标准方法为参考的定量和定性分析关系的校正模型，并使用此数学模型对预测集样品进行预测，以各种评价参数判断所建立模型的优劣，如稳定性、泛化能力等。尽管光谱分析属于间接检测和二次检测技术，但是光谱分析技术在与标准方法关联后所建立的关于物质性质的模型与常规检测方法并无系统偏差，并且其结果更加稳定，精度更高。

本书所研究的光谱检测方法为近红外光谱技术。虽然近红外光谱吸收强度和灵敏度都较低，在痕量分析领域具有挑战性，使用近红外光谱技术检测含量在 10^{-3} 以下物质的近红外光谱检测技术的可行性研究较少，但在理想条件下，只要样品在近红外光谱下可以提取到光谱数据，并且借助于化学计量学在近红外光谱分析中的应用，使用近红外光谱技术分析微量农药残留量是可能的。并且，随着近红外光谱分析技术的不断改进，近红外光谱仪信噪比可高达 10^5 ，利用其性能和精确度的不断提高，近红外光谱分析技术应用于食品农药残留污染方面的检测研究在一定程度上得到了进一步的发展。

本章参考文献 [28] 在近红外光谱分析技术用于菠菜中毒死蜃残留的定量分析研究中，为了使检测更接近于现实农作物的组分，制作的用于后续适用的试验样品是 37 个浓度为 $0.005 \sim 0.1 \text{ mg/kg}$ 的以效仿真实果蔬成分背景的毒死蜃物质的混合液，在此后，又准备了 36 个浓度为 $0.1 \sim 3.65 \text{ mg/kg}$ 的菠菜萃取溶液，在菠菜萃取液中加入了以实验设计浓度为标准的毒死蜃农药物质，采用偏最小二乘（Partial Least Squares, PLS）回归统计方法建立校正模型，从实验的效果来看，两种样品均得到了良好的预测效果。

熊艳梅在短波近红外模拟测定农药乳油中的有效成分含量的研究中，以其试验的初步设计和对样品值的规划，配置了用于校正的检验对象 23 个，样品的平均分布为 8.77，经过计算得到了 4.53 的检测对象标准差。另外，设置了 7 个样品平均分布为 8.37 的用于预测适用的目标集合样品，此集合的标准差为 4.29。对近红外仪器扫描到的样品数据使用中心化方法、极差归一的优化方法进行预处理，评判试验结果的方法使用全交互法进行评价，得到了较优的校正结果： $R^2 = 97.37\%$ ， $\text{RMSEC} = 0.73$ ， $\text{RMSEP} = 1.06$ ， $\text{RPD} = 4.05$ ，表明了短波近红外光谱技术在农药质量检测应用中的优势。

沈飞等对近红外光谱分析技术在辛硫磷农药残留检测中的应用研究中，纵观近红外光谱技术的新发展，综合考虑后采用的优化手段如下：以硅胶为吸附剂来吸附微量的农药，这种方法需要将需要扫描的对象掺入硅胶，然后再扫描其近红外数据，并以此建立 PLS 数学模型，建立模型后的评价手段采取留一法交互验证。结果表明，以 0.5 mg/L 作为间隔选取的

21个待测目标的评价参数如下,交互验证 $R^2 = 0.958$, $RMSECV = 0.872\text{mg/L}$ 。以间隔为 0.25mg/L 选取一组待检测目标集的41个待测目标的评价结果为,交互验证 $R^2 = 0.924$, $RMSECV = 1.15\text{mg/L}$ 。模型的预测能力虽然在间隔为 0.25mg/L 样品中有所损失,但是其相关系数仍是较优的,以硅胶吸附样品中农药的方法对痕量农药残留量的近红外光谱检测技术是有效的。

刘丽丽基于近红外光谱技术研究了白菜乐果残留的分析和检测探索,使用的主要样品制作手段为以超高压对蔬菜进行预处理,压力持续3min,并且保持300MPa的压强,并在其中掺入50mL的萃取液,需要25g农作物。仪器参数的设置情况大致如下,光谱扫描的分辨率: 4cm^{-1} ,按采集32次进行光谱平均处理,光程1mm,以此建立了近红外光谱仪器扫描到的数据的分析模型,结果表明,使用超高压处理过的蔬菜样品的相关系数为87.14%,模型的 $RMSECV = 0.139$,根据研究结果,此方法的分析精度为 10^{-5} ,比同等条件下不进行超高压处理的样品数据的分析精度高100倍,并且,此方法的快捷程度为NY/T761-2004所规定的国标有机磷农药分析快速程度的20倍左右。

由于农产品种类繁多,在进行基于近红外光谱技术的农药残留量的检测技术中样品的处理需要考虑其地域性和所含背景物质的复杂性等繁多因素,针对各品种进行合理处理的最优方法仍未有统一的处理技术标准,所以在样品准备和制备方面还没有形成公认的统一方法。由于使用近红外光谱技术进行农药残留检测仍未完全成熟,如何使在一次实验中处理样品的方法同时适用于不同研究室和研究人员所使用的仪器及不同样品,仍需要大量的研究证实,并且对于样品扫描得到的光谱数据进行预处理方法的标准仍需大量研究实践来制定。

对农药残留近红外光谱原始光谱数据进行优化处理后,定性、定量分析模型的建立和模型检验是农药残留量近红外光谱分析的核心。常用的多元校正分析方法有多元线性回归(Multivariate Linear Regression, MLR)、主成分回归(Principle Component Regression, PCR)、逐步回归分析(Stepwise Regression Analysis, SRA)法、偏最小二乘(PLS)法和人工神经网络(ANN)和拓扑(Topological)等,模式识别在20世纪60年代末被引入到化学领域后,化学计量学专家们研究发展了多种模式识别算法,如聚类分析(Cluster analysis)法、最短距离法、重心法、离差平方和法、K-均值聚类方法等。

文献[44]在近红外光谱技术检测溶液中毒死蜍含量的试验中,对效仿蔬菜成分的试验用试验液37个样品——农药残留含量为 $0.005 \sim 0.100\text{mg/kg}$ 有机磷农药试验液和10个不含农药的蔬菜模拟液分别以PLS和模式识别中的聚类分析法建立分析模型,进行农药残留量的含量和鉴别分析,证明此技术在检测农药残留量为 $0.008 \sim 0.090\text{mg/kg}$ 的氯吡硫磷果蔬模拟液中预测效果好,聚类分析模型的分类正确率高。

周向阳等人在近红外光谱法(NIR)快速诊断蔬菜中有机磷农药残留的研究中,基于近红外光谱技术对诸如十字花科等20余种农作物果蔬产品进行了探索微量农残分析研究,研究表明,此技术的检测分析优势显著,并认为近红外光谱法是一种易用、迅捷并且稳健的农药残留分析技术。文中利用近红外光谱技术分析溶液中微量毒死蜍含量,首先,在配置的过程中将样品分为两类:一类是以甲醇和水为混合背景的氯吡硫磷溶液,另一类是由甲醇、水、维生素C、蔗糖和氯吡硫磷掺杂而效仿农作物果蔬的试验溶液(蔬菜模拟液)。以二阶导数+17点平滑和一阶导数+21点平滑作为光谱数据的优化手段,并利用PLS法建立模型,结果表明,不同样品中氯吡硫磷质量比(大于或等于) 3.2mg/kg 的样品模型效果

近红外光谱技术在食品品质检测方法中的研究

更好。

另外，在农药残留近红外光谱分析的非线性校正技术中，ANN 的应用十分广泛。吴泽鑫等基于近红外光谱的番茄农药残留无损检测方法研究中，建立的基于误差反传理论的神经网络方法数学校正分析模型，试验得到了如下结果：对校正集外的样品的正确判断高达 96%，在模型的建立过程中，对神经网络模型的训练仅有 0.015 的偏差， $R=0.971$ ，表明了 BP 神经网络在近红外光谱法检测农药残留量研究中的可行性。代芬等人在基于近红外光谱分析的龙眼表面农药残留无损检测研究中，对龙眼上存在的 0, 0 - 二甲基 - (2, 2, 2 - 三氯 - 1 - 羟基乙基) 磷酸酯和 0, 0 - 二甲基 - O - (2, 2 - 二氯乙烯基) 磷酸酯残留含量分析进行了新的探索，仪器的扫描波数为 500 ~ 1000nm，扫描用不同浓度的 0, 0 - 二甲基 - (2, 2, 2 - 三氯 - 1 - 羟基乙基) 磷酸酯和 0, 0 - 二甲基 - O - (2, 2 - 二氯乙烯基) 磷酸酯喷洒的龙眼，对扫描到的采样值进行主成分分析、聚类分析并建立 BP 神经网络模型，结果表明，模型判断两种农药物质的微量残留结果高达 93% 和 80%，为 BP 神经网络在近红外光谱检测中应用的可行性提供了实验依据。

农药残留检测技术的需求促进了分析科学特别是仪器分析的飞速发展。同时，仪器分析技术的高度发展，也为农药残留检测能够达到超痕量水平提供了保障。未来农药的发展是以持续发展、保护环境和生态平衡为方向的。目前，农药残留检测的发展趋势有以下几个方面：

1) 总体上，农药残留分析继续朝着安全、高效、经济、环保的方向快速发展，表现为微型化、省时省力、高效、减少溶剂用量、限用或慎用高毒性溶剂和氯代烃类溶剂以及分析成本不断下降，实验室人员的安全意识和环境保护意识比过去有了提高。

2) 在检测方面，农药分析范围扩大到数百种；超临界流体色谱 (SFC)、高效毛细管电泳 (HPCE)、高效薄层色谱 (HPTLC) 和离子色谱 (IC) 等新的分析技术不断涌现；各种联用、自动化技术的广泛应用，近红外光谱分析技术也开始进入农药残留检测研究中。

3) 快速检测技术迅速发展，其中主要以酶联免疫 (ELISA) 技术为代表的快速检测技术已进入市场，满足生产和生活实际中的需要，实现农药残留现场和即时分析。

4) 不断加快农药残留分析方法的标准化步伐，越来越多的农药残留实验室的质得到认可，农药残留分析的技术和结果在国际间的协同化应用已变为可能，这将极大地提高农药残留检测的效率。

5) 世界经济一体化的趋势和人们日益强烈的质量及健康意识，农药残留已经成为食品安全问题的主要部分，正在形成完善的农药残留监测和管理体系。

由上可以看出目前色谱法等主流检测手段虽然能够达到检测要求，但随着农药残留检测朝着安全、高效、经济、环保的趋势发展，选择一种快速、方便、无污染的“绿色”检测方法是十分必要的。

1.2.2 近红外光谱技术在食用油品质检测领域的研究进展

我国是世界第一植物油消费大国。据美国农业部统计，2010 - 2011 年度，中国植物油总消费量为 2475 万 t，占全球消费量的 18.3%。随着我国综合国力的不断增强、人民生活水平的日益提高，食品的卫生和质量问题也备受人们的关注。在人们的日常饮食中，食用油成为不可或缺的重要组成部分，同时其优劣与人们的健康息息相关，劣质食用油所引起的安

全隐患问题被人们广泛重视。长期使用劣质食用油，如地沟油、掺假油等，大量的毒害物质在体内积累，会引起人们身体不适，甚至致癌。近年来，我国普遍加强了对食用油安全的监管力度，如2006年国家质量监督检验检疫总局组织对芝麻油产品进行了国家监督抽查，产品抽样合格率仅为87%。2008年1月甘肃省质量技术监督局对该省获得食品生产许可证的食用植物油生产企业的产品进行了监督抽查，结果显示，食用油的合格率为76.5%。2010年6月，中山市依据《花生油》(GB 1534—2003)标准，巡查食用植物油企业132家，花生油生产合格率仅为68.57%。

劣质食用油主要包括三类：一是狭义的劣质食用油，即将下水道中的油腻漂浮物或剩饭、剩菜经过简单加工、提炼出的油；二是劣质猪肉、猪内脏、猪皮加工以及提炼后产出的油；三是用于油炸食品的油使用次数超过规定要求后，再被重复使用或往其中添加一些新油后重新使用的油。

劣质食用油与地下水泥壁（含有多种微量元素，如As、Pb、Cd、Cr、Hg、Zn、Co、Ni、Tl、Be、Sn、Cu、Sb、Mo和V等）、地下生活污水、废旧铁桶、果蔬腐败物、生活垃圾、细菌毒素、寄生虫及虫卵等接触，所受污染严重，逐渐会发生水解、氧化、缩合、聚合、酸度增高、色泽变深等一系列变化，产生游离脂肪酸、脂肪酸的二聚体和多聚体、过氧化物、多环芳烃类物质、低分子分解产物等，有些物质如醛、酮等是导致劣质食用油特殊酸腐恶臭气味的重要成因。

劣质食用油经过水洗、蒸馏（餐厨废油脂成为粗毛油）、脱色（加入活性炭、白土等成为脱色油）、脱臭（高温真空脱臭成为脱臭油）、调节酸价等精炼工艺后，其色泽、气味、滋味等感官指标，以及酸价和过氧化值等理化指标常会接近或完全达到国标GB 2716—2005《食用植物油卫生标准》，已经很难通过感官分析和检测一些理化指标与合格食用油进行区分。而如果将劣质食用油与合格的食用植物油掺兑以后贩卖，更是进一步加大了检测难度。

目前依据我国制定的各类食用植物油的检验和卫生标准，检测劣质食用油的理化指标包括水分含量、比重、折光率、皂化值、酸值、羰基值、过氧化值、碘值重金属、脂肪酸相对不饱和度、胆固醇、残留检测、氧化产物检测等。常用的检测手段有薄层色谱技术、光谱法、色谱法、电导法、低磁场核磁共振鉴定法、质谱法及其食用油谱图模型的构建等。

1. 薄层色谱法

薄层色谱技术对劣质食用油中含有的而合格食用植物油中不存在的氧化产物醛、酮类等进行分离检测，这类物质在薄层色谱中表现为明显的拖尾斑，并且劣质食用油与合格食用油的薄层色谱特征斑点形状存在明显区别，由劣质食用油计算所得的 R_f 值也与食用植物油相差很大。该方法具有简单、快捷的优点，主要缺点是劣质食用油中所含的极性成分（醛、酮类物质）在油脂的深度精炼过程中大部分可以被去除，而且该方法的灵敏度和定量准确度较低，不适用于低含量掺伪食用油的检测。

2. 水分含量法

水分含量法大多使用水分仪测定待测样品中的含水量，将含水量作为鉴别劣质食用油的特征指标之一。通常劣质食用油的水分含量高于1%，食用油水分含量低于0.2%；但是煎炸老油在反复的加热过程中，水分尽数挥发，因此水分含量不可以作为煎炸老油的鉴别指标，具有一定的局限性。

3. 快速检测法

(1) 便携仪器快速监控法

由于劣质食用油中极化分子含量与正常植物油有很大区别，而极化分子总含量越高，其介电常数越大。康春雨、刘珉恺通过测量劣质食用油电容量，与标准油的参数进行比对，来判定食用油的品质。刘珉凯通过食用油和劣质食用油产生并收集到不同的光散射和光吸收信号，开发了便携式劣质食用油检测装置。

(2) 快速检测试纸法

植物油酸败产生的游离脂肪酸，能与喷溴甲酚绿指示剂作用显色，在试纸条表面的硅胶层中嵌入指示剂，滴入其上的劣质食用油与指示剂反应，通过观察颜色变化，快速检验食用油变质情况。但由于油脂的储藏不当，皆可以使油脂出现氧化、酸败，导致酸价和过氧化值升高，因此该方法的准确性有待商榷。

4. 色谱法

从餐饮业废弃物中提取的劣质食用油成分复杂，往往会含有大量的动物性食品。因而，劣质食用油中不可避免地混有动物油脂。动植物组织中都含有甾醇，动物油脂的特征性甾醇是胆固醇（猪油、猪板油和牛油中胆固醇含量分别0.75mg/g、1.01mg/g和1.45mg/g），而植物油中一般不含或只含有极少量的胆固醇。因此通过测定油脂中胆固醇的量，即可判定植物油中是否含有动物油脂，从而推断该油是否混有劣质食用油。胆固醇的检测手段主要是色谱技术。郭涛等人通过高效液相色谱法测定劣质食用油中胆固醇含量，当待测菜籽油样品中胆固醇的含量大于0.05mg/g时，出现胆固醇吸收峰，判定为掺杂劣质食用油。但由于植物油的基质较为复杂，胆固醇吸收峰受基质干扰影响较大，液相色谱很难完全分离，且灵敏度较低，当掺杂劣质食用油的含量较低时，较难准确地进行定性、定量分析。

气相色谱技术是目前国内检测油脂中胆固醇含量的主要方法。张蕊等人用气相色谱法测定胆固醇含量来鉴别劣质食用油，先利用极性毛细管色谱柱可将油脂中的胆固醇和植物甾醇很好地分离，然后进氢火焰离子化检测器检测，从而鉴别油样中是否混有劣质食用油。但是这种方法有一定的不足，其检测灵敏度较低，当食用植物油中掺入的劣质食用油比例少于10%时，难以检出其中的胆固醇。气相色谱法还可用来针对油样中的脂肪酸组成成分进行测定。黄道平等人对油样中的脂肪酸经甲酯化处理、程序升温分离、氢火焰离子化检测器检测，实验结果表明，由于劣质食用油是混合油脂（植物油和动物油），其脂肪酸的组成和含量与单一植物油有明显的区别，因此劣质食用油同时具有多种油脂的脂肪酸谱图特征，并且单一植物油的脂肪酸谱图与混合油脂的脂肪酸谱图存在差异，可利用气相色谱测得的脂肪酸谱图进行比较分析来鉴别样品是否是掺入劣质食用油的混合油脂。但该方法存在一定的弊端，各种油脂中的脂肪酸保留时间相似，在保留时间误差范围内，并且当掺杂劣质食用油的含量较低时难以进行判断。

5. 电导率法

食用油脂属于非导电物质，电导率低。油在使用过程中（加入食盐等调味剂）和烹调过程中有机物会分解产生可电离物质，以及劣质食用油在收集、提炼、加工过程中由于油脂变质，引入许多酸败、游离产物，这些物质导致其电导率增大。研究表明，酸败、游离产物含量与电导率成正相关关系。在石油醚中用水浸提油脂中的导电物质，测定水相电导率，反映其中的酸败、游离产物含量，从而对劣质食用油进行检测。电导率法的主要技术缺陷是：

通过深度精炼，可以将餐厨废油脂中绝大部分的水溶性物质去除，精炼工艺也会引起电导率的较大波动，不适用于深度精炼劣质食用油的鉴别。

6. 低磁场核磁共振鉴定法

低磁场核磁共振鉴别食用油品质的方法以合格食用油和劣质食用油弛豫时间图谱数据的区别为主要鉴别依据，以合格食用油和劣质食用油的核磁共振信号为主要观察对象，以合格食用油与劣质食用油的横向弛豫时间图谱数据对比分析为主要手段，进行合格食用油与劣质食用油的快速准确鉴别。其特征在于首先针对不同品质的油类样品在相同参数下进行核磁共振信号采集，利用横向弛豫时间谱和横向弛豫时间与扩散系数相关性谱对已知合格食用油进行分析，找出不同品质油类样品核磁共振测试数据的区别，并对同一品质油类样品进行重复性测试验证；然后根据分析的结果将测试数据与油类品质进行对应，确定油类品质与核磁共振数据的相关性；随后，根据不同品质油类的测试数据建立油类品质测试数据库，并形成分析测试标准；最后，通过数据库内的测试数据对未知油类样品进行品质鉴定。

低磁场核磁共振技术对劣质食用油初步研究表明，劣质食用油的弛豫谱显示特征小峰。实验结果表明，掺假比例越大，小峰信号越强。该结果尚需大量实验验证，成峰的化学本质也有待于进一步研究确证。

7. 真菌霉素法和基因鉴定法

黄曲霉素是真菌霉素的一种，已有通过检测黄曲霉素来判定待测油脂是否是劣质食用油的报道。该方法的不足在于：并不是所有的劣质食用油中都含有真菌霉素，并且通过餐厨油脂的深度精炼，可以除去劣质食用油中的大部分真菌霉素。况且，霉变的花生油及其他植物油中也会产生黄曲霉素和出现超标现象。

基因鉴定法是检测人员根据分子生物学基因鉴定的方法，鉴定油脂中的动物基因，来判定食用油中是否含有动物源性成分进而来确定是否掺杂劣质食用油。

传统的检测食用油品质的方法分析仪器设备往往很昂贵，样品需要经过前期处理，并且实验本身的耗时性、破坏性又是许多场合所不允许的。而近红外光谱分析技术是利用近红外谱区包含的物质信息，用于鉴别食用油种类和分析食用油的品质，它克服了传统方法的不足，对食用油样品无破坏性，且操作简便、快速。针对目前食用油市场存在的质量问题，如真伪、掺假和卫生指标超标等，应用化学计量学和计算机技术相结合的方法，建立近红外定性和定量模型用于食用油品质检测，常见的检测指标如下所示：

(1) 检测食用油的脂肪酸含量

脂肪酸，如棕榈酸、油酸和亚麻酸等，是食用油品质的重要指标，其含量直接影响着食用油的品质。花生油、大豆油和棕榈油等，其脂肪酸含量的范围各有其特定的标准，当含量不满足标准的范围时，则有可能是油品变质甚至不法商贩掺假。比如，在花生油中掺入大量的棕榈油来降低生产成本谋取利润，近红外光谱分析技术便可以通过检测油品中的脂肪酸含量的明显差别来判断油品是否纯正，是否含有其他的物质。同样，也可通过近红外光谱分析技术检查单一油品是否变质等，进一步监控食用油的品质。

近年来，在食用油脂脂肪酸等重要指标的含量测定方面有了一定的研究成果。陈嫚等人用0.15%脂酶于60℃恒温水浴下酶解天然棕榈油，配制成不同游离脂肪酸浓度梯度的棕榈油，利用近红外光谱扫描，利用多元线性回归创建校正模型，相关系数达到0.99，成功完成了游离脂肪酸的含量测定。

近红外光谱技术在食品品质检测方法中的研究

于燕波、臧鹏、付元华等人应用近红外光谱技术建立了测定食用油中脂肪酸含量的模型，相关系数达到 0.949，结果表明，使用该方法能够较准确地测定特征含量，便于实际应用。

但是，随着时间的推移，食用油中脂肪酸含量会有所变化，所建立的模型可能不具备良好的重现性，这就需要对模型进行不断地优化，使得该模型能够具有更好的预测能力。另外，建模时需要收集大量的样品，这样才能满足所建立的模型具有较好的样品覆盖性。这些都是现有的研究中需要亟待解决的问题。本书就此做了进一步的优化，尽可能收集更多的样品并且不断地优化模型以克服上述研究中的不足。

(2) 判别食用油种类及掺假种类

大豆油、花生油和菜籽油等在外观上相似，肉眼不容易区分，这样就会给实际的生产、生活带来很多不便，在不破坏食用油样品的前提下，近红外光谱分析技术能够有效地对其进行区别分类，通过建立可靠性较高的模型，检测不同种类的食用油。另外，对于掺假种类的定性分析方面，近红外光谱分析技术也有着一定的优势。通过聚类分析或支持向量机(SVM)等算法建立分类模型，检测待测食用油样品，可以有效地分辨出原纯食用油中掺假了特定的哪一种廉价食用油，这个方面的应用，对于消费者和质检部门有很重要的指导和参考作用。

刘福利、陈华才等人应用近红外分析技术对调和油中大豆油、花生油的含量进行分析检测，有效地判别出了调和油的品质问题。

传统常用聚类分析法来建立检测食用油掺伪的校正模型，但该方法通常无法判别出掺伪油的品种，本书深入研究了应用 SVM 的方法建立定性分析模型，在一定的掺伪油浓度范围内，能够很好地进行鉴别，为进一步的探索提供了可能。

(3) 测定食用油过氧化值和酸价

食用油由于存放不当或者贮存时间过久，会产生氧化而分解成甘油和游离脂肪酸等有害物质，这对油脂品质将是一种无形的损害，而食用油在氧化中将产生氢过氧化物，其含量的多少则用过氧化值来表征。因此，测定食用油的过氧化值对检测食用油的氧化程度有着很重要的作用。

汪旭升等人利用偏最小二乘法建立校正模型，对大豆油过氧化值进行了很好的预测。在过氧化值为 0 的油样中随机混合三苯磷和过氧化物定量作用生成三苯氧化磷，结果表明，该方法较准确地测定了大豆油过氧化值。

王立琦利用近红外光谱分析基于与 BP 神经网络算法对大豆油的酸价进行测定。验证结果预测标准差为 0.1387%，成功地对大豆油的酸价建立了定量模型。

8. 光谱法

(1) 紫外可见分光光度法

王耀等人基于花生油、调和油、葵花籽油和劣质食用油的紫外可见吸收光谱的吸收峰来鉴定各种油脂。该方法具有仪器简单、检测速度快、费用低廉的特点；但灵敏度低，当掺杂含量低时，吸收峰很弱甚至消失，目前尚不能准确地进行定性定量分析。

(2) 荧光分光光度法

刘薇等人研究发现，漂水油中含有洗洁精的主要成分十二烷基苯磺酸钠，实验表明污染物十二烷基磺酸钠在 $\lambda_{ex}/\lambda_{em} = 230\text{nm}/290\text{nm}$ 处有特征荧光峰，且所含十二烷基苯磺酸钠的

浓度与其荧光强度呈线性关系，回归方程为 $Y = 21.553X + 191.27$ ， $R = 0.9941$ ，具有较好的相关性。从而可以根据油脂中荧光的位置和强度鉴别食用油脂中是否掺杂潲水油。

(3) 原子吸收光谱法

王利等人用加热萃取方法将劣质食用油中的 Na^+ 迅速萃取出来，结合原子吸收分光光度法测定 Na 元素的含量，并通过测定合格植物油中的 Na^+ 进行对比实验，作为鉴别劣质食用油的方法。

(4) 近红外光谱法

1.2.3 近红外光谱技术在面粉品质检测领域的研究进展

小麦粉作为我国居民日常生活中不可或缺的主食原料和加工部分食品的基础原料，具有其他粮食作物不可替代的优势，我国每年小麦产量大约为 1 亿 t，占全国粮食总产量的 23% 左右，小麦粉品质的好坏直接影响面粉制品的质量，也直接关系到人们的身体健康，在这个日益注重产品质量、食品安全和身体健康的社会里，人们更加注重小麦粉的品质。

我国小麦粉品质分析基本上是由食用品质检验、物理品质检验及化学品质检验三部分组成的。其中，食用品质包括熟食品质、判断其气味、有害残留物测定、口味等；物理品质包括加工精度、面筋质、粗细度、磁性金属物等；化学品质包括灰分、脂肪酸值、含砂量、水分等。其中，水分、灰分及面筋的含量是影响小麦粉品质的重要因素，也是工厂日常检测的主要工作，三项指标需要实时检测，同时也对指导实时生产起到了重要作用。

目前小麦粉检测方法是以前化学原理为主流的检测方法，以下为小麦粉检测中水分、灰分、面筋三种重要指标的国标检测方法：

1. 水分

小麦粉的水分是指在高温下烘干面粉，所损失的水分占试样的百分比含量。小麦粉的水分直接影响产品的白度及存储情况，水分超过标准时，面粉不宜存放，很容易结块、生虫甚至霉变。所以，要根据不同的天气和季节条件来控制小麦粉的水分。

GB 5497—1985《粮食、油料检验 水分测定法》中规定的常用于小麦粉测定水分的方法有两种：105℃ 衡重法和定温定时烘干法，两种方法均存在操作复杂、所需的检测时间长等问题。此外还有快速水分检测仪检测法、隧道式烘箱法等。测水分所用的仪器用具主要有分析天平、电烘箱、干燥器、铝盒、快速水分检测仪等。

2. 灰分

小麦粉的灰分是指小麦粉经过高温灼烧后遗留下来的残渣即各种矿物质元素的氧化物占面粉的百分比含量。它是衡量小麦粉纯度的重要指标，我国特一粉的灰分含量在 0.75% 以下，面包用粉在 0.6% 以下，标准粉在 1.2% 以下，饺子、面条用粉在 0.55% 以下。

小麦粉的灰分含量可以通过间接的方法来衡量，如通过出粉率的高低、粉色深浅等。准确的方法是进行灰分测定，通常是将小麦粉放在指定高温的电炉中灼烧，燃烧后所剩下的灰烬的含量占样品量的百分比即灰分含量。其检测常用的方法是 550℃ 灼烧法和 850℃ 高温定时法。仪器用具有坩埚、干燥器、高温炉等。通过测定小麦粉灰分可鉴别小麦的加工精度，可以鉴别小麦品种，还可以反映小麦粉的营养价值，并可以进行掺假检验等。

3. 面筋

小麦粉的面筋是经过加水揉制成面团后，在水中揉洗，淀粉和麸皮微粒呈悬浮状态分离

近红外光谱技术在食品品质检测方法中的研究

出来，它水溶性和溶于稀 NaCl 溶液的蛋白质等物质被洗去，最后剩余的有弹性和黏弹性的不溶于水的胶状物质即面筋，用百分比表示。小麦粉的面筋含有丰富的蛋白质，其主要由麦谷蛋白和麦胶蛋白组成，还含有少量的糖分、淀粉、脂肪和其他蛋白质。小麦粉的面筋测定方法有手工洗涤法、仪器设备洗涤法和化学测定法。使用水洗方法测定小麦粉的面筋含量时，有许多因素影响水洗面筋的收率及质量：一是水洗前面团放置时间的长短；二是小麦粉的种类、数量和所用加水量等；三是水的种类。手洗法较费时，并且实验结果会因人而异；机洗法（面筋仪）则采用规范化的标准方法，相对快速且结果准确性较高，目前被广泛用于小麦粉面筋含量的测定来及时了解小麦粉的蛋白品质特性并掌控小麦的存储品质，该方法的主要仪器用具有面筋仪、玻璃棒、天平等。面筋的性质和含量是小麦粉品质优劣的重要指标，也是决定小麦粉用途的重要依据。

目前，小麦粉的品质检测采用的上述传统实验室测定法，虽然能够达到检测的要求，但是这些方法存在检测时间较长、操作复杂及人为因素影响较大等问题，比如小麦粉的水分、灰分定量分析的测定至少需要 3 ~ 4h，即使检测人员全力以赴，每日的检验也只能做 1 ~ 2 次，而面筋的检验不仅耗时而且受人为因素的影响较大，这对保证产品质量的稳定性是远远不够的，特别是生产自动化高度发展的今天，面粉厂的配粉工艺要求品质研发部及时提供品质检验结果，以便及时采取措施，调整小麦粉的生产工艺和搭配，减少不合格产品的生产。

针对目前小麦粉品质检测方法的种种弊端和实际生产的需要，研究一种简便、快速、准确、无污染、无损的检测方法是小麦粉品质检测的重要发展方向。

近年来，近红外光谱技术用于小麦及小麦粉品质检测的研究报道较多，并取得了成功的应用。

彭玉魁等人用近红外分析技术对小麦的组分含量进行了比较测试，结果表明用近红外分析技术测得小麦样品的水分、粗纤维、粗蛋白、赖氨酸含量与常规分析法之间的相关系数较高，均达到了相近的水平，说明能够利用近红外光谱技术进行小麦粉的品质分析。刘继明等人探讨了近红外分析仪在面粉厂的重要应用，可以测定小麦及面粉的水分、灰分及粒度含量等。Feng 等人通过对面包老化特性和货架寿命的研究表明，利用近红外交叉验证来测定货架寿命的结果比质构仪（Texture Profile Analyzer, TPA）的测定值达到更好的效果。

Wesley 等人运用近红外光的漫反射技术对小麦粉的相关蛋白进行了测定，结果得到了较高的相关系数和较低的偏差，进一步证明了近红外光谱技术具有较强的应用性，值得一提的是，澳大利亚 Black 和 Panozzo 利用可见光 - 近红外光漫反射技术测定小麦的水分、蛋白质、面团黄度、出粉率、吸水率、延展性、硬度、最大抗延阻力和峰值黏度 9 项指标，只有最大抗延阻力和延展性的相关程度较低，其他的各项指标与常规分析结果相比均具有较好的相关性。

邓益锋和张志霞分别用近红外分析方法与传统分析方法测定小麦粉中粗蛋白和粗灰分，并对两种方法进行比较，结果表明，两者的 R^2 分别为 0.986 和 0.991，相对误差均小于 5%，这说明可以用近红外光谱分析技术来分析小麦粉粗蛋白和粗灰分的含量。

高居荣利用国标化学法和 DA7200 近红外仪分别对 20 个小麦品种的面筋含量、吸水率、蛋白质含量、面团形成时间和稳定时间进行了检测。结果表明近红外光谱技术与国标法具有良好的重现性和相关性，从而证明可以把 DA7200 近红外仪应用于在小麦品质分析、优质小

麦选育和种质资源评价中。

陈锋、何中虎等人利用近红外的透射光谱对自全国各地 426 份小麦样品的水分、面筋、硬度、蛋白质等含量进行了测定,指出化学值与近红外光谱之间具有较好的相关性,如蛋白质、水分等指标校正集和预测集决定系数分别为 0.96、0.97 和 0.97、0.96,可知误差较小,这说明近红外光谱能够用来检测分析小麦的蛋白质和水分含量。

1.2.4 近红外光谱技术在淀粉品质检测领域的应用

淀粉是以谷类、薯类、豆类为原料,不经过任何化学方法处理,也不改变淀粉内在的物理和化学特性加工而成的。它是日常生活中必不可少的作料之一。常见的产品主要有玉米淀粉、马铃薯淀粉、红薯淀粉和绿豆淀粉等,不同种类的淀粉价格差别较大,有的相差高达 10 倍以上,但是不同种类淀粉颗粒的宏观外观和普通物化指标差别不明显,无法辨认。由于缺乏相应的食用淀粉鉴别检验技术标准,国内淀粉市场严格监管很难执行。一些不法商家为追求利润,常把廉价的、大量的玉米淀粉包装成马铃薯淀粉或绿豆淀粉来销售,从而提高销售价格,从中牟取暴利。为保护合法生产经营者和消费者的利益,有必要进行淀粉种类及淀粉掺杂的鉴别。传统的感官评定方法一方面需要经验,另一方面其检测结果受主观因素影响较大,准确度难以保证。

近红外光谱技术以其快速、多组分的检测优势在淀粉营养和安全品质检测上有极大的应用潜力。

1.3 基于近红外光谱的数据处理分析方法概述

20 世纪 50 年代以前,由于当时理论与技术水平还无法将近红外谱区的信息充分提取出来,近红外光谱的研究只能限于实验室中,无法实现实际应用。50 年代中后期,美国的 Norris 等人开始研究近红外光谱漫反射技术测定谷物、饲料和蔬菜的品质,通过大量的实验研究,从而掀起了该领域应用上的一个小高潮,但这些都是基于传统的光谱定量分析方法。当步入 60 年代中后期时,随着中红外光谱技术的不断发展以及在化学结构分析中的巨大作用,人们似乎渐渐地忘记了近红外光谱技术昔日的光芒和无限的潜力,在此后的近 20 年里,除了传统领域中的应用外,近红外光谱技术被人们束之高阁,因此该技术沉睡了数十年。

近 20 年来,由于光学技术、电子技术与计算机硬件和软件技术的发展,使分析化学中弱信号处理的理论和技术越来越成熟,可用来解决近红外光谱吸收弱的困难;现代数学、统计学在分析中的应用,化学计量学的发展以及计算机的广泛应用,使多组分分析中多元信息处理的理论与技术得到发展,它们被用来解决近红外光谱信息提取和背景干扰方面的困难。

化学计量学作为近红外光谱分析技术的一个重要组成部分是建立定量分析模型的基础。化学计量学综合了统计学、数学和计算机等手段与工具,是进行选择或设计最优化的测量方法,通过在计算机上实现统计学与数学的方法,来优化化学测量的过程,并且最大程度地从测量数据(信号)中提取有用信息。近红外光谱的谱峰比较宽而且样品的各成分吸收峰的重叠较为严重,化学计量方法能够针对微弱的化学信息进行有效地提取和分析,对复杂混合物的定量或者定性分析是一种有力工具,可提高检测的准确度和计算速度。化学计量学分析方法在近红外分析技术中占有非常重要的地位,正是由于计算机和化学计量学的应用,才使

近红外光谱技术在食品品质检测方法中的研究

得近红外分析成为一门实用性的分析技术。

在近红外光谱检测中化学计量学主要应用到以下两个方面：

1) 预处理方法的研究。进行预处理的目的是减弱甚至消除非目标因素对近红外光谱的影响，净化近红外光谱谱图信息，从而为建立准确、稳定的定量分析模型奠定基础。其中一方面是异常样品的处理，另一方面是特征谱区提取即波长优选方法的研究。建模前进行波长优选，可以提取近红外光谱的有效特征信息，从而在很大程度上简化模型并提高模型的预测能力。

2) 建模方法的研究。所测样品的近红外光谱图包含了样品的结构与组成信息，而样品的结构与组成又与去品质参数相关。因此，样品的谱图与品质参数之间存在内在的联系。可以通过选取数学方法对两者进行关联，从而来确定两者间定性或定量的关系。建立模型后，只需测量未知样品的近红外光谱，就可以根据所建立的校正模型来预测样品的品质参数。

使用化学计量学方法的目的在于建立稳定、可靠的定性或定量分析模型，以对近红外光谱进行快速、实时的定性、定量分析。目前，对化学计量学方法与近红外光谱的结合应用，国内外都连续有文献报道。

1.3.1 光谱数据预处理方法概述

光谱预处理在近红外光谱分析技术中是相当重要的。近红外光谱分析预处理可以针对特定的样品体系，剔除异常样品、消除光谱噪声、筛选数据变量、优化光谱范围、净化谱图信息、减弱以至于消除各种非目标因素对光谱的影响（见图 1-1），为建立近红外光谱校正模型和预测未知样品组分浓度或性质奠定基础。

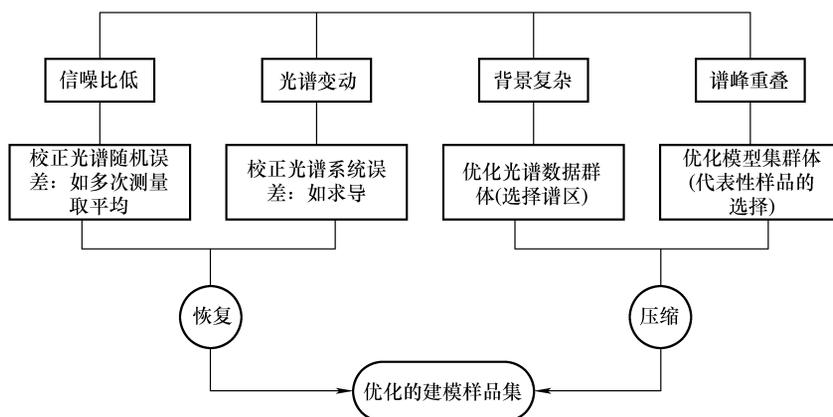


图 1-1 近红外光谱信息预处理

光谱的预处理研究主要包括两个方面：一是噪声和其他不规则影响因素的滤除，如消除随机噪声、样品背景干扰、光程变化、测样器件等因素引起的光谱差异对校正结果产生的影响。近红外光谱定量分析的过程中，分析结果的可靠性很大程度上取决于原始数据的准确性，即获得的样品近红外光谱及其化学值的可靠性。样品的原始数据，即样品的近红外光谱图和化学值的相关性，其直接影响模型的预测能力，而异常样品的干扰是影响分析模型的重要因素，因此异常样品的判别与处理是提高模型预测能力的一个重要步骤。产生异常样品的

原因如下：①测量仪器、测量方法及环境等客观因素的影响；②技术人员主观因素的作用；③样品的复杂性、多样性。所以判断异常数据是近红外光谱检测过程中首要解决的问题，建立定量分析模型时，样品中常常混有异常样品，需要在训练模型之前，把异常样品剔除。奇异点的存在会影响整体数据的分析，从而使校正模型的准确性降低，因此剔除奇异点对校正模型的建立和未知样品预测结果的准确性至关重要。

目前对样品的前期预处理的主要方法有主成分分析法、马氏距离法、杠杆值法、半数重采样法数据规范化处理、光程校正、基线校正、平滑消噪、导数消噪等。二是谱区信息的优化，即近红外光谱波长选择，对反应样品信息突出的光谱区域进行选择，筛选出最有效的光谱区域，提高运算效率。因此，在建模前，对原始光谱进行预处理是非常必要的，用以过滤噪声、提高信噪比、消除光程变化等。常用的滤波方法有积分法、累加/平均法、平滑；背景扣除及信号分离法有差谱法和导数法等；基线扣除和校正方法有线性漂移校正、附加散射校正（MSC），数据统计过程还采用中心化（Mean Center）和归一化处理（如 SNV）等。

1.3.2 校正模型方法概述

近红外光谱分析是通过将样品光谱和其质量参数进行关联，建立校正模型，然后通过校正模型预测样品的组成和性质的方法，使用数学方法对两者进行关联，可确立这两者间的定量或定性关系，即校正模型。建立模型后，只要测量未知样品的近红外光谱，根据校正模型和样品的近红外光谱就可以预测样品的质量参数。

近红外光谱分析技术主要应用在两个方面：一是近红外光谱的定性判别分析；二是近红外光谱的定量分析。相应地，在近红外光谱分析中的建模方法研究也分为两类：一是近红外定性分析方法研究；二是近红外定量校正方法研究。

定量校正方法多采用各种多元线性校正方法：逐步多元线性回归（Step Multi Linear Regression, SMLR）、主成分回归（Principal Component Regression, PCR）、偏最小二乘回归（Partial Least - Squares Regression, PLSR）、稳健偏最小二乘回归（Robust Partial Least - Squares Regression, RPLSR）等方法。在建立近红外光谱定量校正模型时，通常使用逐步回归分析（SRA）、主成分回归（PCR）及偏最小二乘法（PLS）等建立的样品化学标准值与样品光谱参数之间的模型都是线性的，线性算法对具有内在线性关系的体系的校正能力是很强的，但对于存在非线性关系的体系就会出现预测误差远大于校验误差的现象。在近红外吸收光谱中，光谱参数与样品组分含量的化学标准值之间具有一定的非线性，特别是当样品的含量范围较大时，其非线性较明显。非线性校正算法主要有神经网络（Artificial Neural Networks, ANN）、非线性 PLS、局部权重回归（Local Weighted Regression, LWR）等方法。ANN 方法非线性逼近能力很强，在近红外光谱分析方面的应用已越来越广泛，但 ANN 方法也存在一些局限性，如训练速度很慢，容易陷入局部极小，也存在过拟合现象及当输入变量间存在共线性时预测能力较差，同时，关于人工神经网络的类型、结构、训练参数、训练样品数目、学习过程等参数的选择大多还是凭经验，还有待于进一步研究。

模式识别最早出现在 20 世纪 20 年代，60 年代末被引入化学领域。化学模式识别是利用统计学、信号处理、数学算法等工具，从化学量测数据推理出物质类的本质属性，进而对物质进行识别和归类的一门技术。

化学模式识别方法包括聚类分析（Cluster Analysis）、判别分析（Discriminant Analy-

近红外光谱技术在食品品质检测方法中的研究

sis)、特征投影显示 (Latent Projection) 等方法。按照有没有训练集可以划分为无监督的模式识别方法 (Unsupervised Pattern Recognition) 和有监督的模式识别方法 (Supervised Pattern Recognition)。聚类分析属于无监督的模式识别方法, 判别分析属于有监督的模式识别方法, 特征投影显示可以是无监督的, 也可以是有监督的。

本书针对果蔬的农药残留、植物油品质、小麦粉、淀粉的近红外光谱技术检测进行研究, 如何将几种模式识别技术, 即 PLS、模糊模式识别、ANN、SVM 及近红外光谱分析有机地结合, 以实现定量和定性分析, 旨在找到一种更为有效的近红外光谱的建模方法, 为上述农产品的近红外光谱分析提供新思路和新技巧。

1.4 完成的相关科学研究项目概况

北京工商大学计算机与信息工程学院在自动控制、虚拟测试、智能检测与控制技术, 特别是在智能信息处理、测量建模方法、软测量模型校正、软测量工程化实施技术方面已取得了多项创新性、实用性成果。已完成多项“九五”、“十五”、“十一五”国家重点科技攻关项目, 曾荣获 2004 年国家科技进步二等奖、2002 年度北京市科学技术二等奖项等。

自 2005 年创建“光谱技术与品质检测”研究室, 进行专门的基于光谱技术的测量技术及信息处理的研究, 课题组由具有丰富的科研工作经验和实验动手能力的中、青年骨干教师组成, 特别是对该课题的相关知识已经有丰富的积累, 参与了中华人民共和国科学技术部“十五”重点科技攻关项目“建立与完善我国部分农产品近红外光谱库”; “稻谷品质快速检测技术与装置的研制与开发”、“特高含水井油层测量方法的研究”、国际科技合作项目等多项科研项目。

近几年, 课题组主持完成了北京市自然科学基金“基于光谱技术的农药残留量快速智能检测方法研究 (4073031)”、北京市优秀人才基金“基于近红外技术的食用植物油品质检测方法研究 (20081D0500300130)”、“基于多光谱信息技术的面粉品质快速检测方法研究 (2012D005003000007)”、北京市科技创新平台“多光谱信息融合食品无损检测技术创新平台 (PXM_2012_014213_000023)”。在研北京市教育委员会科技计划重点项目“基于多光谱技术的食品安全快速无损检测方法研究 (KZ201310011012)”、北京市自然科学基金项目“基于拉曼及近红外光谱技术的果蔬类农药残留量快速检测研究 (4142012)”、“基于 NIR-Raman 光谱技术的食用植物油综合品质快速诊断机理研究 (4132008)”等多项相关科研课题, 在“光谱学与光谱分析”、“食品科学”、“农业机械学报”、“食品工业科技”等国际、国内学术期刊、会议论文集等公开发表各类学术论文 100 余篇。奠定了较扎实的研究基础, 建立了国内较先进的近红外光谱测量系统实验平台, 并且与德国布鲁克公司的开放实验室、国内的相关研究所有过多次合作, 因此具有先进的实验条件。在基于近红外光谱技术的农药残留量、食用油、淀粉、面粉的智能快速检测方法方面有较深的研究, 取得了一定的结论, 积累了较丰富的研究经验, 写本书愿与同行在该项技术的探索中共同交流, 深入探讨。

1.5 本书的主要内容

近红外光谱分析技术作为一种快速、无损、高效的分析技术近些年来受到越来越广泛的

关注,该技术已在农业、食品、环境科学药物、医学、石化等多种行业中得到了广泛的应用。近红外光谱分析技术在农产品品质检测方面具有很大潜力。近红外光谱分析技术是一种“多快好省”的较为理想的现代分析技术。随着近红外光谱技术软硬件系统的不断改进和提高,近红外光谱技术的应用领域不断扩大,在很多领域得到应用。

本书主要介绍近红外光谱分析技术在果蔬类农药残留量的检测、植物油品质、小麦粉、淀粉的品质检测中的应用。本书的研究内容如下:

第1章首先综述了近红外光谱法的分析特点以及发展应用过程。主要介绍了近红外光谱具有样品无需预处理、多组分同时测定、分析速度快、无污染、低成本、重现性好,效率高,对测试人员的要求不高,易培训、推广等优点。同时介绍了近红外光谱的发展历程,同时近年来,随着近红外光谱仪器的不断更新,光谱预处理、定量校正模式及模型传递等近红外光谱分析技术进一步发展起来。为了更好地了解近红外光谱的应用,本章还重点介绍了近红外光谱技术在食品安全检测领域的研究进展,如近红外光谱技术在农药残留物检测、食用油品质检测、面粉品质检测领域的研究进展。主要介绍了近红外光谱技术在这三大领域的应用现状以及产生的问题,并简单介绍了在这三大领域检测中所采用的预处理以及建模方法。其次介绍了基于近红外光谱的数据处理分析方法概述,主要包括光谱数据预处理方法以及多元校正方法介绍。最后介绍了北京工商大学计算机与信息工程学院在软测量、虚拟测试、智能检测与控制技术,特别是在软测量建模方法、软测量模型校正、软测量工程化实施技术方面取得的多项创新性、实用性成果。

第2章主要介绍了近红外光谱分析技术的理论基础和基本流程。在基本的理论基础介绍中提到了近红外光谱分析检测技术的两大类,即透射技术和反射技术,并通过两个流程图详细地介绍了近红外光谱分析的基本流程,主要包括样品的预处理、模型的建立和预测。然后针对流程中所提到的预处理程序和模型建立,重点介绍了几种常用的预处理方法(如数据规范化处理、光程校正、基线校正、导数法、平滑处理、主成分分析法、马氏距离、杠杆值法、蒙特卡洛交叉验证法、傅里叶变换、正交信号校正、标准正态变量变换等)和模型建立方法(如多元线性回归、PCR、PLS、ANN BP法、聚类分析法、SVM法等)。为了更好地进行预处理和建模,在第2章中详细介绍了农药残留检测的前处理技术,包括超临界流体萃取、固相微萃取、微波辅助萃取技术、凝胶渗透色谱、基质固相分散萃取、固相萃取等。并提出了近红外光谱定标模型的评价指标,主要有测定系数(R^2)、相关系数(R)、交叉验证预测均方差(RMSPCV)、预测标准差(SEP)。最后介绍了近红外光谱仪器的发展和应用,重点介绍了VERTEX 70傅里叶红外光谱仪、MPA多功能傅里叶变换近红外光谱仪、紫外分光光度计的性能指标和优缺点。

第3章主要介绍了农药残留量的快速检测方法的研究。首先针对蔬菜中有机磷农药残留量的现状提出了关于食品检测方面的几项国内、国外标准。然后重点介绍了国内、外蔬菜中农药残留检测技术,主要包括色谱法、生物传感器法、生物化学测定法、发光菌法及免疫分析法等。对这几种方法的分类和检测原理以及应用现状进行了详细的介绍。最后提出了本书进行农药残留量检测方法时所采用的预处理和建模方法,以及使用的仪器。

此章主要探索运用近红外光谱定量分析模拟液体中的微量有机磷农药毒死蜱,通过设计不同的实验背景,逐步探究近红外光谱在测定溶液中微量毒死蜱的应用情况。实验设计按照毒死蜱在溶液中的浓度由高到低,依次为3.13~10mg/kg、3.03~10mg/kg、0.005~

近红外光谱技术在食品品质检测方法中的研究

0.1mg/kg依次分析。背景由简单到复杂,即由甲醇、水、毒死蜱的溶液到甲醇、水、维生素C、蔗糖和毒死蜱的溶液。针对农药残留量痕量分析研究,分别从近红外光谱技术方向进行了综合分析研究,以毒死蜱农药标准物质和炔螨特农药标准物质作为制作样品的研究对象,从农药水溶液的初步研究开始,基于近红外光谱的微量农药残留检测分析,通过对毒死蜱和炔螨特光谱数据进行不同预处理方法优化后建立了PLS数学模型和农药溶液的ANN BP模型,通过对比不经任何优化方法处理的光谱数据模型和其他常用优化方法,确定了光谱数据优化对建立模型的重要贡献:降低样品数据噪声和其他干扰。将其应用于萝卜农残检测试验中,各模型都得到了良好的评价参数,且具有很好的相关性,证明了近红外光谱技术可用于微量农药残留量的初步分析。

第4章主要是探索研究基于近红外光谱分析技术的蔬菜中农药残留量检测方法并建立模型,最终采用PLS完成了模型建立及分析,阐述了应用近红外光谱分析法测定大白菜和菠菜的毒死蜱含量。配制后混合溶液样品,经过采样分辨率、采样次数及采样模式的选择,使用德国布鲁克仪器公司的VERTEX 70型傅里叶红外光谱仪采集混合溶液样品的近红外光谱;其次采用10种不同的光谱预处理方法对样品的近红外光谱进行处理;最后经过波长优化、主成分数目的确定,建立了基于偏最小二乘法的近红外校正模型。另外此章还将基于近红外光谱的微量农药残留检测分析技术应用于萝卜农残检测试验中,各模型都得到了良好的评价参数,且具有很好的相关性。

通过以上实验表明,近红外光谱分析技术对蔬菜中农药残留检测具有巨大的潜力和应用价值,对实际农药残留快速检测具有一定的借鉴意义。

第5章阐述了采用近红外光谱技术结合模式识别方法建立食用油种类、花生油和橄榄油掺伪定性识别模型的过程。实验结果表明,采用近红外光谱与SVM技术建立了性能较好的花生油和橄榄油掺伪定性鉴别模型,对实验中预测集和测试集的预测率均能达到100%;而采用近红外光谱与系统聚类法结合能有效地识别食用油种类。实验表明将近红外光谱-模式识别技术相结合作为一种快速无损鉴别食用油掺伪方法具有较强的可行性和实用性,此章工作的探索有力地推动了近红外光谱技术在食用油品质定性快速识别方面的应用进程。

第6章采用近红外光谱技术结合化学计量学方法建立了食用油脂肪酸含量、过氧化值和酸价的定量分析模型,并且探讨了光谱分辨率的设定对食用油近红外光谱模型性能的影响。实验结果表明,特征谱峰的选取结合适当的光谱预处理可以有效地提高近红外光谱模型的预测性能,使食用油定量分析模型的实用性得到进一步提高。此章工作的探索有力地推动了近红外光谱技术在食用油品质快速监管方面的应用进程。

第7章为了解决传统小麦粉检测时间长、操作复杂等难题,研究提出应用近红外光谱检测小麦粉品质的研究思路。为了排除异常样品的干扰,首先利用马氏距离法和蒙特卡洛采样法剔除异常样品,然后采用K-S法对样品集进行划分,为了提高模型的精度,尝试利用常规区间PLS、联合区间PLS、向后区间PLS和CARS算法进行特征谱区筛选从而对小麦粉的水分、灰分及面筋的定量分析模型进行优化。最后,引入非线性回归方法(SVM和ANN BP)对模型的稳定性及预测能力进行了比较分析。

第8章运用近红外光谱技术快速、无损地鉴别淀粉品质。采用聚类分析法和SVM的定性建模方法,对玉米淀粉和马铃薯淀粉进行了准确分类;优化建立淀粉中水分的含量近红外分析模型;同时此章利用近红外光谱对马铃薯淀粉中掺杂玉米淀粉及淀粉中掺杂滑石粉进行

了定量分析。本章内容表明,运用近红外光谱可以快速、准确地对淀粉品质进行定性和定量检测。

第9章对近红外光谱技术在食品品质快速检测方面的研究结果进行了总结,并对此技术的应用与下一步将开展的研究进行了展望。

参 考 文 献

- [1] 褚小立,许育鹏,田高友.近红外光谱解析实用指南[M].北京:化学工业出版社,2009.
- [2] 陆婉珍,袁洪福,徐广通,等.现代近红外光谱分析技术[M].北京:中国石化出版社,2000.
- [3] 严衍禄.近红外光谱分析基础与应用[M].北京:中国轻工业出版社,2005.
- [4] M Blanco, I Vilarroya. NIR spectroscopy: a rapid-response analytical tool [J]. Trends in analytical chemistry, 2002, 21 (4): 240-250.
- [5] 徐广通,袁洪福,陆婉珍.现代近红外光谱技术及应用进展[J].光谱学与光谱分析,2000,20(2): 134-142.
- [6] 王豪.模型转移技术在白酒品质近红外光谱测定中的应用[D].镇江:江苏大学,2006.
- [7] 王丽杰.快速检测牛奶成分的近红外光谱测量方法及系统研究[D].哈尔滨:哈尔滨理工大学,2006.
- [8] 周向阳,林纯忠,胡样娜,等.近红外光谱法(NIR)快速诊断蔬菜中有机磷农药残留[J].食品科学,2004,25(5): 151-154.
- [9] 夏俊芳,李培武,李小昱.不同预处理对近红外光谱检测检测脐橙VC含量的影响[J].农业机械学报,2007,28(6): 107-111.
- [10] 孙竹生,杨连学.农产品农药残留分析技术技术进展[J].潍坊高等职业教育,2006,2(1): 43-45.
- [11] 朱俊杰,陆利霞,熊晓辉.食品中农残检测及预处理新技术进展[J].食品研究与开发,2008,29(2): 129-132.
- [12] 何永红,高志贤.基于生物学原理的农药残留检测技术研究进展[J].卫生研究,2004,33(1): 112-114.
- [13] 唐勤学,陶小林,离司.有机磷农药残留检测技术的研究进展[J].化工时刊,2008,22(9): 68-72.
- [14] Wang Jingjing, Hou Dexin, Wu Beilei, et al. An effective classification procedure for plastic food packaging based on NIRS [J]. Applied Mechanics and Materials, 2011 (88-89): 399-403.
- [15] Chen Quansheng, Cai Jianrong, Wan Xinmin, et al. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy [J]. LWT - Food Science and Technology, 2011 (44): 2053-2058.
- [16] Nicoletta Sinelli, Ernestina Casiraghi, Stefania Barzaghi, et al. Near infrared (NIR) spectroscopy as a tool for monitoring blueberry osmo-air dehydration process [J]. Food Research International, 2011 (44): 1427-1433.
- [17] Fodor Mariettal, Woller Agnes, Turza Sandor, et al. Development of a rapid, non-destructive method for egg content determination in dry pasta using FT-NIR technique [J]. Journal of Food Engineering, 2011, 107 (2): 195-199.
- [18] Hattori Yusuke1, Otsuka Makoto. NIR spectroscopic study of the dissolution process in pharmaceutical tablets [J]. Vibrational Spectroscopy, 2011, 57 (2): 275-281.
- [19] Zhou Jun, Tsai Yi-Ting1, Weng Hong, et al. Real time monitoring of biomaterial-mediated inflamma-

近红外光谱技术在食品品质检测方法中的研究

- tory responses via macrophage - targeting NIR nanoprobe [J]. *Biomaterials*, 2011, 32 (35): 9383 - 9390.
- [20] 谈爱玲, 毕卫红, 赵勇. 基于稀疏非负矩阵分解和支持向量机的海洋溢油近红外光谱鉴别分析 [J]. *光谱学与光谱分析*, 2011, 31 (5): 1250 - 1253.
- [21] 岳永德, 花日茂, 张承祥. 茶叶农药残留与控制 [M]. 北京: 中国农业出版社, 2001.
- [22] Williams P C, Norris K H. *Near - Infrared Technology in the Agricultural and food industries* [M]. Minneapolis: Cereal Chem, 1987.
- [23] 张学博, 冯艳春, 胡昌勤. 近红外多元校正模型传递的进展 [J]. *药物分析杂志*, 2009, 29 (8): 1390 - 1399.
- [24] 陆婉珍. 近红外光谱仪器 [M]. 北京: 化学工业出版社, 2010.
- [25] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用 [J]. *化学进展*, 2004, 16 (4): 528 - 542.
- [26] 尼珍, 胡昌勤, 冯芳. 近红外光谱分析中光谱预处理方法的作用及其发展 [J]. *药物分析杂志*, 2008, 28 (5): 824 - 829.
- [27] 刘翠玲, 隋淑霞, 吴静珠, 等. 基于近红外光谱的微量成分 (毒死蜱) 检测技术研究 [J]. *农机化研究*, 2008 (6): 156 - 167.
- [28] 刘翠玲, 隋淑霞, 孙晓荣, 等. 近红外光谱技术用于菠菜中毒死蜱残留的定量分析研究 [J]. *食品科学*. 2008, 29 (07): 356 - 358.
- [29] 王惠文. 偏最小二乘回归方法及其应用 [M]. 北京: 国防工业出版社, 1999.
- [30] Svante Wold, Michael Sjöström, Lennart Eriksson. PLS - regression: a basic tool of chemometrics [J]. *Chemometrics and Intelligent Laboratory Systems* 2001, 58: 109 - 130.
- [31] 褚小立, 袁洪福, 陆婉珍. 近年来我国近红外光谱分析技术的研究进展 [J]. *分析仪器*, 2006, 2: 1 - 10.
- [32] 赵环环, 严衍禄. 噪声对近红外光谱分析的影响及相应的数学处理方法 [J]. *光谱学与光谱分析*, 2006, 26 (5): 842 - 845.
- [33] 董守龙, 任芊, 黄友之. 近红外光谱分析技术的发展和应 [J]. *化工生产与技术*, 2004, 11 (6): 44 - 46.
- [34] 熊艳梅, 王冬, 闵顺耕. 短波近红外模拟测定农药乳油中的有效成分含量 [J]. *现代科学仪器*, 2010, 4: 85 - 86.
- [35] 沈飞, 闫战科, 叶尊忠, 等. 近红外光谱分析技术在辛硫磷农药残留应用检测中的应用 [J]. *光谱学与光谱分析*, 2009, 29 (9): 2421 - 2424.
- [36] 刘丽丽. 基于超高压预处理的蔬菜农药残留近红外检测技术的初步研究 [D]. 北京: 吉利大学, 2009.
- [37] 史永刚, 冯新沪, 李子存. 化学计量学在近红外光谱定性分析中的应用 [J]. *光谱实验室*, 1999, 16 (3): 237 - 239.
- [38] 刘涛, 孙旭东, 刘燕德. 农作物品质的近红外光谱无损检测研究进展 [J]. *食品与机械*, 2010, 26 (3): 161 - 166.
- [39] 褚小立, 许育鹏, 陆婉珍. 用于近红外光谱分析的化学计量学方法研究与应用进展 [J]. *分析化学*, 2008, 36 (5): 702 - 709.
- [40] 王迪, 刘潇威, 刘岩, 等. 蔬菜、水果中 12 种限量有机磷农药残留量测定方法 [J]. *分析实验室*, 2010, 29 (7): 37 - 41.
- [41] 李伟, 肖爱平, 冷鹏. 近红外光谱技术及其在农作物中的应用 [J]. *中国农学通报*, 2009, 25 (3): 56 - 59.

- [42] Osborne B G, Fearn T. Applications of near - infrared spectroscopy in food analysis [M]. UK: Longman Scientific & Technical, 1986.
- [43] 熊艳梅, 段云青, 王冬, 等. 近红外光谱技术快速测定农药有效成分的研究 [J]. 光谱学与光谱分析, 2010, 30 (6): 1488 - 1492.
- [44] 刘翠玲, 隋淑霞, 吴静珠, 等. 近红外光谱技术检测溶液中毒死蜱含量试验 [J]. 农业机械学报, 2009, 40 (1): 129 - 131.
- [45] 吴泽鑫, 李小昱, 王为, 等. 基于近红外光谱的番茄农药残留无损检测研究方法研究 [J]. 湖北农业科学, 2010, 49 (4): 961 - 963.
- [46] 代芬, 张昆, 洪添胜, 等. 龙眼表面农药残留的无损检测研究—基于近红外光谱分析 [J]. 农机化研究, 2010, 10: 111 - 114.
- [47] 王耀, 尹平河, 梁芳慧, 等. 紫外可见分光光度法鉴别掺兑泔水油的花生油 [J]. 分析试验室, 2006, 25 (3): 92 - 94.
- [48] 刘薇, 尹平河, 赵铃. 荧光法测定十二烷基苯磺酸钠鉴别泔水油的研究 [J]. 中国油脂, 2005, 30 (5): 24 - 25.
- [49] 王利, 陈晓枫, 王静梅. 原子吸收法光谱法在食品安全检测中的应用—测定钠盐鉴别泔水油 [J]. 现代科学仪器, 2010 (4): 110 - 114.
- [50] 康春雨. 一种用于快速鉴别食用植物油中是否含有地沟油的仪器: 中国, 201020179881.5 [P]. 2010 - 11 - 24.
- [51] 刘珉恺. 一种电容式地沟油快速检测方法: 中国, 201010540165 [P]. 2011 - 04 - 27.
- [52] 刘珉恺. 一种光吸收检测地沟油的方法: 中国, 201010540170 [P]. 2011 - 05 - 04.
- [53] 刘珉恺. 一种光散射地沟油快速检测方法: 中国, 201010540228 [P]. 2011 - 04 - 20.
- [54] 刘珉恺. 手持式地沟油快速检测装置: 中国, 201020602035 [P]. 2011 - 07 - 06.
- [55] 尹建洪, 罗立新. 食用油安全问题分析及对策, 中国食品安全高峰论坛论文集 [C]. 广州: 广东省微生物研究所, 2010: 88 - 90.
- [56] 郭涛, 杜蕾蕾, 万辉, 等. 高效液相色谱法测胆固醇含量鉴别地沟油 [J]. 食品科学, 2009, 30 (22): 286 - 289.
- [57] 张蕊, 祖丽亚, 樊铁, 等. 测定胆固醇含量鉴别地沟油的研究 [J]. 中国油脂, 2006, 31 (5): 65 - 67.
- [58] 黄道平, 彭进, 谢燕湘, 等. 泔水油鉴别检测方法研究 [J]. 中国卫生检验杂志, 2006, 16 (2): 151 - 153.
- [59] 刘薇, 尹平河, 赵玲, 等. 电导率法快速鉴别泔水 [J]. 城市环境与城市生态, 2004, 17 (3): 4 - 6.
- [60] 胡小泓, 刘志金, 郑雪玉, 等. 应用电导率检测泔水油方法的研 [J]. 食品科学, 2007, 28 (11): 482 - 484.
- [61] 徐冲, 刘远洋. 电导率法鉴别食用油中泔水油试验条件的选择 [J]. 农产品加工, 2010 (4): 39 - 42.
- [62] 彭进, 黄道平, 刘吉星, 等. 电导率的测定在鉴别泔水油中的应用研究 [J]. 实用预防医学, 2007, 14 (3): 878 - 879.
- [63] 朱锐, 王督, 杨小京, 等. 电导率测定在鉴别食用植物油掺伪应用研究 [J]. 粮食油脂, 2008 (11): 42 - 43.
- [64] 杨培强, 王欣, 刘宝林, 等. 低场核磁共振鉴别食用油品质的方法: 中国, 201010268825 [P]. 2011 - 02 - 16.
- [65] 刘建学. 实用近红外光谱技术 [M]. 北京: 科学出版社, 2008.

近红外光谱技术在食品品质检测方法中的研究

- [66] Dong Wang, Zhihong Ma, Ping Han, et al. Research on Detection of Lime in Wheat Flour by NIR Micro-Imaging [J]. *Sensor Letters*, 2012, 10 (1/2): 252-257.
- [67] Smita Tripathi, Mishra H N. A rapid FT-NIR method for estimation of aflatoxin B1 in red chili powder [J]. *Food Control*, 2009, 20 (9): 840-846.
- [68] Sirinnapa Saran Wong, Sumio Kawano. Interpretation of near infrared calibration structure for determining the total aerobic bacteria count in raw milk; interaction between bacterial metabolites and water absorptions [J]. *Journal of Near Infrared Spectroscopy*, 2008, 16 (6): 497-504.
- [69] Yongnian Ni, Minghua Mei, Serge Kokot, et al. Analysis of complex, processed substances with the use of NIR spectroscopy and chemo metrics: Classification and prediction of properties-The potato crisps example [J]. *Chemo metrics and Intelligent Laboratory Systems*, 2011, 105 (2): 147-156.
- [70] Vikram J Pansare, Shahram Hejazi, William J Faenza, et al. Review of Long-Wavelength Optical and NIR Imaging Materials: Contrast Agents, Fluorophores, and Multifunctional Nano Carriers [J]. *Chemistry of Materials*, 2012, 24 (5): 812-827.
- [71] 黄艳, 王锡昌. 近红外光谱分析在食品检测中的最新进展 [J]. *食品研究与开发*, 2007, 28 (7): 137-140.
- [72] Paz, Patricia, Sanchez, et al. Evaluating NIR instruments for quantitative and qualitative assessment of intact apple quality [J]. *Journal of the Science of Food and Agriculture*, 2009, 89 (5): 781-790.
- [73] Guchhait A, Rath A K, Pal A J, et al. To make polymer: Quantum dot hybrid solar cells NIR-active by increasing diameter of PbSnanoparticles [J]. *Solar Energy Materials and Solar Cells*, 2011, 95 (2): 651-656.
- [74] Xudong Sun, Xiaoling Dong, Lijun Cai, et al. Visible-NIR Spectroscopy and Least Square Support Vector Machines Regression for Determination of Vitamin C of Mandarin Fruit [J]. *Sensor Letters*, 2012, 10 (1/2): 506-510.
- [75] Jiang H, Liu G, Xiao X, et al. Monitoring of solid-state fermentation of wheat straw in a pilot scale using FT-NIR spectroscopy and support vector datadescription [J]. *Microchemical Journal*, 2012, 102: 68-74.
- [76] Salgo A, Gergely S. Analysis of wheat grain development using NIR spectroscopy [J]. *Journal of Cereal Science*, 2012, 56 (1): 31-38.
- [77] Nisha Shetty, Rene Gislum, Anne Mette Dahl Jensen, et al. Development of NIR calibration models to assess year-to-year variation in total non-structural carbohydrates in grasses using PLSR [J]. *Chemo-metrics and Intelligent Laboratory Systems*, 2012, 111 (1): 34-38.
- [78] Mermelstein N H. A brief look at grain and seed analysis [J]. *Food Technology*, 2010, 64 (11): 65-69.
- [79] 吴静珠, 王一鸣, 张小超, 等. 基于近红外的奶粉品质检测技术研究 [J]. *光谱学与光谱分析*, 2007, 27 (9): 1735-1738.
- [80] Knischka R, Lehmann U, Stadler U, et al. Novel approaches in NIR curing technology [J]. *Progress in Organic Coatings*, 2009, 64 (2/3): 171-174.
- [81] Minasny, Budiman, McBratney, et al. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon [J]. *Geoderma: An International Journal of Soil Science*, 2011, 167/168: 118-124.
- [82] Xie S F, Xiang B R, Yu L Y, et al. Tailoring noise frequency spectrum to improve NIR determinations [J]. *Talanta*, 2009, 80 (2): 895-902.
- [83] 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 [J]. *西北农*

- 业学报, 1996, 5 (3): 31 - 34.
- [84] 刘继明. Perton8600 近红外仪在面粉厂的重要应用 [J]. 粮油食品科技, 2000, 8 (5): 28 - 29.
- [85] Feng X, Dowell E F, Xiuzhi S S. Comparison of near infrared reflectance spectroscopy and texture analyzer for measuring wheat bread changes in storage [J]. CerealChemistry, 2003, 80 (1): 25 - 29.
- [86] Wesley I J, Larroque O. Measurement of gliadin and glutenin content of flour by NIR spectroscopy [J]. Journal of Cereal Science, 2001, 34: 125 - 133.
- [87] Black C K, Panozzo J F. Whole grain quality evaluation in wheat based on near infrared spectroscopy [J], An Australian and Chinese Collaboration on Wheat Quality: Conference Proceedings, 2001, 32: 123 - 125.
- [88] 邓益锋, 张志霞. 近红外线光谱分析与传统方法检测通用面粉中粗蛋白和粗灰分的比较 [J]. 畜牧与兽医, 2005, 37 (7): 20 - 21.
- [89] 高居荣, 韩秀兰, 孙彩玲, 等. DA7200 红外仪在小麦品质分析中的应用研究 [J]. 实验室科学, 2009, (1): 173 - 176.
- [90] 陈锋, 何中虎, 崔党群, 等. 利用近红外透射光谱技术测定小麦品质性状的研究 [J]. 麦类作物学报, 2003, 23 (3): 1 - 4.
- [91] 吴玉田, 柴逸峰, 等. 褶合变换近红外光谱分析系统的研究 [C]. 全国第一届近红外光谱学会议论文集, 2006: 240 - 249.
- [92] 吴海龙, 韩清娟, 等. 化学计量学与近红外光谱相结合研究的若干新进展 [J]. 全国第一届近红外光谱学会议论文集, 2006: 54 - 57.
- [93] 倪永年. 化学计量学在分析化学中的应用 [M]. 北京: 科学出版社, 2004.
- [94] 叶静. 利用近红外光谱分析技术检测绿茶有效成分的研究 [D]. 镇江: 江苏大学. 2009.
- [95] 白英奎, 申铨国, 等. 红外和神经网络无损分析 VC 银翘片中二组分含量 [J]. 激光与红外, 2004, 5 (34): 354 - 356.
- [96] 林敏, 吕进, 等. 基于神经网络与近红外光谱的玉米成分检测 [J]. 红外技术, 2004, 26: 78 - 80.
- [97] 吴新生, 等. 主成分 - BP 算法在近红外光谱法纸浆卡伯值 [J]. 分析测试学报, 2005, 6 (25): 10 - 12.
- [98] 刘木清, 等. 聚类算法用于中药材的近红外光谱分析 [J]. 光谱学与光谱分析, 2007, 10 (27): 1985 - 1988.
- [99] 岳喜军, 曹晓建, 等. 近红外光谱和聚类分析法快速鉴别周围神经束性质的实验研究 [J]. 陕西医学杂志, 2007, 3 (36): 278 - 279.
- [100] 李彦周, 闵顺耕. 近红外化学模式识别方法及应用研究 [J]. 光谱学与光谱分析, 2007, 7 (27): 1299 - 1303.
- [101] 侯瑞锋, 黄岚, 等. 用近红外漫反射光谱检测肉品新鲜度的初步研究 [J]. 光谱学与光谱分析, 2006, 12 (26): 2193 - 2196.

近红外光谱分析技术基础

2.1 近红外光谱分析的理论基础

近红外光是指波长介于可见光与中红外光之间的电磁波，其波长范围为 780 ~ 2526nm，波数范围为 12500 ~ 4000 cm^{-1} 。根据近红外光谱仪器的有效波长范围，近红外光谱区又分为短波近红外光谱区和长波近红外光谱区两个波段，如图 2-1 所示。

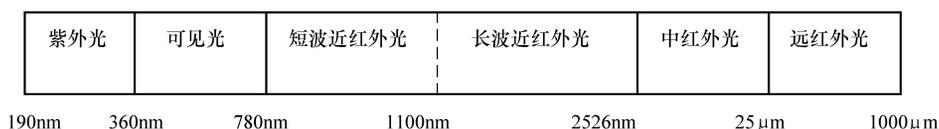


图 2-1 近红外光谱波长范围

近红外光谱主要是有机分子含氢基团（C-H、O-H、N-H、P-H、S-H 等）的伸缩、弯曲、振动的各级倍频和合频吸收光谱，表 2-1 是主要基团合频与各级倍频吸收带的相似位置。不同基团（如甲基、苯环、亚甲基等）或同一基团在不同化学环境中的近红外吸收波长与强度都有很明显差别。本谱区几乎包含了有机物中所有含氢基团的信息，蕴含着分子的结构、组成状态等信息，信息量极为丰富，从而为近红外光谱定量及定性分析有机样品提供了可能。

表 2-1 主要含氢基团合频与各级倍频吸收带的相似位置

单 位	cm^{-1}				nm			
	C-H	N-H	O-H	H ₂ O	C-H	N-H	O-H	H ₂ O
基 团								
合 频	4250	4650	5000	5155	2350	2150	2000	1940
一级倍频	5800	6670	7000	6940	1720	1500	1430	1440
二级倍频	6500	9520	10500	10420	1180	1050	950	960
三级倍频	11100	12500	13500	13300	900	800	740	750
四级倍频	13300				750			

近红外光谱区的吸收主要是分子或原子振动基频在 2000 cm^{-1} 以上的倍频与合频吸收，主要包括 C-H、N-H 和 O-H 等含氢基团的倍频和合频吸收带。因为不同的基团或同一基团在不同的特定化学环境下近红外光所吸收的波长与强度有明显差别，所以近红外光谱具有丰富的结构和物理化学信息。如食用油脂结构中的 -C=C- 和 -C-H- 键的伸缩振动第一组合频在 5800 ~ 5600 cm^{-1} 和 4800 ~ 4400 cm^{-1} 两个光谱范围的吸收最为明显，即反映了油脂脂肪酸的含量差异，这些吸收特征为近红外光谱定量及定性分析样品的物理化学性质提供了前提。

近红外光谱分析技术是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。近红外光谱的常规分析技术有透射光谱和漫反射光谱两大类。透射光谱法是把样品放于光源与检测器之间，用以测定均匀透明的溶液或固体样品。反射光谱分析法是把样品放于光源与检测器的同侧，且用以测定固体或半固体样品。

在透射技术中，若样品是透明真溶液，透射光强与样品中各组分的浓度符合如下朗伯 - 比尔定律：

$$A = -\lg \frac{I}{I_0} = \varepsilon bc \quad (2-1)$$

式中， A 是吸光度； I_0 是波长为 λ 的平行、均匀入射光束强度； I 是透过溶液后的光束强度； ε 是待测组分的摩尔吸光系数； b 是光程； c 是待测物质组分的浓度。

若样品是浑浊的，那么由于光散射的原因，光经过样品时的路径不确定，不符合朗伯 - 比尔定律，将其称为漫透射分析测定法。

物体对光的反射分为规则反射和漫反射。应用漫反射进行的分析测定称为漫反射分析测定法，该方法在现代近红外光谱分析中起着非常重要的作用。当样品发生漫反射时，光线将包含有关样品中的大量的物质成分信息，即不同成分在特定波长下吸收了不同能量的光，对样品成分的分析提供了方便。但漫反射并不遵循朗伯 - 比尔定律，其遵循 Kubelka - Munk 方程，如下：

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

式中， A 是吸光度； R_∞ 是样品厚度无穷大时的相对漫反射率； K 是吸收系数； S 是散射系数。

透射光谱一般用于均匀透明的真溶液或固体样品，仪器测量得到的吸光度与光程及样品的浓度之间符合朗伯 - 比尔定律，如图 2-2 所示。

即一种物质的溶液对光的吸收程度，取决于在光通过的路径中能吸收光子的物质微粒数目和长短，显然，物质的浓度越大，吸收的光就越多，光通过的液体厚度越大，吸收的光也越多。朗伯 - 比尔定律说明了被测物质的吸光程度与吸收介质的厚度有关，和溶液中吸收物质的浓度有关，与入射光的强度无关。

一束强度为 I_0 的光辐射通过一层溶液时，一部分能量 I 被吸收，一部分被透射 I_t ，还有一部分被吸收池散射 (I'_s) 和反射 (I_r)。如果样品浑浊或发荧光，则透射光中还包括样品的散射光 (I_s) 和荧光 (I_f)。如果样品为透明介质，不发射荧光，由于样品测量总是在类似的比色池中相对参比溶液而言的，所以 I_r 、 I_s 、 I_f 和 I'_s 均可被忽略。据朗伯 - 比尔定律有如下的公式：

$$A = klc = \lg \frac{I_0}{I} \quad (2-3)$$

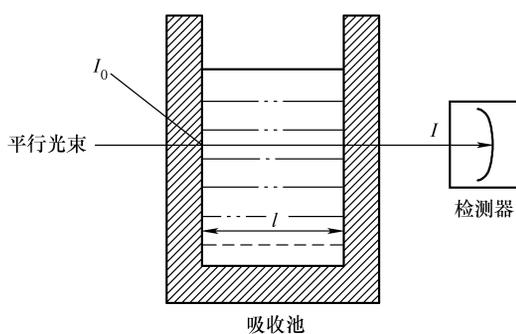


图 2-2 朗伯 - 比尔定律

$$A = -\lg T = \lg \frac{I}{I_0} \quad (2-4)$$

式中, A 是吸光度; k 是吸光系数, 它随不同的介质、环境、波长而异, 它可以表示一个物质的吸收特征, 在数值上等于单位摩尔浓度在单位光程中所测得的溶液的吸光度; c 是溶液浓度; l 是介质厚度, 在这里为常数; T 是透光度, 表示透射辐射能与入射辐射能的比。

近红外光谱漫反射分析技术在近红外光谱分析中占有非常重要的地位, 近红外光谱早期在农业及农副产品分析中的应用大多是通过漫反射完成的。漫反射是从光源出来的光进入样品内部经过多次反射、折射、衍射及吸收后返回样品表面的光。因此, 漫反射光是分析光与样品内部分子发生作用后的光, 负载了样品的结构和组成信息。由于漫反射过程中样品与光的作用有多种形式, 除样品的组成外, 其粒径大小及分布和形状均对漫反射光的强度有一定的影响, 因此漫发射光的强度不符合朗伯 - 比尔定律。漫反射光谱定量的基本原理遵守 Kubelka - Munk 方程:

$$A = \lg (1/R_\infty) = -\lg [1 + K/S - \sqrt{(K/S)^2 + 2(K/S)}] \quad (2-5)$$

式中, R_∞ 为样品厚度无穷大时的相对漫反射率; A 为吸光度; K 为吸收系数; S 为散射系数。

所谓漫反射是光源发出的光投射到所需测量样品后, 经过样品表面及其内部组分多次的反射、折射、衍射及吸收后最终携带样品的组成和结构的相关信息返回到样品表面的过程。一般在测量以下样品的近红外光谱时采用漫反射方式, 分别为不透明、半固体以及固体。

近红外光是具有波粒二重性的电磁波, 其光子能量可以用光子来表示。根据量子力学的原理, 光子能量可表示为 $E_p = h\nu$ (其中 h 为普朗克常数, ν 为光的频率)。根据简正模式理论, 当红外光从光源射出后, 会与样品中的分子相碰撞, 若光粒子与样品分子的振动频率不同, 则光会通过样品而不发生任何状态的变化, 如果光粒子与样品分子的振动频率一样, 则会发生能量方面的变化, 光粒子的能量将会被样品分子吸收而影响样品分子的振幅, 这就是样品的红外活性, 正是有了红外活性, 所以可以从测量样品的出射光或者反射光的衰减和吸收中来得得到样品所包含物质的信息, 从光谱解析的角度来看, 其信息源主要为 O-H、N-H、C-H 等含 H 基团的分子内部原子间振动耦合反映的倍频 (基频的倍数) 与合频 (两个以上基频和)。

不同的含氢基团在近红外光谱区域上的吸收强度与位置均不相同, 此外, 相邻基团的性质、基团的数量及氢键的存在也会影响光谱谱峰的强度和位置。因此, 与化学结构相关的某些性质就可以根据朗伯 - 比尔定律和近红外光谱图来确定, 如确定成分浓度等。

从以上的分析可知, 近红外光谱区含有大量的含氢基团的吸收峰。由于某些无机物的存在导致化学环境的改变影响了有机物的定量测定, 近红外光谱也能间接测定食品中的无机物含量。通过对食品样品进行近红外光谱分析, 并结合多元校正方法建立定量分析模型, 可以得到相应的模型, 从而定量预测食品的有效成分。食品的品质与其所含的各种化学成分直接相关, 如水、蛋白质等的含量及比例。由此可知, 通过对近红外光谱数据进行分析, 并结合相应的化学计量学方法, 可以建立食品品质参数的定量分析模型。

2.2 近红外光谱分析流程

近红外光谱分析技术是随着计算技术的发展而发展的, 近红外光谱技术发展的同时, 欧

洲的 Wold 教授和美国的 Kowalski 教授于 20 世纪 70 年代开始了化学计量学 (Khemometrics) 的研究, 他们使用数学、统计学和计算机科学知识, 避免了因近红外吸收及漫反射光谱谱带都很宽而得不到解析的困难。

光谱化学计量学方法研究在现代近红外光谱分析中占有非常重要的地位。稳定、可靠的近红外光谱分析仪器与功能全面的化学计量学软件相结合标志着现代近红外光谱技术进入了崭新的一页。

近红外光谱分析技术是通过建立校正模型从而对未知样品的组分含量进行定量分析的一种间接的分析技术。通过以下几步来完成分析模型的建立: ①测量样品的近红外光谱, 并选择有代表性的样品作为校正集; ②采用国标法或已被公认的方法对样品的化学成分进行测定; ③根据已测量的近红外光谱图和化学值同时运用化学计量学方法来建立定量分析模型; ④根据定量分析模型来对预测样品的组分含量进行测定。

近红外光谱分析过程分为两步: 校正模型的建立和未知样品的预测, 流程如图 2-3 和图 2-4 所示。

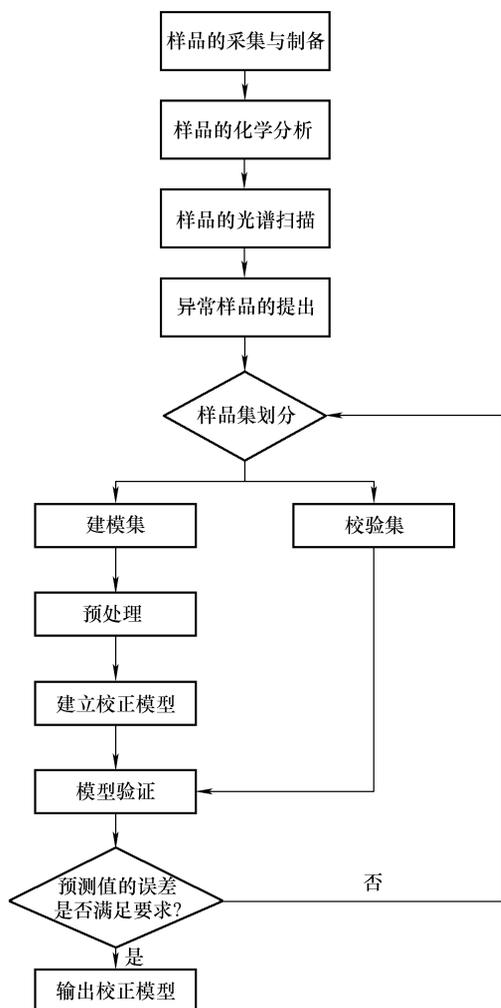


图 2-3 建立近红外光谱校正模型流程

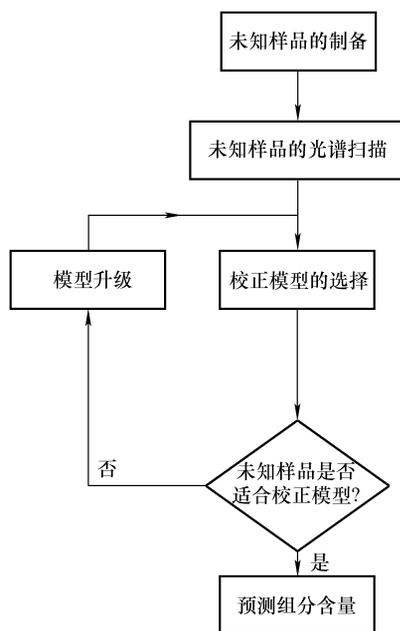


图 2-4 未知样品组分浓度预测流程

1. 样品选择

建立分析模型前，收集有代表性的样品十分重要。样品要能涵盖待测样品成分的全部范围，并且在各个范围内的样品数目分布最好均匀一致，从而保证以后分析模型测量的精度较高，避免出现共线性现象。收集样品代表的范围越宽，则所建立模型的适用性越好，但是分析模型的精度会有所下降；反之，如果所收集的样品范围比较窄，则其适用范围将有所缩小，但建立的分析模型精度会相对提高。通常，对于一个单组分的系统，校正集需要大量样品组成。

2. 标准比对

近红外光谱分析技术是二级分析技术，需要有标准进行比对，标准精度越好，则近红外光谱分析模型越好。因此对于收集到的样品，最好使用国际或者国内标准的测试方法进行化学分析，并且尽量减少人为误差。虽然化学分析试验中不可完全避免地存在一些误差，但是可以通过选用精密度较高的仪器或增加测量的重复次数等方法，来减小实验室误差，从而降低因化学分析而引入的模型误差。

3. 光谱扫描及预处理

采集样品光谱时，首先，测量时间应与化学分析时间一致，避免因时间间隔过长而引起样品内成分的变化。其次，被测样品要能够涵盖该批样品的特性，特别是对装样量较少的近红外光谱仪而言。此外，仪器的稳定性常常受外界环境变化的影响，因此最好选在不同的时间进行测量，这样可以将环境因素造成的光谱数据变化融入以后的定标模型之中。

完成光谱数据采集之后，还要进行适当的光谱预处理。这是为了达到如下目的：针对特定的样品体系，剔除异常样品，消除光谱噪声，筛选数据变量，优化光谱范围，净化谱图信息，减弱以至于消除各种非目标因素对光谱的影响，为建立近红外光谱校正模型和预测未知样品组分浓度或性质奠定基础。

4. 建立校正模型

用标准分析方法测得的数据和经过预处理的光谱数据利用数学方法建立校正模型，用来预测位置样品的成分浓度等。常见的校正方法包括多元线性回归（MLR）、主成分回归（PCR）、人工神经网络（ANN）和偏最小二乘法（PLS）。

5. 校正模型的验证

校正模型建立以后，必须对其进行验证，以确定模型的可靠性。通常用校正模型对预测集样品进行预测，如果预测结果与标准值接近，表明模型预测效果较好，可以用于样品分析；反之，说明校正模型的精度较差，需进一步修正。

6. 样品分析

可靠的校正模型建立之后，便可用于样品分析。分析样品之前，应注意校正模型的测量范围及适用性。浓度范围是指测量样品的成分浓度是否在该校正模型的测量范围之内，如果不在，即为超常样品。那么需要将该样品的近红外光谱收集到原来的校正集样品中，并对该样品作标准的化学分析，将原来的校正模型进行修正，以适用更大范围的样品。

2.3 近红外光谱分析预处理方法

下面介绍几种常用的光谱预处理方法。

2.3.1 主成分分析法

主成分分析 (Principle Component Analysis, PCA) 法可以把多个指标化为几个综合性指标, 被广泛应用到各领域。在实际问题中, 研究多指标 (变量) 问题时, 不同指标之间具有一定相关性, 由于多指标及指标之间有相关性, 从而增加了分析问题的复杂性。PCA 就是将原来具有一定相关性的指标, 组合成的一组新的相互无关的综合指标来代替原来的指标。对近红外光谱进行主成分分析, 能够达到对光谱矩阵进行降维的目的, 不会过多地丢失光谱所含有信息, 同时还能够减少原始近红外光谱的冗余信息。

设样品观测数据矩阵为 \mathbf{X} , 则 PCA 具体计算步骤如下:

$$\text{其中, } \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

该矩阵为 n 个样品 p 个波长上的光谱信息阵, 记 \mathbf{X} 的样品均值、样品协差矩阵以及相关阵分别为 $\bar{\mathbf{X}}$ 、 \mathbf{S} 和 \mathbf{R} , 通过由样品信息阵 \mathbf{X} 出发, 构造原各波长点光谱信息变量的线性组合但不相关, 且具有最大的样品方差, 就是构造样品主成分。

第一样品主成分 = 线性组合 $l'_1 X_j$, 满足 $\|l_1\| = 1$, 且 $l'_1 X_j$ 的样品方差最大 (其中, X_j 为在新坐标系下的第 j 个观测值, l 为其系数);

第二样品主成分 = 线性组合 $l'_2 X_j$, 满足 $\|l_2\| = 1$ 及 $l'_2 X_j$ 与 $l'_1 X_j$ 样品协方差为 0, 且使 $l'_2 X_j$ 的样品方差最大。

一般地, 第 i 个样品主成分 = 线性组合 $l'_i X_j$, 满足 $\|l_i\| = 1$ 得 $l'_i X_j$ 与 $l'_k X_j$ ($k < i$) 的样品协方差为 0, 且使 $l'_i X_j$ 的样品方差最大。

1) 对原始数据进行标准化处理:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{var}(x_j)}} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p) \quad (2-6)$$

式中, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$; $\text{var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ ($j = 1, 2, \dots, p$)。

2) 计算样品相关系数矩阵:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (2-7)$$

为了方便, 假定原始光谱数据标准化后仍用 \mathbf{X} 表示, 则经过标准化处理的数据的相关系数为

$$r_{ij} = \frac{1}{n-1} \sum_{i=1}^n x_{i1} x_{ij} \quad (i, j = 1, 2, \dots, p) \quad (2-8)$$

3) 求 \mathbf{R} 的特征值 ($\lambda_1, \lambda_2, \dots, \lambda_p$) 及其相应的特征矢量 $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{ip})$, $i = 1, 2, \dots, p$ 。

近红外光谱技术在食品品质检测方法中的研究

4) 选择重要主成分, 并列出其主成分表达式。

通过主成分分析能够得到 p 个主成分, 但是, 由于各个主成分包含的信息量随着方差的递减而递减, 所以在实际分析过程中, 一般不选取 p 个主成分, 而是通过分析各个主成分的累计贡献率大小来选取前 m 个主成分, 贡献率是指某个特征值占全部特征值的比重, 如下:

$$\text{贡献率} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (2-9)$$

5) 取前 m 个较大的特征值对应特征矢量来构成变换矩阵 T^T :

$$T^T = (a_1, a_2, \dots, a_m) \quad m < n \quad (2-10)$$

6) 通过

$$Y = TX \quad (2-11)$$

计算前 m 个主成分, 从而达到降低维数的目的, 依据成分得分的数据, 则可以进行进一步的统计分析。

2.3.2 马氏距离法

马氏距离法可以用来剔除异常样品, 计算原理如下所示。

当 C_x 满秩时, 样品与光谱 X_c 重心之间的马氏距离定义为

$$MD^2(i) = (x_i - \bar{x}) C_x^{-1} (x_i - \bar{x})^T \quad (2-12)$$

其中, \bar{x} 表示平均光谱; C_x 是 X 的方差-协方差矩阵。

马氏距离法是度量样品与中心之间距离的一种有效途径, 已被应用到光谱的聚类分析中。计算马氏距离不仅照顾了相同波长点光谱信号的变化(方差), 还考虑了不同波长点光谱间的变化(协方差)。

设光谱数据为 $A(n \times k)$ 矩阵, n 为样品数, k 为选出的主成分数。马氏距离算法分为以下步骤:

1) 计算 n 个样品的平均光谱:

$$\bar{A}_j = \sum_{i=1}^m A_{ij} / n \quad (2-13)$$

式中, A 为光谱矩阵; j 为波长序号; n 为样品数; \bar{A} 为平均光谱。

2) 把近红外光谱矩阵进行中心化处理:

$$A(i, :) = A(i, :) - \bar{A} \quad (2-14)$$

3) 计算得到马氏矩阵:

$$M = \left(\frac{A' \cdot A}{n-1} \right) \quad (2-15)$$

此时的波长数为 k , 所以 M 是 $k \times k$ 维矩阵。

4) 计算校正集各个样品到平均光谱的马氏距离:

$$D^2 = (A_i - \bar{A}) \cdot M^{-1} \cdot (A_i - \bar{A})' \quad (2-16)$$

5) 根据计算出的 n 个马氏距离设置一个阈值来检验异常样品的存在, 计算方法为

$$D_t = \bar{D} + e \cdot \sigma_D \quad (2-17)$$

式中, \bar{D} 为马氏距离平均值; e 为调整闭值范围参数, σ_D 为马氏距离标准差。

如果 $D_i \leq D_i$, 则样品 i 与样品的平均光谱很接近, 则称 i 为平均样品的邻近样品。令 N_i 为 i 的邻近的样品个数, N_i 值越大, 样品 i 的邻近的样品数越多, 在空间上越密集。可以设置不同的阈值范围参数 e , 来调节 N_i 值的大小, e 值越大, N_i 值越大, 阈值的范围越宽, 邻近样品空间上的密集度就会越高; 反之亦然。

2.3.3 杠杆值法

杠杆值体现样品对模型影响的重要程度, 所以位于被测组分浓度范围两端的样品具有较大的杠杆值, 往往认为杠杆值大于 $(2 \sim 3) p/n$ 的样品为奇异点。

当 \mathbf{X} 列满秩时, 帽子矩阵为 $\mathbf{h}_{ii} = \text{diag}(\mathbf{H}) = \mathbf{x}_i (\mathbf{X}_c^T \mathbf{X}_c)^{-1} \mathbf{x}_i^T$, 其对角线上的元素 \mathbf{h}_{ii} 就是第 i 个样品的杠杆值:

$$\mathbf{H} = \mathbf{X}_c (\mathbf{X}_c^T \mathbf{X}_c)^{-1} \mathbf{x}_i^T \quad (2-18)$$

杠杆值与马氏距离之间存在着线性关系:

$$\mathbf{h}_{ii} = \frac{1}{n-1} \mathbf{MD}^2(i) + \frac{1}{n} \quad (2-19)$$

因此马氏距离和杠杆值的效果一样。马氏距离虽然直观, 但是如果样品中存在杂质且其在所选波长区间有吸收, 则将影响判断的效果。另外, 如果 \mathbf{X}_c 存在共线, 则不能求逆矩阵, 因此所选波长点不宜超过 20 个, 否则易产生过拟合。故可以先对光谱矩阵进行 PCA, 然后用其得分参与计算以便于克服共线性。

2.3.4 半数重采样法

半数重采样 (RHM) 法是指从原始光谱矩阵 \mathbf{X} 中随机选择一半样品, 组成新矩阵 $\mathbf{X}(i)$ (表示第 i 个采样矩阵), 计算 $\mathbf{X}(i)$ 的均值 $\mathbf{m}(i)$ 和方差 $\mathbf{s}(i)$, 对原始光谱矩阵 \mathbf{X} 进行标准化处理。第 i 个采样矩阵中的样品的矢量长度 $\mathbf{l}(i)$ 定义为

$$\mathbf{l}(i) = \sqrt{\sum_{k=1}^P \left(\frac{\mathbf{X}_k - \mathbf{m}_k(i)}{\mathbf{s}_k(i)} \right)^2} \quad (2-20)$$

对全部样品的矢量长度进行排序, 其中距离最大的 5% 的样品分布进行奇异点剔除。

2.3.5 蒙特卡洛交叉验证法

蒙特卡洛交叉验证 (Monte Carlo Cross Validation, MCCV) 法是一种近期新提出的筛选异常样品的方法, 该方法基于预测误差对异常样品的敏感特性, 能够在一定程度上降低由掩蔽效应带来的风险, 有效检出光谱阵和性质阵方向的奇异点, 与传统方法相比具有较高的识别奇异样品的能力。

具体算法如下:

- 1) 利用 PCR 或 PLS 确定最佳主成分数;
- 2) 利用蒙特卡洛随机取样 (Monte Carlo sampling, MCS) 法随机选取部分样品作校正集, 剩余样品作预测集, 分别建立 PLS 回归模型;
- 3) 校正集通过第一步确定的主成分数建模, 预测集用来预测每一个样品的预测误差, 进行循环是为了保证每个样品均被预测到, 选择循环多次, 从而得到每个样品的预测误差

分布；

4) 计算每个样品的预测残差的均值 (MEAN) 和方差 (STD), 绘制样品的均值 - 方差分布图。

2.3.6 数据规范化处理

数据中心化: 将光谱矩阵 \mathbf{X} 数据中心化 (Mean - centering) 得 \mathbf{X}' (其均值为零):

$$\mathbf{X}' = \begin{bmatrix} x'_{11} & x'_{12} & \cdots & x'_{1p} \\ x'_{21} & x'_{22} & \cdots & x'_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x'_{n1} & x'_{n2} & \cdots & x'_{np} \end{bmatrix}$$

其中

$$x'_{ij} = x_{ij} - \bar{x}_i, \bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij} \quad (i = 1, 2, \dots, n) \quad (2-21)$$

仅需要变化信息, 也就是说在近红外光谱分析中并不关心光谱数据与浓度数据绝对值之间的直接关系, 而是寻找它们变化之间的相对关系使数据间在标度上具有可比性, 消除单位不同产生的影响。

数据标准化: 将光谱矩阵 \mathbf{X} 数据标准化 (Autoscaling) 得 \mathbf{X}'' (其均值为零, 方差为 1):

$$\mathbf{X}'' = \begin{bmatrix} x''_{11} & x''_{12} & \cdots & x''_{1p} \\ x''_{21} & x''_{22} & \cdots & x''_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x''_{n1} & x''_{n2} & \cdots & x''_{np} \end{bmatrix} \quad (2-22)$$

其中

$$x''_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i}, \bar{x}_i = \frac{1}{p} \sum_{j=1}^p x_{ij}, S_i = \sqrt{\frac{1}{p-1} \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2} \quad (i = 1, 2, \dots, n) \quad (2-23)$$

2.3.7 光程校正

光程校正的目的是消除光程的变化或样品的稀释所产生的影响。

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, 即使相同的样品, 多次测量的光谱也会出现差异, 此即光散射现象。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的光谱变化。散射的程度与光的波长、颗粒度和样品的折射指数有关, 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得尤为明显。

MSC 是由 Geladi 等人提出的, 目的是校正每个样品的散射并获得较“理想”的光谱。MSC 假定与波长有关的散射对光谱的贡献和对成分贡献是不同的, 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。因此每个样品在任意波长点下反射吸光度值与其平均光谱的相应吸光度呈近似线性关系, 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）大小反映样品反射效果而斜率大小则反映样品的均匀性。具体计算公式如下：

$$\text{平均光谱: } \bar{A}_j = \sum_{i=1}^n A_i / n \quad (2-24)$$

$$\text{线性回归: } a_i = m_i \bar{A}_j + b_i \quad (2-25)$$

$$\text{MSC 校正: } A_{i(\text{MSC})} = \frac{a_i - b_i}{m_i} \quad (2-26)$$

上述公式中， \mathbf{A} 是校正集光谱矩阵， A_i 为第 i 个样品的光谱， m_i 和 b_i 是第 i 个样品光谱 A_i 与平均光谱 \bar{A}_j 的线性回归的斜率与截距，都是列矢量。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与化学成分相关的信息。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。

2. 标准正态变量变换 (Standard Normalized Variate, SNV)

SNV 校正也是用来校正样品间因散射引起的光谱差异。但与 MSC 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

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (2-27)$$

消除趋势 (Detrending) 方法如下：对 SNV 校正后的光谱矩阵 $\mathbf{A}_{n \times p}^{\text{SNV}}$ 的每一条样品光谱，拟合出一个二次多项式，然后将 $\mathbf{A}_{n \times p}^{\text{SNV}}$ 的每一条样品光谱减去对应的二次多项式，即可去除趋势。消除趋势一般与 SNV 一起使用，很少单独使用。

3. 归一化 (normalization) 处理

在建立近红外光谱定标分析模型时，需将光谱的特征与待测样品的性质或结构特征相互关联。正是基于如上的特点，在建立 NIR 定量或定性模型前，往往需要采用一些数据增强算法来降低直至消除一些冗余信息，从而在降低样品间相关性的同时，也能够增大样品之间的差异，进而达到提高模型的重现性和预测能力的效果。主要采取归一化 (normalization) 算法等。该方法的基本思路是计算每个样品的平均值，将光谱数据减去这些数值，使所有样品的有关数据都分布在零点两侧，充分反映变化信息，这样，温度或人为操作等客观因素所带来的变化对光谱数据造成的影响可以有效的被去除，并且对以后回归运算有一定的简化作用。具体算法是

$$y'_i = \frac{y_i - \bar{y}}{\sqrt{\sum (y_i - \bar{y})^2}} \quad (2-28)$$

式中， y_i 表示原始光谱； \bar{y} 表示光谱的均值。

这样处理后的光谱数据充分反映了信息变化，所有数据都分布在零点两侧，简化并稳定了以后的回归运算。

2.3.8 基线校正

基线校正的目的是扣除仪器背景或漂移对信号的影响。自动 (Automatic) 基线校正正

一种典型的基线校正方法。

具体算法是：对光谱矩阵 A 在每个波长点平均吸光度和浓度矩阵以最小二乘法拟合出一条直线，然后每条样品光谱减去该直线，即完成对基线的校正。

2.3.9 导数法

当样品的背景颜色等因素导致所测光谱出现位移或漂移时，在近红外光谱分析中常用导数法来消除基线漂移或平缓背景干扰的影响，从而提供比原始光谱分辨率清晰度更高的光谱轮廓变化。因此导数法能用来消除基线漂移及其他背景干扰的影响，分辨重叠峰，提供比原光谱更高的分辨率和更清晰的光谱轮廓变化。近红外光谱测量的光谱是样品振动的 3 级和 4 级倍频吸收，所测光谱常受样品的背景颜色和其他因素的影响，出现明显位移和漂移。微分处理是净化谱图比较常用的方法，可根据需要进行一阶或二阶微分处理。但在微分处理时，应当注意的是光谱的导数变换会将噪声引入光谱，此外根据微分的级数不同，对微分窗口的大小也应做出合适的选择。

对光谱求导一般有两种方法：直接差分法和 Savitzky - Golay 求导法。对于分辨率高、波长采样点多的光谱，应用直接差分法得到的光谱与实际相差不大；但对于稀疏波长采样点的光谱，此方法所求的导数则与实际存有很大的误差，这时通常采用 Savitzky - Golay 卷积求导法计算。导数方法可分辨出重叠峰，快速有效地提高了分辨率和灵敏度。但同时它也会引入噪声信息，从而得不到较高的信噪比。在实际的应用中，差分宽度的选择非常重要：若差分宽度过小，噪声会很大，这样就会影响所建模型的质量；若差分宽度过大，平滑过度，从而丢失大量的有效信息。由经验可知，差分宽度不应大于光谱吸收峰半峰宽的 1.5 倍。

2.3.10 平滑处理

经近红外光谱仪器扫描后，样品的光谱数据中在含有有用信息的同时，也包含了一定量的噪声信息，使得信噪比降低。信号平滑算法的本质是对光谱进行低通滤波，去掉高频成分，有效地保留低频信息。因此，平滑算法是一种有效的降低噪声的办法。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法。主要是去掉高频噪声对信号的干扰。常用的信号平滑方法是窗口移动平均法，由 Savitsky 和 Golay 提出，称为 SG 平滑法。平滑处理涉及处理窗口的大小（或点数），点数多时可使信噪比提高，但同时也会导致信号的失真。点数少时，噪声得不到有效滤除。窗口大小的不同，平滑效果不同。若窗口宽度太大，在做求均值计算时，会忽略掉一些有用信息，使得光谱信号产生一定的失真。若窗口宽度太小，噪声不能被有效去除。因此，便出现 Savitzky - Golay 卷积平滑法这种改进算法，该方法并不是进行简单的求平均计算，而是使用加权平均法通过多项式对移动窗口内的数据进行多项式最小二乘拟合，更加突出了中心点的中心作用。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。因此必须考虑仪器的具体情况，适当选择平滑窗口的大小。Savitzky - Golay 卷积平滑法已经成为最为广泛应用的光谱去噪算法。

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。

2.3.11 多元散射校正

多元散射校正 (MSC) 由 Martens 等人提出, 是基于一组样品的光谱阵进行运算。该方法的基本思想是假定散射系数在所有波长处都是相同的, 将化学物质的吸收信息与光谱中的散射光信号进行有效的分离。多元散射的校正过程如下:

首先, 计算所需校正光谱的平均光谱, 如式 (2-29) 所示:

$$\bar{A} = \frac{\sum_{i=1}^n A_i}{n} \quad (2-29)$$

其次, 对平均光谱作回归, 如式 (2-30) 所示:

$$A_i = m_i \bar{A} + b_i \quad (2-30)$$

最后, 对每一条光谱作多元散射校正, 如式 (2-31) 所示:

$$A_i(\text{校正后}) = \frac{(A_i - b_i)}{m_i} \quad (2-31)$$

多元散射校正可消除近红外漫反射光谱中由于样品的镜面反射及不均匀造成的噪声, 消除光谱的基线漂移现象及光谱的不重复性。但是, 由于多元散射校正假定了散射同波长和样品浓度的变化无关, 所以对于组分性质变化范围较大的样品处理效果并不明显, 甚至导致较大误差。

2.3.12 傅里叶变换

傅里叶变换 (Fourier Transform) 是一种信号处理技术, 有效地滤除高频噪声以及由仪器原因或基线偏移等因素引起的低频噪声, 该方法对光谱数据进行平滑、差值、滤波等运算增加光谱信噪比。傅里叶变换首先将数据进行快速傅里叶变换 (FFT), 在频率空间下与高斯窗函数进行运算, 然后进行反快速傅里叶变换 (IFFT) 计算, 得到经带通滤波处理后的光谱数据。带通滤波器的中心频率和带宽分别取决于高斯函数的均值与标准差, 并且应用数值优化方法来实现滤波的参数, 以获得最佳滤波效果。通过对光谱的平滑去噪, 数据压缩等一系列处理进一步提高模型对测量条件扰动的稳健性。该方法原理清晰, 实现方便, 大大缩短运算时间, 而且样品光谱质量得到有效改善, 校正模型的预测能力和适应能力也得到进一步的增强。

2.3.13 正交信号校正

利用正交信号校正 (Orthogonal Signal Correction) 算法, 对原始光谱数据进行预处理, 可以有效地取出光谱数据中所包含的各种干扰噪声信号。在应用实际的近红外光谱技术进行分析时, 部分系统误差或干扰噪声等与有效信息无关的信息常常会被引入在光谱中, 这时, 用偏最小二乘法建立的定标模型的前几个主因子数对应的光谱载荷经常不是有用的浓度阵信息, 而是与浓度阵无关的噪声信号。因而, 在建立定标模型前, 通过正交的数学运算, 除去浓度阵无关的噪声, 然后将经过数学运算处理后的光谱阵作为新的自变量矩阵, 再利用 PLS 方法建立校正模型。

只要保证除去的部分与预测值矩阵能够正交，则除去部分所含有的信息必然与光谱矩阵没有必然的关系。正是基于这种新颖的思想，S. Wold 等人在 1998 年提出正交信号校正方法。但是这种最初的算法的缺点是每次迭代计算均需要计算一遍 PLS，计算时间较长，且内置 PLS 算法的成分数很难确定。Sjoblom 又改进了该算法，其算法与 Wold 基本一致，只是计算次序略有差别。这种算法在迭代时不用反复计算 PLS，其主成分数也容易通过交叉验证的方法确定。但同样该方法的缺点是无法严格证明其剔除的信息与光谱矩阵正交。经过学者们长时间的不断探索和研究，陆续出现了逐渐完善的算法，如 Fearn 的类 PCA 算法、Andersson 的 DO 算法、Westerhuis 的 DOSC 算法、Trygg 的 0-PLS 算法、Feudale 的 POSC 算法等。这些改良后的预处理方法的基本原理是在建立定标分析模型前，将光谱阵与浓度阵正交，去掉光谱与浓度阵无关的冗余信号，再进行多元校正，从而使所建立的模型更加简化并且提高模型的稳定性和预测能力。

2.3.14 标准正态变量变换

SNV 也用来校正样品间因散射引起的光谱差异。但与 MSC 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该满足一定的分布（如正态分布），借助这一假设对每条光谱进行校正，而无需“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 ，实质是使原光谱数据标准正态化，即

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (2-32)$$

2.3.15 基于分段窗口的特征谱区筛选方法

1. 常规区间法

常规区间偏最小二乘波长筛选法 (iPLS) 是由 Lars Norgaard 在 2000 年提出的一种波长筛选法。该方法主要用于筛选偏最小二乘建模的波长区域，其算法步骤如下：

- 1) 对原始近红外光谱进行预处理；
- 2) 建立全光谱范围待测样品品质的偏最小二乘模型，即全局偏最小二乘模型；
- 3) 将整个近红外光谱区域划分为多个等宽的光谱子区间，设为 n 个；
- 4) 对每个光谱子区间建立 PLS 回归模型，即建立待测样品品质的“局部回归模型”，可以得到 n 个局部回归模型；
- 5) 用交互验证均方差 RMSECV 的值作为各个模型精度的衡量标准，分别比较各个局部模型和全光谱模型的精度，选取精度最高的子区间为入选区间。

2. 联合区间法

联合区间偏最小二乘法 (Synergy interval PLS, SiPLS) 建立在常规区间偏最小二乘法的基础之上，它是将在同一次区间划分中预测精度较高的几个子区间联合起来建立预测模型，称其为联合子区间法。实际应用表明，将几个精度较高局部模型所在的子区间联合在一起建立预测模型是可行的，但目前尚无方法来确定参加联合建模的子区间数目。

3. 向后区间法

向后阈值区间偏最小二乘法 (Backward interval PLS, BiPLST) 是基于 Lars Norgaard 的 iPLS 理论思想的改进，是一种“只出不进”的方法，其算法步骤如下：

- 1) 对原始近红外光谱进行预处理。
- 2) 建立全光谱范围待测样品品质的 PLS 模型, 即全局 PLS 模型。
- 3) 将整个近红外光谱区域划分为多个等宽的光谱子区间, 假设为 n 个。
- 4) 每次去掉 1 个光谱子区间, 将余下的 $(n-1)$ 个区间联合并进行 PLS 回归, 得到 n 个联合区间的 PLS 回归模型。
- 5) 运用交互验证均方差 RMSECV 的值作为各个模型的精度衡量标准, 比较各联合模型的精度, 去掉精度最高的联合模型所去掉的光谱子区间, 即第一去除区间。
- 6) 将剩余的 $(n-1)$ 个光谱子区间逐一去除子区间, 将产生 $(n-1)$ 组联合子区间。在每一个联合区间上进行 PLS 回归, 得到 $(n-1)$ 个联合模型, 其中 RMSECV 值最小的联合模型对应去除的子区间为第二去除区间。依此运行, 直至剩下一个子区间为止。
- 7) 检验第 6) 步中所有联合模型的 RMSECV 值, 找出在所有联合模型中性能最佳者 (即 RMSECV 最小), 其所对应的光谱区间组合即最佳组合。

可以看出, BiPLS 采用的仍然是几个子区间联合建模的方法, 但其区间的搜索方法继承了向后选择变量法“只出不进”的特点, 因此可以确定联合模型的建模区间数, 同时 BiPLS 法能在不同的区间划分中搜寻到最佳的联合区间。

2.4 近红外光谱模型多元校正分析方法

近红外光谱分析是通过对样品光谱和其质量参数进行关联, 建立校正模型, 然后通过校正模型预测样品的组成和性质的方法, 使用数学方法对两者进行关联, 可确立这两者间的定量或定性关系, 即校正模型。建立模型后, 只要测量未知样品的近红外光谱, 根据校正模型和样品的近红外光谱就可以预测样品的质量参数。

近红外光谱分析常用的多元校正方法主要包括: 多元线性回归 (MLR)、主成分回归法 (Principle Component Regression, PCR)、偏最小二乘法 (Partial Least Squares, PLS), 人工神经网络 (ANN) BP 法等。

2.4.1 多元线性回归 (MLR) 法

在近红外光谱分析的建模中, 经常遇到处理多个光谱信息与所研究质量参数的相关性, 多元线性回归 (MLR) 分析是最早使用的统计分析方法。多元线性回归是一元线性回归的扩展, 又称为逆最小二乘法, 用于建立自变量和因变量之间的关系。MLR 只要求已知定标样品待测特征成分的含量值, 选择若干个光谱点的吸光度值与该特征成分含量进行线性回归即可得到定标模型的回归系数。要选择好对应于被测特征成分的数据矢量。

设在近红外光谱分析中校正集样品数为 n , 样品成分数为 p , 选定的近红外光谱数为 m , 在多元线性回归分析中, 总是假设校正集样品数大于光谱数, 即 $n > m$ 。光谱矩阵为 $X_{n \times m}$, 成分含量矩阵为 $Y_{n \times p}$, 测量误差矩阵为 $E_{n \times p}$, 则有关系式:

$$Y_{n \times p} = X_{n \times m} B_{m \times p} + E_{n \times p} \quad (2-33)$$

$$B_{m \times p} = (X_{n \times m}^T X_{n \times m})^{-1} X_{n \times m}^T Y_{n \times p} \quad (2-34)$$

近红外光谱技术在食品品质检测方法中的研究

式 (2-34) 是由已知样品的成分含量矩阵 $\mathbf{Y}_{n \times p}$ 和光谱矩阵 $\mathbf{X}_{n \times m}$, 求 $\mathbf{B}_{m \times p}$ 的过程为校正; 将 $\mathbf{B}_{m \times p}$ 代回式 (2-33) 中, 则式 (2-33) 为该样品集的数学模型。通过校正过程计算得到 $\mathbf{B}_{m \times p}$, 再利用式 (2-33) 通过未知样品的近红外光谱计算未知浓度作为预测。其中 $\mathbf{Y}_{n \times p} = \{y_{ij}\} (i=1, 2, \dots, n, j=1, 2, \dots, p)$; $\mathbf{X}_{n \times m} = \{a_{ik}\} (i=1, 2, \dots, n, k=1, 2, \dots, m)$ 。

对于单成分分析, $p=1$, 则式 (2-33) 可简化为

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (2-35)$$

多元线性回归法比较适合线性较好的数据, 由质量参数之间相互干扰带来的影响可以不考虑, 而且公式含义清楚, 计算简单, 对数据矢量较少的体系回归效果较好。其局限性是不能够解决光谱矩阵的共线性问题以及对线性不好的数据回归效果较差, 所建模型的预测能力较差。而且波长点数不应该超过校正集的样品数目, 这样难免会损失一些有效的光谱信息; 吸光度矩阵存在共线性问题, 导致其成为病态矩阵; 此外, 运算过程中没有考虑到存在的噪声, 这样同样会导致过拟合情况, 从而降低了模型的可靠性。因此, MLR 在近红外光谱早期应用中较为广泛, 为了建立更为稳健的模型, 人们逐渐在此基础上发展出一系列更为有效的算法。

2.4.2 主成分回归 (PCR) 法

主成分回归法是采用多元统计中的主成分分析方法, 对混合物光谱量测矩阵 \mathbf{X} 进行分解, 然后选取其中的主成分来进行多元线性回归分析。其核心是主成分分析, 将原始数据降维, 通过对原变量进行一系列的变换, 使少数几个新变量变成原变量的线性组合, 与此同时, 这些新变量最大限度地表征了被测样品的组成和结构特性, 并且最小限度地包含噪声等干扰因素。经转换得到的新变量是相互彼此正交的, 即互不相关, 以消除大量信息共存时相互重叠而掩盖的信息部分。

如何针对被测物质主要成分进行近红外光谱解析, 找到最优定标波长组合用于建立稳健的定标模型, 常用的波长选择方法为利用主成分分析结合多元线性回归的建模方法。PCR 不是直接考虑 \mathbf{y} 和 \mathbf{X} 的回归, 而是对 \mathbf{X} 中的信息重新进行综合, 提取对系统有最佳解释能力的成分信息。由于对变量系统重新进行了综合, 将会克服多重相关性造成的信息重叠问题, 同时由于对变量系统中的信息进行了筛选, 将有效地区分系统信息和噪声, 提高模型的精确度。

首先对光谱 \mathbf{X} 进行主成分分解, 得到载荷矩阵和得分矩阵:

$$\mathbf{X}_{n \times f} = \mathbf{T}_{n \times f} \mathbf{P}_{p \times f} + \mathbf{E}_f \quad (2-36)$$

式中, $\mathbf{T}_{n \times f}$ 为光谱 \mathbf{X} 的得分, 其列变量之间正交; $\mathbf{P}_{p \times f}$ 为主成分的载荷, 其行变量之间正交; \mathbf{E}_f 为残差; f 为主成分数。

然而此时参考浓度并没有起作用, 主成分回归法在主成分分析的基础上, 将浓度 \mathbf{y} 和 \mathbf{X} 的得分 \mathbf{T} 进行回归: $\mathbf{y} = \mathbf{TB} + \mathbf{E}$, 回归系数的最小二乘解如下:

$$\mathbf{B} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y} \quad (2-37)$$

PCR 通过合理选择参与回归的主成分，去掉了大部分的噪声。由于 \mathbf{T} 的各列互相正交，解决了多元线性回归中的共线问题。然而在得到的一系列主成分中，某些主成分和浓度之间不一定有相关关系，因此回归结果就不一定合理。

主成分回归克服了 MLR 所产生的严重共线性问题，即病态。在最大程度利用光谱信息的同时，通过忽略那些次要成分，起到了抑制噪声对模型的影响，进一步提高了校正模型的稳定性和可靠性。该方法不需要了解干扰组分的存在便可以较为准确地测出待测组分。但其运算速度比 MLR 慢，且不如 MLR 方法对模型的那么直观。

在主成分回归中，对于模型建立至关重要的便是最佳主成分数的选取。若主成分数过少，将会丢失一定量的有用信息，拟合不充分；若选取的主成分数过多，则会因为包含过多的噪声而出现过度拟合现象，使得模型的预测误差增大。因此，要合理选取最佳主成分数，有效地提高模型的预测性和可靠性。

2.4.3 偏最小二乘 (PLS) 法

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，其方法是将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，去掉代表干扰组分和干扰因素主成分，仅选取有用的主成分参与质量参数的回归。较适合分析微量成分。

偏最小二乘法也是一种多元回归算法。类似 PCR，对光谱矩阵 \mathbf{X} 进行分解，去除其中无效的噪声干扰。同时，分解浓度矩阵 \mathbf{Y} ，消除其中的无用信息。并且，在分解光谱矩阵 \mathbf{X} 的同时考虑了浓度矩阵 \mathbf{Y} 的影响，其实质是将矩阵分解和回归并为一。在每计算一个新主成分之前，将 \mathbf{X} 的得分矩阵 \mathbf{T} 与矩阵 \mathbf{Y} 的得分 \mathbf{U} 矩阵进行交换，使得得到 \mathbf{X} 主成分直接与 \mathbf{Y} 关联。这样就弥补了 PCR 方法只针对 \mathbf{X} 进行分解的缺点。

PLS 算法如下所示：

PLS 首先对光谱矩阵 \mathbf{X} 和浓度矩阵 \mathbf{Y} 进行分解，其模型如式 (2-38) 和式 (2-39) 所示：

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} \quad (2-38)$$

$$\mathbf{Y} = \mathbf{UQ} + \mathbf{F} \quad (2-39)$$

其中， \mathbf{T} 是 \mathbf{X} 矩阵的得分矩阵； \mathbf{U} 是 \mathbf{Y} 矩阵的得分矩阵； \mathbf{P} 是 \mathbf{X} 矩阵的载荷矩阵； \mathbf{Q} 是 \mathbf{Y} 矩阵的载荷矩阵； \mathbf{E} 是 \mathbf{X} 矩阵的 PLS 拟合残差矩阵； \mathbf{F} 是 \mathbf{Y} 矩阵的 PLS 拟合残差矩阵。

PLS 第二步是将 \mathbf{T} 和 \mathbf{U} 做线性回归，如式 (2-40) 和式 (2-41) 所示：

$$\mathbf{U} = \mathbf{TB} \quad (2-40)$$

$$\mathbf{T} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y} \quad (2-41)$$

在预测时，首先根据 \mathbf{P} 求出未知样品光谱矩阵 $\mathbf{X}_{\text{未知}}$ 的得分矩阵 $\mathbf{T}_{\text{未知}}$ ，然后求得浓度预测值，如式 (2-42) 所示：

$$\mathbf{Y}_{\text{未知}} = \mathbf{T}_{\text{未知}} \mathbf{BQ} \quad (2-42)$$

PLS 在应用中显示出了其独特的优越性：

1) 灵活地使用光谱数据，可以根据需要使用全部或者部分光谱数据而达到很好的建模

效果。

- 2) 将特征矢量与被测成分相关，即把数据分解与回归同时进行，有效地融合在一起。
- 3) 适用于处理样品数量少，但是变量相对较多的问题。
- 4) PLS 是多元线性回归和主成分分析的完美结合，显著提高了校正模型的预测能力。正因为 PLS 的如上优点，在很多领域中，该方法得到了广泛的应用。

2.4.4 人工神经网络 (ANN) BP 法

ANN BP (Back - Propagation) 是 D. E. Rumelhart 和 J. L. McClelland 及其研究小组在 1986 年研究并设计出来的。ANN BP 是非常著名的多层感知器，它的实现过程使用了 BP 算法得名，其基本算法思想是，学习过程由信号的正向传播与误差的反向传播两个过程组成，ANN BP 通过反向传播来不断调整神经网络的阈值和权值，使得神经网络误差平方和的值达到最小。最典型的多层感知器有 3 层：输入层、隐含层和输出层。输入层代表输入网络的原始数据。隐含层的值由输入层的输出及输入层之间的连接权重确定。隐含层可以是单层，也可以是多层，视问题的复杂程度而定。同样，输出层的值取决于隐含层的输出值及隐含层与输出层的连接权重。隐含层的神经元与输入层及输出层神经元的关系是在网络的学习过程中得到的。

目前 ANN BP 已经成为近红外光谱定量分析中应用最广泛的非线性多元校正方法。其神经网络结构如图 2-5 所示。

ANN BP，即误差反向传播算法的学习过程。输入层的各个神经元负责接收来自外界的输入信息，并将信息传递给隐层的各个神经元；隐层是进行内部信息的处理层，负责进行信息变换，根据信息变化能力的需求不同，隐层的结构可以设计为单隐层或者多隐层；输入信号经隐含层逐层处理，通过最后一个隐层的传递将各个神经元的的信息传递到输出层，经进一步处理后，完成一次神经网络学习的正向传播过程，从而由输出层向外界输出信息处理结果。如果输出层得不到期望的输出，则进入误差的反向传播过程。误差通过神经网络的输出层，按照误差梯度下降的方式来修正各层权值，并向隐层、输入层逐层进行反传。通过不断地进行信息的正向和误差反向传播，来调节网络的权值和阈值，这就是 ANN BP 的学习训练过程，直到网络训练达到之前所设定的学习次数，或是输出误差到达可以接受的程度才结束训练。

ANN 具有很强的非线性映照能力，所以倍受化学计量学专家的关注。目前，ANN 已经成为化学计量学中处理非线性体系的重要方法之一。ANN 法具有传统方法无法比拟的优点：一是它是自变量与因变量间的非线性映射，可以避免因近似处理带来的误差；二是它具有学习的能力，可以通过学习来提高模型分析的精度；三是 ANN 模型的抗干扰能力较强。与其他多元统计方法不同，ANN 的优点之一就是对于样品的质量参数无需大量筛选，可不加选择

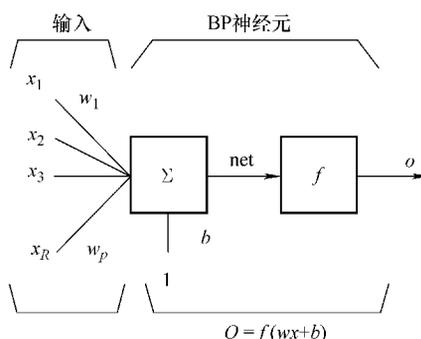


图 2-5 ANN BP 神经网络结构
 $x_1, x_2, x_3, \dots, x_R$ —神经元 R 个输入变量
 $w_1 \dots w_p$ —网络权值和阈值 Σ —求和单元
 f —传递函数

地将所有参数作为输入信息送入网络，进行训练便能得到有意义的结果。这既是 ANN 的优势，又是其不足之处。ANN 的最佳结构以及训练次数都是必须认真考虑的问题。ANN 不需精心设计和整理规则，只要将一组已知的样品作为训练集来“训练”网络，网络自身便可在训练过程中得出与之相应的规律，并且将学习到的知识体现在网络的权重中。但是 ANN 目前还无法对其所得到的预测结果进行解释。

一般情况下，三层的 ANN BP 可以满足大多数问题的需求。由于训练函数自身特点不同，针对特定要求建模的适用性也不同，需多次实验选择出最优训练函数对 ANN 进行训练。

2.4.5 聚类分析法

模式识别方法是 20 世纪 50 年代早期提出的，60 ~ 80 年代在各个学科得到广泛应用。到 80 年代，模式识别方法发展成为一种非常成熟的多元分析方法。

模式识别方法可分为有管理方法、无管理方法和显示方法等几类。有管理模式识别包括线性判别、逐步判别分析、KNN 方法、SIMCA 方法、神经网络等；无管理方法包括最小生成树、聚类分析等；显示方法包括变量 - 变量绘图、线性投影、非线性映射、脸谱法等。

在模式识别中是根据“物以类聚”的原则进行样品分类的，即性质或特征或变量相似的样品聚为一类。化学模式识别中常用相似系数（similarity coefficient）表示相似性。相似系数表示方法有许多形式，而“距离”（distance）便是相似性 [或相异性（dissimilarity）] 的量度。

设样品 \mathbf{x}_i 和 \mathbf{x}_j （测量数据矩阵 \mathbf{X} 中的两个行矢量）之间的距离为 $D(\mathbf{x}_i, \mathbf{x}_j)$ ，作为距离必须满足以下条件：

- 1) 当 $\mathbf{x}_i = \mathbf{x}_j$ 时， $D(\mathbf{x}_i, \mathbf{x}_j) = 0$ ，即同一个样品距离为零；
- 2) $\mathbf{x}_i \neq \mathbf{x}_j$ 时， $D(\mathbf{x}_i, \mathbf{x}_j) > 0$ ，即不同样品距离大于零；
- 3) $D(\mathbf{x}_i, \mathbf{x}_j) = D(\mathbf{x}_j, \mathbf{x}_i)$ ，即距离没有方向性；
- 4) $D(\mathbf{x}_i, \mathbf{x}_j) \leq D(\mathbf{x}_j, \mathbf{x}_i) + D(\mathbf{x}_k, \mathbf{x}_j)$ ，即两点间直线距离最短，这里下角 k 表示第 3 个样品。

化学模式识别中常用的距离指标较多，比较主要的如下：

1. 欧氏距离 (Euclidean distance)

$$D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2-43)$$

式中， p 为变量数，如写成矢量形式有

$$D_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \quad (2-44)$$

在某些场合，对于不同的变量需给予不同的权重，则有加权欧氏距离（weighted Euclidean distance）：

$$D_{ij} = \sqrt{\sum_{k=1}^p w_k (x_{ik} - x_{jk})^2} \quad \left(\sum_{k=1}^p w_k = 1 \right) \quad (2-45)$$

标准欧氏距离（standard Euclidean distance）为

$$D_{ij} = \sqrt{\sum_{k=1}^p [(x_{ik} - x_{jk})/s_k]^2} \quad (2-46)$$

式中, s_k 表示测量数据矩阵 X 中第 k 列的标准偏差:

$$s_k = \sqrt{\frac{1}{n} \sum_{l=1}^n (x_{lk} - \bar{x}_k)^2} \quad (2-47)$$

实际上标准欧氏距离是预先将测量数据矩阵 X 标准化后所确定的矢量间的距离。

2. 马氏距离 (Mahalanobis distance)

$$D_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{V}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \quad (2-48)$$

式中, \mathbf{V} 为协方差矩阵, 其元素 v_{lm} 为

$$v_{lm} = \frac{1}{n-1} \sum_{k=1}^n (x_{kl} - \bar{x}_l)(x_{km} - \bar{x}_m) \quad (2-49)$$

马氏距离实际上可以看作一组样品的中心与某一目标之间的距离, 该距离考虑了样品与样品之间的相关性。欧氏距离相等并不表示马氏距离也相等。

在样品很多的情况下, 对预测样品进行模式识别时, 人们通常事先并不知道样品内在的分类。其中无监督模式识别方法在未知训练集样品的类别的情况下, 同样可以对样品进行分类识别。聚类分析法便是无监督模式识别法的代表。

常见的聚类分析方法有系统聚类法、最小生成树法、最短生成路径法和判别聚类的势函数法等。其中系统聚类法是在近红外光谱分析中应用较为广泛的一类聚类分析方法, 其基本思想是首先定义样品之间和类与类之间的距离。在各自成类样品中, 将距离最近的两类合并, 重新计算新类与其他类之间的距离, 并按最小距离归类, 重复此过程, 每次减少一类, 直到所有的样品成为一类为止。其聚类过程用图表示, 称为聚类图, 如图 2-6 所示。

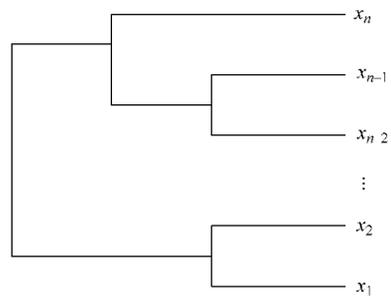


图 2-6 系统聚类图

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。

(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, l 间的距离, D_{ij} 表示类 G_i, G_j 间的距离, 则

$$D_{ij} = \min_{\substack{k \in G_i \\ l \in G_j}} \{d_{kl}\} \quad (2-50)$$

最短距离法的聚类步骤如下:

- 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 中的样品, 为 G_i, G_j 中的全部样品。
- 3) 计算新类 G_r 与其他类 G_s 间的距离 D_{rs} 。

$$\begin{aligned}
 D_{rs} &= \min_{\substack{k \in G_i \\ l \in G_s}} \{d_{kl}\} \\
 D_{rs} &= \min_{\substack{k \in G_i \\ l \in G_s}} \{d_{kl}\} \\
 &= \min \left\{ \begin{array}{ll} \min_{\substack{k \in G_i \\ l \in G_s}} d_{kl}, & \min_{\substack{k \in G_j \\ l \in G_s}} d_{kl} \end{array} \right\} \\
 &= \min \{D_{is}, D_{js}\}
 \end{aligned} \tag{2-51}$$

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

$$D_{ij} = \max_{\substack{k \in G_i \\ l \in G_j}} \{d_{kl}\} \tag{2-52}$$

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

$$D_{rs} = \max \{D_{is}, D_{js}\} \tag{2-53}$$

(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

$$D_{rs} = \sqrt{\frac{1}{2}D_{si}^2 + \frac{1}{2}D_{sj}^2 - \frac{1}{4}D_{ij}^2} \tag{2-54}$$

几何意义如图 2-7 所示。

中间距离法可推广到更一般的形式，即

$$D_{rs} = \sqrt{\frac{1}{2}D_{si}^2 + \frac{1}{2}D_{sj}^2 + \beta D_{ij}^2} \tag{2-55}$$

式中， $-\frac{1}{4} \leq \beta \leq 0$ 。

(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作为两类之间的距离，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

$$\bar{X}_r = \frac{1}{n_r}(n_i \bar{X}_i + n_j \bar{X}_j) \tag{2-56}$$

如果类 G_s 的重心为 \bar{X}_s ，则类 G_s 与 G_r 的距离平方为

$$D_{rs}^2 = \frac{n_i}{n_r} D_{si}^2 + \frac{n_j}{n_r} D_{sj}^2 - \frac{n_i}{n_r} \cdot \frac{n_j}{n_r} D_{ij}^2 \tag{2-57}$$

(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

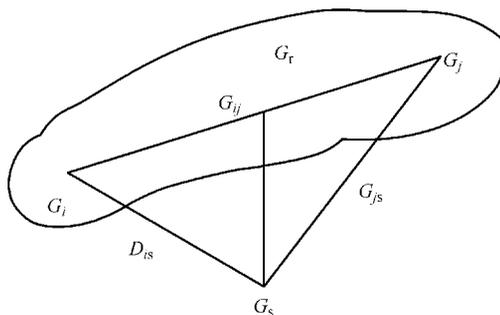


图 2-7 中间距离法图示

$$D_{ij}^2 = \frac{1}{n_i n_j} \sum_{\substack{k \in G_i \\ l \in G_j}} d_{kl}^2 \quad (2-58)$$

如果 G_i 与 G_j 类合并为 G_r , 则新类 G_r 与其他类 G_s 间的距离平方为

$$\begin{aligned} G_{rs}^2 &= \frac{1}{n_r n_s} \sum_{\substack{k \in G_s \\ l \in G_r}} d_{kl}^2 = \frac{1}{n_r n_s} \left[\sum_{\substack{k \in G_s \\ l \in G_i}} d_{kl}^2 + \sum_{\substack{k \in G_s \\ l \in G_j}} d_{kl}^2 \right] \\ &= \frac{n_i}{n_r} D_{si}^2 + \frac{n_j}{n_r} D_{sj}^2 \end{aligned} \quad (2-59)$$

(6) 可变类平均法

在类平均法递推公式中, G_i 和 G_j 之间的距离没有反映进去, 为修改式 (2-59) 得到递推公式:

$$G_{rs}^2 = \frac{n_i}{n_r} (1 - \beta) D_{si}^2 + \frac{n_j}{n_r} (1 - \beta) D_{sj}^2 + \beta D_{ij}^2 \quad (2-60)$$

式 (2-60) 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) 可变法

可变法的递推公式为

$$D_{rs}^2 = \frac{1 - \beta}{2} [D_{si}^2 + D_{sj}^2] + \beta D_{ij}^2 \quad (2-61)$$

式中, $\beta < 1$, 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, \bar{X}_j 表示 G_j 的重心, 则 G_j 中样品的离差平方和为

$$S_j = \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)' (X_{ij} - \bar{X}_j) \quad (2-62)$$

全部 k 类的平方和为

$$S = \sum_{j=1}^k S_j \quad (2-63)$$

离差平方和法先将 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使 S 增加最小的两类进行合并, 直至所有样品为一类为止。

上述 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。Wishart 1969 年提出了统一的递推公式, 即

$$D_{rs}^2 = \alpha_i D_{si}^2 + \alpha_j D_{sj}^2 + \beta D_{ij}^2 + \gamma |D_{si}^2 - D_{sj}^2| \quad (2-64)$$

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 2-2 列出了上述八种聚类方法中 $\alpha_i, \alpha_j, \beta, \gamma$ 在式 (2-64) 中的取值。

表 2-2 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				
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

2.4.6 支持向量机 (SVM) 法

支持向量机 (SVM) 是一种机器学习的方法, 其基础上统计学习理论, 基本思想是在样品空间或特征空间中, 构造一个最优决策的超平面, 使得该超平面到不同类样品集之间的距离最大, 从而使算法的泛化能力得到提高。该方法是一个凸二次优化问题, 能够得到全局最优解。此外, SVM 较传统的神经网络具有收敛速度快、容易训练、不需要预设网络结构等优点。因此 SVM 在模式识别、数据挖掘、函数逼近、图像处理方面都得到了广泛的应用。

SVM 是由 Vapnik 等人 1998 年在统计学习 (Statistic Learn Theory, SLT) 的 VC (Vapnik - Chervonenkis) 维理论和结构风险最小原理基础上提出的一种方法, 根据有限的样品信息在模型的复杂性和学习能力之间来寻求最佳折衷, 以期获得最好的推广能力。用小麦粉的水分、灰分及面筋与仪器响应之间并非简单的线性关系, 而是由近红外光谱的吸收特性提供的物质的物理状态信息得到的值, SVM 可以通过非线性核函数, 将输入样品空间映射到高维线性特征空间, 可以处理高度非线性回归等问题。应用 SVM 进行回归可减少训练时间、避免陷入局部极小, 提高了模型的学习泛化能力。

SVM 算法原理如下:

1) 将原始样品数据进行相应的预处理有利于加快模型的样品训练速度和收敛速度, 提高预测精度。这里采用归一化处理方法:

$$\bar{x}(i) = \frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \quad (2-65)$$

式中, $\bar{x}(i)$ 为归一化后的数据序列值; x_{\max} 、 x_{\min} 分别为数据样品的最大值和最小值。

2) 假定输入样品集为 n 维矢量, 样品及相应输出值可以表示为 $(x_1, y_1), \dots, (x_k, y_k) \in R_n \times R$, 函数回归就是用训练样品进行训练得到一个函数式 $f(x) = (w \cdot x) + b$, 使训练样品外的 x 集合能够通过 $f(x)$ 映射到对应的 y 集合。

3) 设从样品空间到高维特征空间的映射函数为 $\varphi(x)$, 则求解函数 $f(x)$ 参数的问题就

近红外光谱技术在食品品质检测方法中的研究

转化为在 $\begin{cases} y_i - ((w, \phi(x)) + b) \leq \varepsilon \\ ((w, \phi(x)) + b) - y_i \leq \varepsilon \end{cases} i = 1, 2, \dots, l$ 约束条件下求函数 $\min \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \cdot w)$ 的最小值的优化问题。

4) 一般特征空间的维数很高而且目标函数不可微, 所以求解上述的 SVM 回归问题一般

都通过建立拉格朗日函数, 将上述问题转化为在式 $\begin{cases} \sum_{i=1}^l (\alpha_i - \hat{\alpha}_i) = 0 \\ \alpha_i \hat{\alpha}_i \in [0, C] \end{cases}$, 约束下求解式

$$\max = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \hat{\alpha}_i) + \sum_{i=1}^l y_i(\alpha_i - \hat{\alpha}_i) \quad (2-66)$$

式中, α_i 和 $\hat{\alpha}_i$ 为拉格朗日乘子, $\alpha_i \cdot \hat{\alpha}_i = 0$, α_i 和 $\hat{\alpha}_i \geq 0$ 。

在求解二次优化的问题时所得到的拉格朗日乘子, α_i 和 $\hat{\alpha}_i$ 中只有一小部分不为零, 支持矢量即拉格朗日乘子所对应的数据点。

5) 这样 SVM 回归的问题就转化成为一个二次规划问题, SVM 回归的函数就是由这些支持矢量决定。求解后得到回归函数式:

$$f(x) = \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)K(x_i - x_j) + b \quad (2-67)$$

式中, 矢量 x_i 和 x_j 分别为在特征空间 $\phi(x_i)$ 和 $\phi(x_j)$ 中的内积, $K(x_i - x_j)$ 即核函数, N 为支持矢量的个数。

SVM 的算法流程如图 2-8 所示。

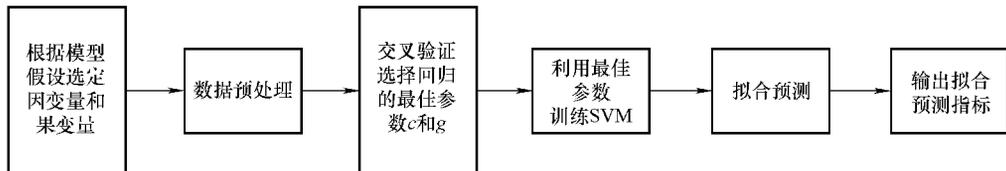


图 2-8 SVM 的算法流程

为了减少运算量, 提高运算速度, 采用主成分分析法对原始的近红外光谱矩阵进行降维, 取累计贡献率达到 99.5% 的特征矢量为 SVM 模型输入的特征子集。惩罚系数 c 、不敏感系数 g 、核函数及相关参数的选择, 对 SVM 的效果有显著的影响。K - CV (K - fold Cross Validation) 是把原始的数据分为 K 组, 将每个自己作为一个验证集, 同时 $K - 1$ 组自己作为训练集, 得到 K 个模型, 用这 K 个模型最终验证集的准确率的平均数作为 K - CV 的性能指标, K - CV 可以有效地避免过学习和欠学习状态的发生, 使用 K - CV 选择参数使 SVM 模型得到较好的预测结果; CV 是在没有先验知识指导的情况下一般选用径向基函数 (RBF) 为核函数, RBF 可以将非线性的样品数据映射到高维特征空间, 降低模型的复杂性, 提高训练速度。这里 SVM 模型参数的选取 RBF 为核函数, 并采用交叉验证算法来确定相关参数。

2.5 近红外光谱定标模型的评价指标

对已经建立的校正模型，必须通过一些方法检验模型预测能力的好坏。检验模型通常有如下指标。

假设 y_i 为样品 i 某一组分浓度标准值， \hat{y}_i 为样品 i 的某一组分浓度预测值， \bar{y} 为 n 个样品的某一组分浓度标准值的平均值。 \bar{y}^P 为 n 个样品的某一组分浓度预测值的平均值：

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2-68)$$

$$\bar{y}^P = \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (2-69)$$

1) 测定系数 R^2 (Coefficient of Determination) 等于相关系数的平方；该评价参数考察样品的预测值与真实值之间的相关程度：

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2-70)$$

2) 相关系数 R (Correlation Coefficient)：

$$R = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}^P)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y}^P)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{R^2} \quad (2-71)$$

3) 交叉验证预测均方差 (RMSPCV) 或称为交叉验证标准差 (SECV)，主要评价某种建模算法的可行性及对内部样品的预测能力。在模型训练过程中通过交互验证的方法来计算。预测方均根误差越小，说明该模型对样品的预测能力越高，反之越低：

$$\text{RMSPCV} = \text{SECV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (n \text{ 为校正集样品数目}) \quad (2-72)$$

4) 预测标准差 (SEP) 或 RMSD 表示模型对预测及采集样品的预测标准差，评价该模型对外部样品的预测能力。该值越小，则表示模型对外部样品的预测能力越高，反之则预测能力越低。

$$\text{SEP} = \text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (n \text{ 为预测集样品数目}) \quad (2-73)$$

2.6 近红外光谱仪器介绍

2.6.1 光谱仪发展介绍

近红外光谱仪是一种测量物质对近红外辐射的吸收率（或透过率）的分析仪器，它可

近红外光谱技术在食品品质检测方法中的研究

以利用每种物质的特征吸收谱对待测物质进行定性分析。它同样可以利用物质总量与吸收总量所呈现的线性关系，对待测物质的特征成分进行定量分析。近红外光谱仪器主要由光源、分光系统、样品池、控制系统、检测器、数据采集和分析软件等部分构成，其结构如图 2-9 所示。

近红外光谱仪经过 50 年的发展历程，从仪器的设计方式、性能和测量方法都发生了很大的变化。从样品光谱信息的获得看，分为可在一个或几个波长范围内测定的专用型滤光仪器，和在近红外波长范围内测定全谱信息的研究型仪器。从光谱测定的波长范围看，由于采用不同的检测器和分光器件，可分专用于短波近红外区域和用于长波近红外区域。从近红外仪器的分光器件看，可分为 4 种主要类型：滤光片、光栅分光、傅里叶变换和声光调制滤光器。从检测器对分析光的响应看，有单通道和多通道两种类型，多通道型又有采用电荷耦合器件和二极阵列器件作检测器的近红外光谱仪。

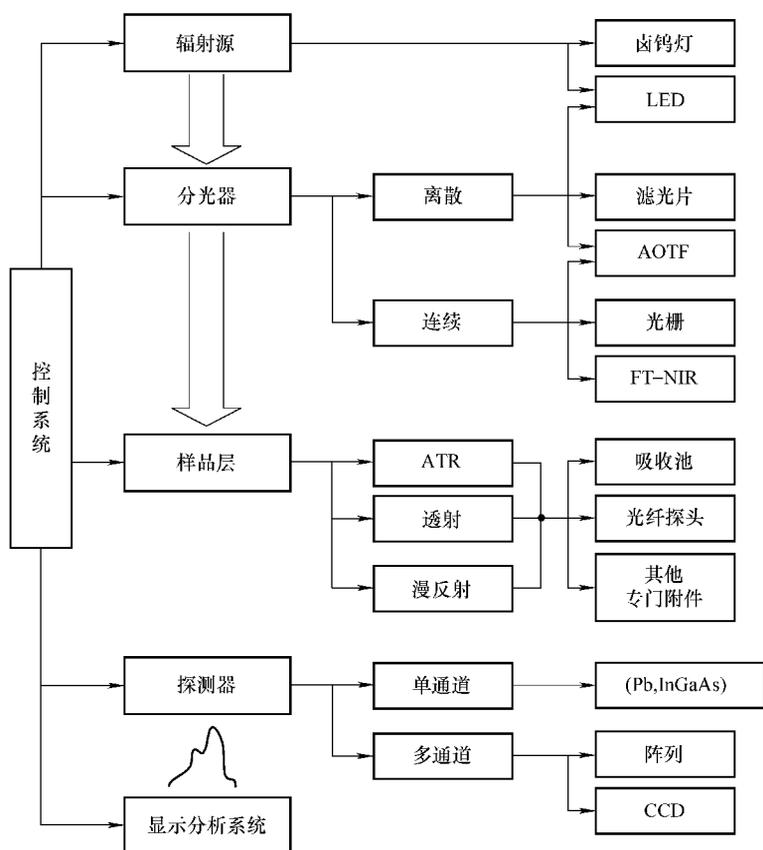


图 2-9 近红外光谱仪的结构

滤光片型近红外光谱仪早在 20 世纪 50 年代就在农副产品的检测分析中得到了广泛的应用。该类仪器采用滤光片作为单色光器件。该类仪器的特点是设计简单、成本低，只能获得几个特定波长处的光谱信息。这类仪器已广泛应用于专用或便携式仪器上。滤光片型近红外光谱仪器需要适当波长的滤光片作为仪器的分光系统，选择滤光片通过分析样品的特征波长确定，

设计简单、成本低、坚固耐用、光通量大、信号记录快,并且可固定波长进行测量,灵活方便。该类仪器设计简单,成本低,采用速度快,比较坚固,可根据需要在固定的波长下进行测量。但仪器的灵活性差,波长的分辨率和重现性较差,若客观测量环境发生较大的变化,常会引起较大的测量误差。因此,此类主要用于对一些成熟的特定项目的分析检测作业。如 Perten 公司的 Inframatic 系列产品和 NDC Infrared Engineering 公司的 Infralab 710 产品。

光栅色散型近红外光谱仪采用棱镜或者光栅作为分光部件,对红外辐射的色散实现分光,是 20 世纪 80~90 年代最常见的近红外光谱仪器,成为第二代近红外分析仪。根据检测器类型不同,又可分为扫描-单通道检测器和固定光路-阵列检测器两种类型。光栅扫描单通道仪器是 60 年代末 70 年代初发展起来的一类仪器,其基本原理是:光源发出的检测光源经准直和入射狭缝照射到光栅上进行分光,光栅不断地转动,由此使单色光依次通过出射狭缝,照射到样品后,与该样品发生作用,进入检测器检测。这类仪器可进行全谱扫描,价格适中,分辨率高,维修也很容易,在粮食和食品领域得到了广泛的应用。如日本岛津公司的 UV-365 型分光光度计、北京普析通用公司 TU-1800 系列和上海棱光技术有限公司的农产品品质分析仪。但其最大的缺点是光栅或反光镜等部件易磨损,部件复杂,抗振性能较差,扫描的速度相对较慢,一般不适合做过程分析仪器。由于测量精度低,需要经常对其进行外部标准校正。值得一提的是,目前已有一些仪器公司针对该类仪器的确定进行了进一步的改进和完善,其同样具有良好的潜力和发展前景。固定光路多通道型仪器是 90 年代发展起来的一类仪器。这类仪器采用固定光栅分光,经光栅色散的光聚焦在多通道检测器的焦面上。多通道近红外光谱仪器是因为仪器的检测器采用多通道光敏器件而得名。这类仪器的色散系统一般采用平面光栅或全息光栅,与光栅扫描型仪器相比,光栅不需要转动即可实现确定波长范围的扫描。这类仪器的最大特点是仪器内部无可移动部件,仪器的稳定性和抗干扰性能好;另外一个特点是扫描速度快,一般单张光谱的扫描速度只有几十毫秒。这两个特点的结合,使该类仪器特别适合作为现场或在线分析仪器使用。

傅里叶变换型近红外光谱仪利用干涉图和光谱图之间的对应关系,通过测量干涉图和对干涉图进行傅里叶变换的方法来测定和研究光谱。它能同时测量、记录所有波长的信号,并以比传统的色散型光谱仪更高的效率采集来自光源的辐射能量,具有更高的信噪比和分辨率。在进入 20 世纪 80 年代以来,便成为近红外光谱仪器的主导产品,其核心部件是迈克尔逊干涉仪。在工作时,需选用合适的光源、分束器与检测器。该类仪器的特点是信噪比高、波长精度高、分辨率好、稳定性好。近年来,国内外涌现出大量较实用的傅里叶近红外光谱仪,如 Thermo-Nicolet 公司的 Antaris 系列、Bruker 公司的 VECTOR 和 MATRIX 产品以及我国瑞利分析仪器公司生产的 WQF-400N 等仪器。

声光可调滤光器型近红外光谱仪是 20 世纪 90 年代近红外光谱技术最突出的进展,采用声光可调滤光器作为分光系统,通过超声射频的变化扫描光谱。该类仪器物可移动部件,测量速度快,滤光器体积小,可做到光谱仪器的小型化,可长期稳定地工作。但此类仪器的分辨率不如传统的近红外分析仪器,并且造价相对较贵,比较适合用于在线过程分析。目前,如 Brimrose 公司的 LUMINAR 系列产品、Maihak 公司的 LIMORL 产品等近红外光谱仪都采用声光可调滤光器分光原理。

我国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方面取得了一定的成绩。如北京第二光学

近红外光谱技术在食品品质检测方法中的研究

仪器厂研制了傅里叶变换近红外辛烷值分析仪，石油化工科学研究院研制了采用电荷耦合器件（CCD）的多通道近红外光谱仪，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，中国农业机械化科学研究院研制了 CA-XP 粮油品质近红外分析系统等。

在仪器研制方面，国外已经有商业化的仪器。如德国 Bruker 光谱仪器公司、美国热电尼高力仪器公司、美国 PE 公司的傅里叶变换近红外光谱仪、瑞典 Foss 公司的透射型近红外分析仪等。由于国内硬件工艺水平所限，以及缺乏功能完善的光谱定量分析软件，国产光谱仪器的使用和推广受到极大的限制，目前我国一些科研院所和企业用户使用的近红外光谱仪器，绝大多数采用的是国外生产的。

2.6.2 本研究应用光谱仪平台介绍

北京工商大学计算机与信息工程学院在 2004 年建立了“光谱技术与品质检测研究室”。该研究室配备了 VERTEX 70 傅里叶红外光谱仪、DXR 光聚焦显微拉曼光谱仪、紫外分光光度计、卓立汉光近红外光谱系统以及一系列制样设备，组成了多光谱信息融合食品无损检测技术创新平台，为基于多光谱的农产品及食品无损检测技术平台的建设提供了良好的硬件基础。

1. VERTEX 70 傅里叶红外光谱仪

VERTEX 70 傅里叶红外光谱仪是 Bruker 公司于 2004 年推出的全数字化红外光谱仪，如图 2-10 所示。该仪器采用多项先进技术，使得 Bruker 公司的红外光谱仪性能更优异，使用更方便。



图 2-10 VERTEX 70 傅里叶红外光谱仪

VERTEX 70 傅里叶红外光谱仪是世界上第一台全数字化的红外光谱仪，主机内置 HTML 服务器通过网卡与计算机进行数据通信。同时 EQUINOX™ 专利技术使光谱仪器“平台化”、“积木化”的理想变为现实；它第一次为广大 FT-IR 用户提供了多种选择，即在同一光学平台上实现：从常规压片到所有的联机试验、从实验室分析到高级研究、从普通应用到特殊领域的专门应用。

该仪器采用的新技术如下：

- 1) RockSolid 干涉仪, Bruker 公司的专利技术;
- 2) DigiTect 技术, 检测器为数字检测器, Bruker 公司的专利技术;
- 3) BRAIN 技术, Bruker 公司的人工智能网络技术, 随时反馈仪器的状态;
- 4) AAR 技术, 附件的自动识别技术;
- 5) ACR 技术, 光学组件的自动识别, 并将参数反馈到 OPUS 红外光谱软件中。

另外, 由于 Bruker 的全新的数字化设计, 用户还能对仪器进行 Validation (校验) 测试, 保证测试数据的准确可靠。同时具有以下特点:

(1) 创新的全数字化设计

全新的数字化检测器系统: 专利技术 DigiTect™, 将检测元件、信号放大器与 24 位的模-数转换器集成在一起, 直接输出数字信号 (模拟信号在传输过程中衰减较大且易受到干扰, 而数字信号则可完全避免), 进一步降低了电子噪声。24bit 模-数转换器则将系统的弱信号检测能力提高了一个数量级。全新的快速傅里叶变换系统: Bruker 公司独家技术——硬件全集成傅里叶变换 (具有单独的 CPU、内存等) 独立地进行快速采样及傅里叶变换, 不占用任何系统资源。因此运算速度快、实时响应好, 干涉仪每次扫描, 均可实时得到一张光谱图; 红外主机与计算机间的数据通信量大大减低, 使采用更加灵活方便的通信方式 (如网络连接) 成为可能。

全新的实时在线监控: 主机所有硬件、工作状态、性能指标, 测量附件的识别、工作状态、测量参数的设定。

(2) 高灵敏度和信噪比

采用 24bit DigiTect™ 检测器系统和 ROCKSOLID™ 干涉仪两项专利技术是优异灵敏度的保证。VERTEX 70 傅里叶红外光谱仪信噪比优于 50000:1 (仪器实际使用环境下测试, 无吹扫; 测试条件: P/P 值, 4cm⁻¹分辨率, 1min 背景/样品扫描, 2200~2100cm⁻¹), 几乎高出同类仪器一倍以上。

(3) 高稳定性

VERTEX 70 傅里叶红外光谱仪采用高抗振、高稳定的 ROCKSOLID™ 专利干涉仪。该专利干涉仪来源于 Bruker 公司独家产品——高分辨 (分辨率高达 0.0035cm⁻¹)、车载型红外光谱仪, 具有极好的稳定性和抗干扰性, 即使在翻转的情况下仍能正常工作, 并且维护、维修十分容易。

(4) 超强的仪器扩展能力

采用 EQUINOX™ 专利技术 (等效的出/入射光光路接口), VERTEX 70 傅里叶红外光谱仪共计有 2 个等效的入光口、5 个等效的出光口。保证了仪器的联机扩展可以随意组合, 互相不受影响, 达到最佳的联机效果。例如可在 VERTEX 70 傅里叶红外光谱仪的右面连接 FT-拉曼, 左边连接红外振动园二色 (PMA37), 前面连接显微镜或近红外探针以及 Bolometer 检测器等; 而且这些连接之间的转换完全由计算机自动控制。测量谱区扩展: 10~30000cm⁻¹, 真正覆盖全部红外谱区。

(5) 独具特色的网络化设计

VERTEX 70 傅里叶红外光谱仪内置网络伺服板与计算机之间通过“网卡”连接, 即插即用。红外主机, 每台红外主机均具有 IP 地址, 可以作为网络工作站实现远程操作、远程

近红外光谱技术在食品品质检测方法中的研究

控制、远程诊断、资源共享。

(6) 仪器密封、防潮性能好

采用上压式顶盖、凹槽、真空胶圈及卡口式螺钉密封，得到最好的密封效果（而采用海绵密封，会因长期使用导致海绵失效，造成密封效果降低，甚至没有密封效果）。光源腔、干涉仪腔、检测器腔互相独立密封，故在更换光源、分束器、检测器时不会破坏整体密封效果，能有效地保障分束器的干燥环境。电子式湿度指示使仪器的维护强度降到最低。

(7) 高采样速率

Bruker 公司独家采用硬件技术，独立地进行快速采样及傅里叶变换，不占用任何系统资源。因此运算速度快、实时响应好。干涉仪每次扫描，均可实时得到一张谱图。连续扫描的采样速率达到 105s，步进扫描的时间分辨率可达 5ns。

(8) 步进扫描技术

Bruker 公司 1988 年发明的技术，目前已经日渐成熟。VERTEX 70 傅里叶红外光谱仪无需对干涉仪及光路做任何升级或调整即可实现步进扫描技术，完成 5ns 分辨率的时间分辨测试、多层复合材料的纵深扫描、逐层剖析、二维相关等研究工作。能否运用此项技术标志着仪器档次的高低。

(9) 高度智能化程度

1) 智能化光学部件：全部光学部件（如激光器、光源、检测器、分束器等）均采用智能化预准直模块设计，自动识别、自动准直技术保证用户可自行更换，且无需进行光路调整。

2) 全程智能化附件：红外应用附件自动识别、自动诊断、测量参数自动设定。

3) 实时、在线监测、PerformanceGuard™ 专利技术，内置校准用标准物质，全自动工作，实时、在线监测、诊断光谱仪的每个部分，保证光谱仪的工作状态及各项性能指标。监测数据及诊断结果实时记录并可给出各种符合要求的报告。

4) 智能化吹扫控制系统。

(10) 人性化设计

1) 仪器面板上有湿度报警、主机工作状态以及激光器工作状态指示灯，随时让用户了解仪器的工作状态。

2) 打开仪器面板，设计有专门的储物空间，用于摆放各种制样工具，方便用户管理和使用。

3) 各种光学器件（如光源、检测器等）均采用燕尾槽、榫式定位设计，更加安全、牢固，用户可以任意插拔和更换，完全不影响光路的准直性。

2. MPA 多功能傅里叶变换近红外光谱仪

MPA 多功能傅里叶变换近红外光谱仪是布鲁克光谱仪器公司集 30 年设计和生产 FT-NIR 光谱仪经验的结晶，如图 2-11 所示。该仪器具有强大的扩展灵活性和优越的性能，最大程度满足不同用户科研和生产的需求。由于其具有高性能附件和灵活性的操作性，成为实验室和过程分析开发各种方法不可缺少的强有力工具，该设备采用了多项先进技术：

1) 采用三维立体角镜的 RockSolid™ 永久准直干涉仪，Bruker 公司的专利技术，光能利用率比 45°角干涉仪提高 41%；

- 2) DigiTECT 技术, 数字化检测器, Bruker 公司的专利技术, 提高了仪器的灵敏度;
- 3) 操作维护简单, 各种易耗部件均采用预准直设计;
- 4) 强大的远程数据交换系统, 可选择 OPC、Modbus、Profibus、RS485、RS232、4 ~ 20mA 等数据交换系统与工业控制系统或其他外围计算机进行数据交换;
- 5) 实时诊断功能使计算机随时诊断仪器的状态, 并自动显示仪器的各种故障。

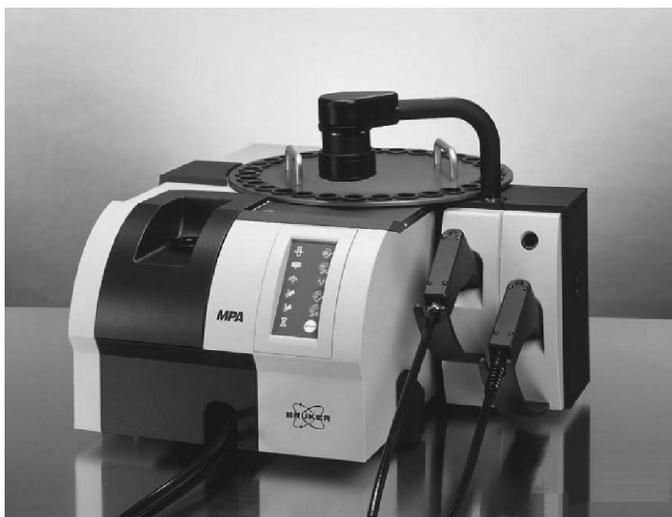


图 2-11 MPA 多功能傅里叶变换近红外光谱仪

MPA 多功能傅里叶变换近红外光谱仪具有以下技术特点:

- 1) 数字化设计: 光谱仪主机内置服务器, 通过 Ethernet 网卡与计算机连接, 光谱仪主机也可以作为网络终端进行网络控制。
- 2) 扩展灵活性: 能够根据用户需要同时或单独安装液体透射、固体光纤探头、液体光纤探头、固体漫反射积分球、固体透漫射主要近红外附件及配套的 3 个检测器, 还可根据需要进行选择配备液体变温、固体自动进样器等附件。
- 3) 固体光纤探头: 可选 BQC 快速接口或 SMA 接口, 由 200 束光纤组成的光纤束, 100 条出, 100 条进, 极大地提高了光的传输性能。
- 4) 镀金大积分球设计: 10cm 内径的镀金积分球, 2cm 的采样窗口, 旋转台通过偏离中心点安装在积分球上, 满足不同大小颗粒度不均匀样品分析, 并可实现背景自动测试。
- 5) 固体透漫射设计: 采用低 OH 石英双聚焦透镜, 在不提高光源功率的前提下, 极大地提高了近红外光的穿透和检测能力, 并且保证了光源的长寿命。
- 6) 操作简单: 中文操作软件在中文 Windows 操作系统下运行, 图形化界面, 简单易用。
- 7) 模型传递: 高分辨率, 优异的灵敏度和波长准确度, 轻松实现模型传递。

2.7 小结

近红外光谱分析技术作为一种快速、无损、高效的分析技术, 近些年来受到越来越广泛

近红外光谱技术在食品品质检测方法中的研究

的关注,该技术已在农业、食品、环境科学药物、医学、石化等多种行业中得到了广泛的应用。近红外光谱分析技术在农产品品质检测方面具有很大潜力。近红外光谱分析技术是一种“多、快、好、省”的较为理想的现代分析技术。随着近红外光谱软硬件系统的不断改进和提高,近红外光谱的应用领域不断扩大,在很多领域得到应用。

本章对近红外光谱技术的基本原理、分析方法、预处理方法、建模方法、仪器设备、在农药残留量测量方面的应用前景等做了基础研究。首先介绍了近红外光谱分析的理论基础和基本流程。在基本的理论基础介绍中提到了近红外光谱分析检测技术的两大类,即透射技术和反射技术。并通过两个流程详细地介绍了近红外光谱分析的基本流程,主要包括样品的预处理、模型的建立和预测。然后针对流程中所提到的预处理程序和模型建立,详细介绍了几种常用的预处理方法(如数据规范化处理、光程校正、基线校正、导数法、平滑处理、主成分分析法、马氏距离、杠杆值法、蒙特卡洛交叉验证法、傅里叶变换、正交信号校正、标准正态变量变换等)和模型建立方法(如多元线性回归、主成分回归法、偏最小二乘法、ANN BP法、聚类分析法、支持向量机等)。重点介绍了近红外光谱定标模型的评价指标,主要有测定系数 R^2 、相关系数 R 、交叉验证预测均方差(RMSPCV)、预测标准差(SEP)。最后介绍了近红外光谱仪器的发展和应用,并重点介绍了本研究中应用的 VERTEX 70 傅里叶红外光谱仪和 MPA 多功能傅里叶变换近红外光谱仪的性能指标和优缺点。

参考文献

- [1] 吴谨光. 近代傅里叶变换红外光谱技术及应用 [M]. 北京: 科学技术文献出版社, 1994.
- [2] 罗庆尧. 分光光度分析 [M]. 北京: 科学出版社, 1992.
- [3] 郭尧君. 分光光度技术及其在生物化学中的应用 [M]. 北京: 科学出版社, 1987.
- [4] 张珩. 仪器分析 [M]. 北京: 冶金工业出版社, 1993.
- [5] 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 [D]. 北京: 石油化工科学研究院, 2000.
- [6] 岳永德, 花日茂, 张承祥. 茶叶农药残留与控制 [M]. 北京: 中国农业出版社, 2001.
- [7] 张学博, 冯艳春, 胡昌勤. 近红外多元校正模型传递的进展 [J]. 药物分析杂志, 2009, 29 (8): 1390 - 1399.
- [8] 周振民, 刘狄. 基于 Matlab 的人工神经网络用水量预测模型 [J]. 中国农村水利水电, 2007, 4: 45 - 49.
- [9] 俞汝勤. 化学计量学导论 [M]. 长沙: 湖南教育出版社, 1991.
- [10] 张德然. 统计数据中异常值的检验方法 [J]. 统计研究, 2003, 5 (5): 53 - 55.
- [11] 崔广. 基于近红外光谱分析的纺织品中羊毛含量检测技术研究 [D]. 镇江: 江苏大学, 2007.
- [12] 李水芳, 单杨, 范伟, 等. 基于 MCCV 奇异样本筛选和 CARS 变量选择法对蜂蜜 pH 值和酸度的近红外光谱检测 [J]. 食品科学, 2011, 32 (8): 182 - 184.
- [13] 许禄. 化学计量学方法 [M]. 北京: 科学出版社, 1995.
- [14] 梁逸曾, 俞汝勤. 分析化学手册 (10) - 化学计量学 [M]. 北京: 化学工业出版社, 2001.
- [15] O Svensson, T Kouti, J F MacGregor. An investigation of orthogonal signal correction algorithms and their characteristics [J]. chemometr - ics, 2002, 16: 176 - 188.
- [16] 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 [J]. 光谱学与光谱分析, 2000, 6: 777 - 780.
- [17] 刘树深, 易忠胜. 基础化学计量学 [M]. 北京: 科学出版社, 1999.

- [18] 吴静珠. 农产品品质检测中的近红外光谱分析技术研究 [D]. 北京: 中国农业大学, 2006.
- [19] 夏柏杨, 任芊. 近红外光谱分析技术的一些数据处理方法的讨论 [J]. 光谱实验室, 2005, 22 (3): 629 - 634.
- [20] 董长虹. Matlab 神经网络与应用 [M]. 北京: 国防工业出版社, 2006.
- [21] J Zupan, J Gasteiger. 神经网络及其在化学中的应用 [M]. 潘忠孝, 陈玲然, 译. 合肥: 中国科学技术大学出版社, 2000.
- [22] 张明祥. 近红外仪器能量变化对模型的影响及 OSC 算法的应用 [D]. 北京: 中国农业大学, 2004.
- [23] 张巧杰. 直链淀粉检测方法与技术研究 [D]. 北京: 中国农业大学, 2005.
- [24] Liqun Han. Two Neural Network Based Methods for Leather Pattern Recognition. Proc. of CAIE'98 [M]. 武汉: 华东理工大学出版社, 1998.
- [25] 芦永军, 张军, 朴仁官. 组合生成算法与多元线性回归相结合用于近红外光谱波长的优选 [J]. 分析化学. 2004, 8. 32 (8): 1116 - 1119.
- [26] 郑咏梅, 张军, 陈星旦. 基于逐步回归法的近红外光谱信息提取及模型的研究 [J]. 光谱学与光谱分析. 2004, 6. 24 (6): 675 - 678.
- [27] Edward V, Thomas J. Adaptable Multivariate Calibration Models for Spectral Applications [J]. Analytical Chemical. 2000, 72: 2821 - 2827.
- [28] 张鹏辉, 孙晚华. 多元线性回归在影响因素分析中的应用 [J]. 光谱实验室, 2000, (7): 61 - 62.
- [29] 王丽杰, 郭建英, 徐可欣. 近红外光谱分析中建模样品优选方法的研究 [J]. 红外技术, 2005, 27 (1): 75 - 78.
- [30] Stark E, Luchter K, Margoshoes M. Near - Infrared Analysis: A Technology for Quantitative and Qualitative Analysis [J]. Applied Spectroscopy Review, 1986, 22 (4): 335 - 339.
- [31] Donald A, Burns, Emil W, et al. Handbook of Near - infrared Analysis [M]. New York: Marcel Dekker, Inc, 1992.
- [32] Chen Quansheng, Zhao Jiewen, Liu Muhua, et al. Determination of total polyphenols content in green tea using FI - NIR spectroscopy and different PLS algorithms [J]. Journal of Pharmaceutical and Biomedical Analysis. 2008, (46): 568 - 573.
- [33] Zou Xiaobo, Zhao Jiewen, Huang Xingyi, et al. Use of FT - NIR spectrometry in non - invasive measurements of soluble solid contents (SSC) of 'Fuji', apple based on different PLS models [J]. Chemometrics and Intelligent Laboratory Systems, 2007, (87): 43 - 51.
- [34] 张华秀. 近红外光谱法快速检测牛奶中蛋白质与脂肪含量 [D]. 湖南: 中南大学, 2010.
- [35] 张华秀, 李晓宁, 范伟, 等. 近红外光谱结合 CARS 变量筛选方法用于液态奶中蛋白质与脂肪含量的测定 [J]. 分析测试学报, 2010, 29 (5): 430 - 434.
- [36] Tuxuyan. Information Structure and Pattern of Neural Systems [M]. Beijing: Invited Speech of International Processing, 1995.
- [37] 许国根, 徐萍萍, 谭宪林. MATLAB 在化学中的应用 [M]. 西安: 西安交通大学出版社, 2005.
- [38] Blanco, MCoello J, Iturriaga H, et al. NIR calibration in non - linear systems: different PLS approaches and artificial neural networks [J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50 (1): 75 - 82.
- [39] 王雷, 乔晓艳, 张妹, 等. 基于 BP 神经网络的荧光光谱法农药残留检测 [J]. 应用光学, 2010, 31 (3): 442 - 446.
- [40] 丁丽, 相玉红, 黄安民, 等. BP 神经网络与近红外光谱定量预测杉木中的综纤维素木质素、微纤维素 [J]. 光谱学与光谱分析, 2009, 29 (7): 1784 - 1787.
- [41] 刘波平, 荣茜, 罗香, 等. PLS - BP 法近红外光谱技术同时测定鲜乳中四种主成分 [J]. 分析实验

近红外光谱技术在食品品质检测方法中的研究

- 室, 2009, 28 (6): 66 - 69.
- [42] 李标, 王磊, 朱金营, 等. BP 网络训练函数选取及其岩土工程应用分析 [J]. 西部探矿工程, 2008, 10: 58 - 60.
- [43] Khanchi A R, Mahani M K, Hajihosseini, M, et al. Simultaneous spectrophotometric determination of caffeine and theobromine in Iranian tea by artificial neural networks and its comparison With PLS [J]. Food Chemistry, 2007, 103 (3): 1062 - 1068.
- [44] Vapnik V. Statistical Learning Theory [M]. New York: John Wiley and sons Inc. , 1998.
- [45] 刘爱华, 傅雪海, 王可新, 等. 支持向量机预测煤层含气量 [J]. 西安科技大学学报, 2010, 30 (3): 309 - 312.
- [46] 史峰, 王小川, 郁磊, 等. MATLAB 神经网络 30 个案例分析 [M]. 北京: 北京航空航天大学出版社, 2010.
- [47] 于晓辉, 张卓勇, 马群, 等. 径向基函数神经网络和近红外光谱用于大黄中有效成分的定量预测 [J]. 光谱学与光谱分析, 2007, 27 (3): 481 - 485.
- [48] 刘萍. 有机磷农药快速检测技术及应用研究 [D]. 重庆: 西南大学, 2006.
- [49] 曾铭, 李树伟, 李小燕. 有机磷农药残留量检测方法研究进展 [J]. 重庆环境科学, 2005 (5): 20 - 22.
- [50] 徐广通, 袁洪福, 陆婉珍. 近红外光谱仪器概况与进展 [J]. 现代科学仪器, 1997, 3: 9 - 11.
- [51] 王宗贤, 高志贤, 马成林. 有机磷检测方法的研究进展 [J]. 中国卫生检验杂志. 2003, 13 (4): 401 - 403.

3.1 简介

随着我们国民经济的不断发展，特别是全球经济一体化和食品贸易国际化，食品安全问题已经成为举足轻重的问题。2009年6月1日，《中华人民共和国食品安全法》的正式实施，更是开启了我国食品安全监管的一个新阶段。因此，加快食品安全快速检测技术的研究和发展，构建食品安全与质量保障体系，对加速我国农产品和食品流通领域的现代化、提高农产品和食品的国际竞争力，具有非常重要的现实意义和深远的战略意义。

目前我国食品安全的主要问题是农药残留、兽药残留、水产品药物残留、食品掺假等食品源头污染，其中农药残留问题最为突出。据统计，我国每年因农药引起的食物中毒事故数量居化学性食物中毒问题之首，而蔬菜农药残留超标更是首当其冲。对于蔬菜等农产品，农业部曾颁布法令条文禁止使用高毒高残留农药，但据农业植保部门按户次调查统计，番茄、黄瓜、菜椒、豆角等主要蔬菜种类，使用过高毒农药的种植户达到32.8%。特别是国家已经明令禁止使用的甲胺磷，因其产量大、价格低、药效好等特点，部分菜农常常偷着使用，在叶菜上的使用率甚至高达50%~60%。1996~1997年广州市分别随机抽查市场中蔬菜181份和951份，超标率分别为16.02%和12.72%。1999年11~12月湖北省药检所对武昌水果湖、大东门、马房山、汉口红桥等集贸市场的22个蔬菜品种进行有机磷农药残留检测，结果超标率高达30%，农药残留量最高达到惊人的 10×10^{-6} 。2000年5月农业部对京、沪、浙、渝、鲁5省市50多个蔬菜品种1293份样品进行抽检，合格率仅仅为78%。2001年第一、第三季度国家质检总局对全国蔬菜质量进行抽查的结果显示：农药残留问题最为突出，抽样合格率分别仅为70.7%和52.5%。2002年孙鑫贵等人在北京市抽检281个蔬菜样品，检测出12种有机磷农药，其中禁用农药占53%，说明禁用农药在蔬菜中的使用情况十分严重。2004年10月~2005年4月郭忠等人先后两次对唐山市内蔬菜市场随机抽取了23个品种的175份新鲜蔬菜，结果显示韭菜和菜花的农药残留超标最为严重，超标率高达80%。2005年1月，湖南桑植县澧源镇第一小学，829名学生因食用的四季豆农药残留超标而出现中毒症状。同年4月，广州石滩镇三江第二中学食堂青菜农药残留超标，造成30多名学生出现急性头晕、呕吐等症状。

尽管受到食品安全相关的法律法规和监督政策的制约和监督，食品生产规模和食品进出口贸易的不断扩大，使得部分以营利为目的的黑心商家顶风做浪，生产并销售有问题食品，严重危害了消费者的健康和利益，降低了人民大众对食品安全的信心。其中，为了消除农产品病虫害问题，某些农户和商家大量使用高毒、剧毒农药，这些农药的过量喷洒，导致瓜果蔬菜和主要日常农作物农药残留严重超标。面对如此严峻的食品安全形势，加大宣传力度、完善法律法规的同时，加强农药残留检测的监督，研究新型检测方法是从源头上解决农药残

留问题的当务之急。

农药作为化学工业的重要成果，为全世界的粮食生产作出了巨大的贡献。农作物在生长的过程中，病、虫、草害严重威胁着其产量，若不加以防治，粮食产量相对于现在的收成水平将大大降低。然而，由于部分农药的高毒和高残留性，以及种植人员和商贩在使用过程中不按标准方法喷洒或过度使用，累积效应使得部分农产品及粮食作物农药残留量严重超标。因食用被残留农药污染的食品导致中毒的重大事件近年来多有报道，2009年4月，沈阳晚报的一篇新闻报道记载了因食用有农药残留的韭菜而使六岁儿童致死事件。我国因农药残留量过多而导致的食物中毒事件方面的记载很多，食用农药残留超标食品引起的轻度中毒表现为头痛、恶心等，中度和重度将会威胁到生命。这些重大的事件为人们敲响了警钟，粮食农药残留超标问题对人体健康的严重危害越来越多被全世界所重视。

由于中毒事件频发，我国政府相关部门正在大力研究和并努力制定更完善的农药残留限量，各大监督管理部门也紧锣密鼓地忙于农药残留检测和防止超标的工作。促进法律法规在农药残留监督中更加完善并且提高监督管理力度是必不可少的防治方法，其中，从实际现状出发，为了从源头上杜绝农药的滥用，使黑心商贩和粮食种植户提高合理、合法、按规定使用农药的意识，必须有针对农药残留量的有效、便捷检测方法。在收购瓜果蔬菜等粮食作物的第一线，如果能同时快速检测所有农产品的农药残留量，拒绝收购农药残留量超标的粮食作物，将大大降低或避免因农药残留量超标而威胁消费者生命安全的事情发生，因此，要同时重点发展限量标准的研究、检测体系的完善和适用于现场农药残留量检测的新方法的研究。

另外，化学工业的发展日新月异，结构各异、毒性不同的农药新品种如雨后春笋般出现在种植户的面前，但是农药残留检测方法的覆盖面却很窄。针对如此花样繁多的农药产品，生物-化学检测方法检测农药残留量的不方便性和耗时长弊端无法使其适用于购买一线现场的检测，因此使农药残留量检测更快速、更简便、更准确是其重要的发展方向。而光谱法恰好在这个方面弥补了化学方法的繁杂检测过程，在一线检测的优点很明显。然而，尽管有如此独特的优势，光谱检测方法的各项标准却没有随着农药残留检测技术的发展而同步更新。农药残留检测的光谱检测方法因光谱数据采集、光谱数据预处理、样品模型建立方法的各项标准选择的不同，可能会导致结果的不同，而且不同仪器建立的模型涉及的数据匹配等模型转移问题也需要进一步研究和完善。面对如此种类繁多的农药品种和农作物粮食品种，建立光谱标准库也是一个巨大的工程。

目前，研究光谱技术检测痕量农药残留的相关文献报道比较少，有很多使用物理-化学方法进行农残检测的文献记载，物理-化学方法的弊端却使其在实时检测和现场大批量检测中受到限制。本书通过研究多光谱检测方法在农药残留量检测中的应用，希望利用实验的分析结果找到光谱法在痕量农药残留检测领域的可行性，并对各光谱检测方法及其不同建模方法的实验结果进行研究分析和对比，旨在为不同样品找到最优的光谱检测方法和最佳的建模方法，本章的研究不仅在理论上有其创新性，在农药残留检测领域应用前景方面也有其现实意义。

3.2 基于近红外光谱的溶液中毒死蜍 PLS 定量分析方法研究

毒死蜍 (Chlorpyrifos), 是美国陶氏化学公司开发研制出来的一种高效、广谱, 中等毒性的有机磷杀虫、杀螨剂, 并获得专利 (USP3244586)。毒死蜍能抑制人体胆碱酯酶的活性, 蓄积于神经系统后会导致恶心、头晕, 甚至神志不清, 高浓度暴露可造成呼吸麻痹或死亡。毒死蜍在许多国家被广泛应用, 反复使用可使毒死蜍经由土壤渗入地下水中, 而毒死蜍在水中较稳定, 消解较慢, 农业灌溉区附近的河水中常可检测出毒死蜍。如果这样的地下水用于饮用, 则有必要对毒死蜍的浓度进行监测。毒死蜍的含量分析常用气相色谱、HPLC 以及 GC-MS 等方法, 但传统的分析方法前处理时间长, 且需消耗大量的溶剂。本节采用化学计量学方法对溶液中毒死蜍在近红外区域的特征吸收进行解析, 得到其特征吸收的波段范围, 并采用 PLS 和 ANN BP 法, 分别建立和比较了测定溶液中毒死蜍含量的近红外光谱分析模型, 为快速准确测定溶液中残留毒死蜍提供新的思路和方法。

3.2.1 样品制备

实验所用的 209 个含毒死蜍溶液样品由化学方法配置而成。

第一组样品范围: 3.23 ~ 10mg/kg 的样品 27 个, 成分: 标准物质浓度为 1mg/mL 的毒死蜍和甲醇混合液、甲醇及水的混合溶液, 甲醇和水作为稀释溶液, 见附表 1。

第二组样品范围: 3.13 ~ 11.1mg/kg 的样品 28 个, 成分: 浓度为 1mg/mL 的毒死蜍和甲醇溶液的标准物质、蔗糖、维生素 C 及甲醇和水混合而成的溶液, 甲醇和水作为稀释溶液, 见附表 2。

第三组样品共 88 个: 使用其中 37 个样品用于建模, 其他作为预测样品。浓度范围为 0.005 ~ 0.1mg/kg, 成分: 标准物质浓度为 1mg/mL 的毒死蜍和甲醇混合液、甲醇、蔗糖、维生素 C 及饮用水的混合溶液, 甲醇和水作为稀释溶液, 见附表 3。

第四组为第三组重现实验样品 66 个: 使用其中 37 个样品用于建模, 其他作为预测样品。其浓度为 0.005 ~ 0.1mg/kg, 成分: 标准物质浓度为 1mg/mL 的毒死蜍和甲醇混合液、甲醇、蔗糖、维生素 C 及饮用水的混合溶液, 甲醇和水作为稀释溶液, 见附表 4。

3.2.2 光谱采集

第一组和第二组样品采用的是 MPA 近红外光谱分析仪 (德国, Bruker 公司) 进行全谱测定。检测器采用铟镓砷 (InGaAs), 液体光纤探头长 2m, 光程池 1mm, 透反射测量有效光程为 2mm。光谱采集与分析软件采用 OPUS5.0。

将所配不同浓度的溶液样品分别装入 25mm 茶色小瓶中, 测量时, 在室温下将样品溶液滴入液体光纤探头的光程池中, 采用透反射采样模式, 对 12500 ~ 4000cm⁻¹ 谱区扫描, 分辨率为 8cm⁻¹, 扫描 32 次。

第三组和第四组的重复实验组样品采用的是 MATRIX™ - F 近红外光谱分析仪 (德国, Bruker 公司) 进行全谱测定。检测器采用数字检测器, 液体光纤探头长 2m, 光程池 1mm, 透反射测量有效光程为 2mm。光谱采集与分析软件采用 OPUS5.0。

近红外光谱技术在食品品质检测方法中的研究

将所配不同浓度的溶液样品分别装入 25mm 茶色小瓶中，测量时，在室温下将样品溶液滴入液体光纤探头的光程池中，采用透反射采样模式，对 $12500 \sim 4000\text{cm}^{-1}$ 谱区扫描，分辨率为 8cm^{-1} ，扫描 32 次。

3.2.3 结果与分析

在采用偏最小二乘适合全谱解析的因子分析技术时一般不需要对波长范围进行选择，但在建模前，如果对建模的光谱谱区进行选择，将有利于减少噪声影响，提高运算效率和模型的稳定性。研究表明，采用全谱建立校正模型时，不仅计算工作量大，而且校正模型的预测精度未必能达到最佳值。波长选择可以从得到的光谱中提取最有效的光谱信息，提高校正模型的精度，增强校正模型的预测能力，且简化了运算。因此，波长选择是建立稳定可靠的校正模型的重要手段。

波长选择的物理依据和意义可以从以下方面来看，从光谱本身看：有些波段样品的有效光谱信息很弱，不能提供与样品结构有关的信息；有些波段光谱信息存在线性相关的现象，光谱信息中有冗余；有些波段光谱与样品的组成或性质之间缺乏相关关系；有些波段光谱数据通常存在严重的重叠现象，使得这些波长处的光谱信息的灵敏度较差，直接影响模型的预测精度。由于仪器或检测器件本身不够稳定，使得扫描的光谱往往带有噪声，且每个谱区的噪声水平不一致，有些波段的信噪比较低，光谱质量差，如果利用全谱建模，不可避免的引入了噪声，降低了模型的精度。从计算量上看，波长选择减少了参与建模的波长点数，降低了建模和预测运算时间，提高了测量速度。从被测对象本身看，每一种组分都会在某一段或者某几段谱区对近红外光谱有较强的吸收，此即该组分的特征谱区，而在特征谱区外，该组分的吸收很弱，因此在建模时，从全谱中挑选出与样品待测组分响应大的光谱区间建模是十分必要的。

图 3-1 所示为第一组毒死蜱浓度为 $3.23 \sim 10\text{mg/kg}$ 的混合溶液（毒死蜱、甲醇及水）样品的近红外全谱图。图 3-2 所示为近红外光谱交叉验证预测值与化学值。

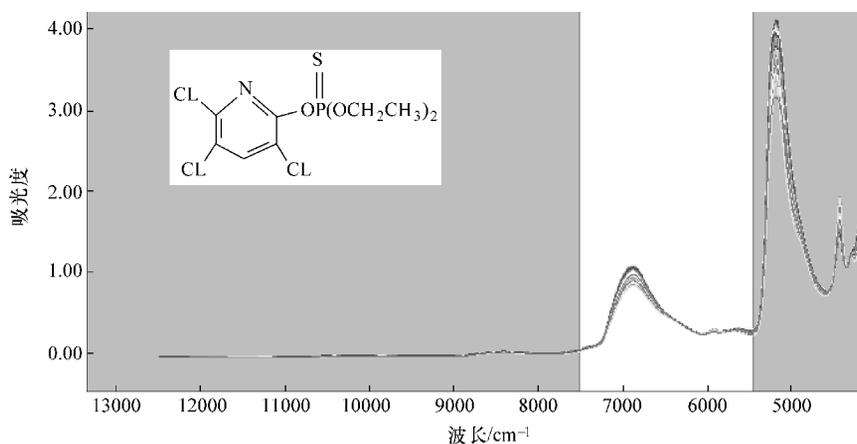


图 3-1 第一组样品毒死蜱、甲醇及水的混合溶液近红外光全谱图

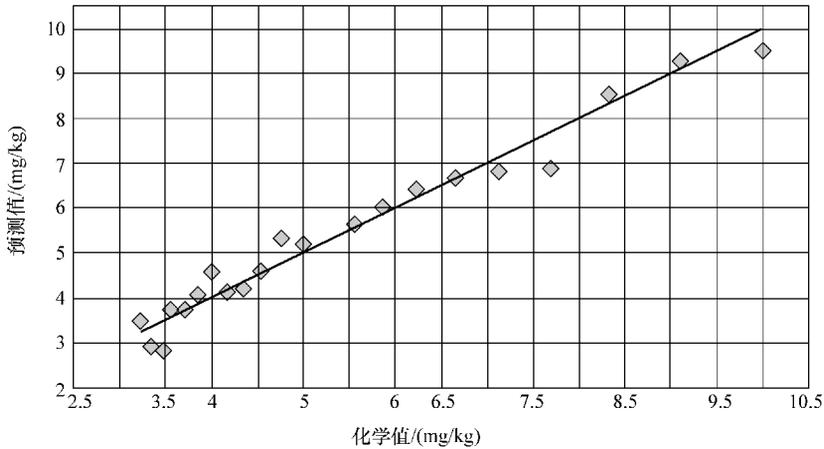


图 3-2 第一组样品近红外光谱交叉验证预测值与化学值

表 3-1 为第一组样品毒死蜱、甲醇及水的混合溶液中毒死蜱含量校正与预测结果。

表 3-1 第一组样品毒死蜱、甲醇及水的混合溶液中毒死蜱含量校正与预测结果

检测成分	校正		交叉验证	
	RMSEE/(mg/kg)	R^2	RMSEV/(mg/kg)	R^2
毒死蜱	0.126	0.9979	0.342	0.9698

通过选择建模的最佳光谱预处理方法、谱区范围，使用交互验证法确定主因子数为 4。对第一组样品数据采用二阶导数 (Second Derivative)，17 点平滑做预处理，毒死蜱对光谱信息贡献量最大的谱区范围是 $7502.2 \sim 5446.3 \text{ cm}^{-1}$ 。

图 3-3 所示为第二组毒死蜱浓度为 $3.13 \sim 11.1 \text{ mg/kg}$ 的混合溶液 (毒死蜱、维生素 C、蔗糖，甲醇及水) 样品的近红外全谱图，图 3-4 所示为近红外光谱交叉验证预测值与化学值，表 3-2 毒死蜱、蔗糖、维生素 C、甲醇及水的混合溶液中毒死蜱含量校正与交叉验证误差结果。

表 3-2 第二组样品毒死蜱、蔗糖、维生素 C、甲醇及水的混合溶液中毒死蜱含量校正与交叉验证误差结果

检测成分	校正		交叉验证	
	RMSEE/(mg/kg)	R^2	RMSEV/(mg/kg)	R^2
毒死蜱	0.2	0.9954	0.89	0.8803

对第二组样本数据采用一阶导数 (First Derivative)，21 点平滑做预处理，毒死蜱对光谱信息贡献量最大的谱区范围是 $9998.3 \sim 7497.9 \text{ cm}^{-1}$ 和 $5778.3 \sim 5442.8 \text{ cm}^{-1}$ ，主因子数为 5。回归统计方法均采用偏最小二乘法。

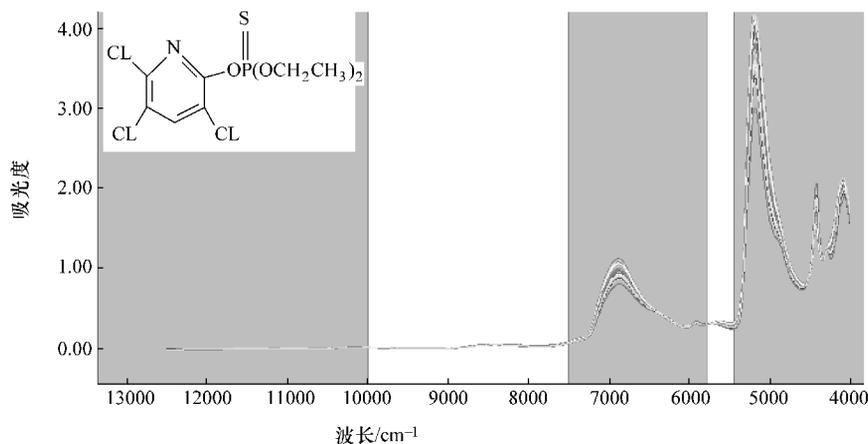


图 3-3 第二组样品毒死蜱、蔗糖、维生素 C 及甲醇和水的溶液的近红外光全谱图

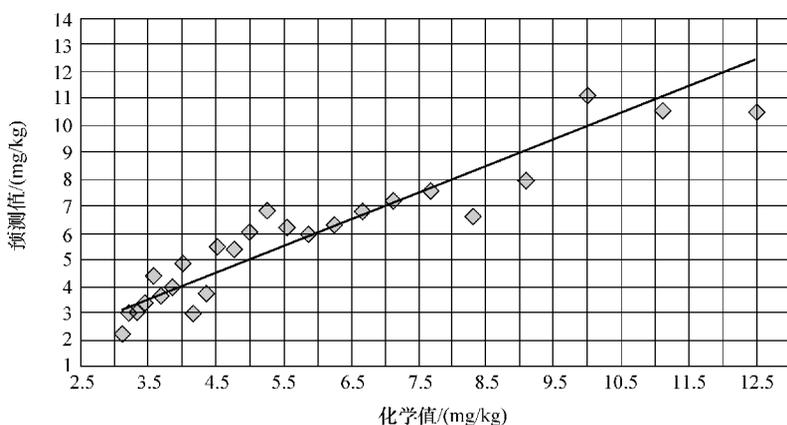


图 3-4 第二组样品近红外光谱交叉验证预测值与化学值

图 3-5 所示为第三组毒死蜱浓度为 0.005 ~ 0.1mg/kg 的混合溶液（毒死蜱、维生素 C、蔗糖，甲醇及水）样品的近红外全谱图，图 3-6 所示为近红外光谱交叉验证预测值与化学值。

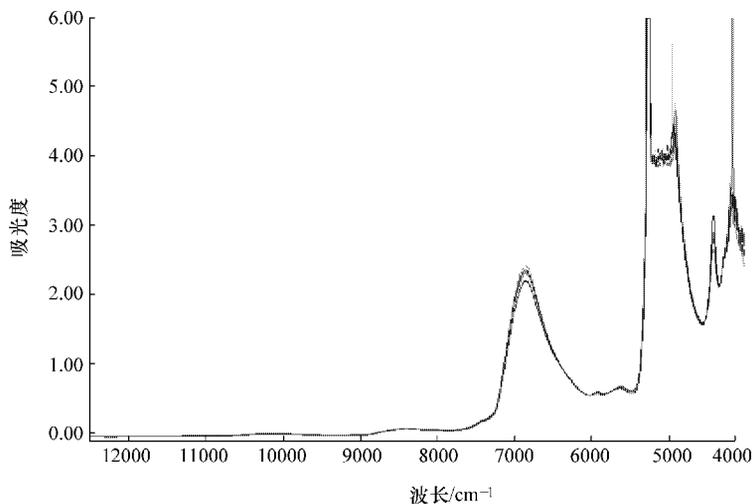


图 3-5 第三组样品毒死蜱、蔗糖、维生素 C 及甲醇和水溶液的近红外光全谱图

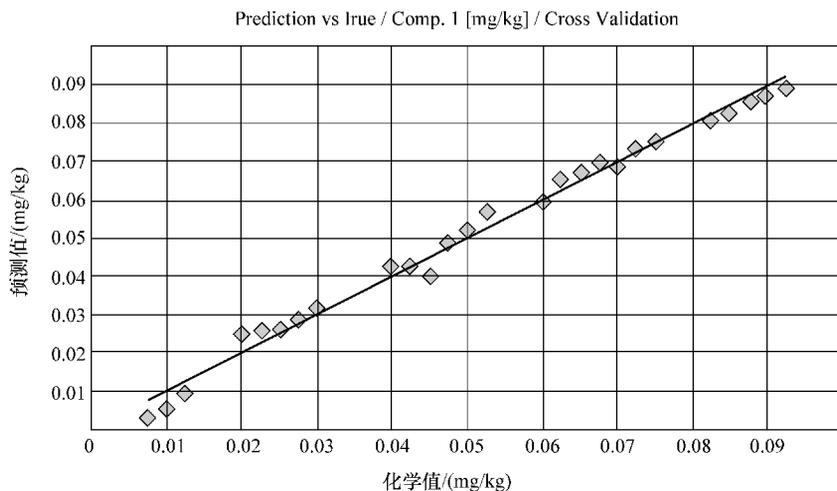


图 3-6 第三组样品近红外光谱交叉验证预测值与化学值

不对数据实行预处理，毒死蜱对光谱信息贡献量最大的谱区范围是 $7240.4 \sim 9088.1 \text{ cm}^{-1}$ 和 $5323.3 \sim 6102.9 \text{ cm}^{-1}$ ，使用交互验证法确定主因子为 8。回归统计方法采用偏最小二乘 (partial least squares, PLS) 法。表 3-3 为混合溶液中毒死蜱含量校正与预测结果。

表 3-3 第三组样品毒死蜱、蔗糖、维生素 C、甲醇及水的混合溶液中毒死蜱含量校正与预测结果

检测成分	校正		交叉验证		外部验证	
	RMSEE/(mg/kg)	R^2	RMSEV/(mg/kg)	R^2	RMSEP/(mg/kg)	R^2
毒死蜱	0.0016	0.996	0.0021	0.9911	0.0023	0.9974

3.2.4 重现性实验结果与分析

在改变了稀释溶液的配比的情况下，配制样品浓度仍然在 $0.005 \sim 0.1 \text{ mg/kg}$ 。其他测量条件相同。

不对数据实行预处理，毒死蜱对光谱信息贡献量最大的谱区范围是 $7853.4 \sim 8913 \text{ cm}^{-1}$ 和 $5415 \sim 6244 \text{ cm}^{-1}$ ，使用交互验证法确定主因子为 8。回归统计方法采用偏最小二乘方法。第三组样品重现性实验数据集 PLS 模型如图 3-7 所示。模型的评价指标参数见表 3-4。

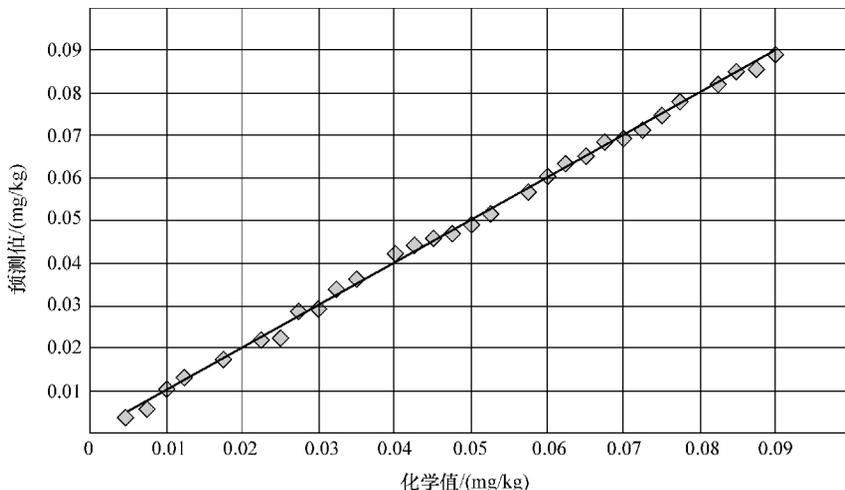


图 3-7 31 个第三组重现样品的 PLS 模型

表 3-4 第三组重现样品的毒死蜱含量校正与预测结果

检测成分	校正		交叉验证		外部验证	
	RMSEE/(mg/kg)	R^2	RMSEV/(mg/kg)	R^2	RMSEP/(mg/kg)	R^2
毒死蜱	0.0007	0.9993	0.0011	0.9981	0.0021	0.9961

重现性实验表明近红外光谱法在测定混合液中的毒死蜱浓度为 0.008 ~ 0.09mg/kg 时, 样品的预测值和化学值之间均有很好的相关性。

3.3 基于近红外光谱的溶液中毒死蜱 BP 网络定量分析方法研究

3.3.1 BP 算法的实验设计

1. 网络结构的设计

一般情况下, 三层 BP 网络即可。因此实验采用三层的网络机构来建立 BP 网。隐层的传递函数采用对数 Sigmoid 型传递函数 logsig。隐层的神经元个数通过编写程序实现, 将神经元个数限制在 3 ~ 20 个, 训练过程隐层节点个数在 3 ~ 20 个之间选择, 最后保留最优网络。对于最后一层神经元如果采用 Sigmoid 型函数, 那么整个网络输出被限制在一个较小的范围内; 如果最后一层神经元采用 purelin 型函数, 则整个网络输出可以取任意值, 根据实验的具体情况 (预测毒死蜱的含量), 输出函数采用 purelin 型函数, 一个输出。

2. ANN BP 的训练和仿真实现工具

为了使 ANN BP 能更快、更好地应用于近红外光谱定量分析, 研究在操作系统 Windows XP 下, 借助于 MATLAB 7.1 工具箱实现仿真。实验分别比较了 6 种算法训练得到的 ANN BP 性能。

3.3.2 样品制备

输入的训练样品为 31 个含毒死蜱的混合溶液, 样品标号为 36 ~ 72# (见附表 3), 浓度为 0.1 ~ 3.6mg/kg。5 个样品作为预测集样品。

3.3.3 结果与分析

实验分别比较了如下 6 种 BP 算法训练得到的 ANN BP 的训练过程及结果:

- 1) 有动量的梯度下降法 (对应 ANN 工具箱函数: traindm);
- 2) 有自适应 lr 的梯度下降法 (对应 ANN 工具箱函数: traingda);
- 3) 能复位的 BP 训练法 (对应 ANN 工具箱函数: trainrp);
- 4) 量化共轭梯度法 (对应 ANN 工具箱: trainscg);
- 5) 一步正割的 BP 训练法 (对应 ANN 工具箱: trainnoss);
- 6) Levenberg - Marquardt 训练法 (对应神经网络工具箱: trainlm)。

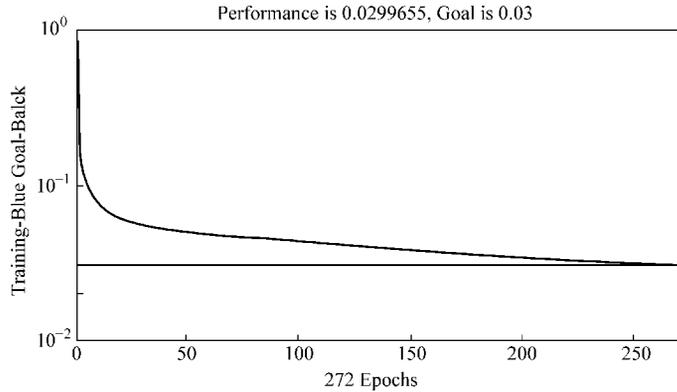
在每个训练方法的训练过程中, 均保留了较好的 (网络) 分析结果图。以下是各算法的训练结果。

图 3-8 所示为动量的梯度下降法的训练过程及分析结果图。

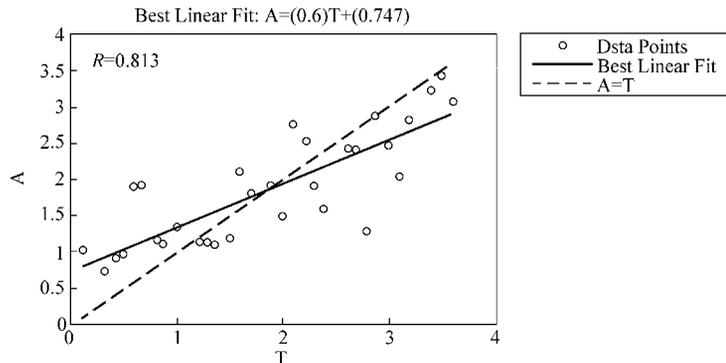
图 3-9 所示为自适应 lr 的梯度下降法的训练过程及分析结果图。

图 3-10 所示为能复位的 BP 训练法的训练过程及分析结果图。

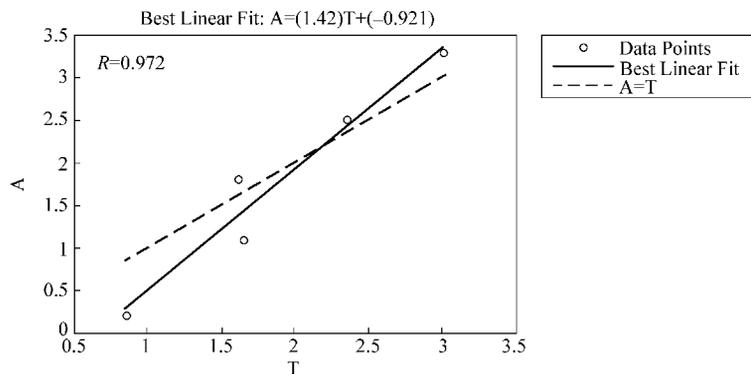
图 3-11 所示为量化共轭梯度法的训练过程及分析结果图。



a) 动量法训练过程



b) 动量梯度法训练集预测值 - 目标值关系图

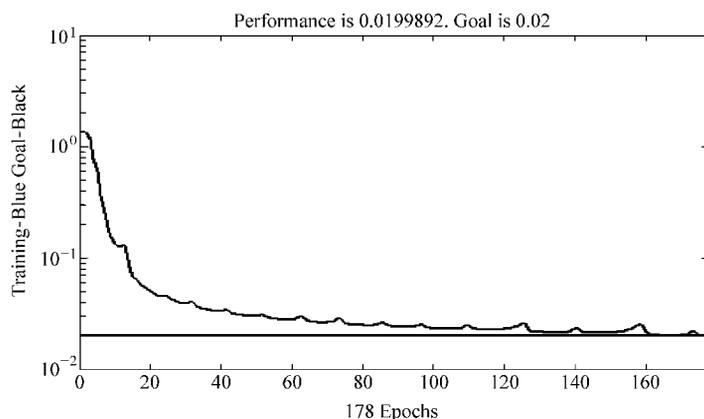


c) 动量梯度法验证集预测值 - 目标值关系图

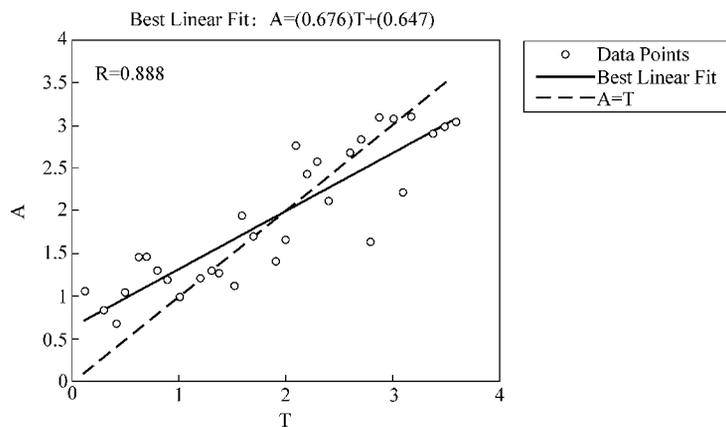
图 3-8 动量的梯度下降法的训练过程及分析结果图

图 3-12 所示为一步正割的 BP 训练法的训练过程及分析结果图。

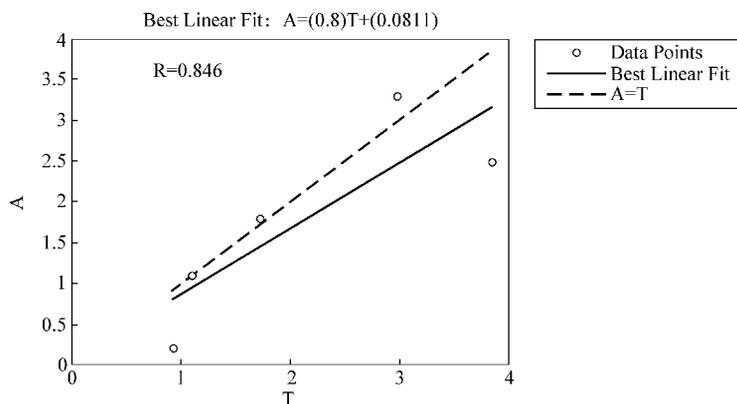
图 3-13 所示为 Levenberg – Marquardt 训练法的训练过程及分析结果图。



a) 自适应lr梯度法训练过程

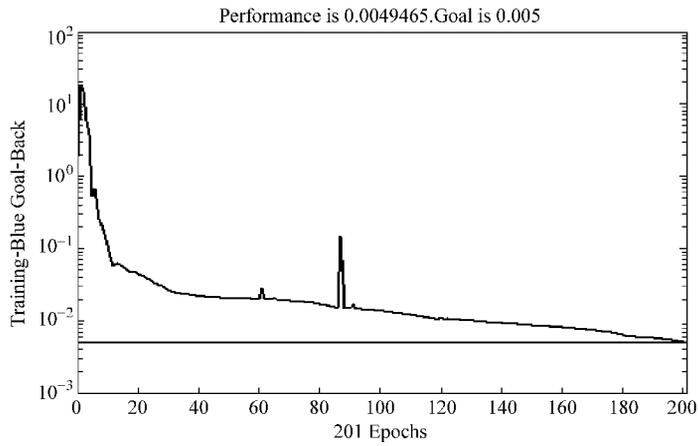


b) 自适应lr梯度法训练集预测值-目标直线性回归分析

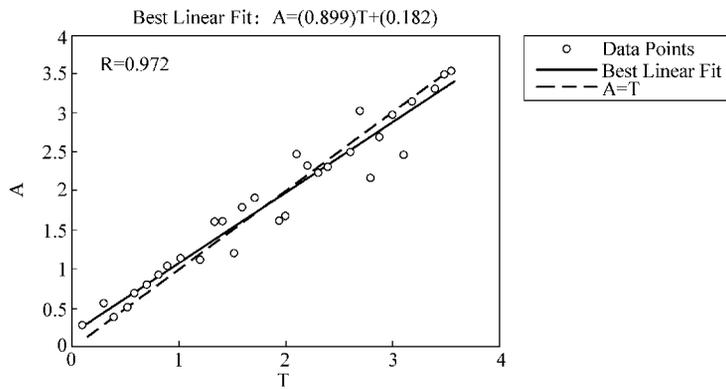


c) 自适应lr梯度法验证集预测值-目标值关系图

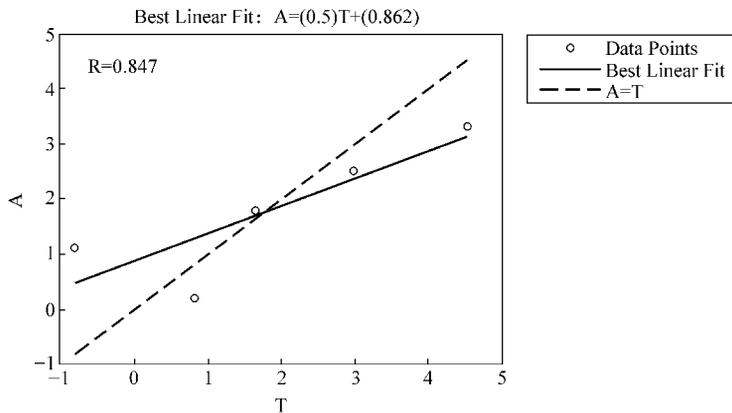
图 3-9 自适应 lr 的梯度下降法的训练过程及分析结果图



a) 能复位的BP训练法训练过程

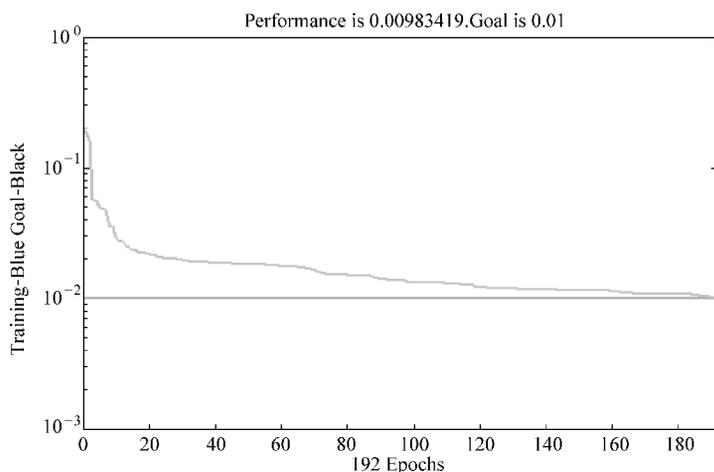


b) 能复位的BP训练法校正集样品预测值-目标值线性回归分析

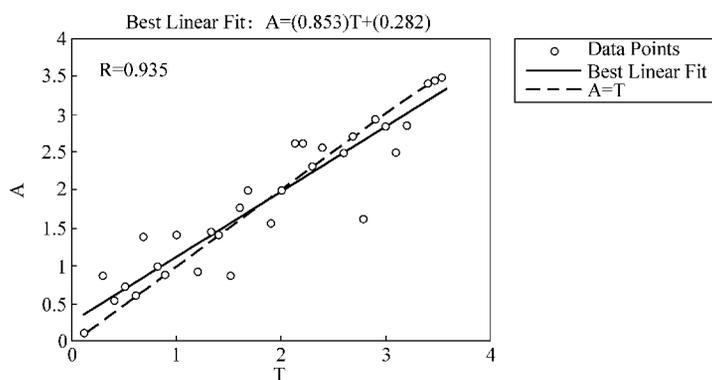


c) 能复位的BP训练法验证集样品预测值-目标值线性回归分析

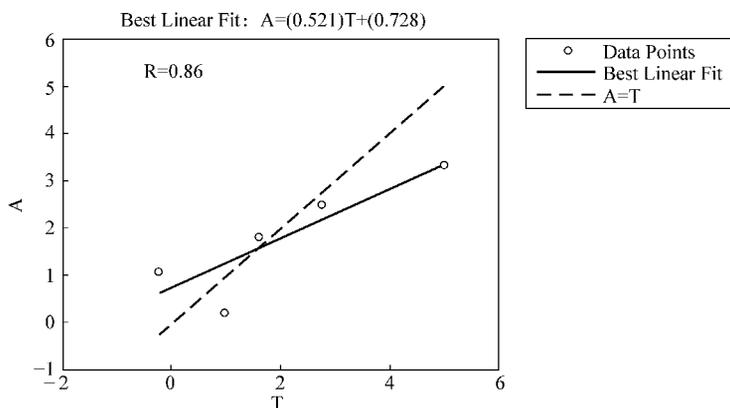
图 3-10 能复位的 BP 训练法的训练过程及分析结果图



a) 量化共轭梯度法训练过程

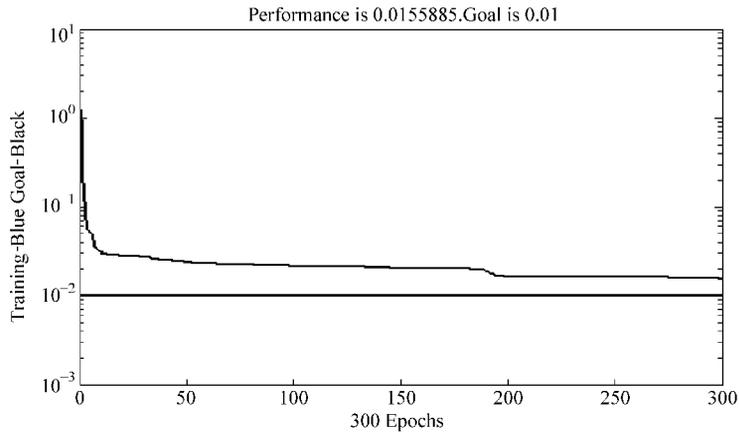


b) 量化共轭梯度法训练集样品预测值-目标值性回归分析

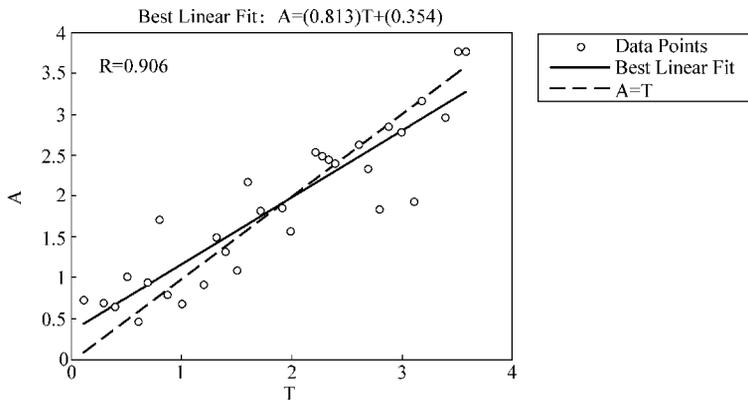


c) 量化共轭梯度法验证集样品预测值-目标值线性回归分析

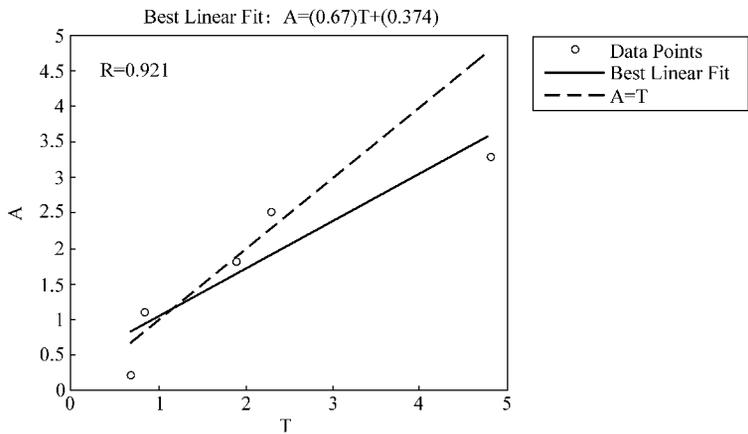
图 3-11 量化共轭梯度法的训练过程及分析结果图



a) 一步正割的BP算法训练过程

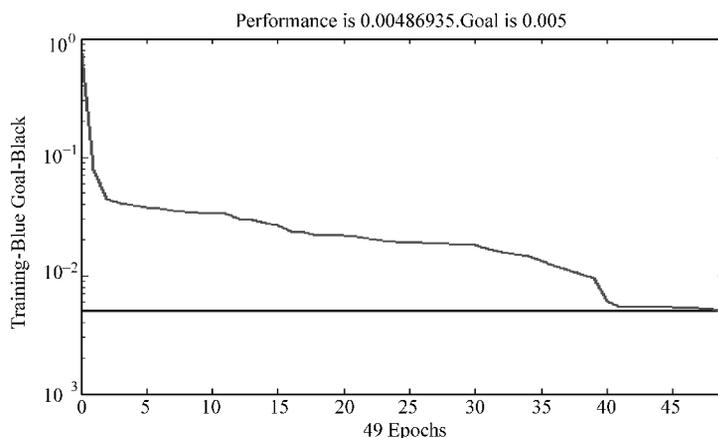


b) 一步正割的BP训练法训练集样品的预测值-目标值线性回归分析

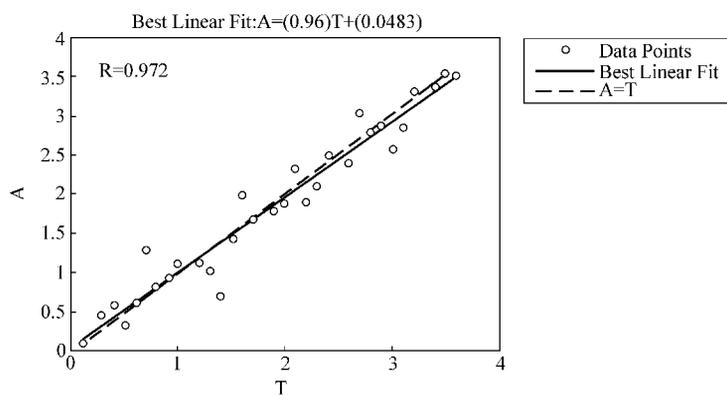


c) 一步正割的BP训练法验证集样品的预测值-目标值线性回归分析

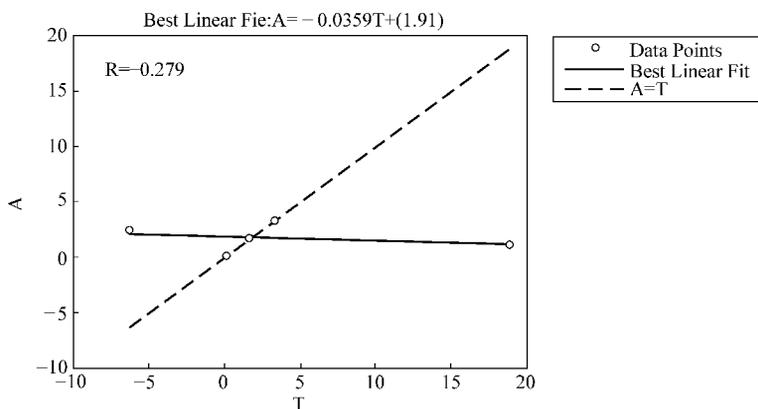
图 3-12 一步正割的 BP 训练法的训练过程及分析结果图



a) Levenberg-Marquardt法训练过程



b) Levenberg-Marquardt法训练集样品预测值-目标值线性回归分析



c) Levenberg-Marquardt法验证集样品预测值-目标值线性回归分析

图 3-13 Levenberg-Marquardt 训练法的训练过程及分析结果图

6种算法下训练得到的 ANN BP 见表 3-5。

表 3-5 6种算法下训练得到的 ANN BP

BP 算法	隐层节点个数	训练步数	R^2	RMSEC	RMSEP
有动量的梯度下降法	11	204	0.813	0.6061	0.8517
有自适应 lr 的梯度下降法	17	178	0.888	0.4948	0.9165
能复位的 BP 训练法	14	201	0.972	0.2462	1.0766
量化共轭梯度法	18	192	0.853	0.3675	1.0395
一步正割的 BP 训练法	14	192	0.86	0.4369	0.7178
Levenberg - Marquardt 训练法	5	49	0.972	0.2443	8.9037

前 5 种算法在一定程度上均提高网络的训练速度，且模型的自预测能力较高。尤其是一步正割的 BP 训练法训练速度最快，模型最稳定。Levenberg - Marquardt 训练法占用内存空间大，收敛很慢，因此训练过程时间长。如果内存有限，可以采用其他收敛速度较快的算法。对于大型网络，就要采用其他训练函数，例如 `trainscg` 或者 `trainrp`。

基于经验风险最小化原则的神经网络学习过程中过分强调得到较小的训练误差，导致泛化能力的下降，因此训练得到的 ANN BP 对验证集样品的预测能力差。

3.4 溶液中毒死蜱的 PLS 和 BP 定量分析模型性能的比较

3.4.1 PLS 法的样品选择

样品采用与 3.3 节相同的样品，即 36 ~ 72# 样品（附表 3）。样品浓度为 0.1 ~ 3.6 mg/kg 的混合液样品（毒死蜱、水、甲醇、维生素 C、蔗糖）。

对样品数据采用一阶导数，17 点平滑做预处理，毒死蜱对光谱信息贡献量最大的谱区范围是 7842.9 ~ 9497.3 cm^{-1} 和 5446.8 ~ 6102.3 cm^{-1} ，主因子数为 5。回归统计方法采用 PLS 法。

3.4.2 BP 法的样品选择

采用 3.3 节的样品分析结果与 PLS 法做比较。

3.4.3 比较分析结果

由比较结果可以看出（见表 3-6），PLS 法在分析基于近红外光谱的微量成分检测模型预测精度时要高于 BP 法。

表 3-6 BP 算法训练的最佳网络与 PLS 模型的性能比较结果分析

算法名称	校正		交叉验证	
	RMSEE/(mg/kg)	R^2	RMSEV/(mg/kg)	R^2
BP 法	0.2462	0.972	1.076	0.847
PLS 法	0.119	0.9870	0.060	0.9981

3.5 基于近红外光谱的溶液中毒死蜍 PLS 定性分析方法研究

3.5.1 样品制备

实验所用的 50 个含毒死蜍溶液样品由化学方法配置而成。

不含毒死蜍的样品：无毒死蜍样品 10 个。成分：甲醇、蔗糖、维生素 C 及饮用水的混合溶液，甲醇和水作为稀释溶液。

含毒死蜍的样品：标准物质浓度为 1mg/mL 的毒死蜍和甲醇混合液、甲醇、蔗糖、维生素 C 及饮用水的混合溶液，甲醇和水作为稀释溶液。浓度为 0.005 ~ 0.0525mg/kg 的样品 20 个，即 1 ~ 20#样品（见附表 3），称为第一组样品。

重现实验样品：标准物质浓度为 1mg/mL 的毒死蜍和甲醇混合液、甲醇、蔗糖、维生素 C 及饮用水的混合溶液，甲醇和水作为稀释溶液，浓度为 0.005 ~ 0.0525mg/kg 的样品 20 个，即 1 ~ 20#样品（见附表 4），称为第二组样品。

3.5.2 光谱采集

所有样品均采用的是 MATRIX™ - F 近红外光谱分析仪（德国，Bruker 公司）进行全谱测定。检测器采用数字检测器，液体光纤探头长 2m，光程池 1mm，透反射测量有效光程为 2mm。光谱采集与分析软件采用 OPUS5.0。

将所配不同浓度的溶液样品分别装入 25mm 茶色小瓶中，测量时，在室温下将样品溶液滴入液体光纤探头的光程池中，采用透反射采样模式，对 12500 ~ 4000 cm^{-1} 谱区扫描，分辨率为 8 cm^{-1} ，扫描 32 次。

3.5.3 结果与分析

采用模式识别法中的聚类分析法进行定性分类鉴别，分别对毒死蜍浓度范围在 0.005 ~ 0.275mg/kg 的 10 个样品和 10 种不含毒死蜍的溶液样品进行分类，及 10 个毒死蜍浓度为 0.03 ~ 0.0525mg/kg 的 10 个样品和 10 种不含农药的样品进行分类。两次分类中 10 个不含农药的样品是同一组样品。以树状图形式表示其分类结果。采集的光谱图经过 9 点平滑后求一阶导数，去除干扰信号，光谱范围选择 9073 ~ 7692 cm^{-1} 和 6527 ~ 5354 cm^{-1} 。使用 OPUS5.0 多元统计定性分析软件进行聚类分析：样品间距离采用欧氏距离算法，类间采用 Ward 算法。分类正确率为 100%。分类结果如图 3-14 和图 3-15 所示。

3.5.4 重现性实验结果与分析

采用聚类分析法进行定性分类鉴别，分别对毒死蜍浓度范围在 0.005 ~ 0.275mg/kg 的 10 个样品和 10 种不含毒死蜍的溶液样品进行分类，及 10 个毒死蜍浓度为 0.03 ~ 0.0525mg/kg 的 10 个样品和 10 种不含农药的样品进行分类。两次分类中 10 个不含农药的样品是同一组样品。以树状图形式表示其分类结果。采集的光谱图经过 9 点平滑后求一阶导数，去除干扰信号，光谱范围选择 8913 ~ 7853 cm^{-1} 和 6244 ~ 5415 cm^{-1} 。使用 OPUS5.0

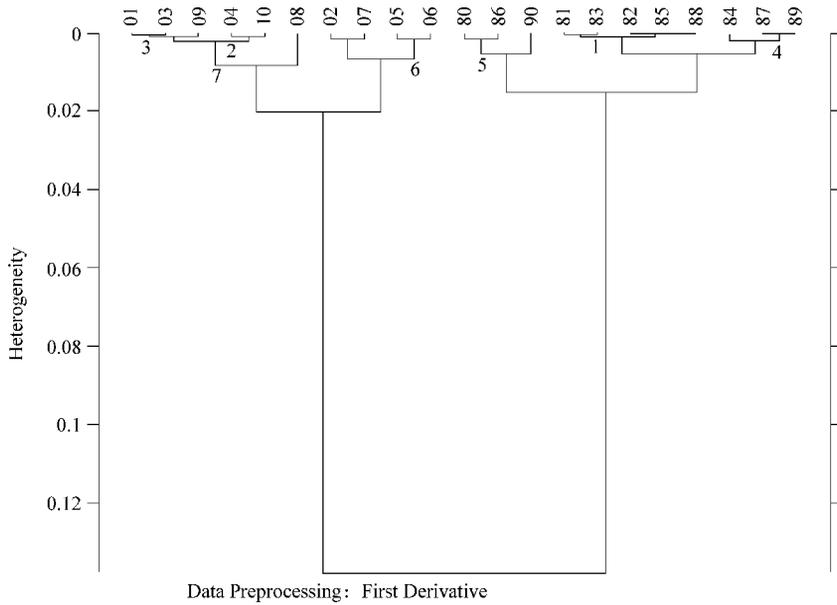


图 3-14 第一组 01 ~ 10#光谱数据样品的聚类分析树状图

注：图 3-14 中 80 ~ 90#为不含毒死蜱的样品，01 ~ 10#样品为毒死蜱浓度 0.005 ~ 0.0275mg/kg 的样品。

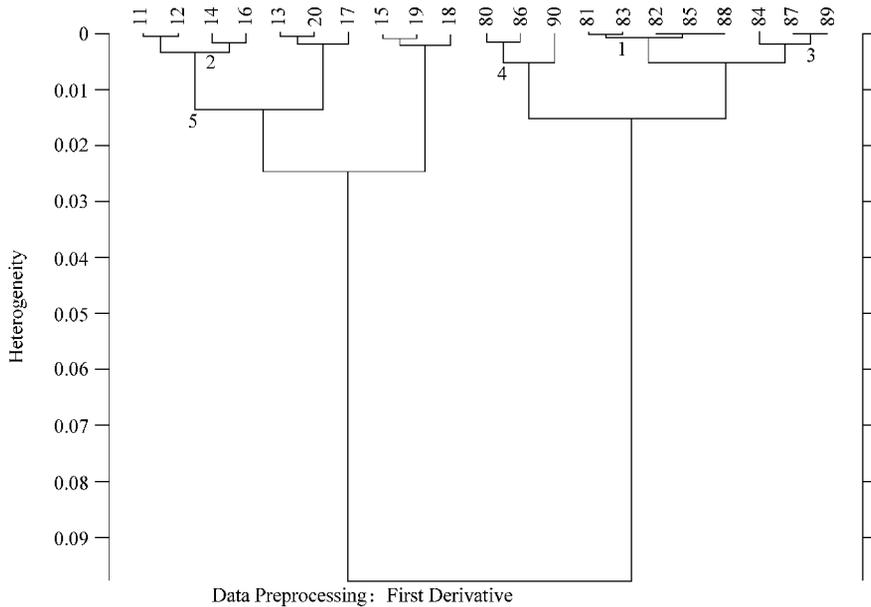


图 3-15 第一组 11 ~ 20#光谱数据样品的聚类分析树状图

注：图 3-15 中 80 ~ 90#为不含毒死蜱的样品，11 ~ 20#样品为毒死蜱浓度 0.03 ~ 0.0525mg/kg 的样品。

多元统计定性分析软件进行聚类分析：样品间距离采用欧氏距离算法，类间采用 Ward 算法。

重复样品毒死蜱浓度为 0.005 ~ 0.0525mg/kg 的光谱数据样品的定性分析结果如图 3-16

和图 3-17 所示。

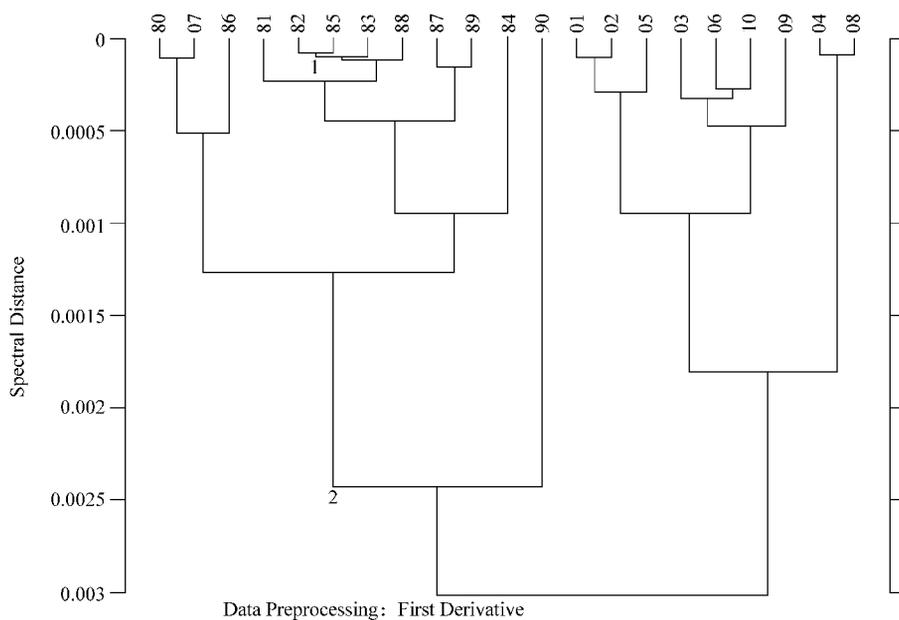


图 3-16 重现样品 01 ~10#光谱数据样品的聚类分析树状图

注：图 3-16 中 80 ~90#样品为无农药样品。01 ~10#样品毒死蜱浓度为 0.005 ~0.0275mg/kg。

聚类分析表明毒死蜱浓度为 0.02mg/kg 的 07#样品未能正确分类。分类正确率为 90%。

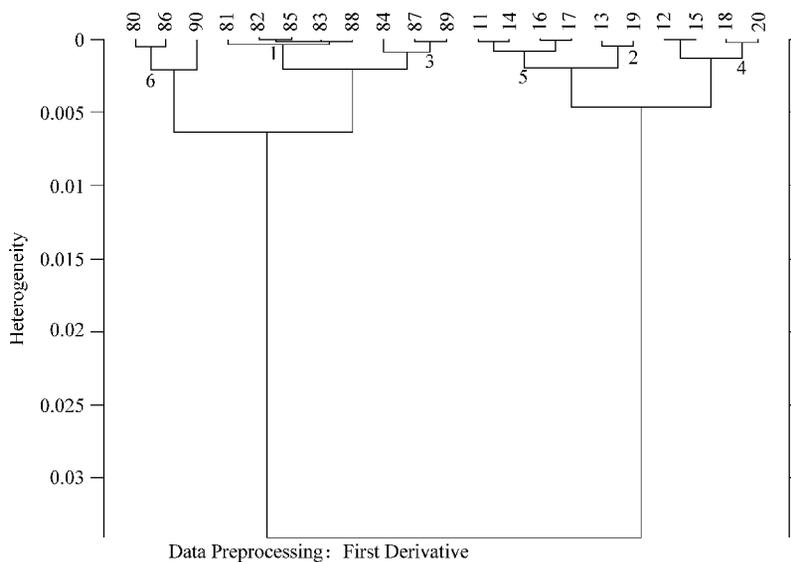


图 3-17 重现样品 10 ~20#光谱数据样品的聚类分析树状图

注：图 3-17 中 80 ~90#样品为无农药样品，11 ~20#为重复样品中毒死蜱浓度为 0.03 ~0.525mg/kg。

聚类分析树状图表明分类正确率为 100%。

3.6 光谱预处理对毒死蜇和炔螨特近红外光谱定量模型性能影响分析

分别使用 10 种常用的预处理方法：一阶导数（17 点平滑）、消除常数偏移量、矢量归一化（SNV）、最小-最大归一化、多元散射校正、减去一条趋势线、二阶导数、一阶导数（17 点平滑）+MSC、一阶导数（17 点平滑）+减去一条趋势线、一阶导数（17 点平滑）+矢量归一化进行优化处理，使用 PLS 法建立模型后与未进行优化的原始光谱数据模型进行对比，选取两种农药样品的最优预处理方法。

3.6.1 样品制备

用购自我国计量科学研究所的甲醇中毒死蜇溶液标准物质（标准值 1.01mg/mL）配置 26 个样品各 15g，制样前，先从冷藏室将样品取出，在室温 20℃ 平衡并摇匀后再打开安瓿瓶，样品浓度为 0.03 ~ 6mg/kg，分布在行业标准规定的最大残留量 1mg/kg 附近，见表 3-7，随机抽取 5 号、9 号、13 号、18 号和 23 号 5 个样品作为预测集。

表 3-7 毒死蜇溶液样品浓度

样品号	浓度值/(mg/kg)	样品号	浓度值/(mg/kg)	样品号	浓度值/(mg/kg)
1	0.03	10	1.3	19	3.45
2	0.09	11	1.54	20	3.78
3	0.18	12	1.62	21	4.2
4	0.28	13	1.72	22	4.56
5	0.36	14	1.84	23	4.88
6	0.5	15	2	24	5.2
7	0.6	16	2.4	25	5.6
8	0.76	17	2.86	26	6
9	1	18	3.2	27	0

用购自我国计量科学研究所的农药炔螨特标准物质（标准值 1.00mg/mL，以乙腈为溶剂）配置 20 个样品各 15g，制样前，先从冷藏室将样品取出，在室温 20℃ 平衡并摇匀后再打开安瓿瓶，样品浓度为 2 ~ 20mg/kg，分布在行业标准规定的最大残留量 5mg/kg 附近，见表 3-8，随机抽取 2 号、5 号、10 号、15 号和 19 号 5 个样品作为预测集。

表 3-8 炔螨特溶液样品浓度

样品号	浓度值/(mg/kg)	样品号	浓度值/(mg/kg)	样品号	浓度值/(mg/kg)
1	2	8	5	15	12
2	2.5	9	5.3	16	12.5
3	3.3	10	5.8	17	13.5
4	3.9	11	6.6	18	15
5	4.3	12	8	19	17
6	4.6	13	9.5	20	20
7	4.8	14	10.8	21	0

所购买标准物质的相关说明参照附录。

3.6.2 光谱采集

把不同质量浓度的毒死蜱和炔螨特样品溶液分别装入 25mL 茶色玻璃小瓶中，如图 3-18 所示，在室温相对恒定的环境中把光纤探头伸入小瓶中进行光谱数据的采集，此操作要保证近红外光纤探头完全浸没在样品中。



图 3-18 制备好的样品溶液

使用德国 Bruker 公司 VERTEX 70 型傅里叶变换红外光谱仪，检测器使用 InGaAs，所使用的液体光纤探头长度为 2m，光程池的有效光程为 1mm，采用的透反射测量的有效光程为 2mm，波长范围为 $12500 \sim 4000 \text{cm}^{-1}$ ，分辨率设置为 8cm^{-1} ，扫描次数为 32 次。光谱数据处理软件使用 OPUS6.5。

3.6.3 不同预处理方法对毒死蜱 PLS 模型性能影响分析

分别使用 10 种预处理方法对毒死蜱溶液样品光谱数据进行优化，建立毒死蜱溶液的 11 个 PLS 数学模型，模型的内部交叉验证结果如图 3-19 所示，使用模型对预测集样品进行预测，模型的优化结果和各评价参数见表 3-9。

利用不同预处理方法对毒死蜱溶液样品近红外光谱数据进行优化处理，并使用 PLS 法对各优化处理后的数据建立数学模型，结果表示，一阶导数（17 点平滑）、消除常数偏移量、SNV、最小-最大归一化、多元散射校正的内部交叉验证相关系数和预测相关因子均达到 0.99 以上，其中，综合考虑各模型的内部交叉验证结果和对预测集的预测结果，SNV 预处理方法为毒死蜱溶液样品数据的最佳优化方法。

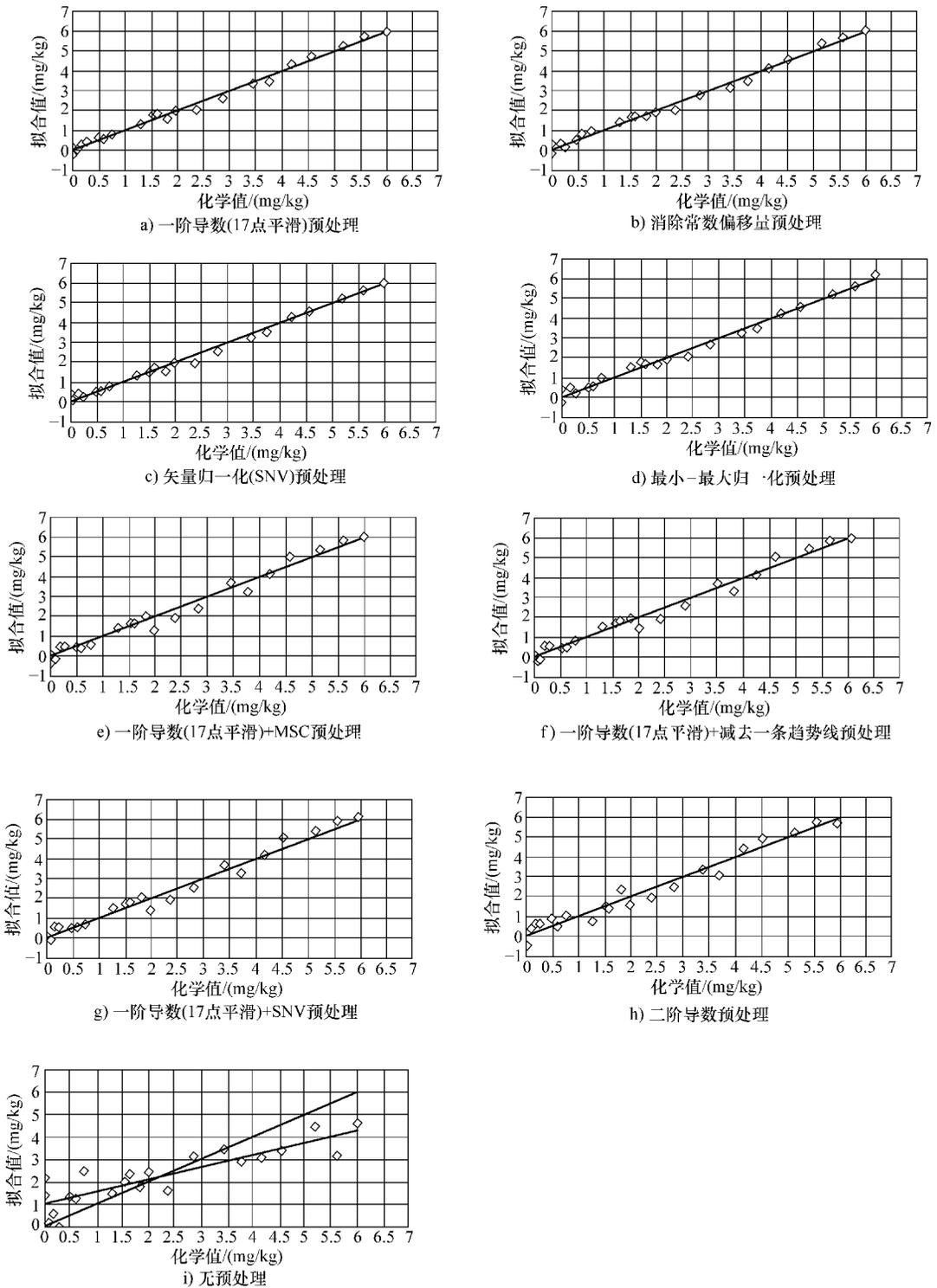


图 3-19 毒死蜱不同预处理方法模型内部交叉验证结果

表 3-9 毒死蜱溶液样品数据优化和建模结果

预处理方法	优化光谱范围/cm ⁻¹	所选数据点	主成分数	内部交叉检验相关系数 <i>R</i>	内部交叉检验标准差 RMSECV	预测相关因子 <i>R</i>	预测标准差 RMSEP
一阶导数 (17 点平滑)	6102 ~ 5446.3	171	3	0.9965	0.16	0.9949	0.164
消除常数偏移量	11995.7 ~ 7498.3 6102 ~ 5446.3	1338	4	0.9959	0.176	0.9938	0.194
SNV	11995.7 ~ 7498.3 6102 ~ 5446.3	1338	3	0.9957	0.182	0.9992	0.0802
最小-最大 归一化	11995.7 ~ 7498.3 6102 ~ 5446.4	1338	3	0.9949	0.198	0.993	0.199
多元散射 校正	6102 ~ 5446.3	171	2	0.9922	0.239	0.9926	0.199
减去一条 趋势线	11995.7 ~ 5446.3	1699	5	0.9902	0.268	0.985	0.296
一阶导数 (17 点 平滑) + MSC	11995.7 ~ 7498.3 6102 ~ 5446.3 4601.6 ~ 4246.7	1431	4	0.9891	0.283	0.9859	0.3
一阶导数 (17 点 平滑) + 减去 一条趋势线	11995.7 ~ 7498.3 6102 ~ 5446.3 4601.6 ~ 4246.7	1431	4	0.989	0.284	0.9877	0.284
一阶导数 (17 点 平滑) + SNV	4601.6 ~ 4246.7	93	4	0.9889	0.287	0.9857	0.303
二阶导数	4601.6 ~ 4246.8	93	9	0.983	0.352	0.9535	0.499
无预处理	11995.7 ~ 4246.7	2010	4	0.8691	1.04	0.9449	0.559

3.6.4 不同预处理方法对吡蚜酮 PLS 模型性能影响分析

剔除 14#异常样品后，分别使用 10 种预处理方法对吡蚜酮溶液样品光谱数据进行优化，建立吡蚜酮溶液的 11 个 PLS 数学模型，模型的内部交叉验证结果如图 3-20，使用模型对预测集样品进行预测，模型的优化结果和各评价参数见表 3-10。

利用不同预处理方法对吡蚜酮溶液样品近红外光谱数据进行优化处理，并使用 PLS 法对各优化处理后的数据建立数学模型，结果表明，10 种预处理方法的内部交叉验证相关系数和预测相关因子均达到 0.99 以上，其中，综合考虑各模型的内部交叉验证结果和对预测集的预测结果，减去一条趋势线预处理方法为吡蚜酮溶液样品数据的最佳优化方法。

近红外光谱技术结合相关预处理方法和 PLS 建模方法可用于初步检测微量农药溶液含量。

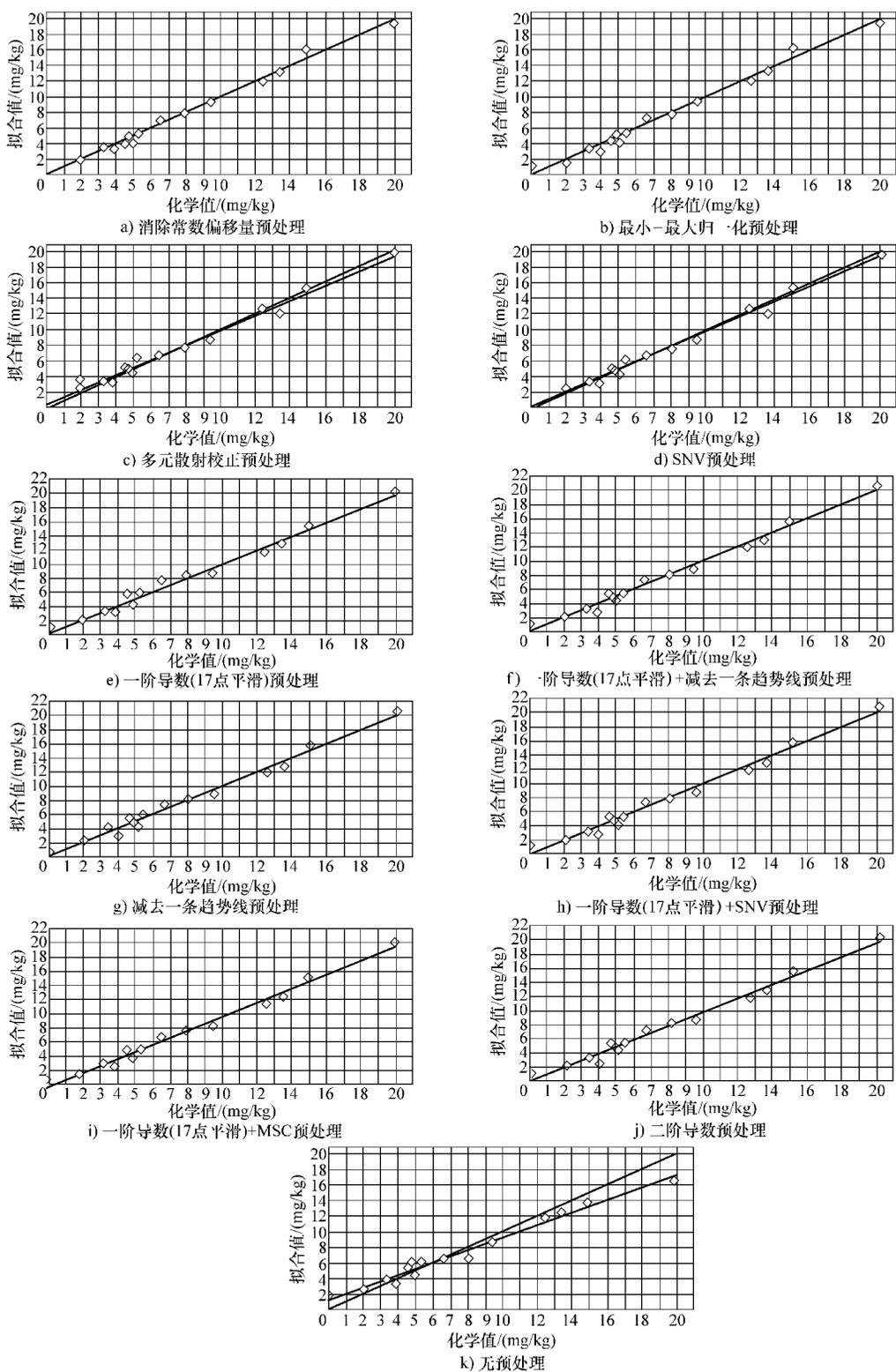


图 3-20 炔螨特不同预处理方法模型内部交叉验证结果

表 3-10 炔螨特溶液样品数据优化和建模结果

预处理方法	优化光谱 范围/cm ⁻¹	所选 数据点	主成 分数	内部交叉 检验相关 系数 <i>R</i>	内部交叉检验 标准差 RMSECV	预测相关 因子 <i>R</i>	预测标准 差 RMSEP
消除常数偏移量	6102 ~ 5774.1	86	3	0.9946	0.553	0.9933	0.856
最小-最大归一化	6102 ~ 5446.3	171	3	0.9943	0.567	0.995	0.79
多元散射校正	7502.1 ~ 5446.3	534	3	0.9943	0.587	0.9933	0.889
SNV	7502.1 ~ 5446.3	534	3	0.9943	0.587	0.9933	0.889
一阶导数 (17 点平滑)	6102 ~ 5446.3	171	1	0.9929	0.63	0.9937	0.749
一阶导数 (17 点平滑) + 减去一条趋势线	6102 ~ 5446.3	171	1	0.9928	0.637	0.9938	0.675
减去一条趋势线	11995.7 ~ 5446.3	1699	3	0.9925	0.649	0.9952	0.646
一阶导数 (17 点平滑) + SNV	6102 ~ 5446.3	171	1	0.9924	0.657	0.9936	0.706
一阶导数 (17 点 平滑) + MSC	6102 ~ 5446.3	171	1	0.9923	0.661	0.9936	0.707
二阶导数	6102 ~ 5446.3	171	1	0.992	0.674	0.9934	0.684
无光谱预处理	11995.7 ~ 4246.7	2010	3	0.9888	1.23	0.9916	0.966

3.6.5 毒死蜥和炔螨特的近红外光谱 ANN BP 检测试验研究

根据 3.6.4 节中的试验研究, 本节以 3.6.4 节的预处理研究方法和结果为基础, 对光谱数据分别进行最佳优化后, 使用 ANN BP 分别为炔螨特和毒死蜥建立校正模型, 并利用此模型对预测样品进行预测, 根据校正模型参数和预测结果, 研究 ANN BP 在近红外光谱法检测农药溶液含量中应用的可行性, 同时, 讨论 ANN BP 的使用方法。

本节 BP 模型的建立都使用 MATLAB7.7 神经网络工具箱中量化共轭梯度 (Scaled Conjugate Gradient, SCG) 反向传播算法训练函数 (trainscg 训练函数), 采用以校正集为训练样品对网络进行训练, 以均方误差 (Mean Squared Error, MSE) 为性能函数, 建模时对 BP 神经网络以 0.01 的目标误差和最大训练次数 3000 次进行模型的训练。

1. 毒死蜥微量溶液的 ANN BP 模型研究

根据 3.6.4 节的研究结果, 在 PLS 模型建立中, 贡献最大的最优预处理方法为 SNV 法, 因此, 首先对毒死蜥样品的近红外光谱数据进行 SNV 预处理, 观察此优化方法对 ANN BP 神经网络建模应用的优化贡献价值, 并找出可应用于近红外光谱法结合 ANN BP 法检测微量农药溶液的最佳优化方法。

(1) SNV 法结合 ANN BP 建立数学模型

当隐含层选为 15 个神经元时, 所建立模型的各个评价参数为 $R^2 = 0.9986$ 、RMSEC = 0.1000、RMSEP = 1.5800。

内部交叉验证的拟合结果如图 3-21 所示。ANN BP 模型对预测样品集的预测结果见表 3-11。

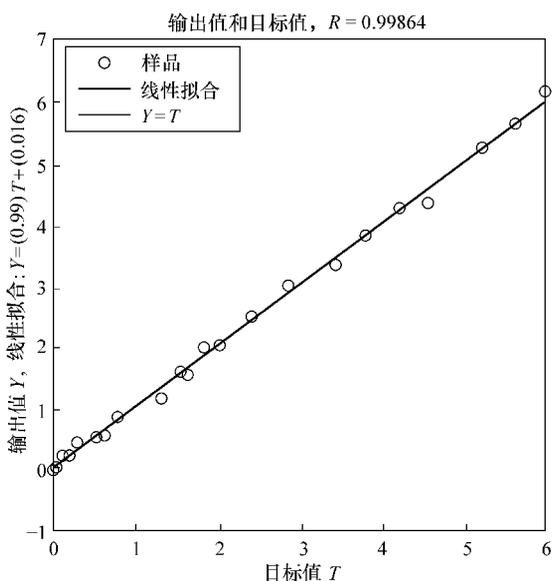


图 3-21 毒死蜱溶液的 SNV ANN BP 内部交叉验证拟合结果

表 3-11 毒死蜱溶液的 SNV ANN BP 模型预测结果

化学值/(mg/kg)	0.3600	1.0000	1.7200	3.2000	4.8800
预测值/(mg/kg)	0.8673	-0.6512	0.9556	5.8252	3.4579
RMSEP	1.5800				

从校正模型的校正结果和对预测样品集的预测结果不难看出，当隐含层神经元选择为 15 个时，所建立的 ANN BP 模型出现了过拟合，相关系数 $R^2 = 0.9986$ 非常接近于 1，内部交叉验证标准差 $RMSEC = 0.1000$ 很小，但是对预测集的预测表示，预测标准差 $RMSEP = 1.5800$ 误差严重。

因此，当隐含层神经元数量为 15 时的模型不可取，根据实验所得数据结果，基于 SNV 优化原始光谱数据的 ANN BP 的最优模型如下：以 SNV 预处理方法结合 ANN BP 所建立的最优数学模型为隐含层神经元数目为 9 的三层神经网络。其模型评价参数： $R^2 = 0.8962$ 、 $RMSEC = 0.8466$ 、 $RMSEP = 0.8104$ 。内部交叉验证的拟合结果如图 3-22 所示。模型对样品集的预测结果见表 3-12。

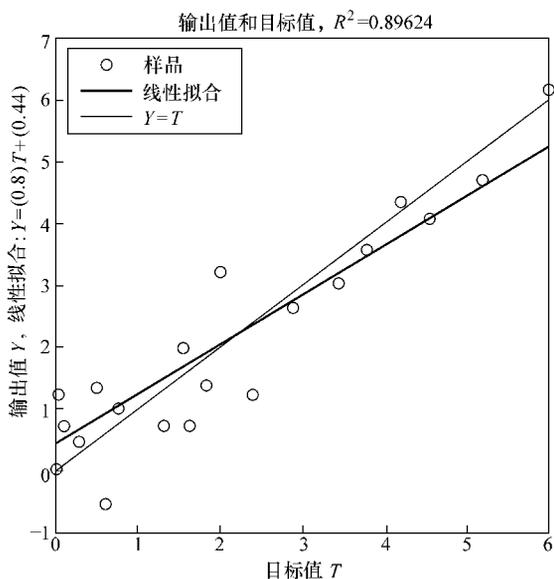


图 3-22 毒死蜱溶液的 SNV 方法下最优 ANN BP 内部交叉验证拟合结果

表 3-12 毒死蜱溶液的 SNV ANN BP 模型预测结果

化学值/(mg/kg)	0.3600	1.0000	1.7200	3.2000	4.8800
预测值/(mg/kg)	0.3135	2.2520	1.6635	4.2267	5.6905
RMSEP	0.8104				

根据 SNV 处理后的数据的 ANN BP 模型结果来看，即使是此条件下最优的模型，其模型的内部交叉验证标准差和预测标准差 $RMSEC = 0.8466$ 、 $RMSEP = 0.8104$ ，误差都较大，因此有必要利用其他优化方法寻找最优模型进行实验研究。

(2) ANN BP 建立数学模型的最优预处理方法

根据以上所研究的 ANN BP 模型中隐含层神经元的数目确定思路和光谱数据最优预处理的研究方法，建立经一阶导数优化后的近红外光谱数据最优 ANN BP 模型，其模型为隐含层神经元数目为 15 的三层神经网络模型。模型的评价参数： $R^2 = 0.9504$ 、 $RMSEC = 0.5960$ 、 $RMSEP = 0.6424$ 。内部交叉验证的拟合结果如图 3-23 所示。模型对预测样品集的预测结果见表 3-13。

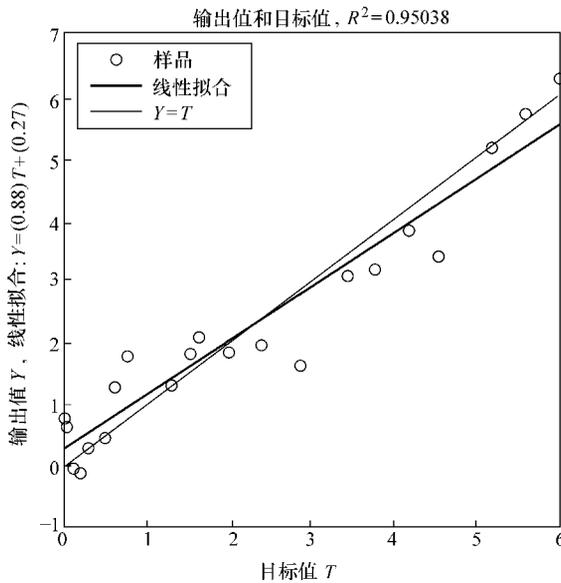


图 3-23 毒死蜱溶液的一阶导数方法下最优 ANN BP 内部交叉验证拟合结果

表 3-13 毒死蜱溶液的一阶导数方法下最优 ANN BP 模型预测结果

化学值/(mg/kg)	0.3600	1.0000	1.7200	3.2000	4.8800
预测值/(mg/kg)	1.0425	1.1092	1.4476	2.9673	3.6727
RMSEP	0.6424				

从以上建模结果可以看到，不同的光谱数据预处理方法对 BP 神经网络模型的建立贡献不同，隐含层神经元数目的选择也对模型的优劣有很大影响。不同的预处理方法对不同建模方法（例如 PLS 法和 ANN BP 法）的模型贡献度也不同，在后续的研究中，需要根据实际情况采用不同的预处理方法和模型参数设置。

2. 呋喃特微量溶液的 ANN BP 模型研究

基于本节前面部分的研究结果,针对呋喃特微量农药溶液的 ANN BP 建模过程,以本节前面部分为基础进行延续研究,确认上面的研究结果在呋喃特农药微量溶液建模中是否有效。

首先对 20 个呋喃特微量溶液样品进行近红外光谱数据采集,得到原始光谱数据后,对数据进行优化,选取最佳的预处理方法:最小-最大归一化法,作为最终的优化方法。建立模型之前,首先剔除 14#异常样品。建立了呋喃特微量农药溶液的近红外光谱法结合 ANN BP 法的数学模型,模型的校正结果和对预测集的预测结果为: $R^2 = 0.9949$ 、 $RMSEC = 0.5361$ 、 $RMSEP = 0.6143$ 。内部交叉验证的拟合结果如图 3-24 所示,模型对预测样品集的预测结果见表 3-14。

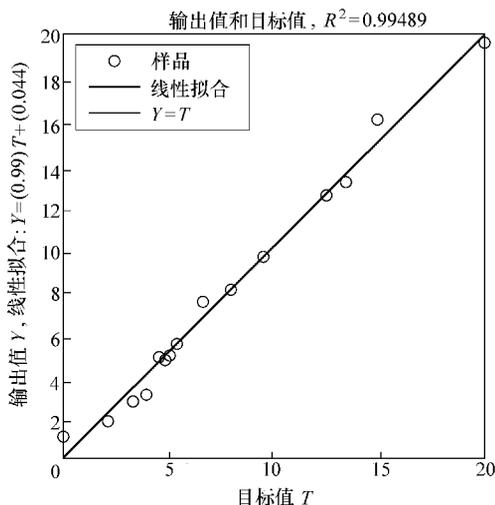


图 3-24 呋喃特溶液的最优 ANN BP 内部交叉验证拟合结果

表 3-14 呋喃特溶液 ANN BP 模型预测结果

化学值/(mg/kg)	2.5000	4.3000	5.8000	12.0000	17.0000
预测值/(mg/kg)	2.7868	4.8710	4.8601	11.8844	17.7626
RMSEP	0.6143				

基于近红外光谱技术结合 ANN BP 分析研究了其在微量农药溶液含量检测中应用的可行性。前半部分以微量毒死蜱溶液为研究对象,以实验结果数据讨论了不同预处理方法和隐含层神经元数目对模型的影响,根据研究结果又在后半部分为微量呋喃特溶液建立了最优 ANN BP 校正模型。从结果来看,微量毒死蜱溶液的 ANN BP 模型的评价参数 $R^2 = 0.9504$ 、 $RMSEC = 0.5960$ 、 $RMSEP = 0.6424$,远不及 3.6.4 节中 PLS 模型的校正和预测结果: $R^2 = 0.9957$ 、 $RMSEC = 0.182$ 、 $RMSEP = 0.0802$ 。微量呋喃特溶液的 ANN BP 模型的评价参数 $R^2 = 0.9949$ 、 $RMSEC = 0.5361$ 、 $RMSEP = 0.6143$ 优于使用 PLS 法所建立的模型: $R^2 = 0.9925$ 、 $RMSEC = 0.649$ 、 $RMSEP = 0.646$ 。对比结果表明,在近红外光谱法检测微量农药含量的建模方法中并没有固定的最优建模方法和数据预处理方法,根据不同的农药物质性质和不同的试验方法,要酌情选择最优的光谱数据预处理方法和所用的建模算法,本节所做实验结果证实,微量毒死蜱农药溶液的 PLS 法所建模型较优,而微量呋喃特溶液的 ANN BP 模型的校正和预测效果最佳。

3.7 小结

本章针对模拟溶液的农药残留量,采用近红外光谱检测方法进行由浅入深的实验设计和研究分析。首先,为了研究光谱法对微量农药在水溶液中的检出敏感程度,将微量的农药以水稀释至所选取的试验样品浓度,扫描光谱数据并加以分析。通过设计不同的实验背景,逐

近红外光谱技术在食品品质检测方法中的研究

步探究近红外光谱在测定溶液中微量毒死蜱的应用情况。实验设计按照毒死蜱在溶液中的浓度由高到低,依次为 3.13 ~ 10mg/kg、3.03 ~ 10mg/kg、0.005 ~ 0.1mg/kg,依次分析。背景由简单到复杂,即由甲醇、水、毒死蜱的溶液到甲醇、水、维生素 C、蔗糖和毒死蜱的溶液。研究表明,毒死蜱的近红外光谱信息较丰富的波段在 9998 ~ 5423 cm^{-1} 。其中使用最多的波段在 7502 ~ 5446 cm^{-1} ,且近红外光谱法在测定混合液中的毒死蜱浓度 $\geq 3.2\text{mg/kg}$ 时,通过选择适当的预处理方法和所分析的谱区范围,在预测以上两种不同溶液样品过程中,样品的预测值和化学值之间均有很好的相关性。比较了同一组样品的 BP 网络和 PLS 模型,从实验分析结果看 PLS 更适合研究基于近红外光谱溶液中微量毒死蜱的定量分析。

在近红外光谱分析中,定量分析是寻找分析光谱数据的共性,而定性分析是寻找分析数据的差异性,同一组数据定量分析效果不错,不一定能够实现其定性分析。因此在之前试验分析的基础上,对甲醇、水、维生素 C、蔗糖和毒死蜱溶液进行定性的分析研究,并研究其分类可能性。通过对有毒死蜱(甲醇、水、维生素 C、蔗糖和毒死蜱)的溶液和无毒死蜱的溶液(甲醇、水、维生素 C、蔗糖)进行聚类分析,并采用比较直观的树状图表示其聚类分析结果。

实验数据样品间均采用欧氏距离,并用 Ward 方法进行聚类,分析过程中对光谱数据选取适当的谱段和预处理方法。两组实验结果表明,毒死蜱浓度在 0.005 ~ 0.0275mg/kg 的样品第一组鉴别率为 100%,重现实验鉴别率为 90%。毒死蜱浓度在 0.03 ~ 0.0525mg/kg 的样品鉴别率均为 100%。

本章针对农药残留量痕量分析研究,从近红外光谱技术方向进行了综合分析研究,以毒死蜱农药标准物质和炔螨特农药标准物质作为制作样品的研究对象,从农药水溶液的初步研究开始,基于近红外光谱的微量农药残留检测分析。通过对毒死蜱和炔螨特光谱数据用不同预处理方法进行优化后建立了 PLS 数学模型和农药溶液的 ANN BP 模型,通过对比不经任何优化处理方法的光谱数据模型和其他常用优化方法,确定了光谱数据优化对建立模型的重要贡献,降低了样品数据噪声和其他干扰。

附 录

附录 A 附表

附表 1 毒死蜱、甲醇及水的样品

样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)
1#	3.23	10#	3.92	19#	5.55
2#	3.27	11#	4	20#	5.88
3#	3.33	12#	4.17	21#	6.25
4#	3.38	13#	4.35	22#	6.66
5#	3.45	14#	4.44	23#	7.14
6#	3.50	15#	4.54	24#	7.69
7#	3.57	16#	4.76	25#	8.33
8#	3.70	17#	5	26#	9.09
9#	3.85	18#	5.26	27#	10

附表 2 甲醇、水、维生素 C 和蔗糖中的毒死蜱

样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)
1#	3.125	11#	3.92	21#	6.66
2#	3.17	12#	4	22#	7.14
3#	3.23	13#	4.17	23#	7.69
4#	3.27	14#	4.35	24#	8.33
5#	3.33	15#	4.54	25#	9.09
6#	3.45	16#	4.76	26#	10
7#	3.50	17#	5.26	27#	11.11
8#	3.57	18#	5.55	28#	12.5
9#	3.70	19#	5.88		
10#	3.85	20#	6.25		

附表 3 甲醇、水、维生素 C 和蔗糖中的毒死蜱

样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)
1#	0.005	31#	0.0825	61#	2.5
2#	0.0075	32#	0.085	62#	2.6
3#	0.01	33#	0.0875	63#	2.7
4#	0.0125	34#	0.09	64#	2.8
5#	0.015	35#	0.0925	65#	2.9
6#	0.0175	36#	0.095	66#	3
7#	0.02	37#	0.1	67#	3.1
8#	0.0225	38#	0.2	68#	3.2
9#	0.025	39#	0.3	69#	3.3
10#	0.0275	40#	0.4	70#	3.4
11#	0.03	41#	0.5	71#	3.5
12#	0.0325	42#	0.6	72#	3.6
13#	0.035	43#	0.7	73#	3.7
14#	0.0375	44#	0.8	74#	3.8
15#	0.04	45#	0.9	75#	3.9
16#	0.0425	46#	1	76#	4
17#	0.045	47#	1.1	77#	4.1
18#	0.0475	48#	1.2	78#	4.2
19#	0.05	49#	1.3	79#	4.3
20#	0.0525	50#	1.4	80#	4.4
21#	0.055	51#	1.5	81#	4.5
22#	0.0575	52#	1.6	82#	4.6
23#	0.06	53#	1.7	83#	4.7
24#	0.0625	54#	1.8	84#	4.8
25#	0.065	55#	1.9	85#	4.9
26#	0.0675	56#	2	86#	5
27#	0.07	57#	2.1	87#	甲醇
28#	0.0725	58#	2.2	88#	甲醇 + 毒死蜱
29#	0.075	59#	2.3		
30#	0.08	60#	2.4		

附表 4 重现实验样品甲醇、水、维生素 C 和蔗糖中的毒死蜱

样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)
1#	0.005	23#	0.06	45#	0.9
2#	0.0075	24#	0.0625	46#	1
3#	0.01	25#	0.065	47#	1.1
4#	0.0125	26#	0.0675	48#	1.2
5#	0.015	27#	0.07	49#	1.3
6#	0.0175	28#	0.0725	50#	1.4
7#	0.02	29#	0.075	51#	1.5
8#	0.0225	30#	0.08	52#	1.6
9#	0.025	31#	0.0825	53#	1.7
10#	0.0275	32#	0.085	54#	1.8
11#	0.03	33#	0.0875	54#	1.9
12#	0.0325	34#	0.09	56#	2
13#	0.035	35#	0.0925	57#	2.1
14#	0.0375	36#	0.095	58#	2.2
15#	0.04	37#	0.1	59#	2.3
16#	0.0425	38#	0.2	60#	2.4
17#	0.045	39#	0.3	61#	2.5
18#	0.0475	40#	0.4	62#	2.6
19#	0.05	41#	0.5	63#	2.7
20#	0.0525	42#	0.6	64#	2.8
21#	0.055	43#	0.7	65#	2.9
22#	0.0575	44#	0.8	66#	3

附录 B 标准物质相关说明

B.1 甲醇中毒死蜱溶液标准物质

1. 来源及用途

国家质量监督检验检疫总局批准号：GBW (E) 081141

样品编号：10002；生产单位：中国计量科学研究院（北京）

主要用途：环境、食品等领域相应农药成分残留检测，农药生产部门和检测实验室质量控制、分析仪器校准、分析方法确认评价等。

2. 样品制备

采用纯度准确定值的毒死蜱纯品为原料，以甲醇为溶剂，采用重量 - 容量法准确配制而成。

3. 溯源性和定值方法

原料纯度采用高效液相色谱法和气相色谱法定值。以配制值作为溶液标准物质浓度的标准值,制备的标准物质和质量控制对照样采用气相色谱法(FID检测器)比对计算,核验配制值。

4. 标准值及抽样检验结果

标准值: 1.01mg/mL 采用玻璃安瓿瓶包装,每支 1mL。

相对扩展不确定度($k=2$): 1%。

根据《一级标准物质》技术规范的随机抽样检验要求,采用气相色谱法(FID检测器)对溶液浓度均匀性、稳定性的考察进行结果表示,均匀性符合F检验规则,稳定性良好,有效期为自定值日期起的12个月。

B.2 农药炔螨特溶液标准物质

1. 来源及用途

标准号: BW 3574; 样品编号: 10001; 生产单位: 中国计量科学研究院(北京)。

主要用途: 环境、食品等领域相应农药成分残留检测,农药生产部门和检测实验室质量控制、分析仪器校准、分析方法确认评价等。

2. 样品制备

采用纯度准确定值的纯品为原料,以乙腈为溶剂,用重量-容量法准确配制而成。

3. 溯源性和定值方法

物质原料纯度采用液相色谱法测定。以配制值作为溶液标准物质浓度的标准值,制备的标准物质和质量控制对照采用液相色谱法(UV检测器)比对计算,核验配制值。

4. 标准值及抽样检验结果

标准值: 1.00mg/mL 采用玻璃安瓿瓶包装,每支 1mL。

相对扩展不确定度($k=2$): 4%。

根据《一级标准物质》技术规范的随机抽样检验要求,采用液相色谱法(UV检测器)对溶液浓度均匀性、稳定性的考察进行结果表示,均匀性符合F检验规则,稳定性良好,有效期为自定值日期起的3.5年。

参考文献

- [1] ATSDR (Agency for Toxic Substances and Disease Registry). Toxicological Profile for Chlorpyrifos [P]. U. S. Department of Health&Human Services, Public Health Service, Atlanta, GA, 1997.
- [2] Durik S E, Collette T W. Degradation of chlorpyrifos in chlorinesolutions: pathways, kinetics, and modeling [J]. Environmental Science& Technology, 2006, 40: 546.
- [3] Gruber S J, Munn M D. Organophosphate and carbamate insecticides in cholinesterase (CHE) inhibition in common Carp (Cyprians carp io) Contam [J]. Toxicol, 1998, 35: 391.
- [4] 董丽华, 董玉莲. 固相萃取-气相色谱法测定水中毒死蜱 [J]. 分析测试技术与仪器, 2008, 14 (1): 30.
- [5] 孟春校. 三唑磷、毒死蜱混剂的高效液相色谱分析 [J]. 浙江化工, 2002, 16.
- [6] 李萍, 闫永利, 王仙琴, 等. 微量毒死蜱的离子阱气-质联用的分析条件研究 [J]. 宁夏农林科技, 2005, 6: 33.

4.1 简介

我国蔬菜产业经过多年发展,已经成为全球最大的蔬菜市场,目前我国蔬菜的产销量占全球市场的50%以上。根据前瞻产业研究院发布的《2014—2020年中国蔬菜种植业市场需求与投资战略规划分析报告》数据显示,近年来,我国蔬菜生产持续稳定发展,种植面积由2007年的2.62亿亩^①增加到2012年的3.08亿亩,产量由2007年的5.47亿t增加到2013年的7.06亿t,已经超过粮食产量,成为我国第一大农产品。但由于生产方式落后、管理松散及菜农的素质和生产技能不能适应蔬菜生产的新要求,因农药使用不合理造成的蔬菜农药残留超标事件逐年增加,这引起了党和国家的高度重视。2007年8月开始,为恢复国际社会对“中国制造”的信心,国务院在全国范围内开展产品质量和食品安全专项整治,在蔬菜质量安全管理上,开展了高毒农药整治和农产品批发市场整治行动,重点打击在蔬菜产品中非法添加甲胺磷等禁限用农药成分的违法行为,并将全国大中城市农产品批发市场全部纳入监测范围。因此,研究不同种类蔬菜农药残留检出率的规律性,对于保障人民群众的饮食安全、改善人们的健康水平、提高我国蔬菜的国际竞争力有着重要的意义。

随着我国国民经济的不断发展,特别是全球经济一体化和食品贸易国际化,食品安全问题已经成为举足轻重的问题。而蔬菜中农药残留超标是目前我国食品安全中非常突出的问题。现存的蔬菜中农药残留检测技术虽然能够达到检测要求,但是费时费力、设备昂贵,需要专业实验人员操作,不适合现场检测,因此开发快速、高效、经济、环保的农药残留检测技术,已经成为当前国内外研究的热点课题。

随着蔬菜结构不断改善和消费者对蔬菜安全质量需求的不断增加,蔬菜农药残留问题已成为各级政府和消费者关注的焦点。由于蔬菜种植业的发展,蔬菜种类、生产管理条件不断变化,导致蔬菜病虫害种类增多,致使农民在蔬菜种植过程中滥用农药,造成蔬菜农药残留严重,引起食物中毒的事件时有发生,使消费者产生一定的恐慌。为了提高蔬菜质量安全,保障消费者身体健康,本章就蔬菜农药残留的现状、成因分析如下:

1. 蔬菜农药残留的现状

目前,由于蔬菜病虫害的不断加剧,农民在实际操作中仍普遍使用化学农药,而且使用高毒、高残留(简称“两高”)甚至伪劣或禁用农药,且使用量巨大,致使蔬菜农药残留问题相当严重,且相当普遍。

2. 蔬菜农药残留的成因

蔬菜农药残留的成因是复杂、多样的;有客观原因造成的,更主要是主观因素造成的。

^① 1亩(a) = 666.6m²。

近红外光谱技术在食品品质检测方法中的研究

蔬菜病虫害严重使农民不得不使用农药，是造成蔬菜农药残留的客观因素。造成蔬菜病虫害不断加剧的原因有如下4个方面：

1) 温室、大棚等蔬菜生产基地迅猛发展，导致适宜害虫在温室中的危害日趋猖獗。随着气候变暖，使原来许多越冬性的害虫，一年四季不断生长繁殖，世代增多，危害加大。长时间大量使用农药使蔬菜害虫的抗药性增强，使主要虫害加重、次要虫害上升。

2) 缺乏科学使用农药的知识，片面追求产量而导致盲目、超量使用农药。缺乏环保意识，为有效防治蔬菜害虫，长期使用“两高”农药。为了使上市之前的蔬菜不再生虫，将安全间隔期缩短，或只要市场价格好，不考虑施药时间就立即采摘上市。

3) 农药市场比较混乱，部分农药存在严重质量问题，加上我国农药类型和品种结构不合理，“两高”农药品种较多。

4) 法律法规不够完善、监督管理体制不明确。

3. 蔬菜农药残留的管理

为了指导和帮助农民科学种植，严格控制“两高”农药，生产绿色蔬菜，保护食用者安全，提出如下管理建议：

1) 尽快完善法律法规保障体系，实行从田头到餐桌全过程的质量控制，加快蔬菜卫生质量安全管理方面法律法规的起草进程，积极采用国际标准，制定国家蔬菜质量安全标准；建立蔬菜生产质量安全检测机构，加强蔬菜流通领域的检测，并积极指导蔬菜生产基地（企业）、批发市场、农贸市场开展自测；全面实施《农产品安全质量无公害蔬菜安全要求》，实现从田头到餐桌全过程的质量控制系统，大力发展绿色蔬菜。

2) 加强蔬菜流通领域的监督检查力度，推行市场准入制度。卫生监督及有关职能部门应依据《中华人民共和国食品卫生法》、《农药管理条例》、《农药合理使用准则》等有关法规，加强蔬菜流通领域的监督检查力度，对销售含农药残留的蔬菜、使用禁用农药和经销“两高”农药的，要依法进行行政处罚，构成犯罪的，应依法追究刑事责任。

3) 积极推行市场准入制度全面贯彻农业部《全面推进“无公害食品行动计划”的实施意见》，建立蔬菜卫生质量安全管理制，创建专销网点，对获得无公害蔬菜认证和经检测合格的蔬菜实行专区销售，并实施标识管理；对销售不合格的蔬菜实行召回、理赔和退出市场。同时可以采取有效措施，减少农药用量，如在生产上，选用抗病品种，改革耕作制度，选择合理的种植模式，实行科学轮作和套种，调节播期，合理用肥，适时收获，提高蔬菜的抗逆性，从而减轻病虫害危害。采用人工捕杀、物理方法防治病虫害危害。积极推广使用植物杀虫、杀菌剂及生物农药。

4) 加强宣传和技术指导，引导农民科学、合理用药，教育消费者提高安全防护意识。我国农药结构应尽快向低毒、高效农药方向调整，要从农药源头上进行控制，蔬菜生产中使用农药应严格执行国家《农药合理使用准则》，严禁使用“两高”农药，要选用无毒或低毒、残效期短的农药。采取各种宣传方式，向农民提供合理用药的知识，重点指导农民；有的放矢使用农药，用药准确、适时；注意用药浓度，正确掌握用药量；改进药效，合理混合作用。

5) 搞好技术培训，做好病虫、农药知识的普及推广。对于消费者，要教育他们增强自我保护意识，掌握安全食用蔬菜的方法“一洗二浸三烫四炒”，从而减少蔬菜农药残留，确保食用安全。

6) 加强科学研究, 积极推广新品种、新技术。采用现代遗传工程技术培育蔬菜抗病新品种, 研制、生产高效、低毒、低残留的无公害农药, 从根本上减少或避免蔬菜农药残留的发生。并抓好新品种、新技术的推广和应用。

本章首先针对蔬菜中有机磷农药残留量的现状提出了关于食品检测方面的几项国内、国外标准。然后重点介绍了国内外蔬菜中农药残留检测技术, 主要包括色谱法、生物传感器法、生物化学测定法、发光菌法及免疫分析法等。对这几种方法的分类和检测原理以及应用现状进行了详细的介绍; 然后采用近红外光谱分析技术, 以毒死蜱为研究对象, 选择大白菜、菠菜和萝卜作为蔬菜代表, 开展了蔬菜中农药残留量检测方法的探索性研究。目前叶菜研究工作主要集中在液体的检测, 对经萃取后菜汁中的毒死蜱含量进行实验研究, 分析不同的预处理方法对模型精度的影响, 采用偏最小二乘 (PLS) 法对叶菜中的毒死蜱建立模型。

4.2 蔬菜中常见的有机磷农药残留量检测方法

我国的蔬菜农药残留问题必须引起高度重视, 这不仅会对国民经济发展和人民身体健康造成不利影响, 还会对我国出口创汇和国家外贸声誉带来极大的伤害。随着我国加入 WTO 和各国农药最大残留限量 (MRL) 标准的实施, 国际上对我国农产品 (特别是蔬菜) 的质量要求也必将越来越严格, 表 4-1 即目前我国规定蔬菜中常见有机磷农药的 MRL 标准。我国出口蔬菜由于农药残留超标, 时常会发生被国外扣留、销毁、退货、索赔和终止合同等现象, 许多传统大宗出口创汇的农产品因此被迫退出国际市场。2002 年因为毒死蜱超标, 我国出口日本的菠菜就一度退出日本市场, 对我国蔬菜出口创汇造成了巨大经济损失。据有关部门统计, 我国每年都会有五六百起因农药超标使农产品出口被退的事件, 因此造成的经济损失超过 70 亿元。所以建立和完善我国蔬菜中农药残留的检测体系, 加快开发农药残留快速检测方法的研究进程, 实现农药残留检测水平与国际接轨, 已经是国内许多专家和政府部门正在关注和着力解决的问题。

表 4-1 我国规定蔬菜中常见有机磷农药的最大残留限量

农药名称	最大残留限量/ (mg/kg)	农药名称	最大残留限量/ (mg/kg)
马拉硫磷	不得检出	倍硫磷	0.05
氧化乐果	不得检出	辛硫磷	0.05
甲拌磷	不得检出	二嗪磷	0.5
甲胺磷	不得检出	杀螟硫磷	0.5
久效磷	不得检出	乙酰甲胺磷	0.2
对硫磷	不得检出	啶硫磷	0.2
毒死蜱	1.0	敌敌畏	0.2
乐果	1.0	敌百虫	0.1

近年来随着近红外光谱分析技术的软硬件发展, 加上它是一种快速、无损、多组分同时分析的“绿色技术”, 与传统的分析方法相比, 具有操作简单、成本低、重现性好等特点, 使得近红外光谱分析技术在食品安全检测中的应用成为热门课题。因此, 针对人们日常生活一直担心的蔬菜中农药残留问题, 本书利用近红外光谱分析技术对蔬菜中农药残留量的检测

方法进行了探索性研究,以便于更好地推动近红外光谱分析技术在农药残留检测中的应用,实现对蔬菜的快速、无损检测,同时也会对提高我国农产品的国际竞争力起到积极的作用。

2013年7月15日,第一届食品安全国家标准审评委员会第八次主任会议在京召开。会议审议通过了《食品安全国家标准审评委员会章程》修订建议,通报食品标准清理工作进展情况,审议通过了71项食品安全国家标准。食品安全国家标准审评委员会秘书长、副秘书长、各专业分委员会主任委员以及工业和信息化部、商务部、农业部、质检总局、食品药品监管总局、国家标准委等部门代表参加会议。本次会议审议通过了《食品中致病菌限量》、《食品添加剂标识通则》、《食品用香料通则》、《特殊医学用途配方食品通则》、《预包装特殊膳食用食品标签》和《特殊医学用途配方食品良好操作规范》以及一批食品添加剂质量规格标准等共计71项食品安全国家标准。目前针对蔬菜等食品的国内、国际标准主要如下:

1. GB/T 5009.20—2003 食品中有机磷农药残留量的测定

该标准规定了水果、蔬菜、谷类中敌敌畏、速灭磷、久效磷、甲拌磷、巴胺磷、二嗪磷、乙嘧硫磷、甲基嘧啶磷、甲基对硫磷、稻瘟净、水胺硫磷、氧化喹硫磷、稻丰散、甲啶硫磷、克线磷、乙硫磷、乐果、啶硫磷、对硫磷、杀螟硫磷的残留量分析方法。

2. GB/T 5009.161—2003 动物性食品中有机磷农药多组分残留量的测定

本标准规定了动物性食品中甲胺磷、敌敌畏、乙酰甲胺磷、久效磷、乐果、乙拌磷、甲基对硫磷、杀螟硫磷、甲基嘧啶磷、马拉硫磷、倍硫磷、对硫磷、乙硫磷等常用有机磷农药多组分残留量的测定方法。

3. GB/T 14553—2003 粮食、水果和蔬菜中有机磷农药测定的气相色谱法

该标准规定了粮食(大米、小麦、玉米)、水果(苹果、梨、桃等)、蔬菜(黄瓜、大白菜、西红柿等)中速灭磷、甲拌磷、二嗪磷、异稻瘟净、甲基对硫磷、杀螟硫磷、溴硫磷、水胺硫磷、稻丰散杀扑磷等作物中有机磷农药的残留量的测定。

4. GB/T 14552—2003 水、土中有机磷农药测定的气相色谱法

本标准规定了地面水、地下水及土壤中速灭磷、甲拌磷、二嗪磷、异稻瘟净、甲基对硫磷、杀螟硫磷、溴硫磷、水胺硫磷、稻丰散、杀扑磷等多组分残留量的测定方法。

5. SN/T 0148—2011 进出口水果蔬菜中有机磷农药残留量检测方法. 气相色谱和气相色谱-质谱法

本标准规定了水果蔬菜中敌敌畏、乙酰甲胺磷、硫线磷、百治磷、乙拌磷、乐果、甲基对硫磷、毒死蜱、嘧啶磷、倍硫磷、丙虫硫磷、辛硫磷、灭菌磷、三硫磷、三唑磷、哒嗪硫磷、亚胺硫磷、敌百虫、灭线磷、甲拌磷、氧化乐果、内吸磷、二嗪磷、地虫硫磷、异稻瘟净、氯唑磷、甲基毒死蜱、对氧磷、杀螟硫磷、溴硫磷、乙基溴硫磷、噻唑磷、丙溴磷、乙硫磷、敌瘟磷、吡啶硫磷、蝇毒磷、甲胺磷、治螟磷、特丁硫磷、久效磷、除线磷、皮蝇磷、甲基嘧啶磷、对硫磷、甲基毒虫畏、异柳磷、稻丰散、杀扑磷、甲基硫环磷、伐杀磷、伏杀硫磷、甲基谷硫磷、二溴磷、速灭磷、甲基乙拌磷、巴胺磷、乙嘧硫磷、磷胺、地毒磷、马拉硫磷、甲基异柳磷、水胺硫磷、啶硫磷、杀虫畏、碘硫磷、硫环磷、威菌磷、苯硫磷、乙基谷硫磷70种有机磷类农药残留量的气相色谱及气相色谱-质谱检测方法。适用于菠萝、苹果、荔枝、胡萝卜、马铃薯、茄子、菠菜、荷兰豆、鲜木耳、鲜蘑菇、鲜牛蒡、鲜香菇、大葱中上述70种有机磷类农药残留量的检测。

6. SN/T 2324—2009 进出口食品中抑草磷、毒死蜱、甲基毒死蜱等 33 种有机磷农药残留量的检测方法

本标准规定了粮谷类食品中 33 种有机磷农药残留量 [抑草磷、毒死蜱、甲基毒死蜱、敌敌畏、敌瘟磷、杀螟硫磷、倍硫磷、地虫硫磷、异柳磷、氧异柳磷、甲基立枯磷、三唑磷、毒虫畏 (E)、毒虫畏 (Z)、丙线磷、特丁磷、二嗪磷、甲基毒虫畏 (E)、甲基毒虫畏 (Z)、苯硫磷、乙嘧啶磷、丰索磷、马拉硫磷、治螟磷、对硫磷、甲基对硫磷、稻丰散、甲基嘧啶磷、啶硫磷、杀虫畏、甲基乙拌磷、完灭硫磷、完灭硫磷砒] 的气相色谱检测方法 & 气相色谱质谱确证方法。本标准适用于进出口大米、糙米、玉米、大麦、小麦中 33 种有机磷农药残留量的测定和确证。

农药残留是指农药使用后残存在生物体、农副产品和环境中的微量农药原体、有毒代谢物、降解物和杂质的总称。从 20 世纪 50 年代以来, 农药残留检测技术发生了巨大的变化和发展。目前, 国内外蔬菜中有机磷农药残留量检测分析方法主要有色谱法、生物传感器法、生物化学测定法、发光菌法及免疫分析法。

4.2.1 色谱法

色谱法也称色层法或层析法, 是一种物理化学分离分析方法。它是利用混合物各组分在两相间分配系数的差异, 达到各组分分离的目的, 并使分析物质的浓度转化为易被测量的电信号, 然后用记录仪记录下来。目前应用于检测有机磷的色谱法主要有薄层色谱法、气相色谱法、液相色谱法、超临界流体色谱法。

1. 薄层色谱法

薄层色谱法是利用被检测物经显色后同标准有机磷农药比较来定性, 用薄层扫描仪来定量的测定方法。它可同时分析多个样品, 多用于复杂混合物的分离和筛选, 具有成本低、可现场操作的特点。薄层色谱法如果和其他技术联用, 可以得到更高的灵敏度和更广的测试范围。若与荧光显色技术结合, 对有机磷的检测限为 $0.01\mu\text{g}$; 与酶抑制技术结合, 对有机磷的检测限可达 0.001Pg 。但是准确度不高、预处理繁琐, 只适用于挥发性农药, 适用范围窄, 选择性和灵敏性不高。

薄层色谱法是一种成熟的、应用也较广的微量快速检测方法。它在农药残留测定技术上有它独特的用处, 它既是重要的分离手段, 又是定性、定量的分析方法。它的特点如下:

1) 薄层上离析的理化过程, 因不受物质的限定, 故在方法的设计、拟定和实际应用上机动灵活, 适合于对多种类型样品、多种农药检测。

2) 薄层的分离能力强, 效果明显, 对复杂混合物的划分、不稳定物质的分离、相似化合物的离析和鉴定特别有用。

3) 薄层色谱的显色反应类型多, 显色剂范围广。同一种农药一般有数种显色方法, 同一种显色剂有时适应多种或一大类农药的显色反应, 而且薄层吸附剂一般不受腐蚀试剂的限制, 可根据样品性状和检测要求选择使用。

4) 薄层层析与仪器配合的联用 (如扫描仪、光密度计等), 则测定迅速、灵敏度高、重现性好、测试范围宽。

5) 薄层色谱原位斑点面积直接定量法, 无需特殊设备和试剂, 方法简易、快速、直观。

检测过程一般先用适宜的溶剂提取有机磷农药，经纯化浓缩后，在薄层硅胶板上分离展开，显色后与标准的有机磷农药比较 R_f 值进行定性测定或用仪器进行定量测定。

2. 气相色谱法

气相色谱法是目前最典型、应用最广的农药检测方法，也是检测有机磷的国家标准方法，检测限可达 0.1ng，具有准确、灵敏度高的特点，并且一次可以测定多种成分。气相色谱法是农药残留检测中比较重要的方法。它具有分析速度快、分离效能高、分离和测试一次完成等优点，被许多国家作为农药残留分析的标准方法。随着毛细管柱的广泛应用，柱处理、检测器以及进样系统的不断优化，大大提高了气相色谱分析的精密度和灵敏度，拓宽了气相色谱在农药残留分析中的应用。

随着现代仪器分析方法的发展，气相色谱法已成为目前典型的，应用最广的仪器分析方法。农药测定技术的应用主要是从 20 世纪 60 年代开始的，可以这样认为，由于气相色谱的应用，特别是高灵敏度的选择性检测仪器的应用，农药残留量的测定水平提高到一个新的台阶。就在各种新的检测方法不断出现的今天，气相色谱法仍占绝对的优势，就是由过去以填充柱为主转为目前的以毛细管柱为主。由于石英毛细管柱的出现和进样系统的不断改进，大大提高了气相色谱法的分析精度、准确度和灵敏度，但气相色谱法对于挥发性差、极性和热不稳定性的农药分析较困难。AOAC 对大部分有机磷农药（如乙拌磷、丰索磷、马拉硫磷、对硫磷）在 80 年代就建立了气相色谱检测方法。我国食品理化检验国家标准方法也采用了气相色谱检测有机磷农药。该方法是利用经提取、纯化、浓缩后的有机磷农药注入气相色谱柱，程序化升温汽化后，不同的有机磷农药在固相中分离，经不同的检测器检测扫描绘出气相色谱图，通过保留时间来定性，通过峰或峰面积与标准曲线对照来定量。一次可同时测定多组分，简便快捷、灵敏度高、准确性好。而色谱条件的最佳设定是气相色谱技术的关键。日本学者 Omura M 研究了 7 种不同极性的固定相来分离检测不同极性的有机磷农药，检出限达 0.05 ~ 10ng。Marble L. K 比较了萃取法和固相净化柱来纯化有机磷农药，认为 OB - 1B 柱能获得最佳效果。而 Manes J 用 C18 柱反相柱从水中分离浓缩有机磷农药得到很好的回收率。其他研究还报道了用凝胶渗透色谱法净化谷物、饲料中的有机磷农药残留，检出限达 0.01 ~ 0.05ng。近年来毛细管气相色谱（CGC）法得到了不断的发展和完善，使毛细管柱基本代替了原来的填充柱，不仅提高了灵敏度，分辨率和检测速度，而且稳定性，使用寿命都比填充柱优良。国外 Miyabara U 和 Jonogai Y 分别用电子俘获检测和火焰光度检测的毛细管气相色谱法测定了食品中几十种有机磷农药。而在国内，杨大进等报道了可同时测定大米中 20 种有机磷农药和 6 中氨基甲酸酯农药残留的毛细管气相色谱方法。余新建等采用 HP1701 毛细管柱分离测定柑橘中 7 种有机磷农药残留量，比较了 7 种有机磷农药在柑橘不同生长时期的残留量情况，7 种有机磷农药的回收率为 83.5% ~ 106.4%，RSD = 3.0% ~ 7.8%，检出限为 0.1 ~ 20Pg/kg。另外陈泉、楼小华等也用毛细管气相色谱法分别检测了 13 种、25 种样品中不同的有机磷农药，回收率和重复性都很好。

目前，随着气相色谱仪的普及，对可能存在有机磷农药残留的样品（如食品、果蔬、谷物、饲料、茶叶、水产品、土壤、植物材料、水体、以及血液、化妆品、羊毛脂、牛奶）都建立了气相色谱检测方法。

3. 高效液相色谱法

高效液相色谱法是一种以流体为流动相的高效、快速的分离技术，与气相色谱法相比，

具有高压、高速、高效、高灵敏度的特点，不需要汽化，不受试样挥发性的限制。但是高效液相色谱法要配备昂贵的检测仪器，试剂消耗也比较大，主要用于检测一些不适用于气相色谱法检测的少数农药。

高效液相色谱法是在液相色谱柱层析的基础上，引入气相色谱理论并加以改进而发展起来的色谱分析方法。高效液相色谱法在农药残留分析领域的应用越来越广泛，是因为高效液相色谱法能适合分析沸点高而不太容易汽化、热不稳定和强极性农药及其代谢产物，且可以与柱前提取、纯化及柱后荧光衍生化反应和质谱等联用，易实现分析自动化，同时一些新型检测器的问世在一定程度上提高了高效液相色谱法的检测灵敏度。与气相色谱法相比，不仅分离效能好、灵敏度高、检测速度快，而且应用面广。AOAC 中有近半数的有机磷农药都建立了高效液相色谱法，研究报道也很多，Branca P、Sharp G. J 和 Brayan J. G 用高效液相色谱法分别测定了食品和水、谷物及其制品、水稻中的有机磷农药。Martindale R. W 比较了高效液相色谱法和气相色谱法测定马铃薯中的有机磷农药。不断改进的高效液相色谱法技术也扩大了高效液相色谱法检测有机磷农药的应用范围。Maillet V N 用 7 种不同的萃取方法提取环境中的有机磷农药，并用高效液相色谱法来检测。Carabias R. 采用紫外高效液相色谱法来检测水果中的有机磷农药。在国内，包宏等采用高效液相色谱法对西红柿中克线磷残留量进行检测，样品用等量无水 Na_2SO_4 吸收水分后，用含 1% 磷酸丙酮对样品进行提取，用 CH_2Cl_2 在样品酸性和碱性溶液中分别进行萃取，以达到净化目的。采用 ODS 柱，以甲醇—水 (30:70) 为流动相，测定波长为 250nm，方法最低检出量 ng ，回收率为 83.9% ~ 91.4%，RSD 为 2.26% ~ 4.37%。

4. 超临界流体色谱

超临界流体色谱是 20 世纪 80 年代的技术，以超临界流体为流动相，综合利用了气相色谱和一些色谱的优点，由于超临界流体具有气体和液体的双重性质，其黏度小、传质阻力小、扩散速度快，其分离能力和速度可与气相色谱相比。超临界流体色谱还可与气相和液相的检测器相匹配，可与红外光、紫外光、质谱联用，成为一种强有力的分离和检测手段。

超临界流体是物质在高于临界压力和临界温度时的一种状态，它具有气体和液体的某些特殊性质，具有气体的低黏度、液体的高密度以及介于气、液之间较高的扩散系数等特征。SFC 是 GC 和 LC 的补充，SFC 可以解决气液色谱分析中的难题，它可以分析气相色谱难气化的不挥发性样品，同时具有比高效液相色谱更高的效率，分析时间更短，因而可以方便地进行定性、定量分析。在众多分析领域已有愈来愈多的应用。

对于某些纯物质来说，具有三相点和临界点，如图 4-1 所示，从图中可以看出，物质在三相点，气、液、固三态处于平衡状态，当处于临界温度和临界压力以上时，则不论施加多大压力，气体也不会液化，此时既非气体，也非液体，而是以超临界流体形式存在。

超临界流体色谱兼有气相色谱和液相色谱的特点。它可分析气相色谱不适应的高沸点、低挥发性样品，且比高效液相色谱有更快的分析速度和条件。操作温度主要取决于所选用的流体，常用的有二氧化

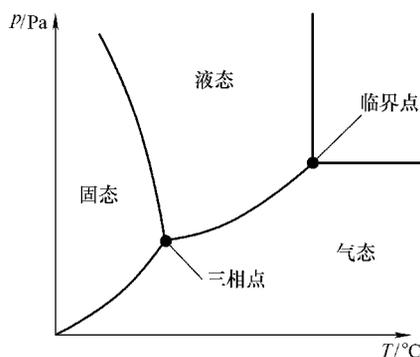


图 4-1 纯物质的相图

近红外光谱技术在食品品质检测方法中的研究

化碳及氧化亚氮。超临界流体容易控制和调节，在进入检测器前可以转化为气体、液体或保持其超临界流体状态，因此可与现有的任何液相或气相检测器相连接，能与多种类型检测器相匹配，扩大了它的应用范围和分类能力，在定性、定量方面有较大的选择范围。还可以用多种梯度技术来优化色谱条件，并且比高效液相色谱法易达到更高的柱效率。

4.2.2 生物化学测定法（胆碱酯酶抑制法）

有机磷农药是哺乳动物中毒作用的基础，通常与它们抑制中枢和周围神经系统的胆碱酯酶的能力有关。酶抑制法是利用有机磷农药的毒理特性建立的一种快速检验方法。由于有机磷农药能抑制乙酰胆碱酯酶的活性，使该酶分解乙酰胆碱的速度减慢或停止，再利用纸片或电极（即纸片法和膜电极法）作为载体将乙酰胆碱酯酶吸附在上面，如果酶的活性没有被抑制，生成了基质水解产物，使用呈色剂或发色的基质而显色。反之，如果被测样品中含有农药残留，则酶的活性被抑制，基质就不被水解，遇显色剂不显色。这样，通过纸片颜色的变化或电极的读数指示变化上可以测定有机磷农药与标准有机磷农药比较则可定量。AOAC 在 1964 年最早公布了该方法检测有机磷农药，美国 1958 年就研制出了称之为酶片（Enzyme Ticket）的商品来检测有机磷农药，灵敏度为 $0.1 \sim 10 \times 10^{-6}$ 。Sigma 公司也有类似的检测试剂盒。国内李发生、朱曜分别报道了纸片法检测食品和出口冻牛肉中的有机磷农药，灵敏度为 $0.1 \sim 3.0 \times 10^{-6}$ 。国内也有生化公司正式生产和销售称之为农药速测卡的产品。Kulys F 研制了胆碱酯酶电流传感器，Bernabei M 用胆碱和乙酰胆碱电化学传感器检测有机磷农药，李发生也报道了利用酶传感器检测 5 种有机磷农药，最低检出浓度为 1×10^{-12} mol/L。

酶抑制法最大的优点是操作简单、速度快，不需昂贵的仪器，特别适合现场检测以及大批样品的筛选，易于推广普及，但灵敏度比仪器法要差一些，重复性、回收率还有待提高。

胆碱酯酶能分解乙酰胆碱或其他酯类，而有机磷能抑制胆碱酯酶，使其失去分解活性，通过检查乙酰胆碱或其他脂类化合物的分解产物，便可判断样品中是否含有有机磷。根据对分解产物检测手段的不同，可分为酸碱指示剂法、速测卡法、分光光度计法、荧光法等。其中速测卡法和分光光度计法已被国家定为标准的分析方法。

1. 速测卡法

农药速测卡是用对农药高度敏感的胆碱酯酶和显色剂做成的酶试纸，可以快速检测蔬菜中有机磷和氨基甲酸酯的残留情况。本法抗干扰性强、操作简便，不需要仪器设备和配制试剂就能单独使用，无需任何专业技术培训。产品容易贮存、携带方便，是现场检测的最佳方法。

胆碱酯酶可催化靛酚乙酸酯（红色）水解为乙酸与靛酚（蓝色），有机磷或氨基甲酸酯类农药对胆碱酯酶有抑制作用，使催化、水解、变色过程发生改变，由此可判断出样品中是否含有有机磷或氨基甲酸酯类农药的存在。将胆碱酯酶和靛酚乙酸酯试剂固化在纸片上，制成速测卡，检出限一般在 $0.3 \sim 3.5$ mg/kg。它具有操作简便、快速、经济等特点，适用于现场筛选，但有时试纸颜色判定不明显，易造成判断与实际不符合；另外，洗脱液（缓冲溶液）对蔬菜农药的提取率在 $50\% \sim 80\%$ 。使用速测卡进行测定时有两种使用方法：

（1）整体测定法

1) 选取有代表性的蔬菜样品，擦去表面泥土，剪成 1cm^2 左右碎片，取 5g 放入带盖瓶中，加入 10mL 纯净水或缓冲溶液，振摇 50 次，静置 2min 以上。

2) 取一片速测卡, 用白色药片沾取提取液, 放置 10min 以上进行预反应, 有条件时在 37℃ 恒温装置中放置 10min。预反应后的药片表面必须保持湿润。

3) 将速测卡对折, 用手捏 3min 或用恒温装置恒温 3min, 使红色药片与白色药片叠合发生反应。

4) 每批测定应设一个纯净水或缓冲液的空白对照卡。

(2) 表面测定法 (粗筛法)

1) 擦去蔬菜表面泥土, 滴 2~3 滴洗脱液在蔬菜表面, 用另一片蔬菜在滴液处轻轻摩擦。

2) 取一片速测卡, 将蔬菜上的液滴滴在白色药片上。

3) 放置 10min 以上进行预反应, 有条件时在 37℃ 恒温装置中放置 10min, 预反应后的药片表面必须保持湿润。

4) 将速测卡对折, 用手捏 3min 或用恒温装置恒温 3min, 使红色药片与白色药片叠合发生反应。

5) 每批测定应设一个洗脱液的空白对照卡。

结果判定: 与空白对照卡比较, 白色药片不变色或略有浅蓝色均为阳性结果, 不变蓝为强阳性结果, 说明农药残留量较高, 显浅蓝色为弱阳性结果, 说明农药残留量相对较低。白色药片变为天蓝色或与空白对照卡相同, 为阴性结果。对阳性结果的样品, 可用其他分析方法进一步确定具体农药的品种和含量。

2. 分光光度计法 (抑制率法)

在分光光度计中, 将不同波长的光连续地照射到一定浓度的样品溶液时, 便可得到不同波长相对应的吸收强度。如以波长 (λ) 为横坐标, 吸收强度 (A) 为纵坐标, 就可绘出该物质的吸收光谱曲线。利用该曲线进行物质定性、定量的分析方法, 称为分光光度法, 也称为吸收光谱法。用紫外光源测定无色物质的方法, 称为紫外分光光度法; 用可见光光源测定有色物质的方法, 称为可见光光度法。它们与比色法一样, 都以比尔-朗伯定律为基础。

近年来, 紫外光及可见光分光光度分析已得到广泛的应用, 它不仅可用于物质的鉴定及结构分析, 还可以用于某些物质含量的测定。

分光光谱技术可用于: ①通过测定某种物质吸收或发射光谱来确定该物质的组成; ②通过测量适当波长的信号强度来确定某种单独存在或与其他物质混合存在的一种物质的含量; ③通过测量某一种底物消失或出现的量同时间的关系, 追踪反应过程。

紫外及可见光分光光度法, 这是一种只在可见光及紫外光光谱应用范围内测量物质吸收辐射线的技术, 应用十分广泛。其中分光光度计可用于精确测量特定波长的吸收值, 而比色计则是一种较简单的测量仪器, 其原理是利用滤光片来测量较宽波段 (如可见光中的绿光、红光或蓝光范围) 的吸收值。

溶液对光的吸收有两个基本法则: ①透过溶液的光的吸收值同吸收溶质的分子数目 (即溶质浓度 C) 呈指数相关; ②透过溶液的光的吸收值同透过吸收溶液的路径长度 l 成指数相关。这两条法则包括在比尔-朗伯关系式中。通常以入射光 (I_0) 和透射光 (I) 的吸收光度来表示:

$$A = \lg(I_0/I) = \varepsilon cl$$

式中, $\lg(I_0/I)$ 称为吸光度, 旧称光密度; ε 为吸光系统或吸收系统; c 为样品浓度, 单位

近红外光谱技术在食品品质检测方法中的研究

为 mol/L 或 g/L; l 为光程。吸收值常用下角表示其波长, 如 A_{550} 表示 550nm 处的吸收值。透过溶液的光的比例称为透光率 (T), 可由出射光和入射光的比值求得。

比色计用于测定颜色明显, 并且是溶液主要组分的待测物, 如血液中的血红细胞, 也可以在待测物中加入一种试剂, 使其形成有色产物 (一种生色团), 如用茚三酮法测定氨基酸含量。定量分析某种物质要做标准曲线, 标准曲线是在测定待测样品的同时测定已知含量的物质来制成的, 而不是使用比尔-朗伯关系。

光源通常为钨丝灯泡, 通过一个凸透镜聚焦后产生一束平行光, 平行光穿过装有溶液的玻璃样品或小池, 然后透过一个有色滤光片到达光电探测器, 检测仪产生一个同落在光电探测器上的光密度成正比的电势, 来自于光电探测器的信号被放大然后传递到电流计或数字读数器。

比色计的使用: ①接通电源使仪器稳定, 使用前至少要让灯预热 5min; ②选择一种同底物颜色互补的滤光器; ③调零 (用空白对照调零); ④调整灵敏度; ⑤分析样品及标准溶液; ⑥由于不同比色杯的吸光特性、杯壁厚度不同, 因此为了提高精确度, 同一试验应用同一比色杯, 且在比色槽中摆放的方位相同; ⑦每次测样前清洗比色杯; ⑧经常重复测定同一溶液检验比色计的可重复性; ⑨用标准溶液绘制标准曲线。

由于大多数过滤器过滤出来的光的波带很宽, 因而比色计既不能用于确定某种复合物, 也无法分辨在混合液中吸收特性非常相近的两种物质。比色计所用光电探测器的变化系数为 0.5% 左右, 因而不适合要求具有高度精确性的工作。使用这种最简单的仪器, 由于仪表上对数测量刻度单位的随意性, 即使是把表上的灵敏度/刻度调节到零控点, 在一个仪器上获得的值不可直接同另一台仪器上测得的值相比较, 同一仪器的不同设置之间也不可直接比较。比色计对于特定波长的量化工作是不合适的。

分光光度计之所以优于比色计就在于使用了一个衍射光栅将光源的复色光转换为单色平行光束。实际上从这种单色平行光束产生的光不是某个波长的光, 而是一段窄的带宽上的光, 带宽是分光光度计的一个重要特性, 这是由于它决定了吸收测量中所用的波长。普通分光光度计的带宽为 5 ~ 10nm, 用于研究的仪器的带宽小是因为光栅夹缝的宽度影响着带宽, 带宽随光栅夹缝宽度的减少而降低, 要获得特定波长下的精确数据, 尽可能使用最小的缝宽度。然而, 减小缝宽也会减少到达监测器的光, 降低了信噪比。缝宽可减小的程度取决于检测放大系统的灵敏度及稳定性。

大多数可见光分光光度计使用的比色杯的光穿过路径为 10mm。一次性塑料杯适合于对水和乙醇溶液在可见光范围内的测量。玻璃比色杯的生产要求更加严格的标准, 因而在精确研究中要使用玻璃比色杯, 尤其当溶液的吸收值很低时 (<0.1), 即使盛对照液与待测样品液的比色杯在光学性质上有稍许不同, 也会导致结果偏差。玻璃和塑料会吸收紫外光, 因此在测波长小于 300nm 的吸收值时要使用石英杯。

进行测量之前, 比色杯要保证干净、无划痕、外表面干燥、盛液到适当高度, 并放在比色槽的正确位置上。生物样品中的蛋白质和核酸可能会在玻璃、石英杯的内表面沉积, 因而要用棉球沾上丙酮擦去比色杯内的沉淀, 或用 1mol/L 硝酸浸泡过夜。腐蚀性及毒性溶液必须使用有盖子的比色杯, 以防止溅出, 破坏仪器。

基本分光光度计使用的光电探测器类似于比色计中所使用的光电探测器。许多情况下, 当波长高于 600nm 和低于 550nm 时必须使用不同的光电探测器, 这是因为它们可在可见光波

长内的灵敏度不同,更精准的仪器中所使用的是具有比光电探测器更高灵敏度和稳定性的检测器。数字显示由于不易产生视觉错误和误读范围的错误,正逐渐代替指针读数。一些仪器可以直接给出所测定物质的浓度。

基本分光光度计只产生单束光。这种仪器首先用空白对照调到零吸收值,然后取出空白液,加入待测液,测定待测液的吸收值。也有一种双束分光光度计,有单色光源产生的光束被分为两束,一束穿过待测液,另一束穿过空白液。吸收值由一个电子电路通过对比透过待测液及空白液的出射光进行测定。双光束分光光度计减少了由于光源输出的不稳定或检测系统灵敏度的变化而导致的测量错误,这是由于待测液与对照液是同时进行测量的。记录式分光光度计是一种双束测定仪,用于记录已知波段下吸收值随时间的变化(如用于酶分析)。

假如已知一种物质在某一波长下的吸光率(通常是该物质的最大吸收值,这时灵敏度最高),这种物质纯溶液的浓度可用比尔-朗伯关系式算出。摩尔吸光系数是指物质在1mol/L的浓度下,比色杯厚度为1cm时的吸收值。该值可以从光谱数据表中查到,也可以用实验方法通过测量一系列已知浓度的物质的吸收值来绘制一条标准曲线。这样,在所要求的浓度范围内,便可确定吸收值与浓度之间存在的线性关系,该直线的斜率即摩尔吸光系数。

比吸光率是指物质质量溶液浓度为10g/L时,比色杯厚度为1cm时测定的吸光值。该值对于未知分子质量的物质如蛋白质核酸的测定很有用,这种情况下溶液中物质的含量以其质量表示而不用摩尔浓度表示。使用公式 $\lg(I_0/I) = \epsilon cl$,比吸光率要除以10才可以得到一个以g/L为单位的浓度值。

正常情况下,酶催化乙酰胆碱水解,其水解产物与显色剂反应,产生黄色物质。在一定条件下,有机磷和氨基甲酸酯类农药对胆碱脂酶的活性有抑制作用,使水解过程受抑制,其抑制率与农药的浓度呈正相关。用分光光度计在412nm处测定吸光度随时间的变化值,计算出抑制率,当抑制率 $\geq 50\%$ 时,表示样品有有机磷或氨基甲酸酯类农药存在(丁酰胆碱酯酶的抑制率一般控制在70%以上为阳性样品)。对检出的阳性样品,可用其他方法进一步确定具体农药品种和含量。检出限一般在0.05~5.0mg/kg,检出时间为30min。该法以数字形式读取数据,较为直观,可避免目视法因显色不明显而判断不准。

3. 速测灵法(金属离子催化显色表面皿法)

农药残留特别是高毒农药残留直接影响人体健康,长期食用含有残留农药的农产品,可导致人体胃肠道疾病、身体免疫力下降,出现慢性腹泻、恶心、经常性感冒、头晕、心悸、盗汗、失眠、健忘等症状。残留农药进入体内,主要依靠肝脏制造酶来吸收这些毒素,进行氧化分解。如果长期食用带有残留农药的食品,肝脏就会不停地工作来分解这些毒素,长时间的超负荷工作会引起肝硬化、肝积水等一些肝脏病变,残留农药中含有的化学物质还可促使人体组织细胞发生癌变。速测灵法(金属离子催化显色表面皿法)依据金属离子催化显色表面皿法研制而成。检测时,食品中的残留农药可迅速与本试剂发生水解反应,水解产物与显色剂反应,使显色剂的紫红色褪去。这种方法为化学试剂检测法,一次可同时检测5个样品,具有快速、简便、成本低等特点,避免了生化试剂需特定的保存和反应条件。

有机磷农药(磷酸酯、二硫代酸酯、磷酰胺)在金属离子催化作用下水解为磷酸与醇,水解产物与显色剂反应,使显色剂的紫红色褪去变成无色。就甲胺磷、对硫磷、敌敌畏而言,检测的灵敏度为0.18~10.0mg/kg。对阳性结果,基本可判定农药残留超标。该法避免

近红外光谱技术在食品品质检测方法中的研究

了使用生化法（酶抑制法）受酶效价的影响，不需使用仪器，操作简便、快速、经济。局限性在于主要针对的是甲胺磷、对硫磷等农药残留定性。该方法适用于食用农产品，主要是叶菜类蔬菜、瓜果、豆类等中的残留高毒有机磷农药。

使用该方法进行检测时的检测步骤如下：①用吸管吸取 1mL 洗脱液，放入洗净的提取杯中，再加洁净水至 20mL；②用吸管吸取提取杯中的洗脱液，反复冲洗待测样品正反页面各 3 次，静置片刻后，其上清液即样品液；③用吸光从提取杯中吸取 1mL 样品液，加入多孔比色管中，建议设对照孔并加等量清洁水（显色管加 5mL，使溶液高度与多孔比色管相同）；④在上述多孔比色管中依次滴加 1 滴检测液 A 和检测液 B，搅拌混匀后，静置片刻，观察样品颜色变化，并对照比较后判断。如果比色管中液体紫红色全部褪去，则表明农药残留超量，不能食用；若不变色或变色不显著，表明无农药残留，或不超量；多点抽样，重复测定，正确判断。必要时留样进一步以仪器法确证、定量。在进行测定时要注意以下 6 个问题：

- 1) 取样时，注意样品的代表性，清洗要均匀到位。
- 2) 检测时，检测液 A、检测液 B 在多孔比色管的孔中或显色管中与样品液混合一定要搅拌均匀。
- 3) 发现超标或有争议的检测结果，可进一步采用定量法检测。
- 4) 每次使用多孔比色管等器皿后，需将孔内液体甩干，并用清水冲洗干净，晾干备用。
- 5) 试剂有轻度腐蚀性，禁入口眼。
- 6) 阴凉处避光密封保存，保存期 12 个月以上。

4.2.3 发光菌检测法

发光菌检测法应用于环境毒物的检测源自 20 世纪 80 年代。由于其检测速度快、灵敏度高、设备简单以及具有极好的可扩展性，在环境样品的毒性检测中得到了迅猛的发展。研究表明，大部分有机污染物对发光细菌毒性与对多种水生生物的毒性呈明显正相关。我国于 1995 年也将这一方法列为环境生物毒性检测的标准方法。通过研究表明，不同种类的发光细菌的发光机理是相同的，是由分子氧作用，胞内荧光酶催化，将还原态的黄素单核苷酸（FMNH₂）及长链脂肪醛氧化为黄素单核苷酸（FMN）和长链脂肪酸，同时释放出最大发光强度在波长为 450 ~ 490nm 的蓝绿光。

发光菌是一类能运动的革兰氏阴性兼性厌氧杆菌，一般将其分为三类：弧菌属、发光杆菌属和异短杆菌属。发光菌的发光现象是其正常的代谢活动，在一定条件下发光强度是恒定的，当与外来受试物（无机、有机毒物，抑菌、杀菌物等）接触后，其发光强度即有所改变，变化的大小与该物质的毒性大小有关，与特定受试物的浓度成相关关系。发光菌检测法是利用灵敏的光电测量系统测定毒物对发光菌发光强度的影响。

发光菌检测法在环境监测中已经得到了广泛的应用，而现在食品当中的大多数污染物也可以用发光菌检测法进行检测。但是由于食品中成分复杂，而且污染物浓度较低，以前用的仪器设备无法达到检测低限。因此发光细菌用于食品安全的检测还不多见，但是随着对该方法的不断深入研究、检测仪器的日益完善、检测范围和检测限的不断增加，作为一种速度快、灵敏度高、成本低廉的生物毒性监测方法—发光菌检测法将在食品安全检测中有着广泛

的应用。

4.2.4 生物传感器法

生物传感器法是利用生物分子的催化作用或特异性结合来检测物质。生物传感器是由分子识别元件（酶、抗体、受体、细胞及组织器官等）与信号转换器（电极半导体、光导纤维、压电晶体及表面声波、热敏电阻等）结合在一起的装置。现阶段测定有机磷农药的生物传感器大多数是以胆碱酯酶（乙酰胆碱酯酶或丁酰胆碱酯酶）为分子识别元件的，按信号转换器的不同可分为电化学生物传感器、半导体器件生物传感器、光导纤维生物传感器、压电晶体传感器和表面声波生物传感器。在有机磷农药检测的应用中，研究得较多的是基于胆碱酯酶作分子识别元件的电化学生物传感器，分为电位型和电流型（安培型）两种，均是基于有机磷对胆碱酯酶活性的抑制作用而设计的。乙酰胆碱在乙酰胆碱酯酶的催化下水解为乙酸和胆碱，生成的乙酸可以电离而使溶液的 pH 值发生改变。有机磷农药可抑制胆碱酯酶活性而影响 pH 值的变化，可用玻璃 pH 值电极或对溶液 pH 值变化敏感的金属氧化物电极（Pd/PdO 或 Ir/IrO₂ 电极）作基底电极，通过检测 pH 值变化来间接测定有机磷农药。

传感器法是将传感技术与农药免疫分析技术相结合而建立起来的，可以说它是免疫分析技术的一种延伸或分支。应用在痕量分析上的有：①生物传感器，生物传感器（Biosensor）是由一种生物敏感部件与转换器紧密配合，对特定化学物质或生物活性物质具有选择性和可逆响应的分析装置；②固相传感器，它通常用压电晶体，当附加电压时，晶体振动的能力取决于晶体的质量和大小。在晶体上包被了能吸附目标物的材料，当晶体与配基接触时就会发生吸附，随着晶体的质量和频率的变化而发生变化，其频率的改变与吸附在晶体上物质的含量有关，依此进行定量测定。这种方法的局限性就是包被的晶体只能使用一周左右。

农药残留分析是一门综合性很高、涉及面很广的分析科学。速测技术更能适应现代高效、快速的节奏和满足社会的需求。仪器分析法可以保证数据的精确性，相对来说，其流程仍比较繁琐，但在残留分析技术中，其地位仍是不可取代的。免疫法的开发过程需投入较多资金和较长时间，但其优点是灵敏度与常规的仪器分析一致，对现场筛选有一定的作用。生物传感器是高科技的产物，它能很快地测定蔬菜水果上的农药及其他化学污染物，操作起来更加方便，且可以重复使用。总之，如何研究出快速、灵敏、简便的农药速测技术，是值得我国相关政府部门、科学研究人员和企业考虑的重要问题之一。

4.2.5 免疫分析法

免疫分析（IA）法是将免疫反应与现代测试手段相结合的超微量测定技术。IA 法被称为使用抗体作为“生物化学检测器”的分析技术，是基于抗原抗体特异性识别和结合反应为基础的分析方法。它的开发过程需要投入较多资金和较长时间，但具有简单、快速、灵敏度高、特异性强、廉价、样品所需量少等优点，其迅速发展成为现代生物化学、医学、临床化学和药物化学等研究领域的重要工具。常见的方法有：①荧光免疫测定法；②酶免疫测试法；③放射免疫测定法；④流动注射免疫分析法。这些方法是通过化学制剂来完成的，由于其样品处理过程繁琐，容易污染样品和环境，它们的应用在某种程度上受到了局限。国内还有深圳天福贸易有限公司生产的农药速测卡，也是根据有机磷农药能强烈抑制胆碱酯酶的活性这一特点而研制开发的。另外，国外一些公司推出了多种酶标试剂盒应用于常规分析及田

间检测的快速筛选，作为仪器分析的辅助方法发挥了一定的作用。但这些速测方法出现“假阳性”引起的纠纷也不少。

目前用于农药残留分析的免疫分析技术主要有放射免疫分析（Radioimmunoassay Immunoassay Analyze, RIA）和酶免疫分析（Enzyme Immunoassay Analyze, EIA）两种。RIA 检测放射性同位素需要昂贵的仪器设备和防辐射设备，而且需要专业技术人员，应用较少。EIA 不存在放射性危害，具有灵敏度高、特异性强、成本低、操作系统简便快速、可自动化等优点。目前国外已研制出几十种农药的酶免疫检测试剂盒，美国使用的 ELISA 试剂盒对有机磷农药最小检出浓度为 $20\mu\text{g}/\text{mL}$ ，免疫检测试剂盒使用起来简便、快速，样品不需净化或只需简单的净化，技术性要求不高，易于推广。

4.2.6 光谱法

光谱法是基于物质与辐射能的作用测量物质内部发生量子化的能级之间因跃迁而产生的发射、吸收或散射辐射的波长和强度进行分析的方法。

光谱法能够根据有机磷农药中的某些官能团或水解、还原产物与特殊的显色剂在特定的环境下发生氧化、磺酸化、脂化、络合等化学反应，产生特定波长的颜色反应来进行定性或定量的测定。也可以根据有机磷农药含氢基团在近红外光谱区的吸收来进行测定。如近红外光谱对有机磷农药毒死蜱的 C-H 键的一级倍频吸收区在 $1600 \sim 1800\text{nm}$ ($5555 \sim 6250\text{cm}^{-1}$)，伸缩振动和弯曲振动的合频吸收带主要在 $2000 \sim 2400\text{nm}$ ($4166 \sim 5000\text{cm}^{-1}$)。

光谱法可分为原子光谱法和分子光谱法。原子光谱法是由原子外层或内层电子能级的变化产生的，它的表现形式为线光谱。属于这类分析方法的有原子发射光谱（AES）法、原子吸收光谱（AAS）法、原子荧光光谱（AFS）法以及 X 射线荧光光谱（XFS）法等。

1. 原子光谱法

(1) 原子发射光谱法

原子发射光谱法，是依据各种元素的原子或离子在热激发或电激发下，发射特征的电磁辐射，而进行元素的定性分析与定量分析的方法，是光谱学各个分支中最为古老的一种。一般认为原子发射光谱是 1860 年德国学者基尔霍夫（Kirchhoff G. R.）和本生（Bunsen R. W.）首先发现的，他们利用分光镜研究盐和盐溶液在火焰中加热时所产生的特征光辐射，从而发现了 Rb 和 Cs 两个元素。其实在更早的时候，1826 年泰尔博（Talbot）就说明某些波长的光线是表征某些元素的特征。从此以后，原子发射光谱就为人们所重视。原子发射光谱分析，一般有光谱的获得和光谱的分析两大过程，具体可分为：①试样的处理，根据进样方式的不同进行处理，做成粉末或溶液等，有些时候还要进行必要的分离或富集；②样品的激发，在激发源上进行，激发源把样品蒸发、分解原子化和激发；③光谱的获得和记录，从光谱仪中获得光谱并进行记录；④光谱的检测，用检测仪器进行光谱的定性、半定量、定量分析。

原子发射光谱法具有以下优点：①同时检出多元素的能力强，可同时检测一个样品中的多种元素。一个样品一经激发，样品中各元素都各自发射出其特征谱线，可以进行分别检测而同时测定多种元素。②分析速度快，样品多数不需经过化学处理就可分析，且固体、液体试样均可直接分析，同时还可多元素同时测定，若用光电直读光谱仪，则可在几分钟内同时做几十个元素的定量测定。③选择性好，由于光谱的特征性强，所以对于一些化学性质极相似的元素的分析具有特别重要的意义。如铈和钽、锆和铪，十几种稀土元素的分析用其他方

法都很困难，而对于原子发射光谱法来说则是毫无困难。④检出限低，用电感耦合等离子体(ICP)新光源，检出限可低至 $\text{ng} \cdot \text{mL}$ 数量级。⑤用 ICP 光源时，准确度高，标准曲线的线性范围宽，可达 4~6 个数量级。可同时测定高、中、低含量的不同元素。因此 ICP - AES 已广泛应用于各个领域之中。⑥样品消耗少，适于整批样品的多组分测定，尤其是定性分析更显示出独特的优势。但是也存在一定的问题：①在经典分析中，影响谱线强度的因素较多，尤其是试样组分的影响较为显著，所以对标准参比的组分要求较高。②含量(浓度)较大时，准确度较差。③只能用于元素分析，不能进行结构、形态的测定。④大多数非金属元素难以得到灵敏的光谱线。

(2) 原子吸收光谱法

原子吸收光谱法是利用气态原子可以吸收一定波长的光辐射，使原子中外层的电子从基态跃迁到激发态的现象而建立的。由于各种原子中电子的能级不同，将有选择性地共振吸收一定波长的辐射光，这个共振吸收波长恰好等于该原子受激发后发射光谱的波长，由此可作为元素定性的依据，而吸收辐射的强度可作为定量的依据。AAS 现已成为无机元素定量分析方法中应用最广泛的一种分析方法。

原子吸收光谱法具有检出限低(火焰法可达 $\mu\text{g}/\text{cm}^{-3}$ 级)、准确度高(火焰法相对误差小于 1%)、选择性好(即干扰少)、分析速度快、应用范围广(火焰法可分析 30 多种/70 多种元素，石墨炉法可分析 70 多种元素，氢化物发生法可分析 11 种元素)等优点。

在温度吸收光程、进样方式等实验条件固定时，样品产生的待测元素相基态原子对作为锐线光源的该元素的空心阴极灯所辐射的单色光产生吸收波峰，其吸光度(A)与样品中该元素的浓度(C)成正比，即 $A = KC$ ，式中 K 为常数。据此，通过测量标准溶液及未知溶液的吸光度，又已知标准溶液浓度，可作标准曲线，求得未知液中待测元素浓度。该方法主要适用于样品中微量及痕量组分分析。

每一种元素的原子不仅可以发射一系列特征谱线，也可以吸收与发射线波长相同的特征谱线。当光源发射的某一特征波长的光通过原子蒸气时，即入射辐射的频率等于原子中的电子由基态跃迁到较高能态(一般情况下都是第一激发态)所需要的能量频率时，原子中的外层电子将选择性地吸收其同种元素所发射的特征谱线，使入射光减弱。特征谱线因吸收而减弱的程度称吸光度 A ，与被测元素的含量成正比：

$$A = KC \quad (4-1)$$

式中， K 为常数， K 包含了所有的常数； C 为试样浓度。

此式就是原子吸收光谱法进行定量分析的理论基础。由于原子能级是量子化的，因此，在所有的情况下，原子对辐射的吸收都是有选择性的。由于各元素的原子结构和外层电子的排布不同，元素从基态跃迁至第一激发态时吸收的能量不同，因而各元素的共振吸收线具有不同的特征。原子吸收光谱位于光谱的紫外区和可见区。

原子吸收光谱法具有以下优越性：

1) 选择性强。这是因为原子吸收带宽很窄的缘故。因此，测定比较快速、简便，并有条件实现自动化操作。在发射光谱分析中，当共存元素的辐射线或分子辐射线不能和待测元素的辐射线相分离时，会引起表观强度的变化。而对原子吸收光谱分析来说：谱线干扰的概率小，由于谱线仅发生在主线系，而且谱线很窄，线重叠概率较发射光谱要小得多，所以光谱干扰较小。即便是和邻近线分离得不完全，由于空心阴极灯不发射那种波长的辐射线，所

以辐射线干扰少，容易克服。在大多数情况下，共存元素不对原子吸收光谱分析产生干扰。在石墨炉原子吸收法中，有时甚至可以用纯标准溶液制作的校正曲线来分析不同试样。

2) 灵敏度高。原子吸收光谱法是目前最灵敏的方法之一。火焰原子吸收法的灵敏度是 $10^{-9} \sim 10^{-6}$ 级，石墨炉原子吸收法绝对灵敏度可达到 $10^{-14} \sim 10^{-10}$ g。常规分析中大多数元素均能达到 10^{-6} 数量级。如果采用特殊手段，例如预富集，还可进行 10^{-9} 数量级浓度范围测定。由于该方法的灵敏度高，使分析手续简化可直接测定，缩短分析周期、加快测量进程；由于灵敏度高，需要进样量少。无火焰原子吸收分析的试样用量仅需试液 $5 \sim 100\mu\text{L}$ 。固体直接进样石墨炉原子吸收法仅需 $0.05 \sim 30\text{mg}$ ，这对于试样来源困难的分析是极为有利的。譬如，测定小儿血清中的铅，取样只需 $10\mu\text{L}$ 即可。

3) 分析范围广。发射光谱分析和元素的激发能有关，故对发射谱线处在短波区域的元素难以进行测定。另外，火焰发射光度分析仅能对元素的一部分加以测定。例如，钠只有1%左右的原子被激发，其余的原子则以非激发态存在。

在原子吸收光谱分析中，只要使化合物离解成原子，不必激发，所以测定的是大部分原子。应用原子吸收光谱法可测定的元素达73种。就含量而言，既可测定低含量和主量元素，又可测定微量、痕量甚至超痕量元素；就元素的性质而言，既可测定金属元素、类金属元素，又可间接测定某些非金属元素，也可间接测定有机物；就样品的状态而言，既可测定液态样品，也可测定气态样品，甚至可以直接测定某些固态样品，这是其他分析技术所不能及的。

4) 抗干扰能力强。第三组分的存在，等离子体温度的变动，对原子发射谱线强度影响比较严重，而原子吸收谱线的强度受温度影响相对来说要小得多。和发射光谱法不同，不是测定相对于背景的信号强度，所以背景影响小。在原子吸收光谱分析中，待测元素只需从它的化合物中离解出来，而不必激发，故化学干扰也比发射光谱法少得多。

5) 精密度高。火焰原子吸收法的精密度较好。在日常的一般低含量测定中，精密度为1%~3%。如果仪器性能好，采用高精度测量方法，精密度为<1%。无火焰原子吸收法较火焰法的精密度低，一般可控制在15%之内。若采用自动进样技术，则可改善测定的精密度。火焰法：RSD <1%，石墨炉3%~5%。

原则上讲，不能多元素同时分析。测定元素不同，必须更换光源灯，这是它的不便之处。原子吸收光谱法测定难熔元素的灵敏度还不太令人满意。在可以进行测定的70多个元素中，比较常用的仅30多个。当采用将试样溶液喷雾到火焰的方法实现原子化时，会产生一些变化因素，因此精密度比分光光度法差。还不能测定共振线处于真空紫外区域的元素，如磷、硫等。标准工作曲线的线性范围窄（一般在一个数量级范围），这给实际分析工作带来不便。对于某些基体复杂的样品分析，尚存某些干扰问题需要解决。在高背景低含量样品测定任务中，精密度下降。如何进一步提高灵敏度和降低干扰，仍是当前和今后原子吸收光谱分析工作者研究的重要课题。

(3) 原子荧光光谱法

原子荧光光谱法是介于原子发射光谱和原子吸收光谱之间的光谱分析技术。它的基本原理是基态原子（一般为蒸气状态）吸收合适的特定频率的辐射，被激发至高能态，而后激发过程中以光辐射的形式发射出特征波长的荧光。

它是测量待测元素的原子蒸气在一定波长的辐射能激发下对发射的荧光强度进行定量分

析的方法。原子荧光的波长在紫外光、可见光区。气态自由原子吸收特征波长的辐射后，原子的外层电子从基态或低能态跃迁到高能态，约经 $8 \sim 10\text{s}$ ，又跃迁至基态或低能态，同时发射出荧光。若原子荧光的波长与吸收线波长相同，称为共振荧光；若不同，则称为非共振荧光。共振荧光强度大，分析中应用最多。在一定条件下，共振荧光强度与样品中某元素浓度成正比。该法的优点是灵敏度高，目前已有 20 多种元素的检出可用原子吸收光谱法和原子发射光谱法；谱线简单；在低浓度时校准曲线的线性范围宽达 3 ~ 5 个数量级，特别是用激光作激发光源时更佳。主要用于金属元素的测定，在环境科学、高纯物质、矿物、水质监控、生物制品和医学分析等方面有广泛的应用。

物质吸收电磁辐射后受到激发，受激原子或分子以辐射去活化，再发射波长与激发辐射波长相同或不同的辐射。当激发光源停止辐照试样之后，再发射过程立即停止，这种再发射的光称为荧光；若激发光源停止辐照试样，再发射过程还延续一段时间，这种再发射的光称为磷光。荧光和磷光都是光致发光。

原子荧光光谱法具有很高的灵敏度，校正曲线的线性范围宽，能进行多元素同时测定。这些优点使得它在冶金、地质、石油、农业、生物医学、地球化学、材料科学、环境科学等领域内获得了相当广泛的应用。

(4) X 射线荧光光谱法

X 射线是一种电磁辐射，其波长介于紫外线和 γ 射线之间。它的波长没有一个严格的界限，一般来说是指波长为 $0.001 \sim 50\text{nm}$ 的电磁辐射。对分析化学家来说，最感兴趣的波段是 $0.01 \sim 24\text{nm}$ ， 0.01nm 左右是超铀元素的 K 系谱线， 24nm 则是最轻元素 Li 的 K 系谱线。1923 年赫维西 (Hevesy G. Von) 提出了应用 X 射线荧光光谱进行定量分析，但由于受到当时探测技术水平的限制，该法并未得到实际应用，直到 20 世纪 40 年代后期，随着 X 射线管、分光技术和半导体探测器技术的改进，X 荧光分析才开始进入蓬勃发展的时期，成为一种极为重要的分析手段。

当能量高于原子内层电子结合能的高能 X 射线与原子发生碰撞时，驱逐一个内层电子而出现一个空穴，使整个原子体系处于不稳定的激发态，激发态原子寿命为 $10^{-14} \sim 10^{-12}\text{s}$ ，然后自发地由能量高的状态跃迁到能量低的状态。这个过程称为驰豫过程。驰豫过程既可以是非辐射跃迁，也可以是辐射跃迁。当较外层的电子跃迁到空穴时，所释放的能量随即在原子内部被吸收而逐出较外层的另一个次级光电子，此过程称为俄歇效应，亦称次级光电效应或无辐射效应，所逐出的次级光电子称为俄歇电子。它的能量是特征的，与入射辐射的能量无关。当较外层的电子跃入内层空穴所释放的能量不在原子内被吸收，而是以辐射形式放出时，便产生 X 射线荧光，其能量等于两能级之间的能量差。因此，X 射线荧光的能量或波长是特征性的，与元素有一一对应的关系。图 4-2 给出了 X 射线荧光和俄歇电子产生过程示意图。

K 层电子被逐出后，其空穴可以被外层中任一电子所填充，从而可产生一系列的谱线，称为 K 系谱线：由 L 层跃迁到 K 层辐射的 X 射线叫 K_{α} 射线，由 M 层跃迁到 K 层辐射的 X 射线叫 K_{β} 射线…。同样，L 层电子被逐出可以产生 L 系辐射，如图 4-3 所示。

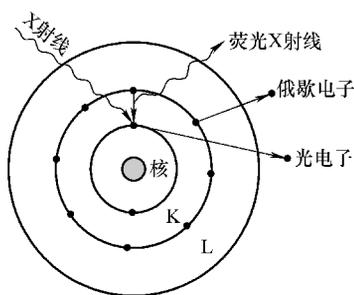


图 4-2 X 射线荧光和俄歇电子产生过程示意图

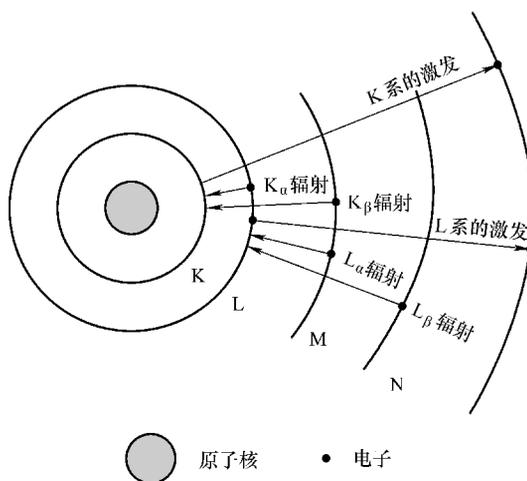


图 4-3 L 系辐射

如果入射的 X 射线使某元素的 K 层电子激发成光电子后 L 层电子跃迁到 K 层，此时就有能量 ΔE 释放出来，且 $\Delta E = E_K - E_L$ ，这个能量是以 X 射线形式释放，产生的就是 K_α 射线，同样还可以产生 K_β 射线、L 系射线等。莫斯莱 (H. G. Moseley) 发现，荧光 X 射线的波长 λ 与元素的原子序数 Z 有关，其数学关系如下：

$$\lambda = K(Z - S) - 2 \tag{4-2}$$

这就是莫斯莱定律，上式中 K 和 S 是常数。而根据量子理论，X 射线可以看成由一种量子或光子组成的粒子流，每个光具有的能量为

$$E = h\nu = hc/\lambda \tag{4-3}$$

式中， E 为 X 射线光子的能量，单位为 keV； h 为普朗克常数； ν 为光波的频率； C 为光速。

因此，只要测出荧光 X 射线的波长或者能量，就可以知道元素的种类，这就是荧光 X 射线定性分析的基础。此外，C 荧光 X 射线的强度与相应元素的含量有一定的关系，据此可以进行元素定量分析。

2. 分子光谱法

分子光谱法是由分子中电子能级、振动和转动能级的变化产生的，表现形式为带光谱。属于这类分析方法的有 紫外-可见分光光度 (UV-Vis) 法、红外光谱 (IR) 法、分子荧光光谱 (MFS) 法和分子磷光光谱 (MPS) 法等。

(1) 紫外-可见分光光度法

紫外-可见分光光度法是根据物质分子对波长为 200 ~ 760nm 这一范围的电磁波的吸收特性所建立起来的一种定性、定量和结构分析方法，操作简单、准确度高、重现性好。波长长 (频率小) 的光线能量小，波长短 (频率大) 的光线能量大。分光光度测量是关于物质分子对不同波长和特定波长处的辐射吸收程度的测量。

运动分子的外层电子通过吸收外来辐射从而产生电子能级跃迁进而形成分子吸收光谱，如图 4-4 所示。

紫外-可见分光光度法是分子光谱方法，它利用分子对外来辐射的吸收特性。UV-Vis 涉及分子外层电子的能级跃迁，光谱区在 200 ~ 800nm。主要用于分子的定量分析，紫外

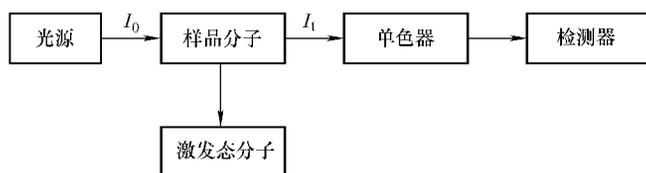


图 4-4 紫外-可见光光谱的形成

(UV) 光谱为四大波谱之一，是鉴定许多化合物，尤其是有机化合物的重要定性工具之一。应用范围包括：①定量分析，广泛用于各种物料中微量、超微量和常量的无机和有机物质的测定；②定性和结构分析，紫外吸收光谱还可用于推断空间阻碍效应、氢键的强度、互变异构、几何异构现象等；③反应动力学研究，即研究反应物浓度随时间变化的函数关系，测定反应速度和反应级数，探讨反应机理；④研究溶液平衡，如测定化合物的组成，稳定常数、酸碱离解常数等。

(2) 红外光谱法

红外光谱法又称“红外分光光度分析法”，简称“IR”，是分子吸收光谱的一种。利用物质对红外光区的电磁辐射的选择性吸收来进行结构分析，及对各种吸收红外光的化合物进行定性和定量分析。被测物质的分子在红外线照射下，只吸收与其分子振动、转动频率相一致的红外光谱。对红外光谱进行剖析，可对物质进行定性分析。化合物分子中存在着许多原子团，各原子团被激发后，都会产生特征振动，其振动频率也必然反映在红外吸收光谱上。据此可鉴定化合物中的各种原子团，也可进行定量分析。具有特征性强、测定快速、不破坏试样、试样用量少、操作简便、能分析各种状态的试样、分析灵敏度较低、定量分析误差较大等特点。

红外光谱具有鲜明的特征性，其谱带的数目、位置、形状和强度都随化合物不同而各不相同。因此，红外光谱法是定性鉴定和结构分析的有力工具。

1) 已知物的鉴定。将试样的谱图与标准品测得的谱图相对照，或者与文献上的标准谱图（例如《药品红外光谱图集》、Sadtler 标准光谱、Sadtler 商业光谱等）相对照，即可定性。使用文献上的谱图应当注意：试样的物态、结晶形状、溶剂、测定条件以及所用仪器类型均应与标准谱图相同。

2) 未知物的鉴定。未知物如果不是新化合物，标准光谱已有收录的，可有两种方法来查对标准光谱：

- ① 利用标准光谱的谱带索引，寻找标准光谱中与试样光谱吸收带相同的谱图；
- ② 进行光谱解析，判断试样可能的结构。

3) 新化合物的结构分析。红外光谱主要提供官能团的结构信息，对于复杂化合物，尤其是新化合物，单靠红外光谱不能解决问题，需要与紫外光谱、质谱和核磁共振等分析手段互相配合，进行综合光谱解析，才能确定分子结构。

4) 鉴定细菌，研究细胞和其他活组织的结构。

(3) 分子荧光光谱法

分子荧光光谱分析也叫荧光分光光度法，是利用某些物质被紫外光或可见光照射后所产生的，并且能够反映出该物质特性的荧光，对其进行定性和定量的分析，是当前普遍使用并

有发展前途的一种光谱分析技术。目前，荧光分析方法已成为一种重要且有效的光谱化学分析手段，具有重大的应用价值和深远的科学意义。

当紫外光照射到某些物质时，这些物质会发射出各种颜色和不同强度的可见光，而当紫外光停止照射时，这种光线也随之很快地消失，这种光线称为荧光。每种物质分子中都具有一系列相隔的能级，称为电子能级，而每个电子能级中又包含一系列的振动能级和转动能级。物质受光照射时，可能部分或全部吸收入射光的能量。在物质吸收入射光的过程中，光子的能量传递给物质分子，于是发生电子从较低能级到较高能级的跃迁。这个过程进行很快，费时 $10 \sim 15 \mu\text{s}$ 。处于激发态的分子是不稳定的，它可能通过辐射跃迁和非辐射跃迁等分子内的去活化过程丧失多余的能量而返回基态。辐射跃迁去活化过程，发生光子的发射，伴随着荧光和磷光现象。激发单重态间的辐射跃迁伴随的是荧光发射。荧光分析法在生物化学、医学、工业和化学研究中的应用与日俱增，其原因在于荧光分析法具有高灵敏度的优点，且荧光现象具有有利的时间表度。当物质吸收紫外光和可见光后，它的电子能级跃迁至激发态，然后将这一部分能量释放出来，接着返回基态。由于物质分子结构不同，所吸收光的波长和发射的荧光波长也不同。利用这一特性，可以定性鉴别物质。研究分子的荧光光谱可为研究分子微观结构、分子的构象特点及变化情况提供帮助。

近几十年来，随着激光、微处理器和电子学的新成就等一些新的科学技术的引入，大大推动了荧光光谱法在理论方面的进展，并且相应地加速了各式各样新型的荧光分析仪器的问世，使荧光分析法不断朝着高效、微量、微观和自动化的方向发展，检测方法的灵敏度、准确度和选择性日益提高，方法的应用范围也得到大大扩展。这朵光学学科中的奇葩必将会有更多、更新的成果奉献于自然科学之林，为 21 世纪的科技和相关产业的发展做出重大贡献。

(4) 分子磷光光谱法

电子由基态单重态激发至第一激发三重态的概率很小，因为这是禁阻跃迁。但是，由第一激发单重态的最低振动能级，有可能以系间窜跃方式转至第一激发三重态，再经过振动弛豫，转至其最低振动能级，由此激发态跃回至基态时，便发射磷光，这个跃迁过程 ($T_1 \rightarrow S_0$) 也是自旋禁阻的，其发光速率较慢，为 $10^{-10} \sim 10^{-4} \text{s}$ 。因此，这种跃迁所发射的光，在光照停止后，仍可持续一段时间，如图 4-5 所示。

荧光是由激发单重态最低振动能层至基态各振动能层间跃迁产生的，而磷光是由激发三重态的最低振动能层至基态各振动能层间跃迁产生的。

分子磷光与分子荧光光谱的主要差别是磷光是第一激发单重态的最低能层，经系间跨越跃迁到第一激发三重态，并经振动弛豫至最低振动能层，然后跃迁回到基态发生的。与荧光相比，磷光具有如下 3 个特点：①磷光辐射的波长比荧光长，分子的 T_1 态能量比 S_1 态低。②磷光的寿命比荧光长。由于荧光是 $S_1 \rightarrow S_0$ 跃迁产生的，这种跃迁是自旋许可的跃迁，因而 S_1 态的辐射寿命通常在 $10^{-9} \sim 10^{-7} \text{s}$ ，磷光是 $T_1 \rightarrow S_0$ 跃迁产生的，这种跃迁属自旋禁阻的跃迁，其速率常数要小，因而辐射寿命要长，为 $10^{-10} \sim 10^{-4} \text{s}$ 。③磷光的寿命和辐射强度对于重原子和顺磁性离子敏感。

1) 低温磷光。由于激发三重态的寿命长，使激发态分子发生 $T_1 \rightarrow S_0$ 这种分子内部的内转化非辐射去活化过程以及激发态分子与周围的溶剂分子间发生碰撞和能量转移过程，或发生某些光化学反应的概率增大，这些都将使磷光强度减弱，甚至完全消失。为减少这些去活化过程的影响，通常应在低温下测量磷光。

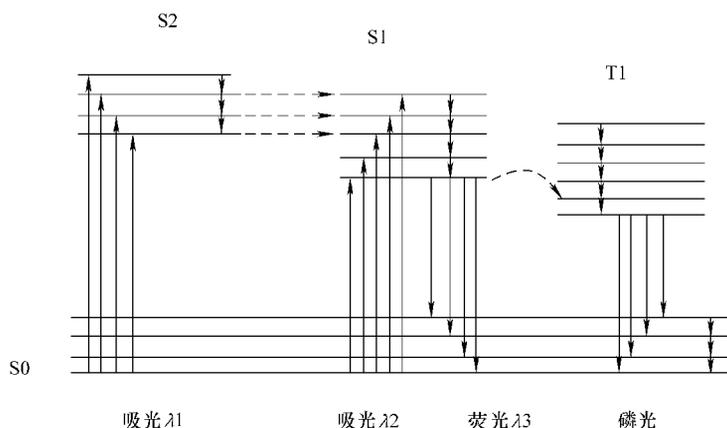


图 4-5 荧光、磷光能级图

低温磷光分析中，液氮是最常用的冷却剂。因此要求所使用的溶剂，在液氮温度（77K）下应具有足够的黏度并能形成透明的刚性玻璃体，对所分析的试样应具有良好的溶解特性。试样的刚性可减少荧光的碰撞猝灭。溶剂应易于提纯，以除去芳香族和杂环化合物等杂质。溶剂应在所研究的光谱区域内没有很强的吸收和发射。最常用的溶剂是 EPA，它由乙醇、异戊烷和二乙醚按体积比 2:5:5 混合而成。使用含有重原子的混合溶剂 IEPA（由 EPA: 碘甲烷 = 10:1 组成），有利于系间跨越跃迁，可以增加磷光效应。

含重原子的溶剂，由于重原子的高核电荷引起或增强了溶质分子的自旋-轨道耦合作用，从而增大了 $S_0 \rightarrow T_1$ 吸收跃迁和 $S_1 \rightarrow T_1$ 系间跨越跃迁的概率，有利于磷光的发生和增大磷光的量子产率。这种作用称为外部重原子效应。当分子中引入重原子取代基，例如，当芳烃分子中引入杂原子或重原子取代基时，也会发生内部重原子效应，导致磷光量子效率的提高。

2) 室温磷光。由于低温磷光需要低温实验装置，溶剂选择等因素的限制，从而发展了多种室温磷光（RTP）法。

① 固体基质室温磷光（SS-RTP）法。此法基于测量室温下吸附于固体基质上的有机化合物所发射的磷光。所用的载体种类较多，有纤维素载体（如滤纸、玻璃纤维）、无机载体（如硅胶、氧化铝）以及有机载体（如乙酸钠、聚合物、纤维素膜）等。理想的载体既能将分析物质牢固地束缚在表面或基质中以增加其刚性，又能减小三重态的碰撞猝灭等非辐射去活化过程，而本身又不产生磷光背景。

② 胶束增稳的溶液室温磷光（MS-RTP）法。当溶液中表面活性剂的浓度达到临界胶束浓度后，便相互聚集形成胶束。由于这种胶束的多相性，改变了磷光团的微环境和定向的约束力，从而强烈影响了磷光团的物理性质，减小了内转化和碰撞能量损失等非辐射去活化过程的趋势，明显增加了三重态的稳定性，从而可以实现在溶液中测量室温磷光。利用胶束稳定的因素，结合重原子效应，并对溶液除氧，是 MS-RTP 法的三个要素。

③ 敏化溶液室温磷光法（SS-RTP）。该法在没有表面活性剂存在的情况下获得溶液的室温磷光。分析物质被激发后并不发射荧光，而是经过系间跨越过程衰减变至最低激发三重态。当有某种合适的能量受体存在时，发生了由分析物质到受体的三重态的能量转移，最后

近红外光谱技术在食品品质检测方法中的研究

通过测量受体所发射的室温磷光强度而间接测定该分析物质。在这种方法中，分析物质本身并不发磷光，而是引发受体发磷光。

3) 磷光仪器。在荧光仪样品池上增加磷光配件：低温杜瓦瓶和斩光片，如图 4-6、图 4-7 所示。斩波片的作用是利用其分子受激所产生的荧光与磷光的寿命不同获取磷光辐射。

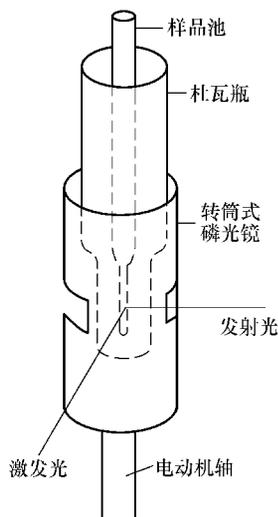


图 4-6 转筒式磷光镜

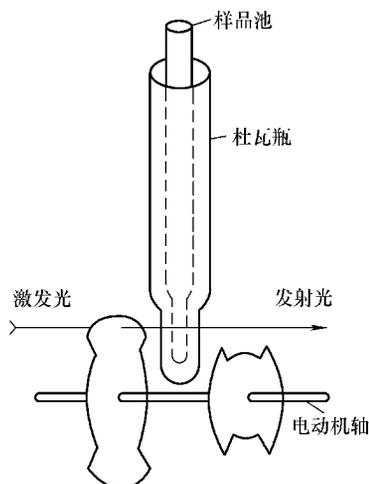


图 4-7 转盘式磷光镜

磷光分析主要用于测定有机化合物，如石油产品、多环芳烃、农药、药物等。

4.2.7 实验室机器人

实验室机器人在农药残留检测方面尚处于起步阶段，原因在于许多情况下没有标准化的测试手段以及修改机器人程序的灵活性。此外，目前机器人行动缓慢，一般要求空间比较宽阔。一旦实验室机器人变得更加方便、快速、灵活、轻巧，并通过制定标准的实验方法，必将在现代化实验室中甚至较为艰苦恶劣的环境中被推广。

4.3 基于近红外光谱的大白菜中毒死蜱残留量检测方法研究

4.3.1 样品制备

实验所需材料见表 4-2。

配制实验样品所需仪器设备见表 4-3。

在超市购买无公害大白菜 500g，配制样品使用 150g，已氰作为萃取液，大白菜与已氰的比例为 1mL 已氰溶液对 0.5g 大白菜。共配制浓度为 0.1 ~ 4mg/kg 的样品 36 个，33 个作为建模集样品，3 个作为预测集样品。在中国农业科学院的协助下配制完成。

样品制备的步骤如下：

- 1) 先将大白菜（或菠菜）在案板上切碎，然后放入多功能食物搅拌器中榨成浆状；
- 2) 天平预热 15min 后，连续 6 次称取 25g 大白菜（菠菜）分别放入 6 个烧杯中；

- 3) 在每个烧杯中加入 50mL 乙腈溶液萃取, 放置 20min 后, 经高速匀浆机粉碎;
- 4) 将粉碎后的物质经过滤分别放入 6 个试管中, 试管提前加入 5g 氯化钠, 进行盐析过程, 此过程持续 30min;
- 5) 将经盐析后的液体取出, 就配成了 0.5g/mL 的蔬菜溶液;
- 6) 根据实验需要, 按照标准配制 36 个浓度范围为 0.1 ~ 0.4mg/kg 的毒死蜱溶液样品, 大白菜样品溶液毒死蜱浓度见表 4-4。

表 4-2 实验所需材料

名称	来源	备注
毒死蜱标准样品	国家标准物质中心	1mg/mL
乙腈溶液	北京化工厂	分析纯
氯化钠 (NaCl)	北京化工厂	
无公害大白菜	甘家口商场	
无公害菠菜	甘家口商场	
蒸馏水	北京工商大学化工学院	

表 4-3 实验仪器和设备

名称	生产厂家	主要参数
多功能食物搅拌机	广州美的生活电器制造有限公司	功率: 250W; 容积: 0.75L
XHF-D 高速分散仪	宁波新芝科技股份有限公司	功率: 360W 调速范围: 5000 ~ 24000r/min
SHZ-D 循环水式真空泵	巩义市英峪仪器厂	功率: 180W; 极限真空: 0.0974MPa
移液器	北京青云卓立精密设备有限公司	范围: 0 ~ 200 μ L; 最小刻度: 1 μ L
电子天平	金洋天平仪器厂	量程: 0 ~ 500g; 最小精度: 0.1g

表 4-4 大白菜样品溶液中毒死蜱浓度

样品号	浓度/(mg/kg)	样品号	浓度/(mg/kg)	样品号	浓度/(mg/kg)
1	0.1	13	0.9	25	2
2	0.2	14	0.95	26	2.2
3	0.3	15	1	27	2.4
4	0.35	16	1.1	28	2.5
5	0.4	17	1.2	29	2.6
6	0.45	18	1.3	30	2.8
7	0.5	19	1.4	31	3
8	0.55	20	1.5	32	3.2
9	0.6	21	1.6	33	3.4
10	0.7	22	1.7	34	3.6
11	0.75	23	1.8	35	3.8
12	0.8	24	1.9	36	4

4.3.2 光谱采集

光谱扫描采用德国 Bruke 公司 MATRIX-F 型近红外光谱仪, 设定分辨率为 8cm^{-1} , 扫

近红外光谱技术在食品品质检测方法中的研究

描次数选择 32 次, 12500 ~ 4000 cm^{-1} 谱区。检测器采用铟镓砷 (InGaAs), 液体光纤探头长 2m。操作步骤如下: 将光纤探头插入装有待测样品的茶色小瓶进行光谱扫描。每次扫描完一个样品都要对光纤探头进行清洗, 以保证不影响下一个样品的光谱采集。

33 个大白菜萃取液样品的近红外光谱图如图 4-8 所示。

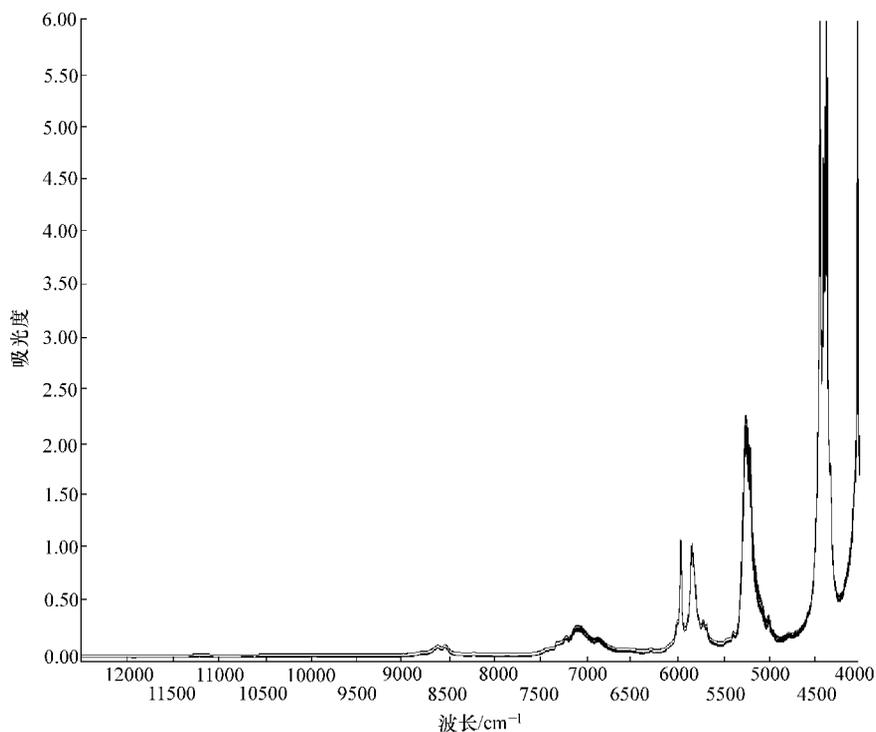


图 4-8 33 个大白菜萃取液样品的近红外光谱图

4.3.3 光谱预处理对模型性能影响分析

用不同的光谱预处理方法对 33 个大白菜萃取液的原光谱进行预处理, 预处理结果见表 4-5。

表 4-5 不同预处理方法的校正结果与预测能力

预处理方法	主成分数	校正集		预测集
		R^2	SECV	SEP
一阶导数 + 基线校正 + 17 点平滑	7	0.993	0.0915	0.1904
无预处理	10	0.9643	0.206	0.213
一阶导数	7	0.992	0.0975	0.1811
矢量归一化	6	0.9587	0.22	0.1897
自动基线校正	7	0.9773	0.164	0.2941
二阶导数	8	0.9702	0.188	0.2064
矢量归一化 + 一阶导数	6	0.9690	0.192	0.2204

以上研究表明, 不同预处理方法能不同程度地影响校正模型的精度, 即影响校正模型的 R^2 、交叉验证标准差 (SECV)、增强校正模型的预测能力, 即影响预测标准差 (SEP)。其中, 一阶导数预处理方法, 校正模型精度较好, 预测能力差也较强。其校正集的预测值与标

准值的相关系数为 0.992, SECV 为 0.0975, SEP 为 0.1811。

4.3.4 近红外光谱波长选择对模型性能影响分析

大白菜萃取液的农残样品毒死蜱含量特别低, 如果全谱处理, 噪声和样品的其他组分对样品光谱的影响较大, 在某些波段, 可能会掩盖住毒死蜱的信息。从图 4-8 可以看出, 大白菜样品的近红外光谱在各个谱区表现为不同的光滑程度, 在某些谱区噪声很大, 因此可对大白菜近红外光谱进行波长选择。

对 33 个大白菜萃取液的原光谱, 根据所建模型的优劣对波长区间进行优化挑选, 对应的谱区为 $7502.7 \sim 6800\text{cm}^{-1}$ 、 $5450 \sim 4594\text{cm}^{-1}$, 如图 4-9 所示。

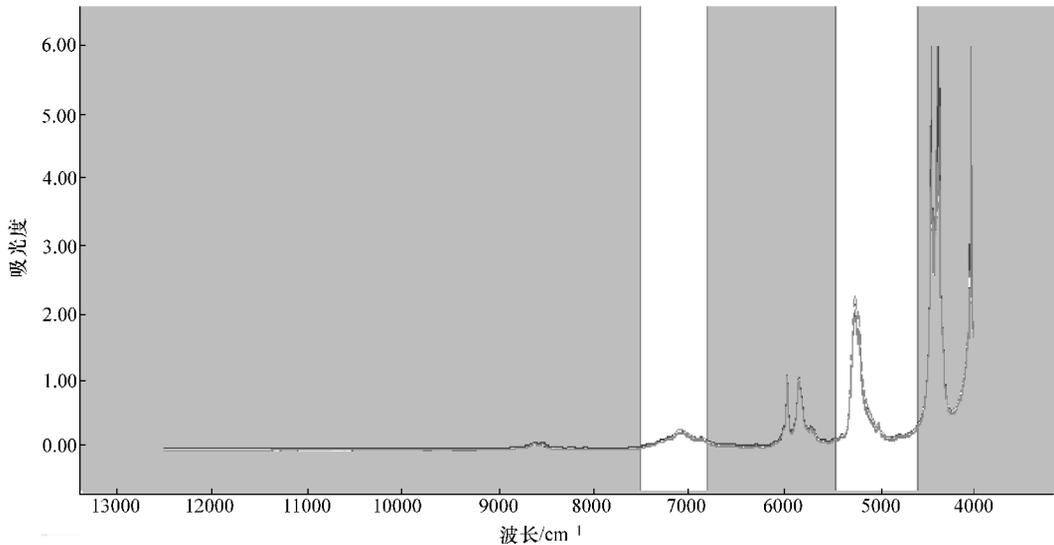


图 4-9 33 个大白菜农药残留样品波长选择光谱图

根据所建模型的优劣挑选的波谱段建立 33 个大白菜样品的校正模型, 交叉验证预测值与标准值的 R^2 为 0.992, SECV 为 0.0975mg/kg , 预测集 3 个样品的 SEP 为 0.1811mg/kg 。

4.3.5 模型性能验证

通过选择波谱段优化后的波段为 $7502.7 \sim 6800\text{cm}^{-1}$ 、 $5450 \sim 4594\text{cm}^{-1}$, 最优的预处理方法: 一阶导数预处理方法, 对 33 个大白菜样品原光谱进行优化, 建立 PLS 定量分析校正模型。交叉验证预测值与标准值的 R^2 为 0.992, SECV 为 0.0975mg/kg , 如图 4-10 所示。

用图 4-10 所示 31 个大白菜样品的校正模型对预测集 3 个大白菜样品进行预测, 预测结果见表 4-6。

本节采用已知毒死蜱含量的大白菜萃取液标样建立测定大白菜中毒死蜱含量近红外模型, 并与标准值进行对比, 对比分析结果表明, 大白菜中毒死蜱含量测定结果比较可靠。

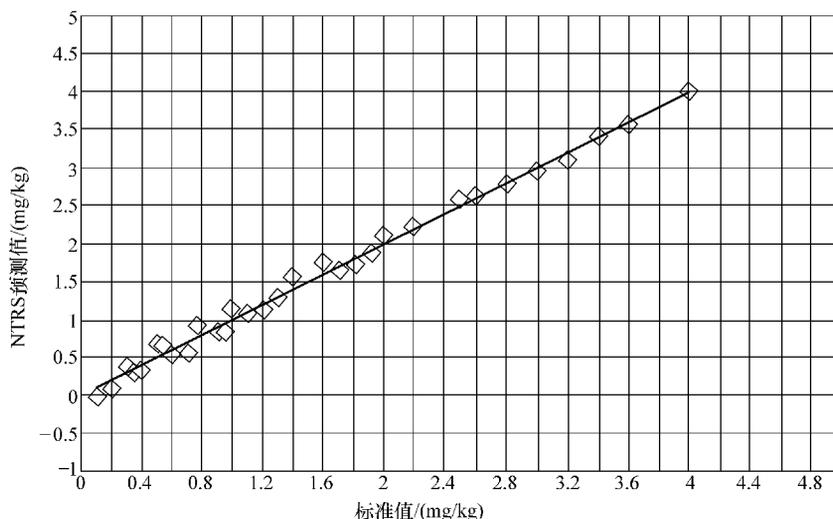


图 4-10 31 个大白菜样品 PLS 结果

表 4-6 预测集 3 个大白菜样品的预测结果

样品号	标准值 (%)	预测值 (%)	绝对误差
12	0.8	1.0523	-0.2523
27	2.4	2.4691	-0.0691
35	3.8	3.6268	0.1732
SEP		0.1811	

4.4 基于近红外光谱的菠菜中毒死蜃残留量检测方法研究

4.4.1 样品制备

在超市购买无公害菠菜 500g，配制样品使用 150g，已氰作为萃取液，菠菜与已氰的比例为 1mL 已氰溶液对 0.5g 大白菜。共配制浓度为 0.1 ~ 4mg/kg 的样品 36 个，31 个作为建模集样品，5 个作为预测集样品。在我国农业科学院的协助下配制完成（见表 4-7）。

表 4-7 菠菜中的毒死蜃

样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)	样品序号	浓度/(mg/kg)
1	0.1	13	0.9	25	2
2	0.2	14	0.95	26	2.2
3	0.3	15	1	27	2.4
4	0.35	16	1.1	28	2.5
5	0.4	17	1.2	29	2.6
6	0.45	18	1.3	30	2.8
7	0.5	19	1.4	31	3
8	0.55	20	1.5	32	3.2
9	0.6	21	1.6	33	3.4
10	0.7	22	1.7	34	3.6
11	0.75	23	1.8	35	3.8
12	0.8	24	1.9	36	4

4.4.2 光谱采集

采集方法同4.3.2节，采集光谱如图4-11所示。

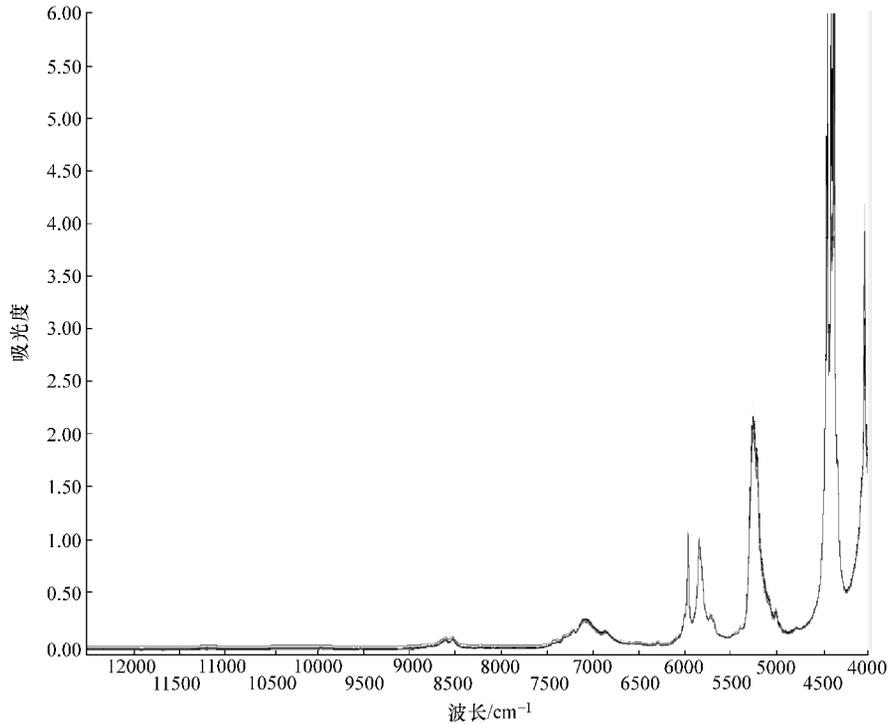


图4-11 31个菠菜萃取液样品近红外光谱图

4.4.3 光谱预处理对模型性能影响分析

用不同的光谱预处理方法，对31个菠菜萃取液的原光谱进行预处理，预处理结果见表4-8。

表4-8 不同预处理方法的校正结果与预测能力

预处理方法	主成分数	校正集		
		R^2	SECV	SEP
一阶导数 + 矢量归一化	6	0.9877	0.124	0.055
无预处理	8	0.9723	0.186	0.056
一阶导数	6	0.9866	0.129	0.071
矢量归一化	6	0.9821	0.149	0.084
自动基线校正	4	0.9802	0.157	0.089
二阶导数	8	0.9308	0.293	0.217
一阶导数 + 基线校正 + 17点平滑	9	0.9859	0.133	0.123

近红外光谱技术在食品品质检测方法中的研究

以上研究表明不同预处理方法能不同程度地影响校正模型的精度，即影响校正模型的 R^2 ，交叉验证标准差 SECV，增强校正模型的预测能力，即影响预测标准差 SEP。其中，一阶导数和矢量归一化 + 17 点平滑相结合的预处理方法，校正模型精度最好，预测能力最强。校正集的预测值与标准值的相关系数由 0.973 提高到 0.9877，SECV 由 0.293 减小到 0.124，SEP 由 1.6154 减小到 1.2803。所以，对于本实验校正集 31 个菠菜样品，选择一阶导数 + 矢量归一化 + 17 点平滑组合的预处理方法最佳。

4.4.4 近红外光谱波长选择对模型性能影响分析

菠菜萃取液的农残样品毒死蜱含量低，如果全谱处理，噪声和样品的其他组分对样品光谱的影响较大，在某些波段，可能会掩盖住毒死蜱的信息。从图 4-12 可以看出，菠菜样品的近红外光谱在各个谱区表现为不同的光滑程度，因此可对菠菜近红外光谱进行波长选择。

对 31 个菠菜萃取液的原光谱，根据所建模型的优劣对波长区间进行优化挑选，对应的谱区为 $6800.7 \sim 6098\text{cm}^{-1}$ 、 $5026 \sim 4598\text{cm}^{-1}$ ，如图 4-12 所示。

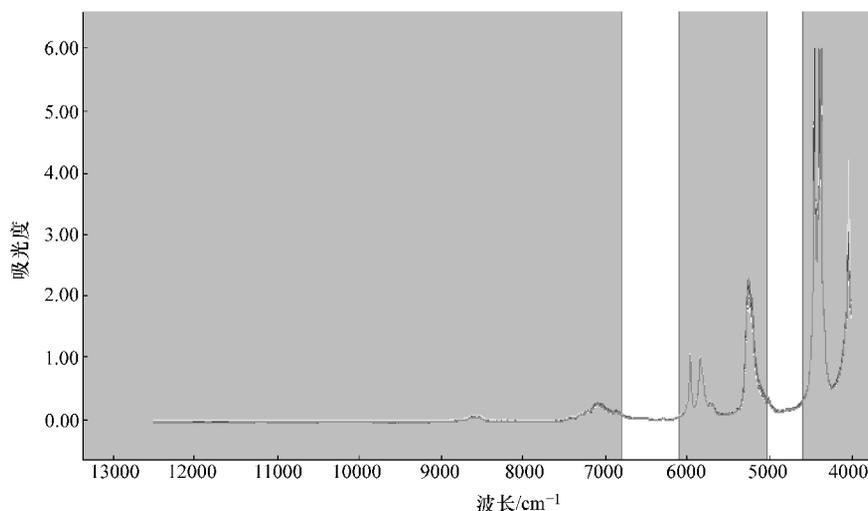


图 4-12 31 个菠菜毒死蜱残留样品波长选择光谱图

根据所建模型的优劣挑选的波谱段建 31 个菠菜样品的校正模型，交叉验证预测值与标准值的 R^2 为 0.9877，SECV 为 0.124mg/kg ，预测集 5 个样品的 SEP 为 0.056mg/kg 。

4.4.5 模型性能验证

采用优化后的波段 $6800.7 \sim 6098\text{cm}^{-1}$ 、 $5026 \sim 4598\text{cm}^{-1}$ ，最优的预处理方法：一阶导数与矢量归一化相结合的预处理方法，对 31 个菠菜样品原光谱进行优化，建立 PLS 定量分析校正模型。交叉验证预测值与标准值的 R^2 为 0.9877mg/kg ，SECV 为 0.124mg/kg ，如图 4-13 所示。

用图 4-13 所示 31 个菠菜样品正模型对预测集 5 个菠菜样品进行预测，预测结果见表 4-9。

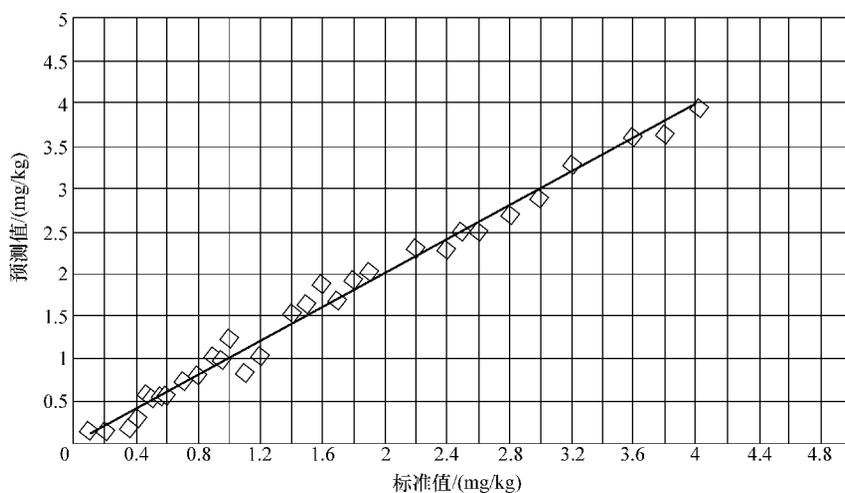


图 4-13 31 个菠菜样品 PLS 结果

表 4-9 预测集 5 个菠菜样品的预测结果

样品号	标准值/(mg/kg)	预测值/(mg/kg)	绝对误差
2	0.2	0.1973	0.0026
11	0.75	0.71	0.04
18	1.3	1.2291	0.0709
25	2.0	2.0465	-0.0465
33	3.4	3.3177	0.083
SEP/(mg/kg)		0.055	
R^2		0.999	

本节采用已知毒死蜱含量的菠菜萃取液标样建立测定菠菜中毒死蜱含量近红外光谱模型，并与标准值进行对比，对比分析结果表明，菠菜中毒死蜱含量测定结果准确可靠，对 5 个待测样品的预测结果表明，其预测能力较强。

4.5 基于近红外光谱的萝卜中毒死蜱残留量检测方法研究

萝卜作为我国主要蔬菜之一，除具有帮助消化等餐用价值外，其种子、叶和根都可入药，且种子具有高出油率。然而，为了防治萝卜的病虫害，往往需要在其生长过程中喷洒毒死蜱等常用农药，不合理使用农药导致的农药残留将影响广大食用者的健康和生命安全。

为了研究近红外光谱技术在微量农药残留检测中应用的可能性和前景，本节分别利用 PLS 法和 ANN BP 对 27 个萝卜样品的毒死蜱残留光谱数据建立数学模型，并用 6 个预测样品对模型的稳健性进行检验。旨在研究近红外光谱技术在农产品农药残留检测应用的可行性和优势，并对两种光谱法进行对比，找到最佳方法。

4.5.1 样品制备

为了降低干扰因素，购买自大型超市出售的无公害萝卜。

甲醇中毒死蜃溶液标准物质（标准值 1.01mg/mL）购自我国计量科学研究院，根据中华人民共和国农业行业标准，NY1500.41.3 ~ 1500.41.6—2009，NY1500.50 ~ 1500.92—2009 规定，萝卜中毒死蜃最大残留量为 1mg/kg。

将萝卜洗净后在清水中浸泡 24h 以消除萝卜表面上其他物质的影响，使用榨汁机将萝卜榨成液体状，并在 33 个茶色螺口小瓶中各装 10g 萝卜汁。在 33 个萝卜汁中滴入不同量的毒死蜃标准物质配制成 33 个浓度分布在 1mg/kg 附近的待测样品。浓度分布表见表 4-10。

表 4-10 萝卜中毒死蜃样品浓度

样品号	浓度值 /(mg/kg)	样品号	浓度值 /(mg/kg)	样品号	浓度值 /(mg/kg)
1	0.1	12	1	23	3.8
2	0.15	13	1.2	24	3.95
3	0.21	14	1.35	25	4.18
4	0.26	15	1.5	26	4.5
5	0.34	16	1.74	27	4.8
6	0.41	17	1.89	28	5
7	0.5	18	2.2	29	5.29
8	0.57	19	2.56	30	5.48
9	0.66	20	2.88	31	5.77
10	0.75	21	3.3	32	6
11	0.88	22	3.49	33	6.5

随机选取 3 号、7 号、12 号、17 号、25 号、31 号作为预测样品集。

4.5.2 光谱采集

光谱采集方法与 4.3.2 节相同。

4.5.3 萝卜中毒死蜃残留量的 PLS 模型建立与分析

根据典型的 5 个近红外光谱频率范围进行所有交叉组合优化检验，优化后选择最佳的预处理方法：最小-最大归一化法，特征波范围：11995.7 ~ 7498.3cm⁻¹，6102 ~ 5446.3cm⁻¹。

近红外光谱法检测萝卜中毒死蜃残留量的 PLS 模型校正结果如图 4-14，回归方程为： $Y=0.951X+0.101$ 。校正结果显示，相关系数 $R^2=0.9901$ ，RMSECV=0.289。

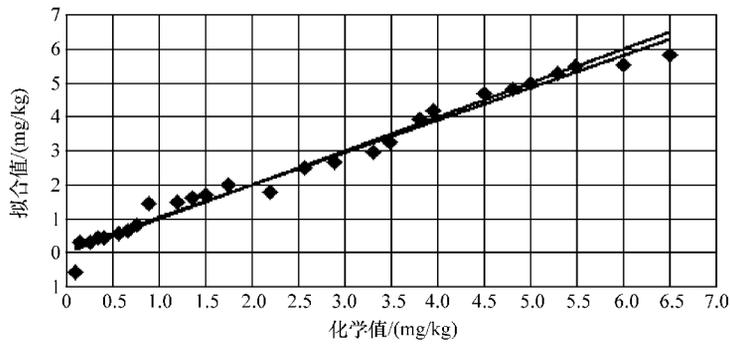


图 4-14 近红外光谱法 PLS 模型校正结果

使用此模型对 6 个预测样品的预测结果见表 4-11。

表 4-11 近红外光谱法 PLS 模型预测结果

化学值/(mg/kg)	0.2100	0.500	1.0000	1.8900	4.1800	5.7700
预测值/(mg/kg)	0.0993	0.6114	1.5235	2.1188	4.3175	5.2176
RMSEP	0.335					

4.5.4 萝卜中毒死蜱残留量的 BP 模型建立与分析

建立模型使用 MATLAB 7.7 神经网络工具箱中量化共轭梯度 (Scaled Conjugate Gradient, SCG) 反向传播算法训练函数 (trainscg), 以校正集为训练样品对网络进行训练, 以方均误差 (Mean Squared Error, MSE) 为性能函数, 隐含层选为 4 个神经元, 建模时对 ANN BP 以 0.01 为训练误差训练 3000 次。

近红外光谱法的 ANN BP 模型校正结果如图 4-15 所示, 回归方程为 $Y = 0.95T + 0.13$ 。校正结果显示, 相关系数 $R^2 = 0.9761$, $RMSECV = 0.436$ 。

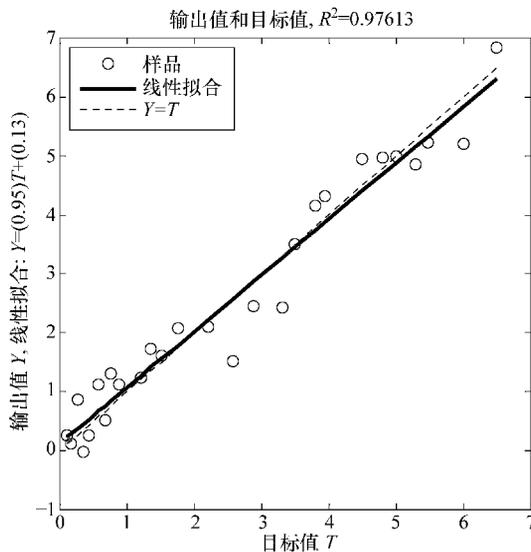


图 4-15 萝卜样品的 ANN BP 模型校正结果

对 6 个预测样品的预测结果见表 4-12。

表 4-12 萝卜样品的 ANN BP 模型预测结果

化学值/(mg/kg)	0.2100	0.5000	1.0000	1.8900	4.1800	5.7700
预测值/(mg/kg)	0.7888	0.9856	2.1461	1.7970	4.5593	5.3248
RMSEP	0.610					

采用 PLS 和 BP 两种校正方法建立萝卜中毒死蜱残留量定量分析模型，都取得了较好的结果，就实验结果而言，PLS 法优于 ANN BP 法。

4.6 小结

本章首先对蔬菜中有机磷农药残留量的现状提出了关于食品检测方面的几项国内、国外标准。然后重点介绍了国内外蔬菜中农药残留检测技术，对这几种方法的分类和检测原理以及应用现状进行了详细的介绍。然后采用近红外光谱分析法测定大白菜、菠菜、萝卜的毒死蜱含量。首先配制 40 个毒死蜱浓度范围在 0.05 ~ 4mg/kg 的混合溶液样品，经过采样分辨率、采样次数及采样模式的选择，使用德国 Bruker 仪器公司的 VERTEX 70 型傅里叶变换红外光谱仪采集混合溶液样品的近红外光谱。其次采用 10 种不同的光谱预处理方法对样品的近红外光谱进行处理，分析比较后发现一阶导数 17 点平滑与标准正态变量变换 (SNV) 相结合的方法光谱预处理效果最好。最后经过波长优化、主成分数目的确定，建立了基于偏最小二乘法的近红外校正模型，校正集样品预测值与真值相关系数为 0.9272，SECV 为 0.3，预测集样品预测值与真值测定系数为 0.9747，SEP 为 0.31，表明该模型具有良好的预测效果和较高的预测精度。

对测定大白菜中毒死蜱含量，建立大白菜萃取液近红外光谱模型，交叉验证预测值与标准值的 R^2 为 0.992，SECV 为 0.0975mg/kg，预测集 3 个样品的预测标准差 SEP 为 0.1811mg/kg。对测定菠菜中毒死蜱含量，通过建立菠菜萃取液近红外光谱模型，交叉验证预测值与标准值的 R^2 为 0.9877，SECV 为 0.124mg/kg。五个验证集样品的预测误差为 0.055mg/kg。通过以上实验表明，近红外光谱分析技术对蔬菜中农药残留检测具有巨大的潜力和应用价值。本章虽为探索性研究，但对实际农药残留快速检测具有一定的借鉴意义。其次本章基于近红外光谱的微量农药残留检测分析技术应用于萝卜农残检测试验中，各模型都得到了良好的评价参数，且具有很好的相关性，证明了近红外光谱技术可用于微量农药残留量的初步分析，但是近红外光谱法难以克服的缺点是试验样本需求量较大、光谱仪器的检测精度很难突破。

本研究主要以毒死蜱含量为分析目标，以菠菜、大白菜和萝卜为蔬菜代表，通过实验设计、实验参数选择和各种光谱预处理方法的比较讨论，最终采用偏最小二乘法完成了模型建立及分析，本章所使用的蔬菜样品均是经过简单的化学前处理的，经过建立近红外校正模型分析，表明近红外光谱分析技术用于蔬菜中农药残留量检测是可行的，希望能够尝试直接使用未经处理的蔬菜样品进行研究，以便于更加接近实际应用。

参考文献

[1] 路凯, 林少彬, 江树人. 中外蔬菜农药残留标准比较与展望 [J]. 解放军预防医学杂志, 2005, 23

- (6): 460 - 462.
- [2] 陈菁菁, 李永玉, 王伟, 等. 微量有机磷农药残留近红外光谱快速检测方法 [J]. 农业机械学报, 2010, 41 (10): 134 - 137.
- [3] Hee Jin Kim, Chul Jae Lee, Mohammad RezaulKarim. Surface - enhanced Raman spectroscopy of Omethoate adsorbed on selver surface [J]. SpectrochimicaActa Part A, 2011, 78: 179 - 184.
- [4] I Lavagnini, A Urbani, F Magno. Overall calibration procedure via a statistically based matrix - comprehensive approach in the stir bar sorptive extraction - thermal desorption - gas chromatography - mass spectrometry analysis of pesticide residues in fruit - based soft drinks [J]. Talanta, 2011, 83: 1754 - 1762.
- [5] H Tang, D Fang, Q Li, et al. Determination of tricyclazole content in paddy rice by surface enhanced Raman spectroscopy [J]. Food Sci, 2012, 77: 105 - 109.
- [6] 王亚飞, 张金艳. 蔬菜中有机磷类农药残留的快速检测 [J]. 黑龙江八一农垦大学学报, 2010, 22 (2): 1 - 4.
- [7] 陆小磊, 吴慧明, 金绍强, 等. 毒死蜱在 4 种作物中残留量的气相色谱分析方法 [J]. 食品工业科技, 2009, (1): 168 - 172.
- [8] 赵明富, 汤斌, 全晓莉, 等. 基于光纤光谱技术的农药残留快速检测实验研究 [J]. 激光杂志, 2010, 31 (2): 48 - 49.
- [9] 吴玉杰, 莫伟媛, 黄志强, 等. 使用微量化学法测定茶叶中有机磷农药残留量的研究 [J]. 食品安全质量检测学报, 2010, 27 (1): 30 - 35.
- [10] 董静, 吕健霞, 宫小明, 等. 农药残留分析质量控制程序 [J]. 分析科学学报, 2009, 25 (1): 46 - 56.
- [11] 苏建峰, 林立峰, 钟茂, 等. 山药中 146 种农药残留的气相色谱 - 质谱分析方法快速研究 [J]. 药物分析杂志, 2010, 30 (2): 339 - 345.
- [12] 高杨菲, 江南平, 汪海. 有机磷农药残留的检测方法研究进展 [J]. 粮油仓储科技通信, 2010 (1): 44 - 46.
- [13] 朱国念. 农药残留快速检测技术 [M]. 北京: 化学工业出版社, 2008.
- [14] 张鑫, 冯利青, 刘岩. 我国果蔬农药残留快速检测技术综述 [J]. 郑州轻工业学院学报 (自然科学版), 2005, 20 (4): 10 - 22.
- [15] 白艳红, 周玲, 王江, 等. 陕西省部分地区蔬菜与水果中有机磷农药残留的调查报告 [J]. 西安交通大学学报 (医学版), 2005, 26 (1): 86 - 88.
- [16] Musleh Uddina, Emiko Okazakia, Moin Uddin Ahmadb, et al. NIR spectroscopy: A non - destructive fast technique to verify heat treatment of fish - meat [J]. Food Control, 2006, 17 (8): 660 - 664.

近红外光谱技术在食用油品质定性分析中的应用研究

5.1 简介

目前国内外检测食用油品质的方法主要有气相色谱法、高效液相色谱法及筛析色谱法等常规化学分析方法，这些方法具有较高的精确度和可靠性，但需借助于昂贵的设备和严格的实验室条件对样品进行复杂的前处理，且分析速度慢，不能满足市场快速检测的需要。近红外光谱技术具有快速、高效、无污染、无需前处理、无损分析及多组分同时测定等优点，在快速检测领域有较大的潜力和发展空间。近红外光谱分析技术用于食用油品质分析检测目前也有一些研究报道，主要用于油脂的品质检测。西方国家最早将近红外光谱技术应用于油脂检测。近红外光谱技术结合判别分析及偏最小二乘分析等方法，在橄榄油、坚果油及其他植物油的检测中已得到成功应用。我国刘福莉以 8 种食用油纯油的 43 个样品为对象，采用近红外透射光谱结合聚类分析法快速鉴别食用油种类，判别模型对预测集样品的精确度达到 100%。范璐和吴娜娜等人利用气相色谱和傅里叶变换红外吸收光谱，对 21 种花生油、20 种棕榈油及两者的 4 种调和油进行分析，对花生油和棕榈油做了识别分析。

5.2 模式识别方法

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，通过肉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区分和识别的综合方法。光谱定性分析中常用的模式识别方法分为两大类：一是有监督的模式识别，如支持向量机（SVM）、判别分析（Fisher 判别、Bayes 判别、逐步判别）、线性学习机（LMD）、KNN、SIMCA、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。

5.2.1 系统聚类分析法

详见 2.4.5 节。

5.2.2 支持向量机分类器

SVM 是 Vapnik 等人于 1992 年首先提出的一类新型机器学习方法，是近年来机器学习研究的一项重大成果。根据 Vapnik 和 Chervonenkis 的统计学习理论，如果数据服从某个（固定但未知的）分布，要使机器的实际输出与理想输出之间的偏差尽可能小，则机器应当遵

循结构风险最小化原理 (SRM), 而不是经验风险最小化原理 (ERM), 就是说应当使错误概率的上界最小化。SVM 正是这一理论的具体实现。与传统的人工神经网络相比, SVM 不仅结构简单, 而且能较好地解决小样品、非线性、高维数和局部极小点等实际问题, 各项技术性能尤其是泛化能力明显提高。

目前已将 SVM 引入到近红外光谱分析中建立近红外光谱定性分析模型, 且都取得了满意的实验结果。如张录达等人以 103 个中药大黄样品为实验材料, 通过 SVM 近红外光谱法建立了大黄样品真伪识别模型, 对学习集中 33 个样品模型识别精确度为 100%, 对 70 个预测样品的识别精确度为 96.77%。刘沐华等人以白芷和丹参为例, 用近红外漫反射光谱数据、应用多类 SVM 方法对中药材产地及生长条件进行自动鉴别, 并进一步用递归 SVM 等方法选择分类最有效的特征谱段, 结果在仅用 5 个和 8 个特征谱段的情况下, 在独立的测试集上识别精确度达到了 92%。

近红外光谱分析中会经常遇到如下问题: 小样品, 建模样品量太少导致模型的稳定性和可靠性都较差; 非线性, 近红外光谱在测定时有非线性因素的干扰, 导致近红外光谱不能真实地反映待测物质的组成和结构; 高维, 近红外光谱全谱参与建模时, 由于全谱的波长点数是成百上千的, 则光谱维数给计算带来了很大的问题。SVM 出色的学习性能正适合解决此类问题, 因此如何将 SVM 与近红外光谱相结合, 利用 SVM 出色的学习性能和近红外光谱技术的分析特点来拓展近红外光谱技术的应用范围, 解决实际生产或生活中遇到的定性判别问题, 是一项值得研究的内容。

5.2.2.1 SVM 的基本思想

SVM 的基本思想如下:

- 1) 通过事先确定的非线性映射将输入矢量映射到一个高维特征空间中:

$$\Phi(x): R_n \rightarrow F \tag{5-1}$$

式中, F 是高维内积空间, 称为特征空间; $\Phi(x)$ 称为特征映射。

- 2) 在此高维空间 F 中构建最优分类面。

在构造 SVM 时, 不必知道 $\Phi(x)$ 的确切表达式, 只要知道如何由输入 x_i 、 x_j 计算内积 ($\Phi(x_i) \cdot \Phi(x_j)$) 就够了, 即

$$\Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j) \tag{5-2}$$

这对构造 SVM 有重要意义。当特征空间的维数巨大时, 直接计算内积的复杂度可想而知。式 (5-2) 表明高维特征空间中的内积运算, 可以转化为低维输入空间上的一个简单的函数计算 $K(x_i, x_j)$ 。该函数被称为内积核函数 (简称核函数)。因此在最优分类面中采用适当的内积核函数 $K(x_i, x_j)$, 就可以实现某一非线性变换后的线性分类, 但计算复杂度却没有增加。此时 SVM 类似于一个神经网络, 输出是中间节点的线性组合, 每个中间节点对应一个支持矢量, 如图 5-1 所示。

从图 5-1 中可以看出, SVM 完全由输入量和核函数构成。SVM 研究最多的核函数主要有 4 类

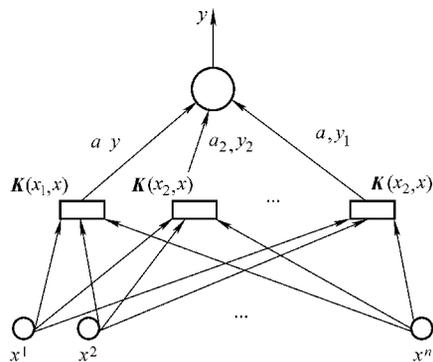


图 5-1 SVM 结构图

(下述表达式中 γ 、 d 、coef 均为核参数):

1) 线性 (Linear) 核函数: $\mathbf{K}(x_i, x_j) = (x_i \cdot x_j)$ 。

2) 多项式 (Poly) 核函数: $\mathbf{K}(x_i, x_j) = [\gamma(x_i \cdot x_j) + \text{coef}]^d$, $d = 1, 2, \dots$, 所得到的 d 阶多项式分类器, $\gamma > 0$ 。

3) 高斯径向基函数 (RBF) 核函数: $\mathbf{K}(x_i, x_j) = \exp \{-\gamma |x_i - x_j|^2\}$, $\gamma > 0$ 。

4) Sigmoid 核函数: $\mathbf{K}(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + \text{coef})$, 这时 SVM 实现的是包含一个隐含层的多层感知器, 隐含层节点数由算法自动确定, 而且算法不存在困扰神经网络方法的局部极小点问题。

现有的应用 SVM 的大量实验表明, SVM 有以下特点: 结构简单; 性能优良, 尤其是泛化能力好; 适合处理高维数据 (计算复杂性与输入模式的维数没有直接关系, 避免了维数灾难); 优化问题有唯一的极小点; 学习速度快; 更换核函数, 就可以得到各种不同的广义超平面。

5.2.2.2 SVM 分类器原理

SVM 是一种新的通用的机器学习方法, 以其小样品下良好的推广能力而被广泛用于各种模式分类问题。下面简单介绍一下 SVM 分类器。

设训练集为 $\{x_i, y_i\}, i = 1, \dots, n, x_i \in R^n, y_i \in \{-1, 1\}$, 则 SVM 分类器的一般形式为

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i \mathbf{K}(x_i, x) + b \right\} \quad (5-3)$$

式中, $\mathbf{K}(x_i, x)$ 是核函数; a_i^* 通过在约束条件 $\sum_{i=1}^n y_i a_i = 0$ 和 $0 \leq a_i \leq c, i = 1, \dots, n$ 下最大化式 (5-4) 求得, 其中 c 为惩罚参数:

$$Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j \mathbf{K}(x_i \cdot x_j) \quad (5-4)$$

Vapnik 证明, 如果训练集中的样品能被 SVM 建立的最优超平面完全划分, 则测试未知样品的最大出错概率, 即 SVM 期望风险的上界为

$$E(\text{Pr}(\text{error})) \leq \frac{E(\text{支持矢量个数})}{\text{训练矢量的个数} - 1} \quad (5-5)$$

式 (5-5) 表明, 支持矢量的数目越少, SVM 期望风险的上界越小, 该 SVM 泛化能力越强。

5.2.2.3 LIBSVM 软件包

1. LIBSVM 软件包简介

目前与 SVM 计算相关的软件有很多, 如 LIBSVM、mySVM、SVMLight 等, 大部分软件免费下载的地址和简单介绍都可以在 <http://www.kernel-machines.org/> 上获得。其中 LIBSVM 是我国台湾大学林智仁副教授等人开发设计的一个简单、易于使用、快速有效的 SVM 模式识别与回归的软件包, 可在 <http://www.csie.ntu.edu.tw/~cjlin/libsvm> 免费下载, 目前的最新版本为 libsvm-2.91。

LIBSVM 软件包可以解决分类问题 (包括 C-SVC、n-SVC)、回归问题 (包括 e-SVR、n-SVR) 及分布估计 (one-class-SVM) 等问题, 提供了线性、多项式、径向基和 S 形函数 4 种常用核函数供选择, 可以有效地解决多类问题, 如交叉验证选择参数、对不平衡样品加权、多类问题的概率估计等。

该软件包不但提供了编译好的可在 Windows 操作系统中执行的文件，还提供了 LIBSVM 的 C++ 语言的算法源代码，以及 Python、Java、R、MATLAB、Perl、Ruby、LabVIEW 及 C#.net 等语言的接口，可以方便地在 Windows 或 UNIX 操作系统下使用，也便于科研工作者根据自己的需要进行改进。另外，该软件包还提供了 Windows 操作系统下的可视化操作工具 SVM-toy，并且在进行模型参数选择时可以绘制出交叉验证精确度的等高线图。该软件还有一个特点，就是对 SVM 所涉及的参数调节相对比较少，提供了很多默认参数，利用这些默认参数可以解决很多问题，并且提供了交叉检验、Cross-SVM 回归等问题，包括基于一对一算法的多类模式识别问题。

2. LIBSVM 的使用方法步骤

LIBSVM 是以源代码和可执行文件两种方式给出的。如果是 Windows 系列操作系统，可以直接使用软件包提供的程序，也可以进行修改编译；如果是 UNIX 类系统，必须自己编译，软件包提供了编译格式文件。LIBSVM 提供的 Windows 操作系统下的可执行文件包括：进行 SVM 训练的 svmtrain.exe，根据已获得的 SVM 模型对数据集进行预测的 svmpredict.exe，以及对训练数据与测试数据进行简单缩放操作的 svmscale.exe。它们都可以直接在 DOS 环境中使用。

LIBSVM 使用的一般步骤如下：

- 1) 按照 LIBSVM 软件包所要求的格式准备数据集；
- 2) 对数据进行简单的缩放操作；
- 3) 考虑选用 RBF 核函数；
- 4) 采用交叉验证选择最佳参数 C 与 g ；
- 5) 采用最佳参数 C 与 g 对整个训练集进行训练获取 SVM 模型；
- 6) 利用获取的模型进行测试与预测。

3. LIBSVM - MBTLAB 接口使用

LIBSVM 提供了多种语言的使用接口。其中 MATLAB 在矩阵运算、二维和三维图形的绘制、数值拟合等方面均有极强的功能，它具有丰富的函数资源供用户直接调用，非常适合于近红外光谱分析。本书中涉及的大部分算法都是通过 MATLAB 来实现，为了便于读者阅读和使用，这里侧重介绍一下 LIBSVM - MATLAB 的使用方法。

根据 LIBSVM 使用的一般步骤，首先要按照 LIBSVM 软件包所要求的格式准备数据集。在 MATLAB 中就无需如此复杂地准备数据集。将 n 个样品光谱文件 (*.txt) 通过“load”函数导入，通过简单组合即可构成一个 n 行 m 列训练样品集，其中， n 为样品数， m 为一个样品的光谱点数。再通过 MATLAB 提供的归一化函数对数据进行简单的缩放操作，使数值在 $[0, 1]$ 或 $[-1, 1]$ 中。

在 LIBSVM 中主要用到两个程序：svmtrain（训练建模）和 svmpredict（使用已有的模型进行预测）。下面分别对 svmtrain 和 svmpredic 的使用方法、各参数意义及设置方法做一个简单介绍。

```
matlab > model = svmtrain (training_label_vector, training_instance_matrix, [, 'libsvm_options'])
```

- training_label_vector: 训练集目标值（即近红外光谱分析中的建模集浓度含量矩阵）

- training_instance_matrix: 训练样品集（即近红外光谱分析中的建模集光谱矩阵）

- libsvm_option: SVM 相关参数设置

```
matlab > [predicted_label, accuracy] = svmpredict (testing_label_vector,
```

近红外光谱技术在食品品质检测方法中的研究

```
testing_instance_matrix, model [, 'libsvm_option']);
- testing_label_vector: 预测集目标值 (即近红外光谱分析中的待测集浓度含量矩阵)
- testing_instance_matrix: 预测样品集 (即近红外光谱分析中的待测集光谱矩阵)
- model: svmtrain 建立的模型
- libsvm_option: SVM 相关参数设置
  Options: 可用的选项, 即表示的含义如下
- ssvm 类型: SVM 设置类型 (默认 0)
0 -- C-SVC
1 -- v-SVC
2 -- 一类 SVM
3 -- e-SVR
4 -- v-SVR
- t 核函数类型: 核函数设置类型 (默认 2)
0 -- 线性: u'v
1 -- 多项式:  $(\gamma * u'v + coef0) ^{degree}$ 
2 -- RBF 函数:  $\exp(-\gamma |u - v|^2)$ 
3 -- sigmoid:  $\tanh(\gamma * u'v + coef0)$ 
- d degree: 核函数中的 degree 设置 (默认 3)
- g gamma: 核函数中的函数设置 (默认 1/k)
- r coef0: 核函数中的 coef0 设置 (默认 0)
- c cost: 设置 C-SVC 和 -SVR, SVR 的参数 (默认 1) 惩罚参数
- n nu: 设置 SVR -SVC, 一类 SVM 和的参数 (默认 0.5)
- p e: 设置 -SVR 中损失函数的值 (默认 0.1)
- m cachesize: 设置 cache 内存大小, 以 MB 为单位 (默认 40)
- e: 设置允许的终止判据 (默认 0.001)
- h shrinking: 是否使用启发式, 0 或 1 (默认 1)
- w weight: 设置第几类的参数 C 为 weight, C (C-SVC 中的 C) (默认 1)
- v n: n-fold 交互检验模式
- b 概率估计: 是否计算 SVC 或 SVR 的概率估计, 可选值 0 或 1 (默认 0)
- q: 退出
```

其中 -g 选项中的 k 是指输入数据中的属性数。Option -v 随机地将数据剖分为 n 个部分并计算交互检验精确度和方均根误差。以上这些参数设置可以按照 SVM 的类型和核函数所支持的参数进行任意组合, 如果设置的参数在函数或 SVM 类型中没有也不会产生影响, 程序不会接受该参数; 如果应有的参数设置不正确, 参数将采用默认值。training_set_file 是要进行训练的数据集; model_file 是训练结束后产生的模型文件, 文件中包括支持向量样品数、支持向量样品及 lagrange 系数等必需的参数; 该参数如果不设置将采用默认的文件名, 也可以设置成自己惯用的文件名。

4. 交互校验与网格搜索

RBF 径向基函数是首选的核函数, 有以下原因:

- 1) RBF 核函数能够解决非线性问题, 而线性核函数却只适用于线性问题, 线性核函数是 RBF 核函数的特例, 另外 Sigmoid 核函数的功能与 RBF 核函数的参数在取某些特定值时类似;
- 2) 核函数的参数影响了模型的复杂度, RBF 的参数较少, 而 poly 核函数的参数则较多;
- 3) RBF 核函数的数值计算范围较小, 根据 RBF 核函数的表达式, 它只能在 $0 \sim 1$ 取值, 而 poly 核函数, 在 degree 足够大时, 可以在 $0 \sim \infty$ 取值。

在选用 RBF 作为核函数时需要考虑两个重要参数: 最佳惩罚参数 c 和 RBF 核参数 γ 。SVM 回归或分类模型的精度取决于这两个参数的组合。通常可以采用网格搜索的方式来找出一个较好的组合, 推荐的初步搜索范围为 $c = [2^{-10}, \dots, 2^{10}]$, $\gamma = [2^{-10}, \dots, 2^{10}]$ 。初步搜索时步长可以根据具体情况适当大一些, 在找到一个较好的参数组合的前提下, 减小搜索范围, 减小步长, 以找到更优的参数组合。这就是所谓的网格搜索 (作者采用 MATLAB 6.5 编写)。

那么如何判断这组参数的好坏呢? 这就要看 SVM 回归模型的预测误差指标。交叉校验可以有效地防止模型过拟合, 并且提高模型的推广能力。SVM 中常用的交叉校验方法有 K 重交叉验证和留一法。

K 重交叉验证法: 该方法是最为普遍的计算推广误差的方法之一, 其过程为将训练样本集随机分为 K 个集合, 通常分为 K 等分, 对其中的 $K-1$ 个集合进行训练, 得到一个决策函数, 并用决策函数对剩下的一个集合进行样品测试。该过程重复 K 次, 取 K 次过程中测试错误的平均值作为推广误差。

留一法: 该方法可以说是 K 重交叉验证法的极端情况, 即 $K=L$, L 为整个训练样品集的大小。该过程为对于第 i 个训练样品, 将其取出, 对剩下 $L-1$ 个样品进行训练, 得到决策函数, 并用其测试第 i 个训练样品, 该过程重复 L 次, 用此方法求出的误差对于实际测试误差来说几乎是无偏的。

5.3 基于近红外光谱的食用油种类定性识别分析

食用油因其种类不同、营养价值不同而价格差异很大。为保护合法生产经营者和消费者的利益, 对食用植物油品种进行准确、快速鉴别是非常必要的。传统的食用油品种鉴别方法为感官评定, 该方法的检测结果受主观因素影响较大, 不适用于大批量样品的检测。本节重点研究基于近红外光谱的食用油种类 (橄榄油、花生油和芝麻油) 的定性识别。

5.3.1 样品制备

食用油品种鉴别实验样品: 从超市购买 25 个不同品牌不同批次的食用植物油样品, 其中纯花生油 7 种, 纯芝麻油 6 种, 纯橄榄油 9 种, 纯大豆油、纯菜籽油、纯棕榈油各 1 种。

5.3.2 光谱采集

实验采用德国 BRUKER 公司的 VERTEX 70 红外光谱仪。设定光谱波数范围为 $12500 \sim 4000\text{cm}^{-1}$, 分辨率为 16cm^{-1} , 每个样品重复扫描 32 次, 采样点为 1102。使用液体光纤探头采样, 光程为 2mm。

全部样品未经任何化学处理, 将光纤探头插入装有样品的小瓶中, 逐一扫描样品, 每次

近红外光谱技术在食品品质检测方法中的研究

测量前均用石油醚清洗探头，避免样品间交叉污染。植物油样品近红外光谱如图 5-2 所示。

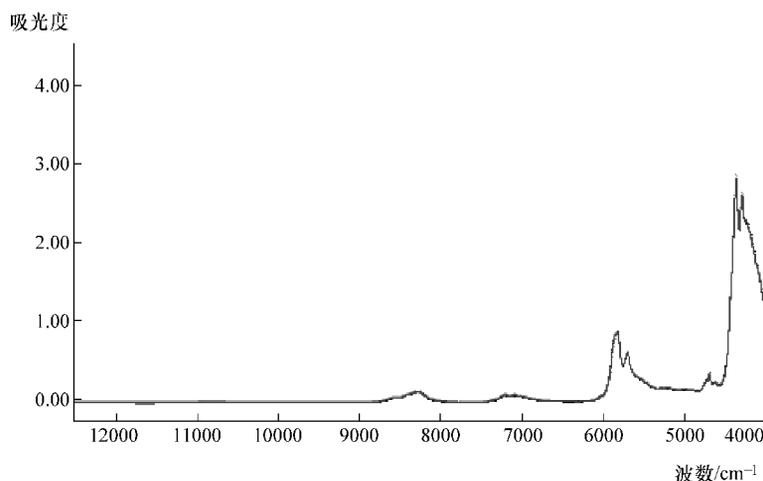


图 5-2 食用植物油样品近红外光谱图

5.3.3 结果与分析

本节研究的食用油种类定性分类是基于样品的近红外光谱特征的。近红外光谱能反映被测样品的组成和结构性质，相同或近似的样品有着相同或相近的光谱；反之，则其近红外光谱存在较大的差异。聚类分析属于模式识别方法中的无管理方法之一。它可以根据“物以类聚”的道理，在没有先验知识的情况下，能合理地按样品各自的特性来进行合理的分类。实验将样品的近红外光谱与系统聚类法相结合，来建立橄榄-芝麻-花生定性识别模型。其中，hs01.0 ~ hs06.0 为纯花生油样品，gl01.0 ~ gl08.0 为纯橄榄油样品，zm0.0 ~ zm05.0 为纯芝麻油样品。光谱预处理采用 - 阶导 (9 点) + 矢量归一化。系统聚类方法较多，本实验采用 OPUS 软件提供的 7 种不同的聚类方法：单连接法、平均连接法、权重连接法、中间算法、质心算法和离差平方和法。经过比较分析结果，采用平均连接方法聚类效果较为理想，聚类结果如图 5-3 所示。

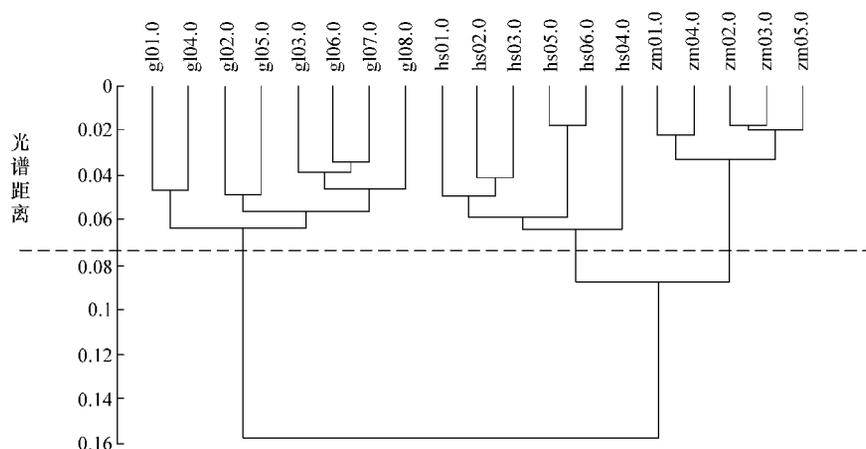


图 5-3 3 种纯植物油聚类结果

从图 5-3 中分析可得，只要令类间距大于 0.062，小于 0.088，则样品集可以聚类为三类，即芝麻油、花生油和橄榄油，识别率为 100%。实验将另外 6 个不同的植物油样品，即纯芝麻油 (zm06.0)、纯花生油 (hs07.0)、纯橄榄油 (gl09.0)、纯菜籽油 (c0.0)、纯大豆油 (d0.0)、纯棕榈油 (z0.0) 作为预测集。预测结果如图 5-4 所示。

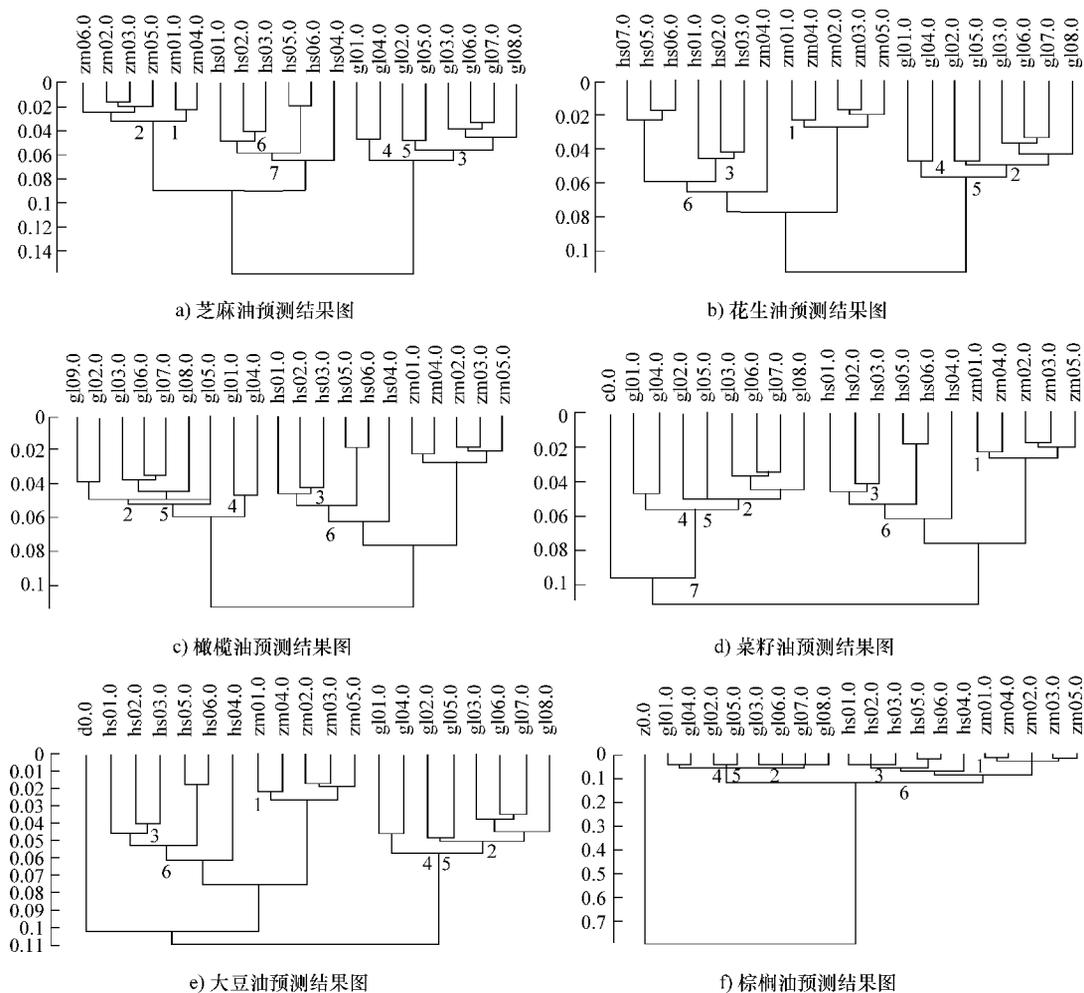


图 5-4 6 种纯植物油预测归类结果图

从预测集样品的预测结果可以得出，当聚类模型类间距确定在 0.062 ~ 0.076 时，纯芝麻油 (zm06.0)、纯花生油 (hs07.0)、纯橄榄油 (gl09.0) 均能被正确识别和归类，纯大豆油 (d0.0)、纯菜籽油 (c0.0)、纯棕榈油 (z0.0) 则可以确认为不属于芝麻油、花生油和橄榄油的任一种，所有样品的预测率达到 100%。因此综合建模和预测两者的结果，可将聚类模型的类间距确定在 0.062 ~ 0.076，则模型的识别率和预测率均可达 100%。

因此本实验中基于近红外光谱的纯橄榄 - 芝麻 - 花生油聚类模型的建立，可以为定性识别植物油的类别提供一个快速、简便的检测方法。

5.4 基于近红外光谱的纯花生油掺假鉴别分析

花生油是我国第三大食用油，以其营养丰富著称，属于市场上价格较高的食用油之一。目前市场上频频出现不法商贩为牟取暴利，将价格相对较低的油脂，如菜籽油、大豆油和棕榈油等掺入花生油进行销售的现象。为了保护消费者权益及维护市场秩序，在加强对花生油生产和销售监管的同时，研究快速识别与检测花生油掺伪的方法在目前的形式下显得尤为重要和迫切。

目前国内将低价油掺入高价油的现象频频出现，屡禁不止。因此本实验研究如何将近红外光谱与 SVM 技术相结合来鉴别花生油是否掺伪，探索其作为一种快速无损鉴别食用油是否掺伪方法的可行性。

5.4.1 样品制备

综合市场上目前出现的在花生油中掺伪的主要是大豆油、菜籽油、棕榈油等植物油，因此购买了如下主要材料：花生油（超市）、大豆油（超市）、菜籽油（超市）、棕榈油（市场）、调和油（超市）。

分别将大豆油、菜籽油、棕榈油、调和油以一定比例掺入纯花生油。以大豆油为例，掺伪比例见表 5-1。菜籽油、棕榈油、调和油的掺伪比例同大豆油，样品编号分别以 hcXX、hzXX、htXX 命名。实验共配置掺伪样品 40 个。纯花生油样品以 hsXX 编号，共采集 5 份纯花生油样品。

表 5-1 大豆油掺入花生油的配比

编号	花生油/mL	大豆油/mL	掺伪比例
hd01	45	5	10%
hd02	40	10	20%
hd03	35	15	30%
hd04	30	20	40%
hd05	25	25	50%
hd06	20	30	60%
hd07	15	35	70%
hd08	10	40	80%
hd09	5	45	90%
hd00	0	50	100%

5.4.2 光谱采集

将制备的样品分别装入 50mL 的实验用白色塑料瓶内，并且放置于温度为 18 ~ 20℃ 的实验室中，静置样品一周左右，使配置的食用油混合物得以充分混合。

采集样品光谱前，将样品逐一摇动。由于棕榈油凝固点很低，在与花生油混合时出现分层现象。因此，需要对其进行外部加热使其成液态后与花生油充分混合，以便光谱扫描，从

而得到精确的光谱图。光谱仪及采集方式同 5.3.2 节。以表 5-1 中的样品为例，其近红外谱图如图 5-5 所示。

5.4.3 结果与分析

实验采用 MATLAB 7.0 和 LIBSVM 模式识别与回归的软件包 (<http://www.csie.ntu.edu.tw/~cjlin/libsvm> 免费下载) 实现 SVM 的分类算法，建立起基于近红外光谱 - SVM 的花生油掺伪鉴别模型。目前的最新版本为 libsvm - 2.91。LIBSVM 使用的一般步骤如下：

- 1) 按照 LIBSVM 软件包所要求的格式准备数据集；
- 2) 对数据进行简单的缩放操作；
- 3) 考虑选用 RBF 核函数；
- 4) 采用交叉验证选择最佳参数 c 与 γ ；
- 5) 采用最佳参数 c 与 γ 对整个训练集进行训练获取 SVM 模型；
- 6) 利用获取的模型进行测试与预测。

1. 归一化

归一化是分类、预测中常用到的基础数据处理方法。归一化有助于提高 SVM 的性能，可以避免数值范围较大的属性控制数值范围较小的属性，并且可以避免计算时的数值困难问题。采用如下归一化的计算公式：

$$z_{ij} = (\text{OUT}_{\max} - \text{OUT}_{\min}) * \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} + \text{OUT}_{\min} \quad (5-6)$$

式中， $x_{j,\max}$ 、 $x_{j,\min}$ 是光谱矩阵 \mathbf{x} 中第 j 列的最大值和最小值； OUT_{\max} 、 OUT_{\min} 是归一化后输出范围设定的最大值和最小值； x_{ij} 、 z_{ij} 为归一化前后光谱矩阵的数。

实验设定归一化后光谱矩阵的数值在 $[-1, 1]$ 内，则归一化公式为

$$z_{ij} = 2 * \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} - 1 \quad (5-7)$$

需要注意对预测集也要进行与训练集相同的归一化，即对预测数据归一化时用的最大值、最小值是训练数据的最大值、最小值。实验采用 MATLAB 提供的 `premnmx` 函数进行归一化。

2. 核函数和参数的选择

要判断参数的好坏需要根据 SVM 分类模型的预测率指标。交叉校验可以有效地防止模型过拟合，并且提高模型的推广能力。 K 重交叉验证是 SVM 中常用的交叉校验方法。

本实验采用 RBF 径向基函数为 SVM 的核函数，通过网格搜索和交叉校验的方式搜索最佳惩罚参数 c 和 RBF 核参数 γ 的最佳组合，设定初步搜索范围 $c = [2^{-10}, \dots, 2^{10}]$ ， $\gamma = [2^{-10}, \dots, 2^{10}]$ 。网格搜索过程在 Matlab 6.5 中编程实现。实验设定参数如下：

- 1) 纯花生油样品用 1 标记，各种掺伪样品用 -1 标记。
- 2) 根据掺伪浓度的梯度选择纯花生油 hs02.0、hs04.0，掺大豆油样品 hd02.0、

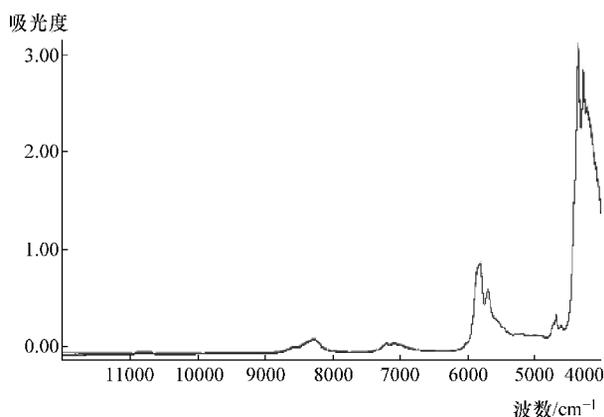


图 5-5 掺伪 (大豆油) 花生油样品近红外谱图

hd04.0、hd06.0、hd08.0，掺菜籽油、棕榈油、调和油样品也做类似选择，作为测试集样品，其他样品作为训练集样品。

3) SVM 类型采用 C - SVC，交叉校验采用 K 重交叉验证， $K = 5$ 。在初步搜索范围内，网格搜索结果如图 5-6 所示。

根据设定的搜索范围，计算得到如图 5-6 所示结果。图 5-6 中横纵坐标分别为 $\log_2(\gamma)$ 和 $\log_2(c)$ ，图 5-6 中所示曲线为其参数组合下 SVM 分类模型的交叉校验预测率的等高线图。选取 SVM 模型交叉校验预测率为 100% 时的一组参数 $c = 1024$ 、 $\gamma = 1$ 。

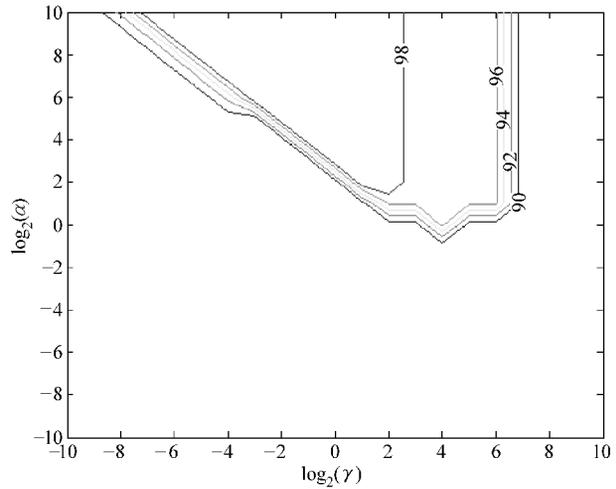


图 5-6 SVM 模型交叉校验预测率等高线图

3. 模型建立与预测

在 MATLAB 7.0 中调用 SVM 模式识别与回归的软件包 LIBSVM。“svmtrain”和“svmpredict”是软件包的两个重要函数，分别用来实现 SVM 的训练和预测过程。根据网格搜索得到的最佳参数训练 SVM 模型，在 MATLAB 中调用形式为

```
model = svmtrain (ytrain, xtrain, '-c 1024 -g 1');
```

模型参数如下：支持矢量数为 11，rho 为 0.6839。

采用模型预测训练集和测试集，调用形式为

```
[predicte, accuracy] = svmpredict (ytrain, xtrain, model);
```

注意这里的训练集或测试集的光谱数据都是经过归一化后的数据，实验结果为

Accuracy = 100% (27/27) (classification) (训练集)

Accuracy = 100% (18/18) (classification) (测试集)

实验结果表明，采用近红外光谱 - SVM 建立花生油掺伪定性鉴别模型的预测效果较好。

4. 与系统聚类的比较

目前食用油研究中采用聚类分析法快速鉴别食用油种类，本实验也采用了系统聚类法对训练集样品进行聚类分析，探索聚类分析在掺伪鉴别中的可行性。光谱矩阵经标准化预处理后，采用 5.3.3 节中所述的 7 种不同的聚类方法，聚类结果均不理想，以离差平方和法为例，聚类结果如图 5-7 所示。因此对于掺伪鉴别，SVM 较系统聚类法具有更好的性能。

5. 食用油掺伪种类鉴别研究

上述实验中，采用 SVM 进行纯花生油掺伪鉴别取得了较为满意的结果，但是对于判断为掺伪的样品，并没有给出掺的是何种油，也就是说不能识别掺伪种类。在后续实验中，采用 SVM 工具包 (<http://see.xidian.edu.cn/faculty/chzheng/bishe/index.htm>) 来进一步实现对食用油掺伪种类鉴别研究。

选取上述实验中配置的 17 个样品，其中包括 6 个掺杂棕榈油样品、6 个掺杂菜籽油样品和 5 个掺杂大豆油的样品。利用 SVM 来检测在花生油中掺杂其他食用油的种类。在判断

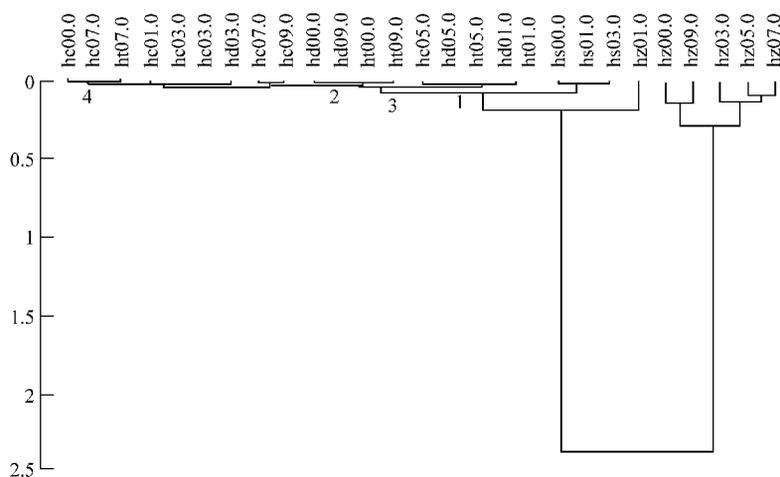


图 5-7 Ward's Algorithm 聚类结果

掺杂种类时，样品的选取是有条件的：①调和油是由不同种类的食用油根据一定的比例配比而成，因此其光谱特性会具有和其他种类食用油类似的吸光度特性，光谱数据的质量较差，使分辨难度加大。因此在本实验中，只研究在花生油中掺杂棕榈油、大豆油和菜籽油三类食用油时的分类情况。关于调和油的分类判别还需要再进一步的研究。②当纯花生油中掺杂其他食用油的比重较小时，样品数据的光谱特性主要体现的是花生油的光谱特性，这样会使样品数据的可判别性变差。因此在本实验中所选择的原始样品数据要求其他种类的食用油所占的比例 $\geq 40\%$ ，有关小比重的食用油掺假还需要再进一步的研究。判别花生油样品掺杂种类流程如图 5-8 所示。在检测样品掺杂的食用油类别时，先判别预测样品中是否掺杂棕榈油，然后再判别样品中是掺杂了菜籽油还是大豆油。

选取 11 个样品作为训练集，剩下的 6 个样品作为测试集。为了简化判别过程，提高判别算法的准确率和鲁棒性，在整个波谱段的高、中、低 3 段上取 4 个波数段来进行训练学习。选取不同的核函数和惩罚因子 c ，可以得到判别结果见表 5-2，掺伪种类的预测率达到 100%。

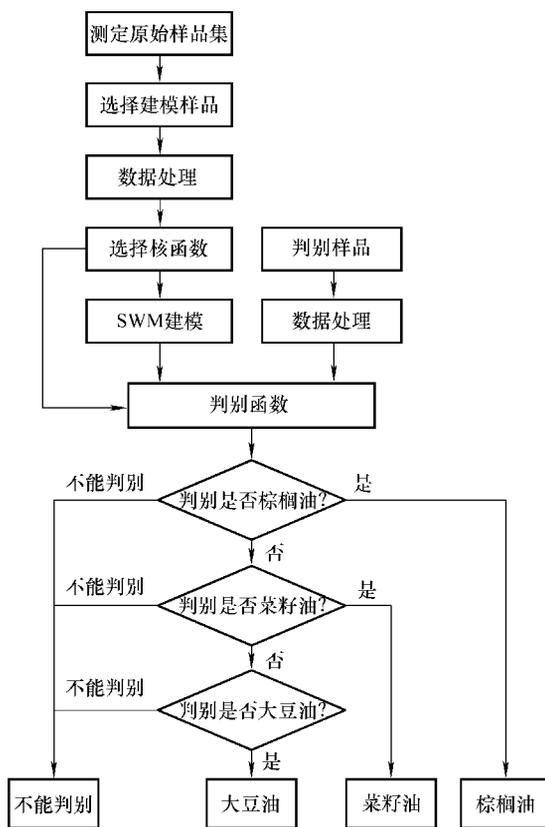


图 5-8 判别花生油样品掺杂种类流程

表 5-2 SVM 分类结果

惩罚因子 c	Linear		poly ($p = 1$)		RBF ($\sigma = 1$)	
	支持矢量数	预测率	支持矢量数	预测率	支持矢量数	预测率
100	9/1052	6/6	8/1057	6/6	10/1049	6/6
500	9/984	6/6	8/999	6/6	10/986	6/6
1000	9/980	6/6	8/981	6/6	10/969	6/6
2000	9/969	6/6	8/970	6/6	10/950	6/6
5000	9/963	6/6	8/963	6/6	10/925	6/6

本实验根据市场上花生油主要的掺假类型，配置了 4 类不同的花生油掺伪样品，采用近红外光谱与 SVM 技术建立了性能较好的花生油掺伪定性鉴别模型，对实验中预测集和测试集的预测率均能达到 100%。实验表明将近红外光谱与 SVM 技术相结合作为一种快速无损鉴别食用油掺伪的方法具有较强的可行性和实用性。实验还进一步对鉴别食用油掺伪种类进行了深入研究，在特定实验条件下也取得了较为满意的鉴别结果。

5.5 基于近红外光谱的纯橄榄油掺假鉴别分析

橄榄油在西方被誉为“植物油皇后”，可供食用的高档橄榄油是世界上唯一以自然状态的形式供人类食用的木本植物油，其价格也较为昂贵。目前国内市场上频频出现不法商贩为牟取暴利，将低价油掺入高价油进行销售的现象，如将菜籽油、大豆油等掺入橄榄油进行销售。橄榄油质量优劣对人体健康有着重要的影响，为了保护消费者权益以及维护市场秩序，在加强对橄榄油生产和销售监管的同时，开发和利用快速鉴别橄榄油掺伪的检测仪器和手段是当前十分迫切的任务。

目前鉴别橄榄油掺伪的方法有：硝酸法、冷冻法、测定油脂的克雷司美尔值、测定角鲨烯含量等方法。但有的实验需借助于昂贵的设备或是严格的实验室条件对样品进行复杂的实验前处理，有的分析速度慢，不能满足市场快速检测的需要。基于近红外光谱的快速定性检测技术目前被逐渐应用于质量控制中，但是样品少以及近红外光谱的高维数给近红外光谱定性识别模型的建立带来很大困难。本节提出应用模式识别与近红外光谱技术相结合快速鉴别橄榄油掺伪，探索其作为一种快速无损鉴别橄榄油掺伪方法的可行性。

5.5.1 样品制备

综合市场上目前出现的在橄榄油中掺伪的主要是大豆油、菜籽油等植物油，因此从超市购买了纯橄榄油、大豆油、菜籽油，分别将大豆油、菜籽油以一定体积比掺入橄榄油，配制两类掺伪橄榄油样品。掺入单一类别（菜籽油）的样品，其橄榄油含量为 71% ~ 90%，相同含量的掺伪样品平行配制两个，共配置 40 个掺伪样品。以同样方式配制掺入混合类别（大豆油和菜籽油）的样品，每个样品中大豆油和菜籽油的体积比为 1:1，相同含量的掺伪样品平行配制两个，掺伪样品共 50 个，橄榄油含量为 71% ~ 90%。

5.5.2 光谱采集

同 5.3.2 节。选取纯橄榄油样品，橄榄油含量为 71% 掺伪样品（掺入单一类别），橄榄

油含量为 71% 掺伪样品（掺入混合种类）的近红外光谱的二阶导图如图 5-9 所示，从图中可以看出，三者的二阶导在一些特征峰处有着较为明显的差异，这些差异为用近红外光谱进行定性识别掺伪样品奠定了理论基础。

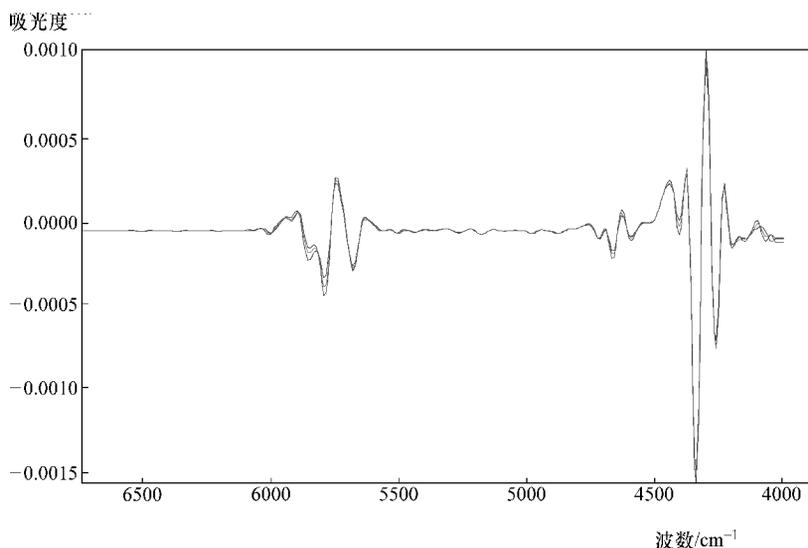


图 5-9 样品近红外光谱二阶导图

5.5.3 结果与分析

1. 基于系统聚类法的橄榄油掺伪鉴别模型的建立

尝试采用系统聚类法对混入菜籽油的掺伪样品和纯橄榄油样品的近红外光谱进行鉴别分析。实验样品集共有样品 50 个，随机选择其中 25 个样品作为训练集，25 个作为测试集。采用 OPUS 软件提供的聚类分析模块，分别使用 5.3.3 节中所述的 7 种不同的聚类方法分析。以离差平方和法为例，聚类结果如图 5-10 所示。其中，标号 101.0 ~ 105.0 的样品为纯橄榄油样品，其他标号样品为掺伪样品。从图中可以分析得出，掺伪样品和未掺伪样品的鉴别效果并不理想。采用其他聚类分析方法的聚类结果均不理想。

2. 基于 SVM 的橄榄油掺伪鉴别模型的建立

鉴于上述实验中建立的基于系统聚类法的橄榄油掺伪鉴别模型的识别率效果较差，尝试采用 SVM 建立掺伪样品的定性识别模型。

实验采用 MATLAB7.6.0 (R2008a) 中自带的 SVM 工具箱，采用如图 5-11 所示流程建立定性识别模型，并对测试集进行预测。图中每个步骤中标注了 MATLAB 中采用的函数。

本实验中实验样品集，采用标准化对数据集进行预处理，实用交互校验方式随机选择训练集和测试集（样品比 1:1），采用核函数为线性函数建立 SVM 模型，最后对测试集进行分类。本实验中建立的 SVM 模型中支持矢量数为 5，支持矢量结构参数为 $a_i^* = \{-0.00727; -0.04459; -0.06262; 0.09295; 0.02153\}$ 、 $b = 2.82933$ 。本实验中对训练集和测试集的分类正确率均达到了 100%。

实验还对掺入混合种类的样品及纯橄榄油样品组成的样品集进行同样的分析，建立的

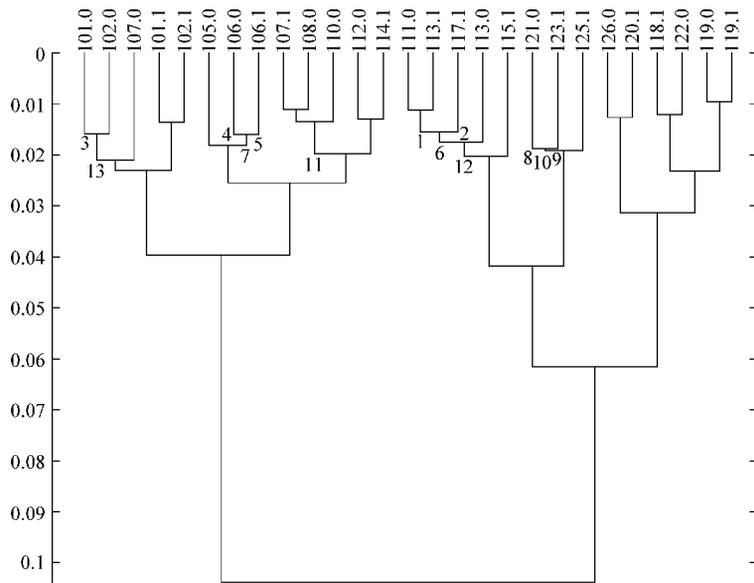


图 5-10 基于系统聚类法的橄榄油掺伪识别

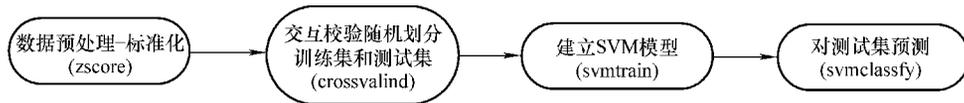


图 5-11 SVM 建模流程

SVM 模型中支持矢量数为 2，支持矢量结构参数 $a_i^* = \{-0.00294, 0.00294\}$ 、 $b = 2.22013$ 。本实验中对训练集和测试集的分类正确率均达到了 100%。

本实验根据市场上橄榄油主要的掺假类型，配置了不同类型的橄榄油掺伪样品，采用模式识别方法中的系统聚类法和 SVM 分别建立基于近红外光谱的橄榄油掺伪鉴别模型。实验中 SVM 建立的定性识别模型优于系统聚类法，模型的识别率和预测率均达到了 100%，结果表明基于近红外光谱的橄榄油掺伪快速检测技术具有较好的可行性和实用性。

5.6 小结

本章采用近红外光谱技术结合模式识别方法建立了食用油种类、花生油和橄榄油掺伪定性识别模型。实验结果表明，采用近红外光谱与 SVM 技术建立了性能较好的花生油和橄榄油掺伪定性鉴别模型，对实验中预测集和测试集的预测率均能达到 100%；而采用近红外光谱与系统聚类法结合能有效地识别食用油种类。实验表明将近红外光谱与模式识别技术相结合作为一种快速无损鉴别食用油掺伪方法具有较强的可行性和实用性，本章工作的探索有力地推动了近红外光谱技术应用于食用油品质定性快速识别的进程。

参考文献

[1] 冯利辉. 食用植物油掺伪检测与定量分析的近红外光谱法研究 [D]. 南昌: 南昌大学, 2010.

- [2] 于燕波, 臧鹏, 付元华, 等. 近红外光谱法快速测定植物油中脂肪酸含量 [J]. 光谱学与光谱分析, 2008, 7 (28): 1554 - 1558.
- [3] 刘福莉, 陈华才, 姜礼义, 等. 近红外透射光谱聚类分析快速鉴别食用油种类 [J]. 中国计量学院学报, 2008, 19 (3): 278 - 282.
- [4] 范璐, 吴娜娜, 霍权恭, 等. 气相色谱分析和傅里叶变换红外吸收光谱识别花生油中棕榈油 [J]. 河南工业大学学报 (自然科学版), 2008, 29 (1): 12 ~ 16.
- [5] 徐云. 农产品品质检测中的近红外光谱分析方法研究 [D]. 北京: 中国农业大学, 2009.
- [6] Nello Cristianini, John Shawe - Taylor. 支持向量机导论 [M]. 李国正, 王猛, 曾华军, 译 [M]. 北京: 电子工业出版社, 2004.
- [7] 张学工. 关于统计学理论与支持向量机 [J]. 自动化学报, 2000, 26 (1): 32 - 34.
- [8] 崔伟东, 周志华, 李星. 支持向量机研究 [J]. 计算机工程与应用, 2001, 37 (1), 58 - 61.
- [9] 张录达, 苏时光, 王来生, 等. 支持向量机在傅里叶变换近红外光谱分析中的应用研究 [J]. 光谱学与光谱分析, 2005, 25 (1), 33 - 35.
- [10] 刘沐华, 张学工, 中药材产地的近红外光谱自动鉴别和特征谱段选择 [J]. 科学通报, 2005, 50 (4), 393 - 398.
- [11] 吴静珠, 刘翠玲, 李慧, 等. 近红外光谱技术在食用油种类鉴别及脂肪酸含量检测中的应用 [J]. 北京工商大学学报 (自然科学版), 2010, 28 (5), 56 - 59.
- [12] 吴静珠, 刘翠玲, 李慧, 等. 基于近红外光谱的纯花生油掺伪快速鉴别方法研究 [J]. 北京工商大学学报 (自然科学版), 2011, 29 (1), 75 - 78.
- [13] 李慧. 基于近红外光谱的食用油品质检测技术研究 [D]. 北京: 北京工商大学, 2011.
- [14] B Sholkopf, K Sung, C J C Burges, et al. Comparing support vector machine with Gaussian kernels to radial basis function classifiers [J]. IEEE Trans, Signal Processing, 1997, 45: 2758 - 2765.

近红外光谱技术在食用油品质定量分析中的应用研究

6.1 简介

脂肪酸含量是食用油营养品质的重要指标，其含量直接影响着食用油的品质，而食用油中分子在近红外光谱区的吸收主要由分子中含氢基团（C-H、N-H、O-H、S-H等）的组合频和倍频吸收组成，食用油脂结构中的 $-\text{CH}_2-$ 和 $-\text{C}=\text{C}-$ 对近红外光谱中的合频信息贡献较大，其数量反映了食用油中脂肪酸的含量。

食用油由于存放不当或者贮存时间过久，会产生氧化而分解成甘油和游离脂肪酸等有害物质，这对油脂品质将是一种无形的损害，而食用油在氧化中将产生氢过氧化物，其含量的多少则用过氧化值来表征。因此，测定食用油的过氧化值对检测食用油的氧化程度有着很重要的作用。

食用油变质通常分为水解变质和氧化变质。水解变质时游离脂肪酸产生不良气味，同时产生二酰甘油酯等一些有很强乳化作用的副产物，水解变质程度的衡量指标通常用酸价来表示。因此，酸价是评价油酸败变质程度的重要指标之一。

近红外光谱分析技术不但可用于食用油品质的定性检测，也可用于食用油常规理化指标的定量分析。本章主要介绍采用近红外光谱技术进行食用油指标（脂肪酸含量、过氧化值和酸价等）定量分析的可行性探索，从而进一步监控食用油的品质。

6.2 光谱分辨率的选取对食用油近红外模型性能影响分析

近红外光谱仪的类型、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。根据采用的分光技术不同，近红外光谱仪的类型可划分为滤光片型、光栅扫描型、阵列检测型、傅里叶变换型和声光过滤调制型等。其中，傅里叶变换型光谱仪器已经成为目前实验室红外光谱仪器的主导产品。它的特点是扫描速度快、波长精度高、分辨率高，加之光能利用率高、输出能量大、不受杂散光影响，因此仪器的信噪比和测定灵敏度较高。一般情况下，近红外光谱仪类型的选择局限于实验室所购买的仪器。而光谱仪的工作状态（如波长准确度、信噪比等）一般在仪器上都带有硬件工作状态的自诊断软件来告知用户当前仪器是否工作在正常状态，因此无需过多考虑。但是光谱扫描时需要人为设定的一些参数，如光谱分辨率，特别需要用户引起注意，因为光谱分辨率影响样品光谱的质量，最终影响分析的准确性。这类参数若在实验过程中设置得当，会有助于用户得到准确、可靠的近红外分析结果。反之，则会严重影响近红外分析结果。如何设定这类参数，需要用户反复测试摸索一些经验值，并结合具体测试对象状态（如粉末、籽粒、均匀液态、悬浊液等）而定。

本节重点研究在傅里叶变换型光谱仪上设定不同的光谱分辨率对食用油近红外定量模型性能的影响,探索对于食用油这类均匀液态物质在做近红外检测时光谱分辨率的参考值。

6.2.1 样品制备

60个植物油样品,包括花生油、大豆油、橄榄油、芝麻油、葵花籽油、玉米胚芽油、芥花油等,均购自当地超市。

采用气相色谱法测定60个食用油样品中棕榈酸的含量。

6.2.2 光谱采集

采用德国Bruker公司VERTEX 70型红外光谱仪采集样品近红外光谱,使用液体光纤探头采样方式,光程2mm。仪器参数设定如下:波数范围 $12500 \sim 4000\text{cm}^{-1}$,每个样品重复扫描次数32次。

一般要求仪器的分辨率比测量峰宽小 $1/10$ 左右。样品在近红外区域的吸收多为宽峰且重叠严重,进行定量分析时不要求高的仪器分辨率。在实际应用中,通常 16cm^{-1} 或 10nm (在 2500nm 处)的分辨率就可满足大部分分析对象的应用要求。但对于结构特征十分相近的复杂样品(如农产品等),要得到准确的分析结果,就要对仪器的分辨率提出一定的要求,一般也不会超过 4cm^{-1} 。综上所述,本试验设定分别在光谱分辨率为 4cm^{-1} 、 8cm^{-1} 和 16cm^{-1} 时扫描样品,相应的波长点个数分别为4407、2203和1102。显然,分辨率设得越高,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。图6-1所示为分辨率 16cm^{-1} 时样品的近红外谱集。

全部样品未经任何化学处理,将光纤探头插入装有样品的小瓶中,逐一扫描样品,每次测量前均用石油醚清洗探头,避免样品间交叉污染。

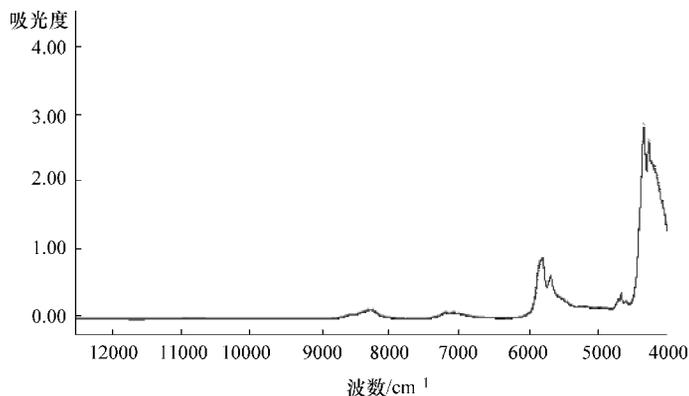


图6-1 食用油样品近红外谱集(16cm^{-1})

6.2.3 结果与分析

实验中食用油棕榈酸近红外定量分析模型的建立采用近红外分析的常规流程,包括异常样品剔除、校正集样品选择以及模型的建立与测试。近红外校正模型的预测精度和稳健性采用主成分数(nf),决定系数 R^2 、交叉校验标准差(Root Mean Square Error of Cross Validation, RMSECV)、预测方均根误差(Root Mean Square Error of Prediction, RMSEP)指标评价。

1. 异常样品剔除

建立性能较好的近红外模型，进行异常样品的剔除是非常必要的。这里采用蒙特卡洛采样法进行异常样品的剔除，在 MATLAB R2008a 中实现。图 6-2 所示是在分辨率为 4cm^{-1} ，随机采样 1000 次的条件下，各个样品的 RMSEP 的统计信息（均值 MEAN 和标准偏差 STD）分布图。设定均值为 2、标准偏差为 0.5 的条件下可将图 6-2 划分成 A、B、C、D 四个区域。

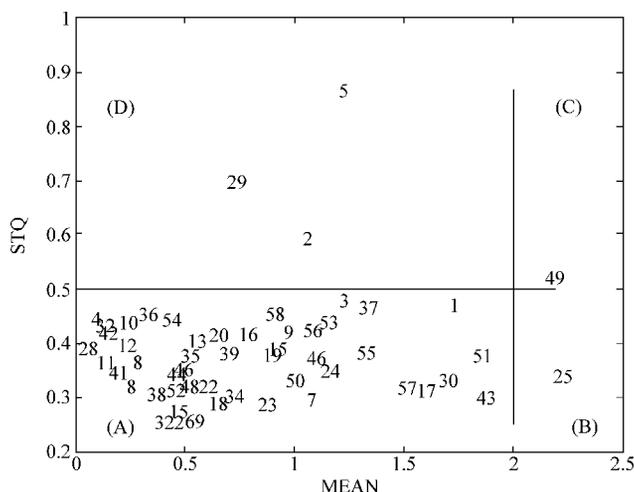


图 6-2 60 个样品的 RMSEP 统计信息分布图

其中，A 区域样品的 RMSEP 的均值和标准偏差较小，因此属于正常样品；B 区域样品 RMSEP 的均值较大，属于浓度异常样品；C 区域样品的误差均值和标准偏差较大，属于异常样品；D 区域样品的标准偏差较大，属于光谱异常样品。因此，B、C、D 区域的样品均可作为异常样品剔除。光谱分辨率为 4cm^{-1} 、 8cm^{-1} 和 16cm^{-1} 时，采用蒙特卡洛法剔除异常样品信息见表 6-1。从表中可以看出，除样品 5 外，在不同分辨率条件下，剔除的异常样品大部分都是不同的。因此这也从一个侧面反映出在不同光谱分辨率条件下，样品所含的光谱信息或者说光谱质量是有明显不同的。

表 6-1 剔除异常样品相关信息

分辨率/ cm^{-1}	均值 MEAN	标准偏差 STD	异常样品序号
4	2	0.5	2、5、25、29、9
8	2.4	1	2、5、17、49、51
16	1.2	0.8	5、6、10、37、51

2. 校正集样品选择

Kennard - Stone 法将光谱差异较大的样品选入校正集，而其余较相近的样品进入校验集，这样可使有代表性的样品全部进入校正集，从而在一定程度上避免了校正集样品分布的不均匀。将剔除异常样品后的样品集采用 Kennard - Stone 法划分校正集和校验集，在 MATLAB R2008a 中实现，结果见表 6-2。表 6-2 中，经过 Kennard - Stone 法划分得到校正集的浓度范围涵盖了校验集的浓度范围，因此校正集所建模型可用于校验集的预测。在不同分辨率条件下，基于光谱信息的 Kennard - Stone 法挑选出的校验集样品明显不同，由此可见，相同样品在不同分辨率下采集得到的光谱信息是有很大差异的。

表 6-2 Kennard - Stone 法划分的样品集信息

分辨率/ cm^{-1}	校正集		校正集	校验集		校验集
	样品个数	浓度范围 (%)	标准偏差	样品个数	浓度范围 (%)	标准偏差
4	40	7.6 ~ 16.8	2.39	15	7.6 ~ 14.4	2.26
8	40	7.6 ~ 16.8	2.36	15	8.5 ~ 15.3	2.17
16	40	7.6 ~ 16.85	2.49	15	10.8 ~ 14.9	1.34

3. 模型建立与预测

采用 5 折交叉校验的偏最小二乘回归法 (Partial Least Squares regression method, PLS) 建立食用油棕榈酸的近红外定量分析模型, 并对校验集做预测。上述建模和预测方法均在 MATLAB R2008a 中实现。结果见表 6-3。从表中可以看出, 采用全谱建模时, 模型指标不是很好, 但是仍旧能得出分辨率为 16cm^{-1} 时所建模型, 且预测结果远优于 8cm^{-1} 和 4cm^{-1} 时所建模型。

表 6-3 5 折 PLS 建模分析结果

分辨率/ cm^{-1}	nf	R^2	RMSECV	RMSEP
4	7	0.8167	1.0111	0.8280
8	15	0.7912	1.0655	1.0317
16	14	0.9384	0.6094	0.6672

本实验中, 对于食用油这类均匀液体做近红外检测, 分析其棕榈酸的常规指标时, 在分辨率为 16cm^{-1} 时扫描样品建模即可获得较好的模型指标。实验表明, 并非近红外光谱分辨率越高, 模型性能越好。因为傅里叶变换型近红外光谱仪的分辨率由动镜移动距离决定, 分辨率越高, 扫描速度越慢, 单位时间内的信噪比也会降低。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。

本工作在 Vertex 70 光谱仪上, 设定 3 种光谱分辨率 (4cm^{-1} 、 8cm^{-1} 和 16cm^{-1}) 分别采集 60 份食用油样品的近红外光谱, 重点研究了不同光谱分辨率对食用油近红外定量模型性能的影响。实验结果表明, 光谱分辨率的高低与模型性能之间并无绝对的关系, 对于食用油这类均匀液态物质中的常规组分做近红外检测时, 光谱分辨率的参考值可首选 16cm^{-1} 。

6.3 基于近红外光谱的食用油脂肪酸定量分析

我国居民食用植物油主要有大豆油、菜籽油、花生油、葵花籽油、橄榄油和芝麻油等。植物油的成分主要是高级脂肪酸, 常见的高级脂肪酸主要有棕榈酸、硬脂酸、油酸和亚油酸等。不同植物油的脂肪酸组成有很大差异, 因此其功效和营养价值也随之不同。随着营养学研究发展, 人们对植物油中脂肪酸的认识不断增加。中国营养学会也推荐膳食中必需脂肪酸摄入理想比值, 市场上也出现大量调和油, 为了评估食用植物油营养价值, 研究分析常用食用植物油中脂肪酸组成是很有必要的。

本节重点研究基于近红外光谱的食用油脂肪酸 (棕榈酸、硬脂酸、油酸和亚油酸) 含量的定量模型优化。高模型预测能力始终是近红外分析研究中的热点和难点。近红外光谱本身的有效信息量很弱, 有些波段与所测样品成分之间缺乏相关关系, 各谱区的噪声水平也不一致, 即使做相关的谱图预处理也可能导致模型不稳定, 降低模型的精度。在近红外光谱结

合 PLS 建模中,传统观点认为 PLS 具有较强的抗干扰能力,可全波长参与多元校正模型的建立。随着对 PLS 方法的深入研究发现,通过特定方法筛选特征波长或波长区间有可能得到更好的定量校正模型。有研究报道,有效的变量筛选不仅可以简化模型,更主要的是可以剔除不相关变量,建立更好的校正模型,从而加强模型预测能力与稳健性。

6.3.1 常见谱区优化算法

由于农产品成分复杂,无法凭借主观经验根据化合物的近红外光谱特征有效地找到相关成分的特征波长,一般需要借助数学手段从近红外光谱中提取出组分的特征波段。

本节主要介绍常见优化光谱波长区间的算法。

1. 分段排序法

近红外光谱谱区选择方法中,一般有经验法和“分段排序”法。经验法就是根据专家经验,凭借图示法,选取组分特征波长谱区,但主观性很大。

“分段排序”法的主要思想是,预先将光谱总区间分割为多个子区间(可以均分,也可以不均分),然后取每一个子区间的波长点数据建模,根据所建模型在最佳主成分时的 PCR/PLS-CV 预测值与标准值的相关系数 R 从大到小, RMSECV 从小到大的顺序,对所有子区间综合排序,给出了各个谱区对某组分建模影响的重要程度,最后还是要凭专家经验选取哪些谱区参与建模。

2. 相关系数法

样品在有些光谱区域的光谱信息很弱,与样品的组成或性质间缺乏相关关系。为了找出最有效的光谱区域,可以将测定的组成或性质数据与样品的光谱数据进行关联,求出每个波长处的相关系数。

相关系数法是将校正集光谱矩阵中每个波长对应的吸光度矢量 \mathbf{x}_i 与浓度矩阵中的待测组分浓度矢量 \mathbf{y}_i 进行相关性计算,相关系数越大的波长其信息越多。因此,可结合经验知识给定一初始阈值,选取相关系数大于该阈值的波长参与建模。然后根据模型的精度调整阈值,从而确定最优波段。相关系数 r 由式(6-1)计算:

$$r_j = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6-1)$$

式中, $\bar{x}_j = \sum_{i=1}^n x_{ij}/n$, $\bar{y} = \sum_{i=1}^n y_i/n$, $j = 1, 2, \dots, m$, m 为波长点数, $i = 1, 2, \dots, n$, n 为样品数。

3. 方差分析法

在理想的情况下,校正集样品光谱信息的最大变化是由被测样品的组成或性质的变化引起的。模型的建立是根据光谱的这种变化,而不是根据光谱的绝对强度。因此光谱变化最明显的区间也应当对应光谱信息最丰富的区域。在光谱仪器测量精度较高时,样品光谱的标准差将直接反映该波长处光谱的信息量。所以,除了通过对样品组成或性质与光谱进行相关处理进行波长选择之外,在不清楚组成或数据性质时,也可以对校正集样品的光谱进行方差处理来优选波长。

方差分析谱区选择方法的流程如图 6-3 所示。

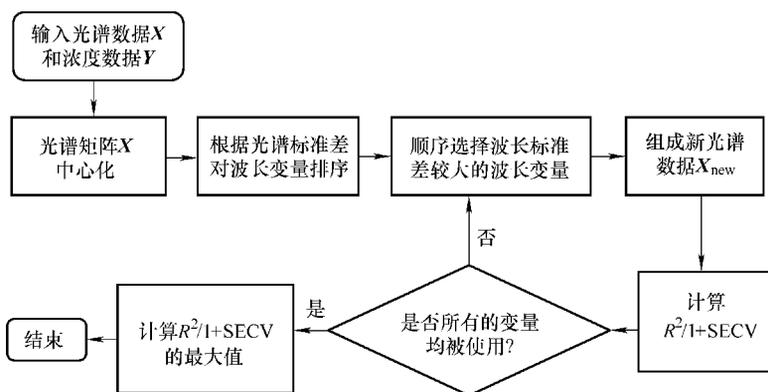


图 6-3 方差分析法谱区优化流程

4. 相关成分分析法

相关成分分析法是从近红外光谱与其组分含量之间的相关信息量的角度提出的一种简单、有效的波长选择方法。

近红外光谱在各个波长点上光谱数据的方差越大，离散程度越大，则相应的样品差异也越大，因此反映离散度的方差值可以作为考察该数据点上信息差异的指标，能从整体上较好地反映各波段信息量的多少。从近红外光谱的角度，虽然考虑了波长点所含信息变化量的大小，但是这些变化可能是干扰因素引起的，与样品的被测成分差异之间并无联系，属于不相关的冗余信息，对近红外光谱的定量分析精度会产生影响。因此还需要考虑近红外光谱与样品的被测组分含量之间的相关关系，从而挑选出与被测组分信息相关性大的区域。为此，构造 $n \times n$ 阶方阵 $Y = [Y_1, Y_2, \dots, Y_n]^T$ ，其中， $Y_i = [0, \dots, 0, y_i, 0, \dots, 0]^T$ ， y_i 为经过标准化之后的样品的浓度数据。进而构造相关成分矩阵 $S = XC$ ，其中， X 为光谱矩阵，相关成分矩阵 S 既体现了光谱矩阵的信息，同时也反映了浓度矩阵的信息。因此只需考察相关成分矩阵 S 各行矢量的方差值，就可以反映出各个波长点上光谱数据所体现的样品被测成分浓度差异的信息量。即可以根据方差的大小判断各波长点上光谱数据的离散度，以及光谱数据与浓度的相关性，从而挑选出与样品被测组分含量相关性较强的近红外光谱区间。

算法的实现步骤如下：

1) 对光谱矩阵 $X_{p \times n}$ 、浓度矩阵 $Y_{1 \times n}$ 分别做中心化处理，得到 $X'_{p \times n}$ ， $Y'_{1 \times n}$ ，即

$$X'_{ij} = X_{ij} - \bar{X}_j, \bar{X}_j = \frac{1}{n} \sum_{j=1}^p X_{ij}, Y'_i = Y_i - \bar{Y}, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (6-2)$$

式中， p 为波长点数； n 为样品个数。

2) 对 Y' 进行标准化得 Y'' ，即 $Y'' = Y' / \text{range} = [y''_1, y''_2, \dots, y''_n]$ (6-3)

式中， range 为浓度矩阵 Y 的极值。

3) 构造 $n \times n$ 阶方阵 Y ：

$$Y = \begin{bmatrix} y_{11} & 0 & \cdots & 0 \\ 0 & y_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{nn} \end{bmatrix}, y_{ij} = y''_j, i = j = 1, 2, \dots, n \quad (6-4)$$

4) 构造相关成分矩阵 $S = XY = [S_1, S_2, \dots, S_p]^T$ ，其中， $S_i = [s_{i1}, s_{i2}, \dots, s_{in}]$ ，

$i = 1, 2, \dots, p$, 是相关成分矩阵 S 的行矢量。

5) 计算相关成分矩阵 S 各行矢量的方差值, 即

$$\text{var}(S) = [\text{var}(s_1), \text{var}(s_2), \dots, \text{var}(s_p)]^T \quad (6-5)$$

6) 设定初始方差阈值, 根据校正模型的精度调整阈值, 从而确定最优波段。

5. 基于小波变换的波长选择方法

小波分解低频系数是对除去高频噪声系数后原光谱的替代, 能够最大限度地表征原数据的结构特征。基于小波变换的波长选择方法首先将原光谱用小波变换进行数据分解, 根据小波分解低频系数与浓度矩阵所建模型的质量, 找出最佳小波分解低频系数。然后将最佳小波分解低频系数与原光谱数据进行关联, 求出最佳小波分解低频系数与原光谱数据的列相关系数 R , 取与原光谱数据相关系数较大的波长组合来建模。最佳小波分解低频系数是根据所建模型的质量选出的, 最后参与建模的波长是最佳低频系数与原光谱相关性大的波长组合, 不仅考虑了浓度矩阵对波长选择的影响, 而且由于已将小波分解的高频系数全部滤除, 避免了高频噪声的干扰。选择与最佳小波分解低频系数相关系数大的波长组合来建模, 因为相关系数大的光谱区间携带原光谱矩阵的信息比较多, 对样品的组成或性质影响也较大。

采用预置的相关系数阈值, 自动选择那些相关系数大于阈值的波长点建立校正模型。阈值的选择根据所建模型的精度来确定。取不同的阈值对波长选择并建模, 模型精度最高时的阈值对应的波长组合即所求。

算法的实现过程如下:

1) 分别对原始光谱矩阵 $X_{n \times p}$ 的每一行 (代表一个样品的光谱数据) 进行小波分解, 分解到第 J 层, 得到 J 个小波分解低频系数矩阵, 找出最好的一个低频系数矩阵 $Y_{n \times pj}^{opt}$, 对应的分解层为 j^{opt} 。

2) 计算最佳小波分解低频系数矩阵 $Y_{n \times pj}^{opt}$ 与原始光谱矩阵 $X_{n \times p}$ 列矢量的相关系数。

3) 设定初始相关系数阈值, 根据所建模型的精度调整阈值, 从而确定最优波长组合。基于小波变换的波长选择算法流程如图 6-4 所示。

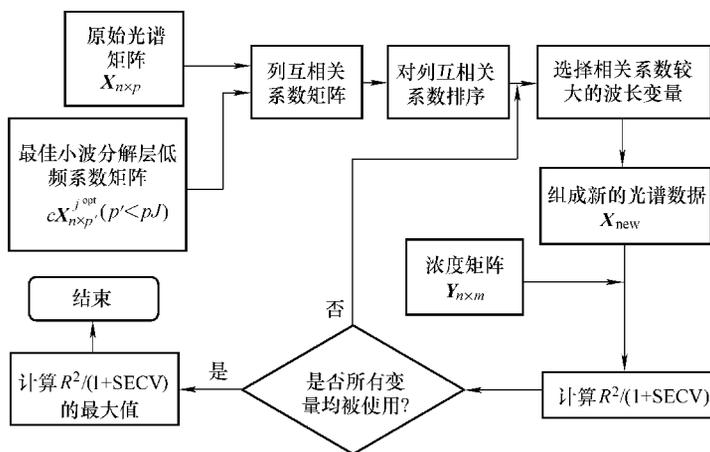


图 6-4 小波变换法波长选择流程

6. 基于遗传算法的波长选择方法

1975 年, Holland 提出了遗传算法 (Genetic Algorithms, GA), 遗传算法是一种模拟生物进化过程的优化方法, 主要是通过选择 (Selection)、交叉 (Crossover) 和变异 (Mutation) 等模拟生物基因操作的步骤来进行“优胜劣汰”, 在算法的运行过程中, 种群逐代优

化而趋近于问题的最优解。

在光谱分析中,遗传算法对于多组分分析的波长选择(Wavelength Selecting by Genetic Algorithms, WSGA)非常有效,是提高分析结果准确性的一个重要途径。Lucasius 等人将遗传算法用于 UV 法测定 4 种 RNA 核苷酸时的波长选择,使用 36 位的二进制串表示 36 个波长,“1”表示该波长被选择,“0”表示该波长不被选择,并分别以体系对波长的选择性、结果的准确性及最小均方误差为评价指标,将遗传算法与模拟退火法、逐步消元法进行比较,预测结果表明遗传算法的效果最佳。褚小力等人利用遗传算法对重整汽油中各组成的近红外光谱的波长变量进行筛选(以相关系数作为评价指标),再用偏最小二乘方法建立分析校正模型,不仅简化优化了模型,而且增强了模型的预测能力。王宏等人利用遗传算法对人体血糖近红外光谱的波长进行了优化选择(评价指标为预测标准偏差 RMSEP),提高了血糖浓度的测量精度。

6.3.2 CARS 波长变量选择法

CARS (Competitive adaptive reweighted sampling, 竞争性自适应重加权采样)法是近年提出的一种新的变量选择方法。此算法能够在一定程度上克服变量选择中的组合爆炸问题,筛选出优化的变量子集,提高模型的预测能力和降低预测方差。

近红外光谱通常由大量数据点构成,建模时波长点数远多于样品数,因此光谱共线性非常严重,利用变量筛选可简化模型,并提高模型的预测能力。CARS 方法模仿达尔文进化理论中的“适者生存”原则,每次通过自适应重加权采样技术筛选出 PLS 模型中回归系数绝对值大的波长点,去掉权重小的波长点,利用交互检验(CV)选出模型交互验证均方差(RMSECV)值最低的子集,可有效选择与所测性质相关的最优波长组合。具体算法如下:矩阵 $X_{m \times p}$ 为所测样品的光谱阵, m 为样品数, p 为变量数, $Y_{n \times 1}$ 表示组分矢量, T 为 X 的得分矩阵,是 X 与 W 的线性组合, W 为组合系数。 c 表示 Y 和 T 建立的 PLS 校正模型的回归系数矢量, E 为预测残差。则有下列关系式成立:

$$T = XW \quad (6-6)$$

$$Y = Tc + E = XWc + E = Xb + E \quad (6-7)$$

$b = Wc = [b_1 \ b_2 \ \dots \ b_p]^T$, 是一个 p 维的系数矢量。 b 中第 i 个元素的绝对值 $|b_i|$ ($1 \leq i \leq p$) 表示第 i 个波长对 Y 的贡献, $|b_i|$ 值越大则该变量越重要。为评价每个波长的重要性,定义权重 w_i 为

$$w_i = \frac{|b_i|}{\sum_{j=1}^p |b_j|}, i = 1, 2, \dots, p \quad (6-8)$$

通过 CARS 法去掉的变量,其权重 w_i 均设为 0。

1) 采用蒙特卡罗采样(MCS)法采样 N 次,每次从样品集中随机抽取 80% 的样品作为校正集,分别建立 PLS 回归模型。

2) 利用指数衰减函数(Exponentially Decreasing Function, EDF)强行去掉 $|b_i|$ 值相对较小的波长点。

第 i 次 MCS 采样时,波长点的保留率为 $r_i = ae^{-ki}$ 。其中, a 与 k 为常数,其值可根据下列两种情况求出:①第一次 MCS 采样时, p 个变量均被用于建模,故 $r_1 = 1$;②第 N 次 MCS

采样时，仅两个波长被使用，故 $r_N = 2/p$ 。 a 和 k 的计算公式为

$$a = (p/2)^{1/(N-1)} \quad (6-9)$$

$$k = [\ln(p/2)] / (N-1) \quad (6-10)$$

3) 通过 N 次自适应重加权采样技术筛选出 PLS 模型中回归系数绝对值大的波长点，用每次产生的新变量子集建立 PLS 回归模型，计算各模型的 RMSECV，选择 RMSECV 值最小的变量子集，即为最优变量子集。

6.3.3 基于 CARS-PLS 的食用油脂肪酸定量模型的优化

1. 样品制备与光谱采集

60 个植物油样品，包括花生油、大豆油、橄榄油、芝麻油、葵花籽油、玉米胚芽油、芥花油等，均购自当地超市（同 6.2.1 节）。采用气相色谱法测定 60 个食用油样品中 4 种脂肪酸（棕榈酸、硬脂酸、油酸和亚油酸）的含量。

光谱采集方法同 6.2.1 节，分辨率 16cm^{-1} 。

2. 结果与分析

(1) 样品集划分

根据预测浓度残差法分别剔除 4 种脂肪酸样品中的 3 个脂肪酸异常样品后，样品集共由 57 个样品组成，样品信息见表 6-4。校正集样品的选取直接影响所建模型的适用性和准确性。为了充分考虑校正样品集的浓度代表性，这里采用含量梯度法以校正集和校验集样品个数为 3:1 的比例进行划分，校正集样品 43 个，校验集样品 14 个。

表 6-4 食用油样品集统计信息

脂肪酸	最小值 (%)	最大值 (%)	标准偏差
棕榈酸	5.6	16.8	2.478
硬脂酸	1.8	5	0.747
油酸	22.7	77.9	14.461
亚油酸	8.6	65.9	14.848

(2) 软件优化建模

采用 Bruker 公司的 OPUS6.5 软件自动优化建模功能，采用常规的光谱预处理方法及其组合进行消噪，采用软件中设定的用于近红外光谱的 5 个典型的谱段范围来优化谱区范围，采用 PLS 建立最优校正模型，结果见表 6-5。

表 6-5 软件自动优化建模及预测结果

脂肪酸	预处理	nf	R^2	RMSECV	RMSEP	保留波长数
棕榈酸	一阶导 (17 pts.) + 矢量归一化	10	0.977	0.383	0.319	844
	12489.4 ~ 7498.3 cm^{-1} , 6102 ~ 4597.7 cm^{-1} 常数偏移量消除					
硬脂酸	7506 ~ 4597.7 cm^{-1} 常数偏移量消除	13	0.961	0.15	0.143	378
	7506 ~ 5446.3 cm^{-1} 最小最大归一化					
油酸	7506 ~ 5446.3 cm^{-1} 最小最大归一化	12	0.991	1.27	0.966	268
	7506 ~ 5446.3 cm^{-1}					
亚油酸	7506 ~ 5446.3 cm^{-1}	11	0.992	1.37	1.46	268

(3) CARS-PLS 模型优化

对校正集光谱进行标准化处理后，采用 CARS 方法分别对 4 种脂肪酸的波长变量进行多次筛选，以棕榈酸为例说明变量筛选过程，其筛选过程如图 6-5 所示。

图 6-5a 所示为筛选过程中选出变量的变化趋势，变量数随着运行次数增加由快到慢递减。运行次数和保留的变量数之间是指数关系。

图 6-5b 所示为经过 5 折交互检验所得残差 RMSECV 的变化趋势图，在 1~32 次间残差呈现递减趋势，表明筛选过程剔除与样品性质无关的变量，32 次后开始递增，则可能剔除了关键变量，从而导致残差增大。

图 6-5c 中“*”所对应点即 RMSECV 最低点，图中各线表示随着运行次数增加各变量回归系数的变化趋势。

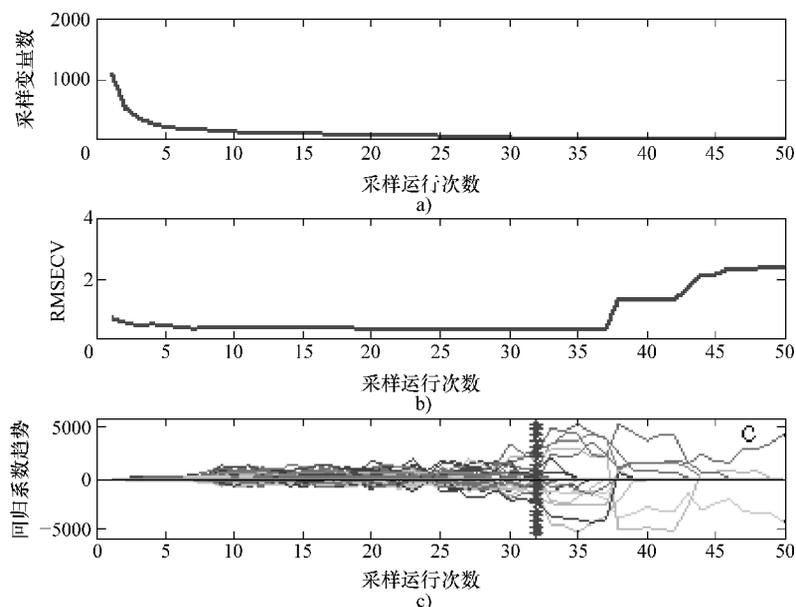


图 6-5 棕榈酸光谱波长变量筛选图

利用所选波长建立 PLS 回归模型，并通过校验集校验，结果见表 6-6。由表可知，经波长筛选，模型质量明显提高，优选波长后建模得到的 4 种脂肪酸模型的 RMSECV 值分别是 OPUS 自动优化建模 RMSECV 值的 79.3%、80.7%、58.3% 和 49.6.3%。根据 RMSEP 指标，CARS-PLS 模型的预测能力明显优于软件自动优化所建模型。

表 6-6 CARS-PLS 优化建模及预测结果

脂肪酸	nf	R^2	RMSECV	RMSEP	保留波长数
棕榈酸	12	0.986	0.304	0.248	23
硬脂酸	12	0.975	0.121	0.110	23
油酸	13	0.997	0.740	0.868	23
亚油酸	12	0.997	0.859	0.833	20

4 种脂肪酸的分子结构中都含有亚甲基 CH_2 和羧酸 COOH ，这类含 H 基团在近红外区有比较明显的吸收峰。其中，线性脂肪族分子的亚甲基在一级倍频区的主要吸收峰约在 5800cm^{-1} 和 5680cm^{-1} 处。羧酸中的羟基 OH 伸缩振动和 $\text{C}=\text{O}$ 伸缩振动的组合频在 5290cm^{-1} 也有吸收峰。CARS 方法挑选的 4 种脂肪酸的特征波长变量如图 6-6 所示。从图 6-6 中可以看出，CARS 法对 4 种脂肪酸挑选出的特征波长在 $6000\sim 5000\text{cm}^{-1}$ 中有公共区域，而亚甲基

CH_2 和羧酸 COOH 正是在该区域有主要吸收峰。另外，与表 6-5 中的保留波长数相比，CARS 法挑选的波长个数显著降低。

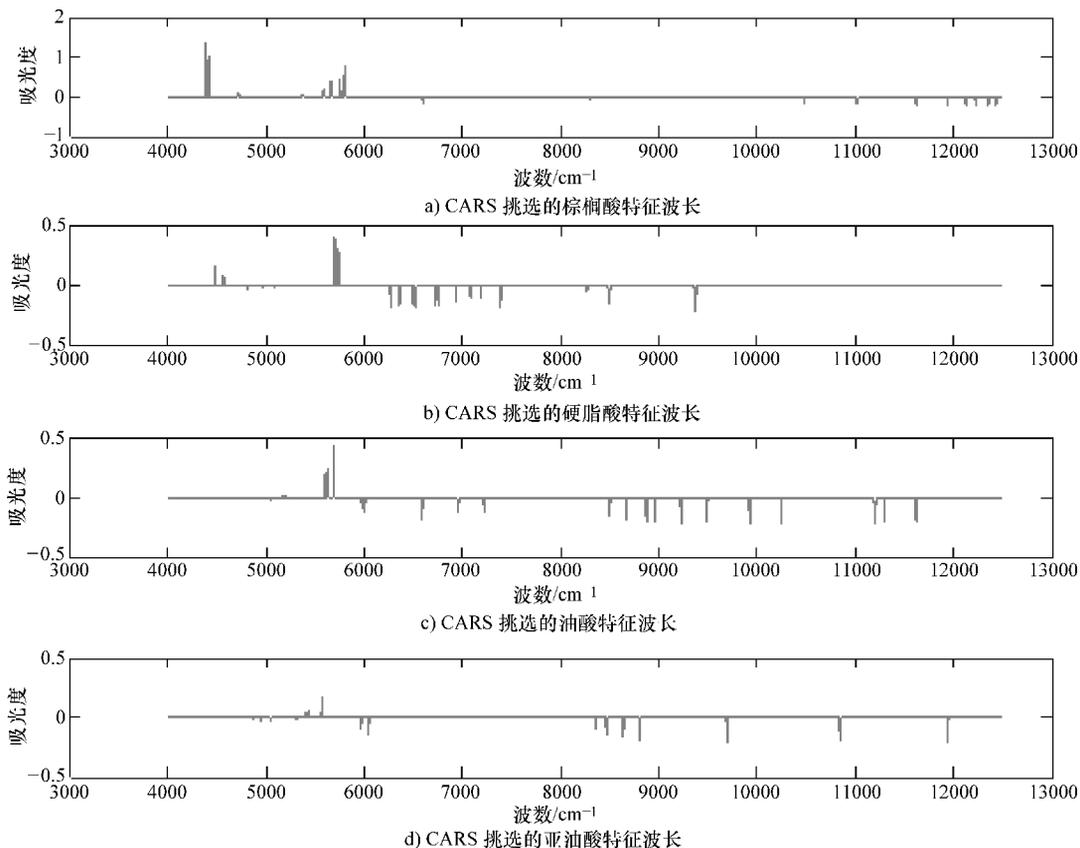


图 6-6 CARS 挑选的食用油脂脂肪酸近红外光谱的特征波长

采用 CARS 分别挑选出食用油中 4 种主要脂肪酸（棕榈酸、硬脂酸、油酸和亚油酸）特征波长变量，实验结果表明 CARS 法不但能有效减少特征波长个数，而且挑选出的波长与物质在近红外区的特征吸收峰相近，能真实地反映出物质所含基团，可为滤光片式食用油近红外分析仪的波长选取作参考。实验中基于 CARS - PLS 的食用油脂脂肪酸近红外定量分析模型的优化，大大提升脂肪酸近红外模型的预测能力。

6.4 基于近红外光谱的食用油过氧化值定量分析

油脂品质的优劣，可用碘值、过氧化值、游离脂肪酸含量等指标来反映。废弃油脂含有在高温下反复使用的煎炸油，反复收集的烹调油，在高水分、高金属离子浓度、高温的复杂环境下迅速酸败，其过氧化值远远高于合格油脂。因此通过测定油脂过氧化值含量也能作为快速鉴别废弃油脂的方法之一。

我国毕艳兰等人应用傅里叶近红外光谱技术测定了常用食用植物油（大豆油、花生油、葵花籽油）的过氧化值，在特定实验温度（ 40°C ）条件下获得较好的定量分析结果。崔晓君等人采用近红外光谱测定了合格花生油样品中的过氧化物含量。目前针对废弃油脂的过氧

化值是否可以在常温下用近红外进行检测的研究甚少。

本节的工作拟在常温下，采用红外光谱技术，针对模拟的废弃油脂中过氧化值建立近红外快速定量分析模型，探索其作为一种快速鉴别废弃油脂技术的可行性。

6.4.1 样品制备

北京市超市购买：纯花生油。山东农贸市场购买：散装压榨纯花生油。

表6-7 模拟制备废弃油脂中高温反复使用的煎炸油。从样品的制备过程中可以得出，虽然电磁炉加热时间远低于水浴锅加热时间，但油品酸败速度远快于后者。且样品随着加热时间越长，颜色逐渐变深发黑，且散发出异味。共制备废弃油脂样品32个，收集未经高温加热样品4个，组成样品集。

表6-7 样品制备

加热方式	加热时间/min	样品个数	样品来源
水浴锅加热	60	4	超市
100℃	100	4	超市
电磁炉加热	5	2	市场
150℃	10	2	市场
	15	2	市场
	20	2	市场
	25	2	市场
	30	2	市场
电磁炉加热	5	2	市场
150℃	10	2	市场
加油大量搅拌	15	2	市场
	20	2	市场
	25	2	市场
	30	2	市场

按照 GB/T 5009—2003 测定样品过氧化值。

6.4.2 光谱采集

采用德国 Bruker 公司 VERTEX 70 红外光谱仪扫描样品近红外光谱。仪器参数设定如下：光谱波数范围：12500 ~ 4000 cm^{-1} ，分辨率：8 cm^{-1} ，每个样品重复扫描次数：32 次。使用液体光纤探头采样，光程：2mm。

全部样品未经任何化学处理，将光纤探头插入装有样品的小瓶中，逐一扫描样品，每个样品分别取两次进行扫描，每次测量前均用石油醚清洗探头，避免样品间交叉污染。样品近红外光谱如图 6-7 所示。

6.4.3 结果与分析

由于分析对象是经过长时间高温加热后的油品，样品的颜色差异较大。从图 6-7 可以看

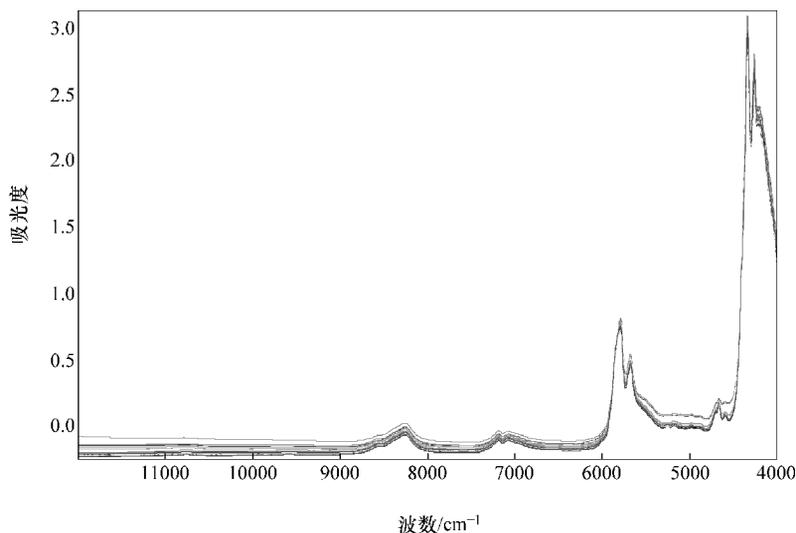


图 6-7 样品集近红外光谱图

出样品颜色差异对其近红外光谱在短波段 $9000 \sim 12500\text{cm}^{-1}$ 的影响尤为明显。因此要建立性能较好的过氧化值的近红外模型，进行光谱数据预处理是非常必要的。数据预处理步骤主要包括：校正集和校验集的划分；消除光谱噪声和其干扰的影响；优化光谱范围，净化谱图信息等。

1. 校正集样品的选择

校正集样品的选取直接影响所建模型的适用性和准确性。为了充分考虑校正集样品浓度的代表性，这里采用含量梯度法以校正集和校验集样品个数为 2:1 的比例进行划分。样品集划分信息见表 6-8，从表中可以看出，校正集样品的浓度范围涵盖了校验集样品的浓度范围。

表 6-8 样品集统计信息

样品集	样品数	最小值/(mmol/kg)	最大值/(mmol/kg)	标准偏差
校正集	24	2.6731	33.6735	6.873042
校验集	12	4.1936	33.5306	7.708292

2. 校正模型的建立

实验中采用 OPUS6.5 软件分析光谱数据，采用 PLS 建立定量回归分析模型。过氧化值校正模型的预测精度和可靠性采用决定系数 R^2 、RMSECV、RMSEP、RPD 来综合评价。其中，RPD 用来验证模型的稳定性和预测能力。RPD > 3，则模型具有较高的稳定性和良好的预测能力。

采用 Bruker 公司的 OPUS6.5 软件优化建模，采用常规的光谱预处理方法及其组合进行消噪，采用软件中设定的用于近红外光谱的 5 个典型的谱段范围来优化谱区范围，采用 PLS 建立最优校正模型，结果见表 6-9。

表 6-9 校正模型指标

指标	预处理方法	最佳主成分数	R^2	RMSECV /(mmol/kg)	RPD
过氧化值	一阶导数(17pts) + MSC 6800.1 ~ 7733.5 cm^{-1} 4933.3 ~ 5866.7 cm^{-1}	10	0.9554	1.44	4.73

表 6-9 中建立最优校正模型采用了一阶导数和 MSC 光谱预处理组合。其中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。实验中由于高温加热导致油的成分发生明显变化,样品更趋于黏稠,而 MSC 可以校正因样品形态不均匀性(粒度分布)导致所测的样品光谱的差异性。优选的谱区范围分别在 6800.1 ~ 7733.5 cm^{-1} 、4933.3 ~ 5866.7 cm^{-1} ,避开了由于样品颜色导致光谱差异的短波段区域。过氧化值本质上衡量的是油脂氧化生成的初级产物之一:氢过氧化物。氧化脂肪酸酯产生的氢过氧化物的特征吸收峰在 6850 cm^{-1} 和 4850 cm^{-1} 处。前者在处于实验所选取的 6800.1 ~ 7733.5 cm^{-1} ,后者在 4933.3 ~ 5866.7 cm^{-1} 区域附近。

表 6-9 中过氧化值的校正模型的决定系数 R^2 为 0.9554,说明化学分析值与近红外光谱预测值之间具有良好的线性关系,且 RPD > 3 时,具有良好的定标效果。

3. 校正模型的外部验证

为了检验校正模型的预测能力,采用外部校验样品集。校验集样品过氧化值预测结果如图 6-8 所示(每个校验集样品分别扫描两次,作为待测样品,共 $12 \times 2 = 24$ 个)。其中决定系数 R^2 为 0.98843, RMSEP 为 0.966mmol/kg, PRD 为 8.01。RPD > 3 时,预测效果良好。因此实验中建立的过氧化值近红外光谱定标模型可以用于实际预测。

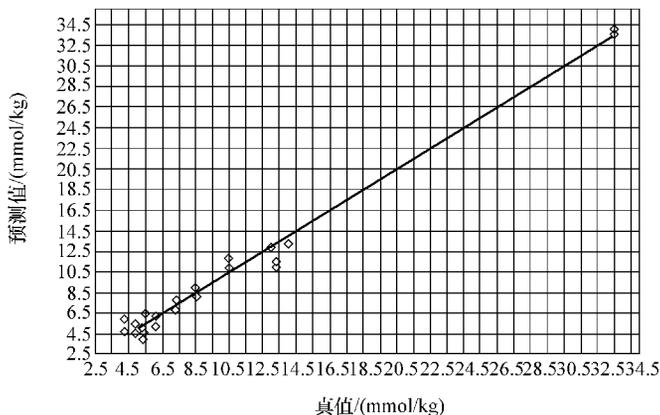


图 6-8 校验集真值 - 预测值相关图

本节工作尝试采用近红外光谱分析技术建立了常温下模拟废弃油脂中的过氧化值定量分析模型。初步实验结果表明,实验中建立的过氧化值定量分析模型具有较高的质量,并通过外部校验法验证了模型的准确性及其稳定性。因此,近红外光谱技术用于废弃油脂中过氧化值的快速检测具有较好的可行基础,但是实验中的废弃油脂样品是模拟配制的,对于真实样品中过氧化值是否可测及测定精确度如何,还需进一步实验探索。

6.5 基于近红外光谱的食用油酸价定量分析

食用油在长期贮存、运输过程中，在温度改变、接触空气等客观因素的影响下会发生一系列的氧化反应，产生有异味的高氧化物质，即油脂酸败。食用油变质通常分为水解变质和氧化变质。水解变质时游离脂肪酸产生不良气味，同时产生二酰甘油酯等一些有很强乳化作用的副产物，水解变质程度的衡量指标通常用酸价来表示。因此，酸价是评价油酸酸败变质程度的重要指标之一。传统用来测定油酸酸败程度的方法有碘量滴定法和分光光度法等，但其操作较为复杂，并且检测周期长，人为因素影响颇大。

6.5.1 样品制备

从超市购买不同厂家不同批次的食用油样品作为测定酸价含量的样品。样品种类包括大豆油、花生油、棕榈油、橄榄油、菜籽油和调和油等多种食用油，所选的样品能够基本覆盖市场上提供的常用食用油，具有一定代表性。

采用3种方法制备待测样品：①150℃加热，最短加热时间为5min，最长加热时间为30min，在不同时间段上取样；②150℃加热，但在加热的同时加入未经处理同类油不断搅拌，从中取样；③采用微波中高火加热30min处理，取样。共制备样品24份。

在对样品进行加热氧化的过程中，可以明显地看到食用油物理色味的变化，随着加热时间的增加，样品颜色逐步加深，并且气味逐步异常。个别样品在冷却后，在17~20℃的室温下凝固了，微温后，颜色深黑，气味难闻。经过氧化处理后，应用国标法测定其酸价。样品的酸价范围为0.091~10.4276mg/g。

6.5.2 光谱采集

光谱采集方法同6.3.3节。

6.5.3 结果与分析

按照浓度分布选择16个样品作为校正集样品，8个样品作为校验集样品。采用OPUS6.5软件进行建模分析，并对预测集食用油样品进行预测。

1. 波长范围的选择

从食用油样品的光谱图可以看出，样品的主要吸收峰在6102~4597.7 cm^{-1} ，该波段内含有丰富的样品特征信息，为校正模型的建立提供可能，因此选此波段范围作为建立校正和预测模型的波长范围，其交互式选择范围的光谱如图6-9所示。

2. 光谱数据的预处理

食用油样品经采集光谱后，需要做一些数学处理以消除基线漂移、降低噪声干扰等无用信息，经过合适的预处理后使得决定系数 R^2 趋近1，RMSECV的值尽量小，建立预测能力较高的校正模型，从表6-10可知：食用油酸价模型，经过MSC法预处理后，其建立模型评价指标优于其他预处理方法建立的模型。决定系数 R^2 为99.73，RMSECV为0.259。

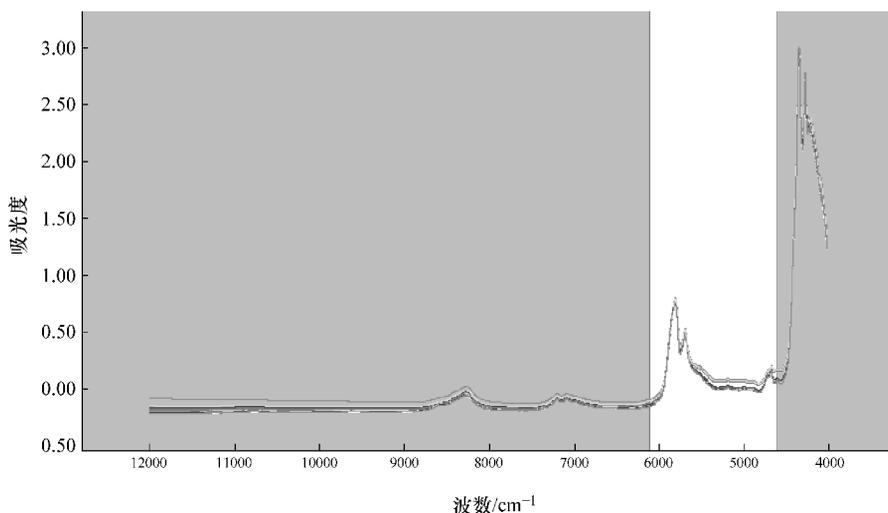


图 6-9 选择光谱范围

表 6-10 光谱的预处理方法及模型的决定系数和 RMSECV

预处理方法	酸价模型	
	R^2	RMSECV
未经任何预处理	99.49	0.359
矢量归一化	99.28	0.428
一阶导数	99.57	0.324
二阶导数	98.18	0.678
MSC	99.73	0.259
消除常量偏移法	99.28	0.428

3. 主因子数的确定

在所选的光谱波长范围和已确定的数据预处理方法后，主因子数的确定对模型的建立同样起着重要的作用。通常采用 RMSECV 值来确定最佳主因子数：选择合适的主因子数；不能过大，否则出现“过拟合”；同样不能选取过小，否则出现“欠拟合”现象。在一个最佳的主因子数下建立的模型，其预测能力才能达到较高的效果。当 RMSECV 值达到最小时，即为最佳主因子数。RMSECV 与主因子的关系如图 6-10 所示，最佳主因子数为 9。

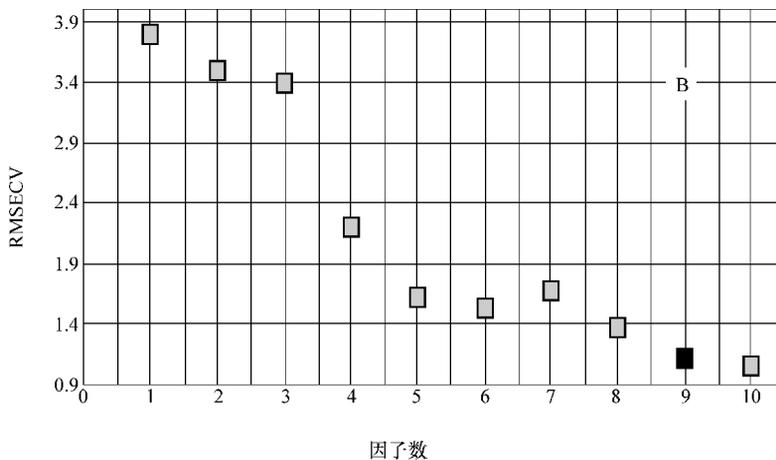


图 6-10 RMSECV 随主因子数的变化

4. 酸价近红外光谱模型的建立

食用油酸价校正模型，光谱预处理方法为 MSC 法，主因子数为 9。所建的模型，酸价模型中决定系数 R^2 为 99.73，RMSECV 为 0.259，如图 6-11 所示。

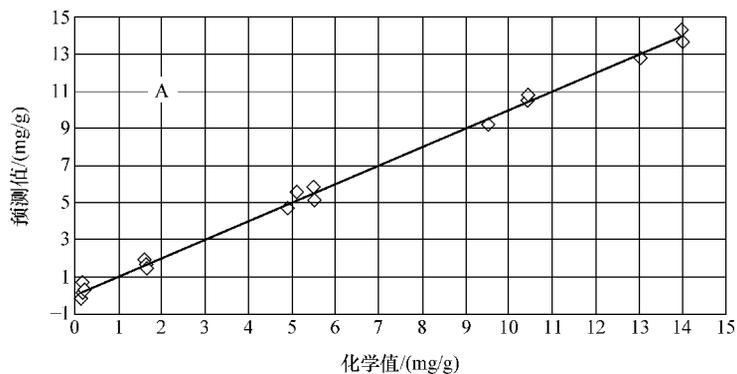


图 6-11 酸价预测值与化学值相关曲线

5. 酸价近红外光谱模型的校验

将选取的 8 个预测集样品放入模型中进行酸价的预测。酸价模型中决定系数 R^2 为 99.47，正交标准差 RMSEP 为 0.567。其预测值与标准值如图 6-12 所示，预测曲线与标准曲线夹角很小，可见预测值趋近于真实值，并且样品散点密集的分布在曲线两侧，预测样品均在校正模型的可预测范围内，无超出预测范围样品，该模型具有较好的适用性。

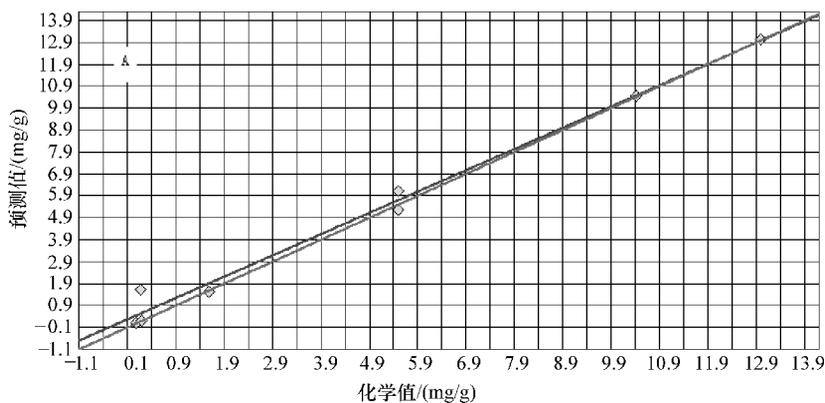


图 6-12 酸价预测值与化学值相关曲线

预测集样品酸价的真实值与预测值结果见表 6-11。

表 6-11 8 个预测集食用油样品的酸价预测结果

样品编号	酸价/(mg/g)	
	真实值	预测值
L07.0	5.5000	6.1011
L10.0	5.5000	5.2783
P1.1	0.1831	1.6471
P3.0	0.0910	0.0981

(续)

样品编号	酸价/(mg/g)	
	真实值	预测值
P5.0	0.1893	0.1957
P7.0	1.5945	1.5496
P12.0	10.4276	10.4780
N01.0	13.0000	13.0450
平均偏差	0.24	

本节工作尝试采用近红外光谱分析技术建立了常温下食用油的酸价定量分析模型。初步实验结果表明,建模时波数范围的选择以及预处理方式的不同对模型预测能力有较大的影响,实验中根据模型评价参数分析优选了最佳的数据预处理方法,结合 PLS 分别建立的校正模型能够有效地进行样品酸价预测。

6.6 小结

本章采用近红外光谱技术结合化学计量学方法建立了食用油脂脂肪酸含量、过氧化值和酸价的定量分析模型,并且探讨了光谱分辨率的设定对食用油近红外光谱模型性能的影响。实验结果表明,特征谱峰的选取结合适当的光谱预处理可以有效地提高近红外光谱模型的预测性能,使食用油定量分析模型的实用性得到进一步提高。本章工作的探索有力地推动了近红外光谱技术应用于食用油品快速监管的进程。

参考文献

- [1] 陆婉珍. 现代近红外光谱分析技术 [M]. 2 版. 北京: 中国石化出版社, 2007.
- [2] Dongsheng Cao, Yizeng Liang, Qingsong Xu, et al. Toward better QSAR/QSPR modeling: simultaneous outlier detection and variable selection using distribution of model features [J]. *Journal of Computer - Aided Molecular Design*, 2011, 25: 67 - 80.
- [3] 吴静珠, 王一鸣, 张小超, 等. 近红外光谱分析中定标集样品挑选方法研究 [J]. *农业机械学报*, 2006, 37 (9): 80 - 82.
- [4] LiHongdong, Liang Yizeng, Xu Qingsong, et al. Key wavelengths screening using competitive adaptive re-weighted sampling method for multivariate calibration [J]. *Anal Chim Acta*, 2009, 648 (1): 77 - 84.
- [5] 陆婉珍, 袁洪福, 褚小立. 近红外光谱仪器 [M]. 北京: 化学工业出版社, 2010.
- [6] 吴静珠, 刘翠玲, 邢素霞, 等. 光谱分辨率的选取对食用油近红外模型性能影响分析 [J]. *北京工商大学学报 (自然科学版)*, 2012.
- [7] Miralbes C. Discrimination of European wheat varieties using near infrared reflectance spectroscopy [J]. *Food-Chemistry*, 2007.
- [8] J H Holland. *Adaptation in Natural and Artificial Systems* [M]. Michigan: University of Michigan Press, 1975.
- [9] 周明, 孙树栋. 遗传算法原理及应用 [M]. 北京: 国防工业出版社, 1999.
- [10] 何险峰, 周家驹. 遗传算法及其在化学化工中的应用 [J]. *化学进展*, 1998, 10 (03): 312 - 318.
- [11] 姚芳莲, 李维云, 邓联东, 等. 遗传算法及其在化学领域中的应用 [J]. *天津化工*, 2000 (04):

- 1-3.
- [12] Lucasius C B, Beckers M L M, Kateman G. Genetic algorithms in wavelength selection: a comparative study [J]. *Analytical Chemical Acta*, 1994, 286 (02): 135-153.
- [13] 褚小力, 袁洪福, 王艳斌, 等. 遗传算法用于偏最小二乘方法建模中的变量筛选 [J]. *分析化学*, 2001, 29 (04): 437-442.
- [14] 王宏, 李庆波, 刘则毅, 等. 遗传算法在近红外无创伤人体血糖浓度测量基础研究中的应用 [J]. *分析化学*, 2002, 30 (08): 779-783.
- [15] 吴静珠, 徐云. 基于 CARS-PLS 的食用油脂脂肪酸近红外定量分析模型的优化 [J]. *农业机械化学报*, 2011.
- [16] Jerry Workman, Jr Lois Weyer. 近红外光谱解析实用指南 [M]. 褚小立, 许育鹏, 田高友, 译. 北京: 化学工业出版社, 2009.
- [17] 王乐, 胡健华. 食用油掺伪餐饮业废油脂鉴别检测方法研究进展 [J]. *中国油脂*, 2007, 32 (9), 75-77.
- [18] 李昌, 王兴国. 食用油掺假检测方法概述 [J]. *农产品工程技术: 农产品安全*, 2007 (5), 30-35.
- [19] 徐冲, 刘远洋. 食用油掺入废食用油脂检测技术研究现状 [J]. *粮食与油脂*, 2009 (4), 34-37.
- [20] 毕艳兰. 利用傅里叶近红外技术快速测定食用植物油的过氧化值 [J]. *中国油脂*, 2009, 34 (3), 71-74.
- [21] 鲍丹青, 毕艳兰, 邓德文, 等. 利用近红外光谱技术快速测定大豆油的过氧化值 [J]. *中国粮油学报*, 2008, 23 (6), 206-209.
- [22] 崔晓君. 花生油中过氧化物的近红外光谱分析 [J]. *应用化学*, 2008, 25 (3), 375-377.
- [23] 张虹, 高霞. 高温加热对植物油脂品质的影响 [J]. *粮油加工*, 1998 (6), 8-11.
- [24] 吴静珠, 王一鸣, 张小超, 等. 近于近红外光谱的奶粉品质检测技术研究 [J]. *光谱学与光谱分析*, 2007, 27 (9), 1735-1738.
- [25] 李桂华. 油料油脂检验与分析 [M]. 北京: 化学工业出版社, 2006.
- [26] Jingzhu Wu, Suxia Xing, Hongbin Xiao, et al. Rapid Determination of Peroxide Value in Simulated Waste Oil by Near Infrared Spectroscopy [J]. *Sensor Letters*, 2012, 511-513.

基于近红外光谱的小麦粉品质检测方法研究

7.1 简介

小麦是世界三大重要粮食作物之一，全世界有 35% ~ 40% 的人以小麦作为主要的粮食。小麦富含蛋白质，被作为很多主食和副食的主要加工原料，它也是一种营养价值较高、比较容易贮藏的国家重要商品粮食。世界上最主要的小麦生产国集中在亚洲、北美、欧洲和俄罗斯，分别是我国、美国、印度、法国、澳大利亚、加拿大、俄罗斯。这 7 个国家的小麦种植面积占世界小麦种植总面积的 57.88%，小麦产量占世界小麦总产量的 60.7%。

我国是世界上最大的小麦生产、消费国，以 2012 年为例，我国小麦播种面积达到 24400 千 hm^2 ，单产为 4816 kg/hm^2 ，总产量为 1.175 亿 t，小麦产量占世界小麦总产量的 17% ~ 19%，我国所产小麦几乎全部用于加工面粉以满足居民的日常饮食需求，除此之外每年还需进口几百万吨的小麦用于各种面粉的生产。小麦粉是我国居民日常生活中不可或缺的主食原料，也是加工部分食品的基础原料，具有其他的粮食作物不可替代的优势，我国每年小麦产量在 1 亿 t 左右，占全国粮食总产量的 23% 左右，小麦粉品质的好坏直接影响面粉制品的质量，也直接关系到人们的身体健康，在这个注重产品质量、食品安全和身体健康的社会里，人们更加注重小麦粉的品质，因此，面粉厂的安全生产检验越来越受到人们的关注，面粉厂想要市场中立于不败之地，必须提高产品质量。

我国小麦粉品质分析基本上是由物理品质检验、化学品质检验、食用品质检验三部分组成的：物理品质包括加工精度、粗细度、面筋质、磁性金属物等；化学品质包括灰分、含砂量、脂肪酸值、水分等；食用品质包括熟食品质，有害残留物测定，判断其气味、口味等。其中，水分、灰分及面筋的含量是影响小麦粉品质的重要因素，也是工厂日常检测的主要工作，三项指标需要实时检测，同时也对指导实时生产起到重要作用，目前小麦粉的品质检测采用传统的实验室测定法，存在检测时间较长、操作复杂及人为因素影响较大等问题，比如小麦粉的水分、灰分定量分析的测定至少需要 3 ~ 4h，即使检测人员全力以赴，每日的检验也只能做 1 ~ 2 次，而面筋的检验不仅耗时长而且受人为因素的影响较大，这对保证产品质量的稳定性是远远不够的，特别是生产自动化高度发展的今天，面粉厂的配粉工艺要求品质研发部及时提供品质检验结果，以便及时采取措施，调整小麦粉的生产工艺和搭配，减少不合格产品的生产。

针对目前小麦粉品质检测方法的种种弊端和实际生产的需要，研究一种简便、快速、准确、无污染、无损的检测方法是小麦粉品质检测的重要发展方向。近红外光谱检测技术作为近些年发展起来的检测方法，在石化、医药、食品和农业方面的检测领域得到了卓有成效的研究和使用，并且在某些领域得到了产业化应用。通过近些年的研究证明，近红外光谱检测技术是无损、无污染、快速、低成本的检测方法。这项技术的推广，被认为是 20 世纪末仅

近红外光谱技术在食品品质检测方法中的研究

次于计算机推广的一次革命。这一方法可以解决传统小麦粉检测时间长、操作复杂等难题,利用近红外光谱检测技术测定小麦粉的品质指标的技术和方法逐渐得到应用和推广。

本研究与北京古船食品有限公司合作从工厂实际生产需求出发,以小麦粉的水分、灰分及面筋三项指标为检测对象,采用近红外光谱技术法,旨在建立实用性较强的小麦粉品质检测模型,以期能指导公司实际生产。利用实验分析结果验证近红外光谱检测方法在小麦粉品质检测领域的可行性,该研究与实际工厂生产的需求相结合,具有现实应用意义。

7.2 基于近红外光谱的小麦粉 PLS 模型的建立与分析

7.2.1 样品制备

从北京古船食品有限公司的面粉厂采集所需样品 304 个,其中,标准 15 个、低筋 2 个、富强 87 个、裹粉 23 个、家用 23 个、饺子 18 个、军供 30 个、麦心 46 个、馒头 2 个、面包 32 个、雪花 17 个、鸭饼 5 个、营养强化 4 个。为了建立适用性广、准确的定量分析模型,需要保证样品多样性、均匀分布及一定的浓度范围,所以从中选取了 150 个样品作为试验样品。

7.2.2 光谱采集

由于小麦粉为不透明的粉末状固体,透射光谱不能取得较好的效果,本研究采用漫反射的方式采集近红外光谱。所用设备为 Bruker 公司生产的傅里叶变换近红外光谱仪 VERTEX 70,设备配有漫反射积分球附件,试验采用积分球的漫反射测样方式。

除了测样方式的选择,选择适合的仪器参数和试验条件对测量结果和分析结果也有一定的影响,参数设置对近红外分析结果有一定影响,下面就光谱参数选取进行讨论:

1) 检测温度的确定。在光谱采集的过程中,会因温度的改变对光谱造成一定的影响,为了减少环境因素改变造成的光谱误差,这里整个检测过程的环境温度保持 20℃。

2) 光谱区间选择。选择合适的谱段可以缩短计算时间,还可以消除一些干扰因素的影响,由于全光谱包含的信息量大,而且样品待测成分的具体吸收峰不能确定,因此对光谱图进行分析时,常常采用全光谱。这里采用选取 12000 ~ 4000 cm^{-1} 谱段进行光谱采集。

3) 扫描分辨率的选择。扫描分辨率是指采集光谱数据时采样点之间相隔的波数,扫描分辨率直接影响光谱图的结构,分辨率越高,光谱图的结构越细,但采样时间相应变长,为了保证光谱包含信息量并节省光谱采集时间,这里均在分辨率 8 cm^{-1} 下采集光谱。

4) 扫描次数的选择。扫描次数不同也会对光谱的确定性造成影响,扫描次数过少,会造成光谱精度不够,扫描次数过多不仅会增加试验时间,而且在对同一样品进行长时间扫描的过程中可能会出现不稳定的现象,如导致光谱发生变化,为了在保证较小的标准差和较大的信噪比的情况下,节省采集时间,这里中扫描次数选择 64 次。

5) 样品测量次数的选择。如果样品分布不均匀,或者试验过程中几次装样的一致性不好,会存在随机误差,为了避免这种情况,以提高测量的精度,这里采用对样品测量 3 次求光谱平均值的方式。

综上,本试验采用积分球的漫反射测样方式,样品杯旋转采样,仪器参数设置如下:波

数范围 $12000 \sim 4000\text{cm}^{-1}$ ，分辨率 8cm^{-1} ，扫描次数 64 次；环境温度 20°C ；测量 3 次求平均值。将小麦粉样品放置在漫反射样品台的样品杯中，进行近红外光谱采集。小麦粉样品的近红外漫反射光谱如图 7-1 所示。

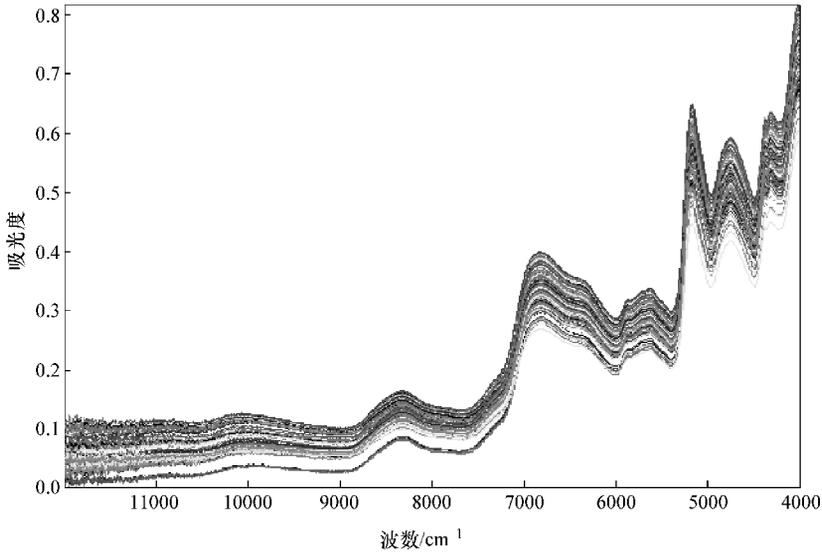


图 7-1 小麦粉的近红外漫反射光谱图

7.2.3 结果与分析

近红外光谱分析技术是一种快速、高效、无损的分析方法，通过建立校正分析模型对未知样品进行分析。因此，近红外光谱分析的关键在于建立能够准确预测、抗干扰能力强的校正分析模型。近红外光谱定量分析过程如图 7-2 所示。

近红外光谱分析技术的应用通过以下 9 个步骤完成：①采集具有代表性的样品；②测得样品的近红外光谱图；③剔除异常样品；④选择有代表性的样品作为校正集；⑤采用国家标准或权威的方法测得样品的组分浓度；⑥对校正集样品的近红外光谱数据进行预处理，并进行波长优选；⑦选择合适的化学计量方法建立校正模型；⑧对模型进行改进，得到准确、可靠、稳定的校正模型；⑨对预测样品的光谱图进行预处理，并利用校正模型对未知样品的组分参数进行预测。

对小麦粉的品质参数进行定量分析，用相关系数 (R^2)、交互验证方均根误差 (RMSECV)、预测方均根误差 (RMSEP) 作为评价指标。

本研究也尝试利用 PLS 法建立小麦粉的近红外光谱校正模型，把样品集中的 150 个样本按照 4:1 浓度法选择 120 个样品作为校正集，剩余 30 个样品作为预测样品，采用典型的 PLS 方法进行建模，用 R^2 、RMSECV 和 RMSEP 作为评价指标，建模结果见表 7-1。

由表 7-1 可知，用 PLS 法建立小麦粉的近红外光谱校正模型是可行的，但是可能由于存在人为误差或系统误差，模型的精度有待进一步提高。分析误差存在的原因可能是由于人为或系统误差造成的大量异常样品的存在，或是近红外光谱中含有大量的噪声信息，可以通过剔除异常样品或筛选特征谱区的方法提高分析模型的精度。

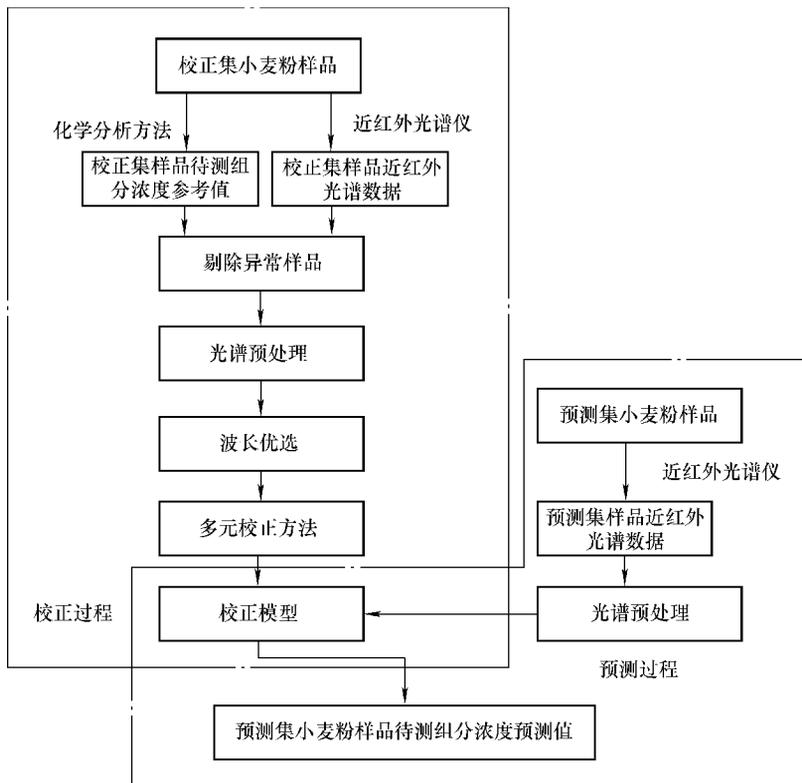


图 7-2 近红外光谱定量分析过程

表 7-1 PLS 法建模结果

小麦粉成分	内部交叉检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	0.8424	0.0646	0.8455	0.0405
水分	0.7176	0.4456	0.7058	0.3931
面筋	0.6923	1.2370	0.8378	0.6699

7.3 基于近红外光谱的小麦粉异常样品剔除方法研究

近红外光谱定量分析的过程中，分析结果的可靠性首要取决于原始数据的准确性，即获得的近红外光谱及化学值的可靠性。小麦粉样品的原始数据，即样品的近红外光谱图和化学值的相关性直接影响模型的预测能力，而异常样品的干扰是影响分析模型的重要因素，因此异常样品的判别与处理是提高模型预测能力的一个重要步骤。产生异常样品的原因如下：①测量仪器、测量方法及环境等客观因素影响；②技术人员主观因素的作用；③样品的复杂性、多样性。所以判断异常数据是近红外光谱检测过程中首要解决的问题。建立定量分析模型时，样品中常常混有异常样品，需要在训练模型之前，把异常样品剔除。

本研究中可能由于客观因素及主观因素的影响造成异常样品的存在，因此，对马氏距离

法和蒙特卡洛交叉验证法剔除异常样品进行了尝试,通过试验来验证两种方法的有效性,并对两种方法进行比较,从而提高近红外光谱小麦粉品质定量分析模型的精确性和可靠性。

7.3.1 基于马氏距离法的异常样品判别分析

这里采用马氏距离法提出异常样品,原理参见2.3.2节。

阈值范围参数设定为3,使用PLS方法建立校正模型判定剔除异常样品前后建立的校正模型对精确度的影响。图7-3所示为近红外光谱的马氏距离分布图,由图可知,马氏距离剔除的异常样品为3、5号。

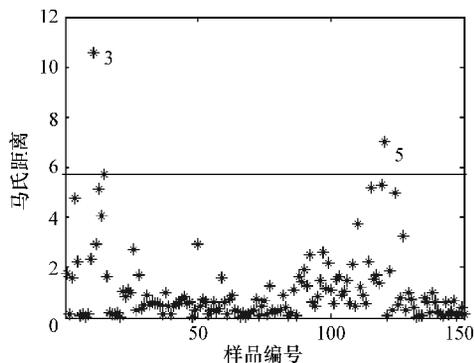


图7-3 马氏距离分布图

7.3.2 基于蒙特卡洛交叉验证法的异常样品判别分析

蒙特卡洛交叉验证(MCCV)法的异常样品筛选法是一种新近提出的筛选异常样品的方法,原理参见2.3.5节。

对150个小麦粉样品用MCCV法剔除异常样品,结果如下:

1) 经MCCV法计算,小麦粉灰分的均值-方差分布图如图7-4所示。

由图7-4可知,对于小麦粉的灰分这一品质参数而言应剔除的样品为8、79、147、148、149、150共6个样品。

2) 经MCCV法计算,小麦粉水分的均值-方差分布图如图7-5所示。

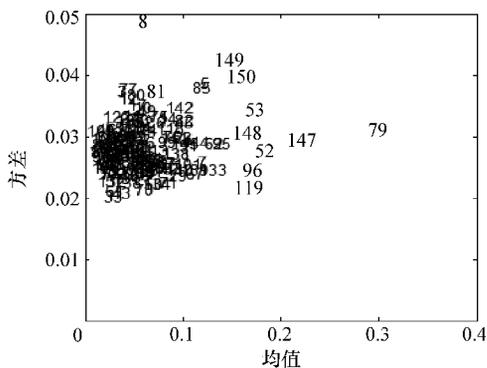


图7-4 灰分的均值-方差分布图

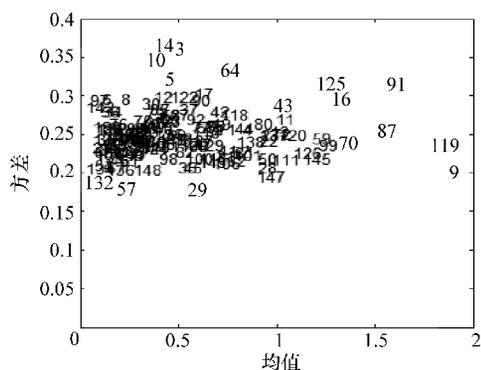


图7-5 水分的均值-方差分布图

由图7-5可知,对于小麦粉的水分这一品质参数而言应剔除的样品为9、16、87、91、119、125共6个样品。

3) 经MCCV法计算,小麦粉面筋的均值-方差分布如图7-6所示。

由图可知,对于小麦粉的面筋这一品质参数而言应剔除的样品为1、7、8、16、17、91共6个样品。

7.3.3 剔除异常样品后的模型评价

经过马氏距离的运算可知剔除的样品数为 2，剔除的样品为 3 号、5 号样品；经过 MCCV 法运算后剔除的样品数为 6，剔除的样品分别为 8 号、79 号、147 号、148 号、149 号、150 号（灰分），9 号、16 号、87 号、91 号、119 号、125 号（水分），1 号、7 号、8 号、16 号、17 号、91 号（面筋）。从剔除样品后的样品集中按照 4:1 浓度

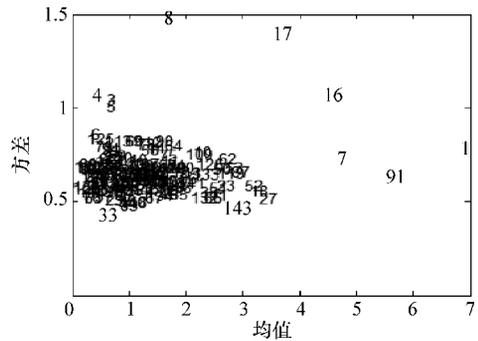


图 7-6 面筋的均值 - 方差分布图

法选择 30 个样品作为预测样品（保证预测集样品相同）来比较剔除异常样品前后模型的预测精度，这里采用典型的 PLS 方法进行建模，用剔除样品前后的 RMSECV 和 RMSEP 作为评价指标，试验结果见表 7-2。

表 7-2 剔除异常样品后的建模结果

小麦粉成分	剔除方法	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	未剔除	0.8424	0.0646	0.8455	0.0405
灰分	马氏距离法	0.8468	0.0629	0.8380	0.0414
	MCCV 法	0.9430	0.0363	0.8808	0.0355
	未剔除	0.7176	0.4456	0.7058	0.3931
水分	马氏距离法	0.7775	0.3818	0.7494	0.3628
	MCCV 法	0.8774	0.2761	0.8394	0.2900
	未剔除	0.6923	1.2370	0.8378	0.6699
面筋	马氏距离法	0.7879	0.9644	0.8308	0.6842
	MCCV 法	0.8453	0.7151	0.8487	0.6469

由表 7-2 可知，经过马氏距离和 MCCV 法两种算法筛选后，RMSECV 和 RMSEP 都有所减小，预测精确度得到了提高。可知，两种算法对剔除异常样品都具有一定的作用，为增强模型的稳定性提供了确实可行的方法。而由整体情况可知 MCCV 法相对马氏距离法剔除异常样品后 RMSECV 和 RMSEP 变化较大，更能提高模型的预测精确度，这可能是由于马氏距离剔除异常样品只需要对近红外光谱数据进行计算而不需要样品的化学值，剔除的样品可能是由于测量近红外光谱时人为误差造成光谱存在异常的情况，而 MCCV 法不仅需要光谱数据而且需要样品的化学值，并经过 2000 次的 PLS 建模交叉验证得到每个样品的均值 - 方差分布，从而得到剔除异常样品的结果，这种方法剔除异常样品后，不仅能降低由于人为误差造成的近红外光谱的异常情况，还可以避免由于人为误差或客观因素造成小麦粉品质参数真值测量误差的情况，所以得到的结果更为精确。

7.4 基于近红外光谱的小麦粉特征谱区筛选方法研究

由于近红外光谱仪所采集的光谱数据除样品自身的的信息外，还包含了其他噪声和无关信息，如样品背景、电噪声等，这些信息很难在光谱预处理中全部消除；其次由于某些区域样品的信息较弱，与样品的组成或性质间缺乏相关关系。如果把这些光谱数据都参与建模，不

但模型复杂、计算量大，而且精度也不一定高。近红外光谱分析一般采用线性模型，通过波长选择能够对自变量进行优选，一方面可以简化模型，更主要是通过剔除非线性或不相关变量，可以得到稳健性好、预测能力强的校正模型。

在近红外光谱结合 PLS 方法的建模中，传统观点认为 PLS 具有较强的抗干扰能力，可以全波长参与多元校正模型的建立。但是随着对 PLS 方法的深入研究和应用可知：通过特定方法进行特征波长或波长区间的筛选有可能得到更好的定量校正模型。目前，建立近红外光谱模型中常用的波长优选方法有：①波谱解析法，即凭借基团的性质来选定波长范围；②数学优化选择法，如无信息变量消除法（Uninformative Variable Elimination, UVE）、相关系数法、逐步回归法、遗传算法（Genetic Algorithm, GA）等；③有效信息区间定位法，即分别考察所划分的各个波长区间，挑选较优者，主要有间隔偏最小二乘法、移动窗口偏最小二乘法等。为了提高分析模型的精度，本研究尝试应用区间偏最小二乘法、联合区间偏最小二乘法、向后区间偏最小二乘法及 CARS 法进行谱区筛选和模型优化。

由于小麦粉的内部品质指标并不是以某一种化学成分存在，所以单独的一个小区间是不能提供足够的特征信息用来预测小麦粉的品质指标的。基于这种情况提出联合区间 PLS 法，通过联合几个子区间来建立预测模型，精确度会有所提高。

7.4.1 小麦粉样品集划分方法比较

近红外光谱分析必须借助相应的数学分析模型，而模型的建立又依赖于样品集的试验数据，试验样品被分为校正集和预测集两个部分。具体的样品分类要求做到：校正集的样品需要包括测量样品的全部背景信息，即所选样品要具有代表性、预测集的样品需要落在校正集样品空间内，只有这样才能对模型进行有效性检验。为了做到以上两点需要了解样品的分布情况来确定模型的适用范围，从而挑选校正集和预测集，使两个集合的样品都有代表性。Kennard - Stone (K - S) 法是一种均匀设计的方法，其核心思想是在试验空间内，根据不同样品之间的最大欧式距离来选择最具代表性的分布样品。本节对用 MCCV 方法剔除后的异常样品集采用 K - S 法进行划分，选择 120 个样品作为校正集，其余 30 个样品作为预测集。按 K - S 法挑选的校正集和预测集与 4:1 浓度法挑选的校正集与预测集作对比，以 R^2 、RMSECV 和 RMSEP 作为评价指标，对比结果见表 7-3。

表 7-3 挑选样品前后建模结果

小麦粉成分	挑选样品前后	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	4:1 浓度法	0.9430	0.0363	0.8808	0.0355
	K - S 法	0.9469	0.0357	0.9258	0.0379
水分	4:1 浓度法	0.8774	0.2761	0.8394	0.2900
	K - S 法	0.8799	0.2739	0.8616	0.2897
面筋	4:1 浓度法	0.8453	0.7151	0.8487	0.6469
	K - S 法	0.8295	0.7558	0.8867	0.5916

由表 7-3 可知，K - S 法挑选样品建立定量分析模型相比 4:1 浓度法，模型的预测精确度有所提高。

7.4.2 基于常规区间偏最小二乘法的小麦粉近红外光谱模型优化

应用常规区间偏最小二乘 (iPLS) 法对小麦粉的近红外光谱进行筛选时, 首先对 150 个小麦粉的原始光谱进行中心化处理, 划分多少个区间需要通过试验确定, 本节选择将整个光谱区域 $11995.67 \sim 3999.843\text{cm}^{-1}$ (共 2074 个波数点) 划分为 20 个子区间, 在划分出的各个子区间上进行 PLS 回归。以下以灰分为例说明, 图 7-7 所示为各个局部模型的 RMSECV 的比较图。横轴为波数点值, 纵轴为 RMSECV 的值, 条形图的高度代表每个子区间的 RMSECV 的值, 条形图内的数字表示局部模型的最佳主成分因子数, 由图 7-7 可知, 灰分的 iPLS 模型中, 最佳波段为第 20 个子区间, 最佳主成分因子数为 5。

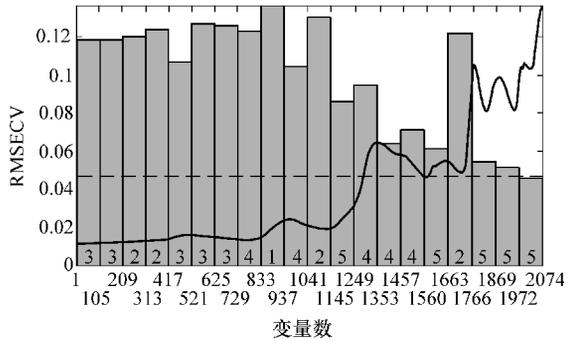


图 7-7 局部模型 RMSECV 比较图

对于小麦粉的灰分、水分、面筋 3 个品质参数 iPLS 法选择的最佳波段与最佳主成分因子数为: 灰分选择第 20 个子区间为最佳波段, 最佳主成分因子数为 5; 水分选择第 17 个子区间为最佳波段, 最佳主成分因子数为 4; 面筋选择第 19 个子区间为最佳波段, 最佳主成分因子数为 4。波段选取结果如图 7-8 所示。

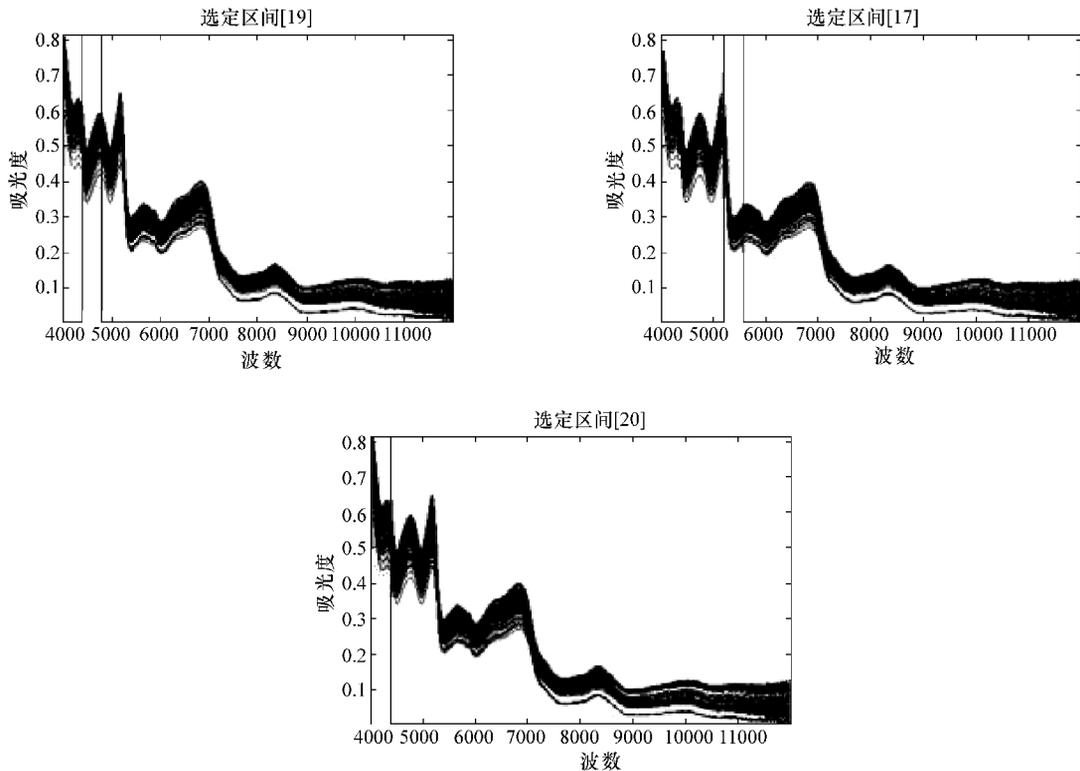


图 7-8 iPLS 法灰分、水分、面筋的波段选择图

对上一节按 K-S 法挑选的校正集与预测集进行建模，iPLS 法建模结果如图 7-9 ~ 图 7-11 所示。

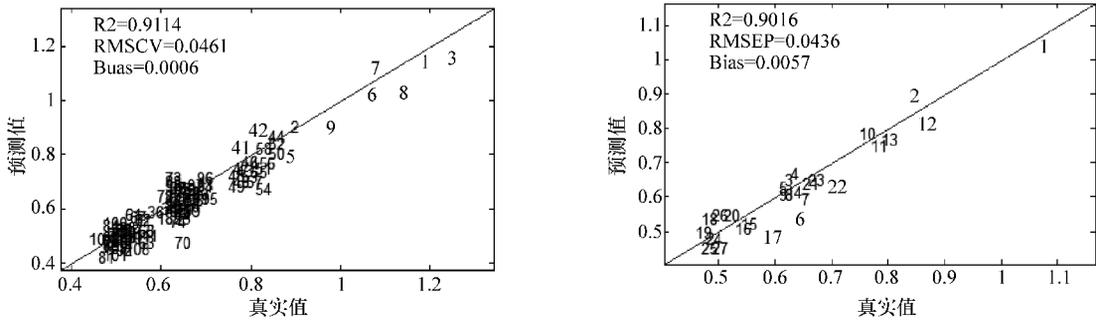


图 7-9 iPLS 法灰分的模型分析结果

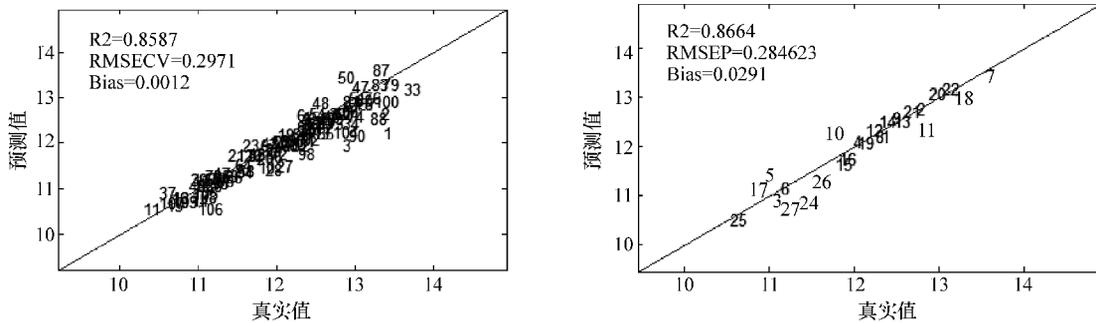


图 7-10 iPLS 法水分的模型分析结果

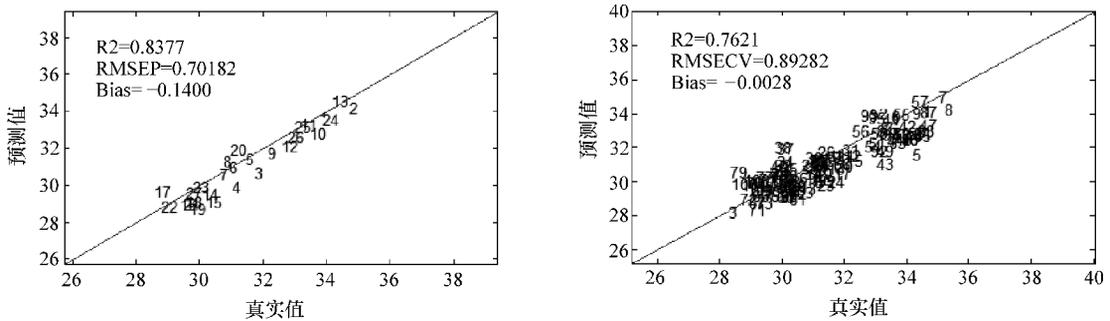


图 7-11 iPLS 法面筋的模型分析结果

7.4.3 基于联合区间最小二乘法的小麦粉近红外光谱模型优化

由于小麦粉的内部品质指标并不是以单一的化学成分存在，单独的一个小光谱子区间能不能提供足够的有效特征信息来预测小麦粉的品质指标。联合区间偏最小二乘 (SiPLS) 法就是在这种前提下提出的，联合几个子区间来建立预测模型，精确度会有所提高。首先对 150 个小麦粉的原始光谱进行中心化处理，将全光谱划分多少个区间及联合多少个子区间需要通过试验确定，本节选择将整个光谱区域 $11995.67 \sim 3999.843 \text{ cm}^{-1}$ (共 2074 个波数点) 划分为 20 个子区间，并联合 3 个子区间，确定主成分因子法与 iPLS 法相同，在各个联合区间上进行偏最小二乘回归。

对于小麦粉的灰分、水分、面筋 3 个品质参数 SiPLS 法选择的最佳联合波段与最佳主成分因子数为：灰分选择第 9、10、16 子区间为最佳联合波段，最佳主成分因子数为 7；水分

近红外光谱技术在食品品质检测方法中的研究

选择第 9、13、14 子区间为最佳联合波段，最佳主成分因子数为 6；面筋选择第 8、17、20 子区间为最佳联合波段，最佳主成分因子数为 6。波段选取结果如图 7-12 所示。

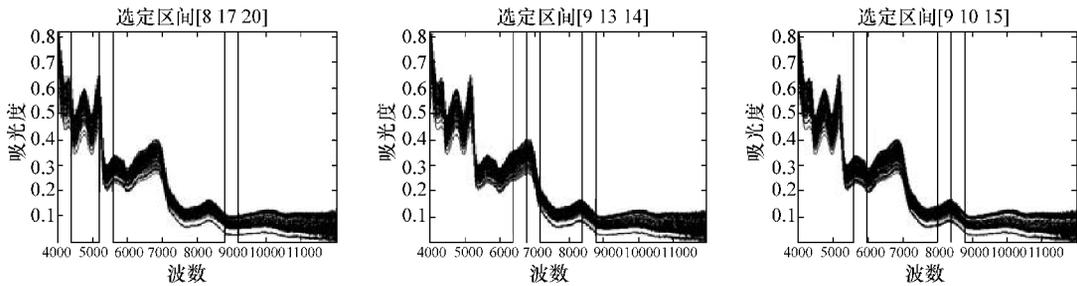


图 7-12 SiPLS 法灰分、水分、面筋的波段选择图

对上一节按 K-S 法挑选的校正集与预测集进行建模，SiPLS 法建模结果如图 7-13 ~ 图 7-15 所示。

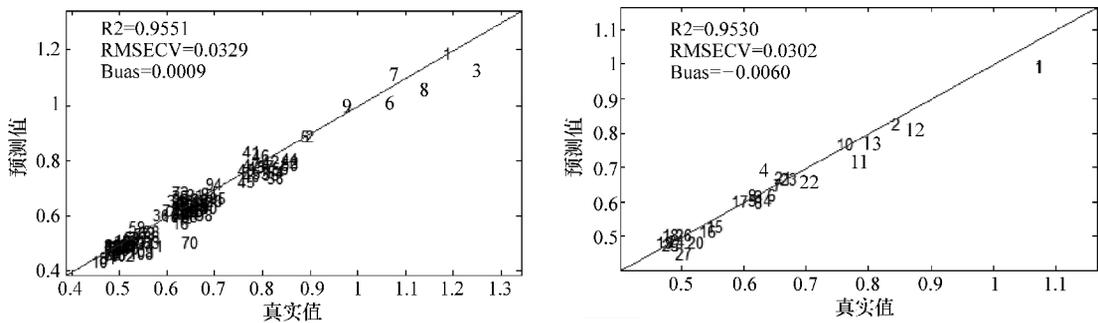


图 7-13 SiPLS 法灰分的模型分析结果

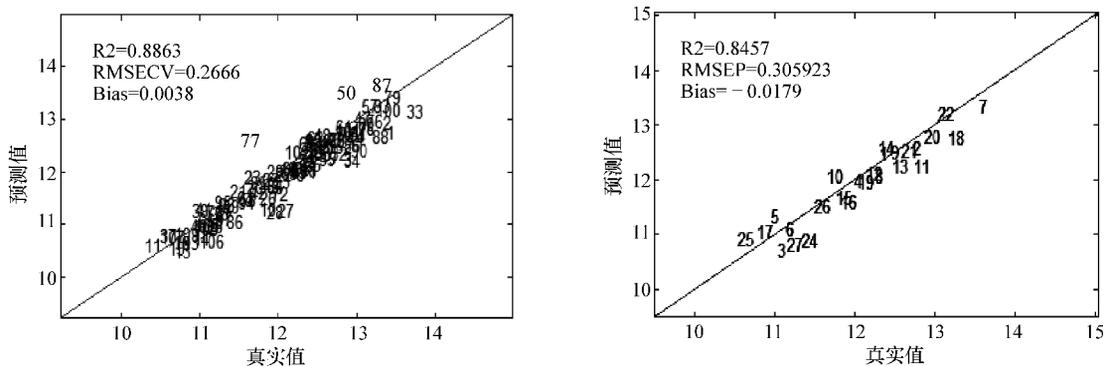


图 7-14 SiPLS 法水分的模型分析结果

7.4.4 基于向后区间最小二乘法的小麦粉近红外光谱模型优化

由于目前尚无方法来确定参加联合建模的子区间数目，所以在 SiPLS 法的基础上提出了向后阈值区间偏最小二乘 (BiPLS) 法，BiPLS 法采用的仍然是几个子区间联合建模的方法，但其区间的搜索方法继承了向后选择变量法“只出不进”的特点，因此可以确定联合模型的建模区间数，同时 BiPLS 法能在不同的区间划分中搜寻到最佳的联合区间。对 150 个小麦粉的原始光谱进行中心化处理，将全光谱划分多少个子区间需要通过试验确定，区间数

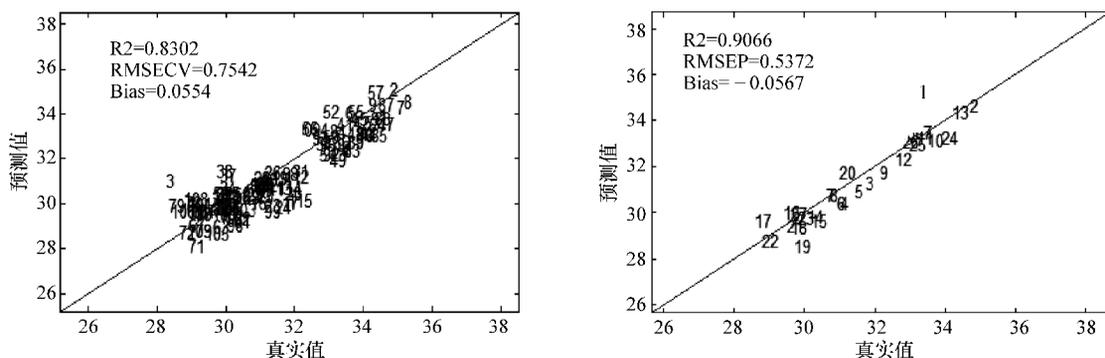


图 7-15 SiPLS 法面筋的模型分析结果

越多运算速度越慢，结合精度和运算速度综合考虑，本节选择将整个光谱区域 $3999.843 \sim 11995.67\text{cm}^{-1}$ （共 2074 个波数点）划分为 20 个子区间，确定主成分因子法与上述方法相同，在各个联合区间上进行 PLS 回归。

对于小麦粉的灰分、水分、面筋 3 个品质参数 BiPLS 法选择的最佳联合波段与最佳主成分因子数为：灰分选择第 6、8、11、12、16、17、19、20 子区间为最佳联合波段，最佳主成分因子数为 7；水分选择第 2、3、13、14、17 子区间为最佳联合波段，最佳主成分因子数为 6；面筋选择第 8、17、20 子区间为最佳联合波段，最佳主成分因子数为 6。

对上一节按 K-S 法挑选的校正集与预测集进行建模，BiPLS 法建模结果如图 7-16 ~ 图 7-19 所示。

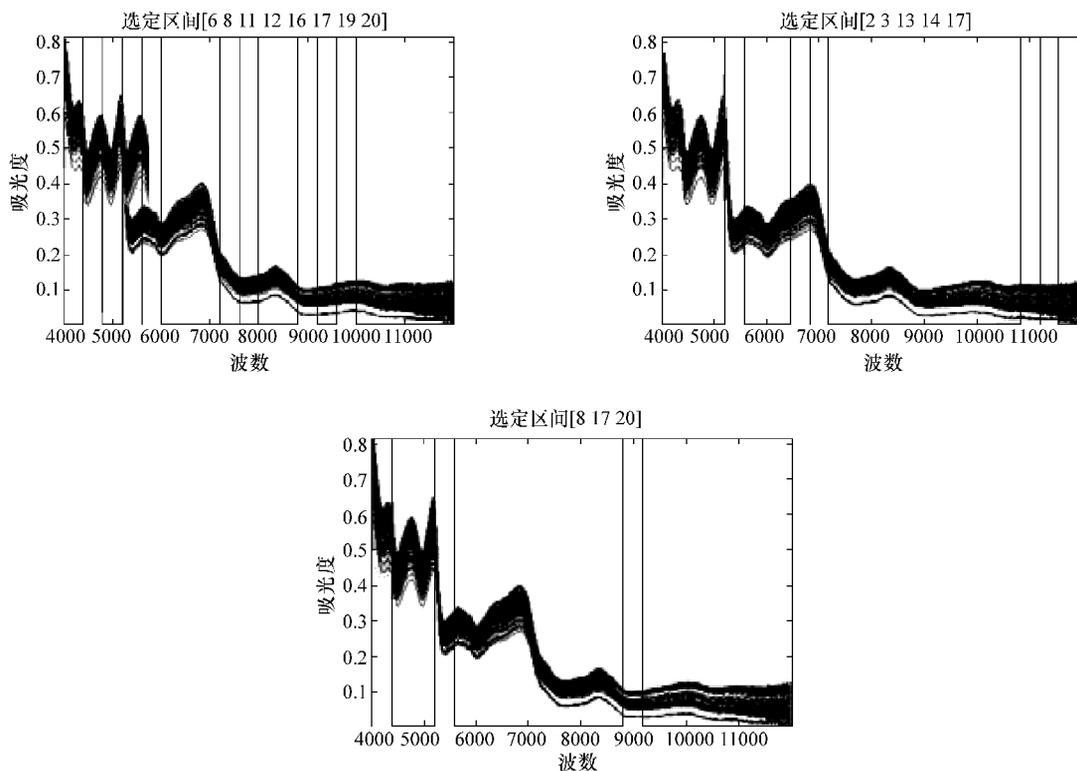


图 7-16 BiPLS 法灰分、水分、面筋的波段选择图

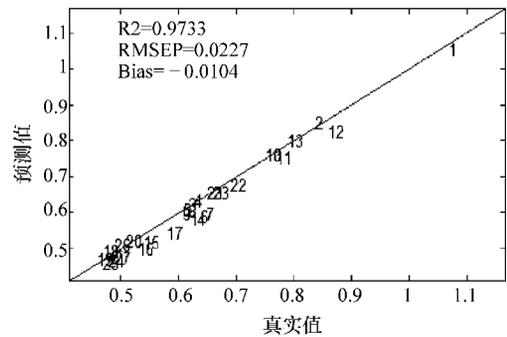
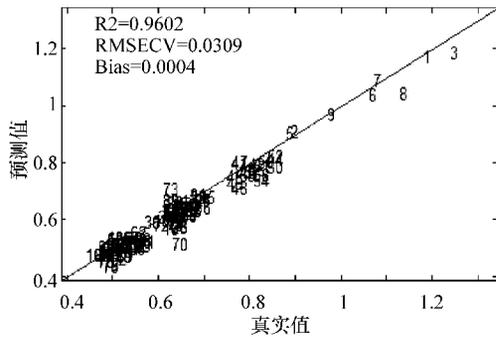


图 7-17 BiPLS 法灰分的模型分析结果

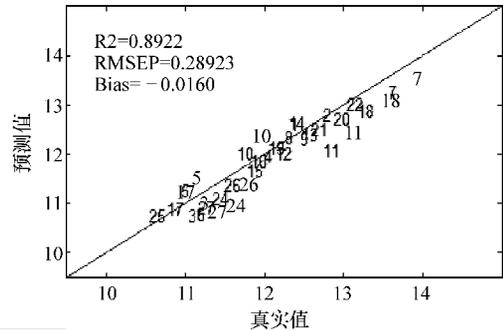
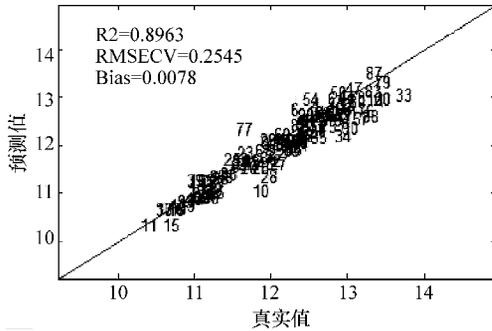


图 7-18 BiPLS 法水分的模型分析结果

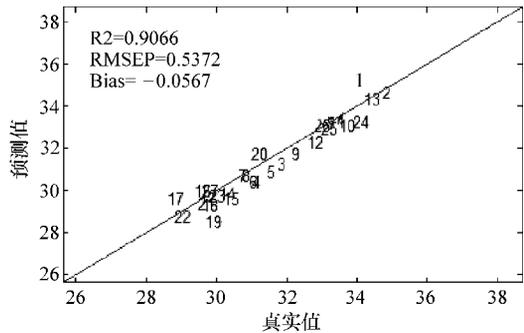
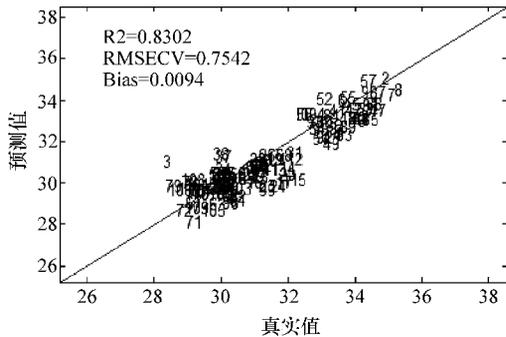


图 7-19 BiPLS 法面筋的模型分析结果

7.4.5 基于 CARS 谱区筛选法的小麦粉近红外光谱模型优化

CARS 算法每次通过 ARS 技术筛选出回归模型中回归系数绝对值较大的波长点，去掉权重较小的波长点，利用交互检验的方法选出 RMSECV 值最低的子集，筛选对象为各个波长点，所以能解决 BiPLS 方法尚无方法来确定子区间数目的困扰，并且可有效选择与所测性质相关的最优波长组合。

下面以灰分为例说明变量的筛选过程，筛选过程如图 7-20 所示。

图 7-20a 表示在筛选过程中被选中变量数目的变化趋势，变量数随着运行次数的增加由快到慢递减。保留的变量数目和运行次数呈指数关系。

图 7-20b 表示随着运行次数的变化 RMSECV 的变化情况，在 1~26 次 RMSECV 呈现递减趋势，证明筛选过程剔除的变量为与样品性质无关的变量，26 次后 RMSECV 开始递增是

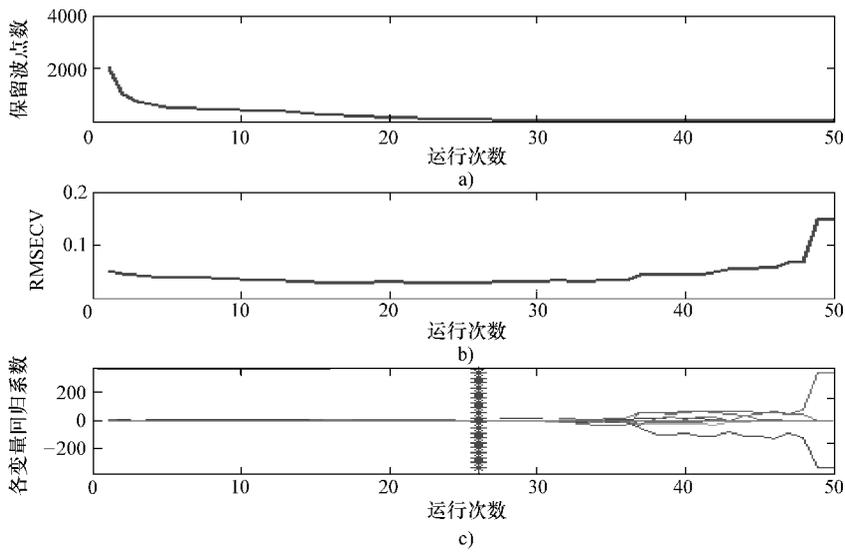


图 7-20 灰分光谱波长变量筛选图

a) 保留波点数 b) RMSECV c) 各波长变量回归系数

由于剔除了关键变量，导致 RMSCEV 增大。

图 7-20c 中“*”所对应的点为 RMSECV 最低的点，各线表示随着运行次数的增加各个变量回归系数的变化情况。

经过 CARS 法变量筛选，保留波点数为 60，利用所选的波长建立 PLS 回归模型，通过校验集进行校验。建模结果如图 7-21 所示。

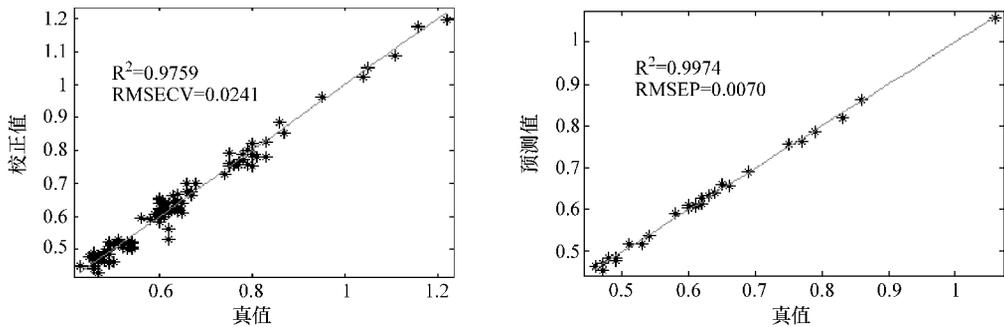


图 7-21 小麦粉灰分的校验结果

以上以灰分为例说明 CARS 谱区筛选法，对小麦粉的水分、面筋也重复以上试验，可得校验结果见表 7-4。

表 7-4 特征谱区筛选后建模结果比较

小麦粉成分	保留波点数	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	60	0.9759	0.0241	0.9974	0.0070
水分	69	0.9792	0.1141	0.9971	0.0419
面筋	52	0.9356	0.4645	0.9940	0.1358

近红外光谱技术在食品品质检测方法中的研究

由表 7-4 可知, 经过波长筛选模型质量有显著提高, 与以上 3 种方法相比降低了运算的波点数, 提高了运算速度。

7.4.6 特征谱区筛选法的比较

由表 7-5 可知, 常规区间偏最小二乘法 iPLS 把整个光谱区域分为 20 个子区间, 选择了最优的区间, 能够减少计算量, 但是由于小麦粉的内部品质指标并不是以某种纯的化学成分存在, 单独的一个小区间不能提供足够的信息来预测小麦粉的品质指标。所以建模结果与全光谱建模结果相比精确度降低, 所以在 iPLS 模型的基础上提出 SiPLS 法, 联合几个子区间来建立预测模型, 精确度相对 PLS 和 iPLS 方法都有所提高, 但是由于目前尚无方法来确定参加联合建模的子区间数目, 所以在 SiPLS 算法的基础上提出了 BiPLS 法, BiPLS 法采用的仍然是几个子区间联合建模的方法, 但其区间的搜索方法继承了向后选择变量法“只出不进”的特点, 因此可以确定联合模型的建模区间数, 同时 BiPLS 法能在不同的区间划分中搜寻到最佳的联合区间, 模型精确度相对 SiPLS 方法都有所提高。由于 BiPLS 还是需要确定子区间数目, 有可能联合的每个子区间内有部分波点属于无效信息。而 CARS 逐一筛查选取了信息量较大、影响权重较大的波数点, 建立了 PLS 模型, 模型预测结果显著优于基于区间选择建立的 PLS 模型。

表 7-5 特征谱区筛选后建模结果比较

小麦粉成分	波长筛选方法	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	PLS	0.9469	0.0357	0.9258	0.0379
	iPLS	0.9414	0.0461	0.9016	0.0436
	SiPLS	0.9551	0.0329	0.9530	0.0302
	BiPLS	0.9558	0.0326	0.9604	0.0277
	CARS - PLS	0.9759	0.0241	0.9974	0.0070
水分	PLS	0.8799	0.2739	0.8616	0.2892
	iPLS	0.8587	0.2971	0.8664	0.2846
	SiPLS	0.8863	0.2666	0.8457	0.3059
	BiPLS	0.8963	0.2545	0.8622	0.2891
	CARS - PLS	0.9792	0.1141	0.9971	0.0419
面筋	PLS	0.8295	0.7558	0.8867	0.5916
	iPLS	0.7621	0.8928	0.8377	0.7082
	SiPLS	0.8302	0.7542	0.9066	0.5372
	BiPLS	0.8302	0.7542	0.9066	0.5372
	CARS - PLS	0.9356	0.4645	0.9940	0.1358

7.5 SVM 和 ANN BP 在小麦粉品质定量分析中的应用研究

近红外光谱的定量分析方法主要有两大类: 线性和非线性。其中应用最为广泛的线性多

元校正方法为 PLS 法, 关于这种方法已经在前面章节中介绍, 并且应用到小麦粉品质检测的定量分析中。非线性多元校正方法有人工神经网络 (ANN), 以及支持向量机 (Support Vector Machine, SVM) 等方法。ANN 既可解决线性问题, 也可解决非线性问题。由于 ANN 有较强的非线性映射能力, 在非线性模型的处理上有很强的优势, 目前已经成为化学计量学常用于非线性体系的校正方法, 并迅速应用于近红外光谱分析技术的定性和定量分析中。SVM 目前已经成功地推广到了信息融合、函数逼近等领域。SVM 回归是基于最小化结构风险, 而不是传统意义上的经验风险最小化, 建模的基本思想是通过一个非线性映射, 将数据样品映射到高维特征空间, 并在这个空间进行线性回归, 模型结构简单, 具有良好的应用前景。目前, SVM 应用于实际较少, 大多数研究仅仅局限于理论和仿真。因此将理论应用于解决实际问题具有重要的意义。

本节利用 ANN 和 SVM 回归法应用于近红外光谱定量分析小麦粉品质含量检测, 分别建立定量分析模型, 并且比较所建模型的稳定性、预测能力等方面的优缺点。研究分别应用 SVM 和 AVV BP 两种非线性方法来建立小麦粉品质的定量分析模型。

7.5.1 小麦粉近红外光谱的主成分分析

小麦粉中许多有机物的含氢基团都能在近红外区域产生倍频与合频的吸收, 并且小麦粉样品的近红外光谱数据间也存在大量的相关性, 会造成大量的信息冗余。把原始的近光谱数据直接作为非线性模型的输入, 会使模型产生过拟合现象。主成分分析 (Principle Component Analysis, PCA) 可以把多个指标化为几个综合性指标, 被广泛应用到各领域。在实际问题中, 研究多指标 (变量) 问题时, 不同指标之间具有一定相关性, 由于多指标及指标之间有相关性, 从而增加了分析问题的复杂性。PCA 就是将原来具有一定相关性的指标, 组合成的一组新的相互无关的综合指标来代替原来的指标。对近红外光谱进行 PCA, 不仅能够达到对光谱矩阵进行降维, 而且不会过多地丢失光谱信息, 同时还可以减少原始近红外光谱的冗余信息。

试验中每个样品的光谱数据点为 2074, 如果把原始的光谱矩阵作为模型的输入, 训练复杂, 运算速度慢, 还可能会出现过拟合现象, 影响模型的稳定性和预测精确度。因此, 需要对原始的光谱矩阵做降维处理。本节中用 PCA 将原始的光谱矩阵进行分解, 取得累积贡献率达到 99.5% 的主成分来作为非线性模型的输入。小麦粉 3 个品质参数的前 5 个主成分的贡献率见表 7-6。

表 7-6 特征谱区筛选后建模结果比较

小麦粉成分	PC1	PC2	PC3	PC4	PC5
灰分	80.3370	18.8587	0.4083	0.1998	0.0329
水分	79.4227	19.7165	0.4245	0.2589	0.0344
面筋	80.5512	18.6292	0.4225	0.2272	0.0306

7.5.2 基于近红外光谱的小麦粉 ANN BP 模型

3 层的 ANN BP 即可满足大多数问题的要求, 因此本节试验建立 3 层的 ANN BP 模型。由于训练函数自身特点不同, 针对特定要求建模的适用性也不同, 需多次试验选择出最优训

近红外光谱技术在食品品质检测方法中的研究

练函数对 ANN 进行训练，本节采用正切 Sigmoid 函数作为隐含层的传递函数，输出层采用 Purelin 型函数。

ANN BP 通过校正集进行建模，通过预测集来评价所建模型的准确性，用 R^2 、RMSECV 和 RMSEP 作为网络训练的评价指标。

模型的稳定性是 ANN BP 在近红外光谱分析应用中最重要的问题之一。应该尽量减少网络的节点数，从而提高模型的稳健性。通过主成分分析可知，灰分、水分、面筋的前 3 个主成分已经包含了光谱 99.5% 以上的有效信息。因此，将各参数的前 3 个特征变量作为 ANN BP 的输入变量，即输入层节点数为 3，经多次训练确定隐含层节点数为 8，输出层节点数为 1。经反复训练确定学习速率、目标误差、网络学习次数和动量因子分别为 0.15、0.0001、1000 和 0.70。小麦粉品质的 ANN BP 预测结果如图 7-22 ~ 图 7-24 所示，ANN BP 建模结果见表 7-7。

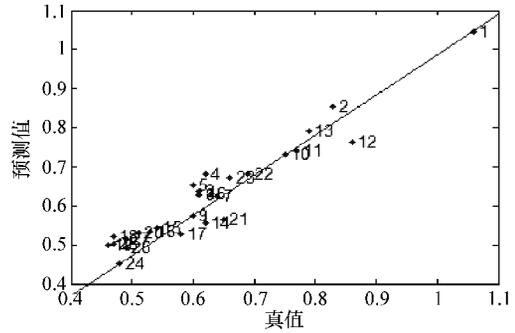


图 7-22 灰分的 ANN BP 预测结果

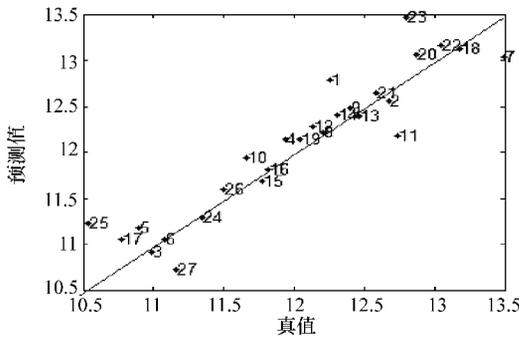


图 7-23 水分的 ANN BP 预测结果

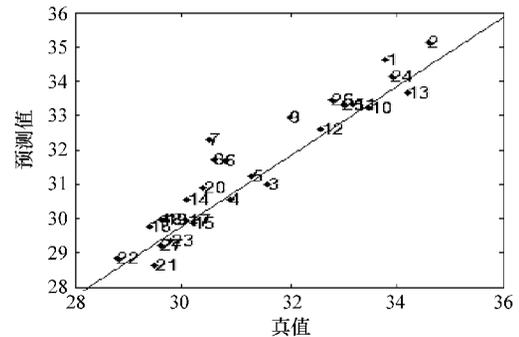


图 7-24 面筋的 ANN BP 预测结果

表 7-7 ANN BP 建模结果

小麦粉成分	主成分数	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	3	0.9359	0.0393	0.9203	0.0393
水分	3	0.9136	0.2323	0.8585	0.2930
面筋	3	0.8361	0.7410	0.8396	0.7040

7.5.3 基于近红外光谱的小麦粉支持向量回归模型

为了减少运算量，提高运算速度，本节采用 PCA 法对原始的近红外光谱矩阵进行主成分分解，取累计贡献率达到 99.5% 的主成分作为 SVM 模型输入的特征子集。惩罚系数 c 、不敏感系数 g 、核函数及相关参数的选择，对 SVM 的效果有显著的影响。K-fold Cross Validation (K-CV) 是把原始的数据分为 K 组，将其中一组作为一个验证集，同时将其余 $K-1$ 组作为训练集，得到 K 个模型，用这 K 个模型最终验证集的准确率的平均数作为 K-CV 的

性能指标, K - CV 可以有效地避免过学习和欠学习状态的发生, 使用 K - CV 选择参数使 SVM 模型得到较好的预测结果; CV 是在没有先验知识指导的情况下一般选用径向基函数 (RBF) 为核函数, RBF 可以将非线性的样品数据映射到高维特征空间, 降低模型的复杂性, 提高训练速度。本节 SVM 模型参数的选取 RBF 为核函数, 并采用交叉验证算法来确定相关参数。

SVM 在 MATLAB R2008a 的运行环境中使用 libsvm 工具箱实现。选取能够达到最高验证分类准确率的参数 c 和 g , 本节先在大范围内寻找 c 和 g , 再利用 SVMcgForClass 进行精细的参数选择。以下以灰分为例说明寻找回归最佳参数的过程, 首先进行粗略的寻找, 如图7-25所示。观察粗略寻找的结果后再进行精细选择, 如图7-26所示。

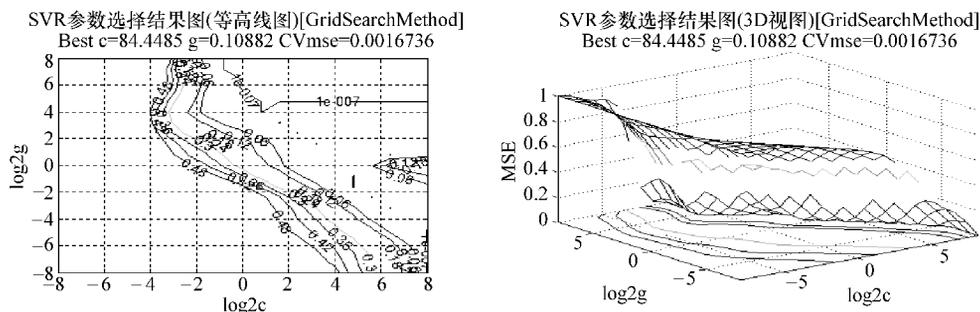


图 7-25 参数粗略选择结果

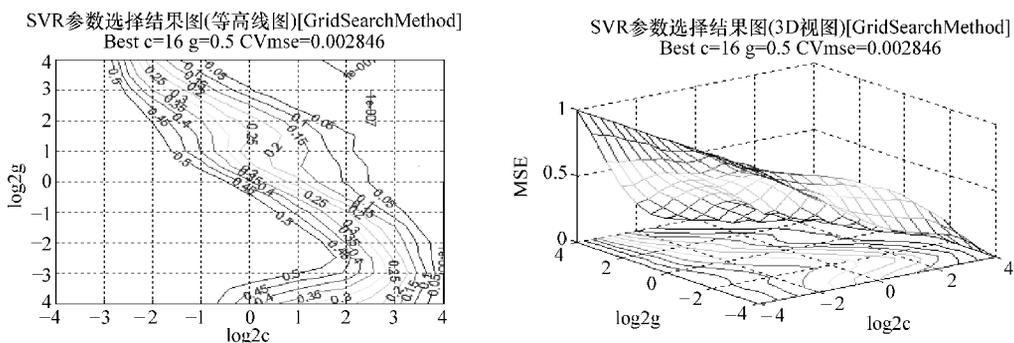


图 7-26 参数精细选择结果

图7-25、图7-26的意义是 x 轴表示 c 取以 2 为底的对数后的值, y 轴表示 g 取以 2 为底的对数后的值, 等高线表示取相应的 c 和 g 后对应的 K - CV 方法的准确率。

SVM 模型的预测结果, 如图7-27 ~ 图7-29所示。

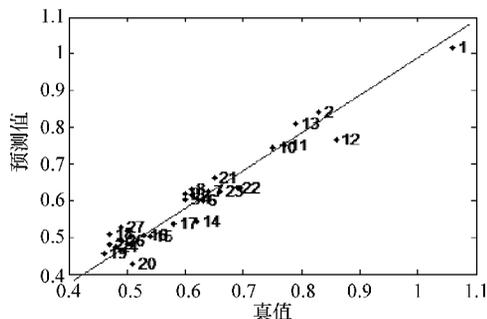


图 7-27 灰分的 SVM 预测结果

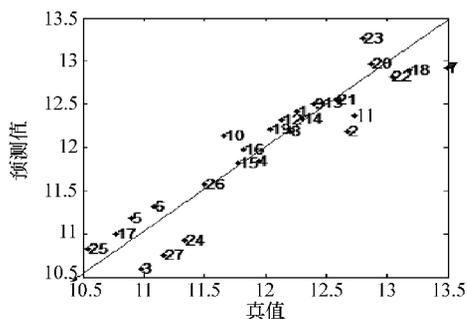


图 7-28 水分的 SVM 预测结果

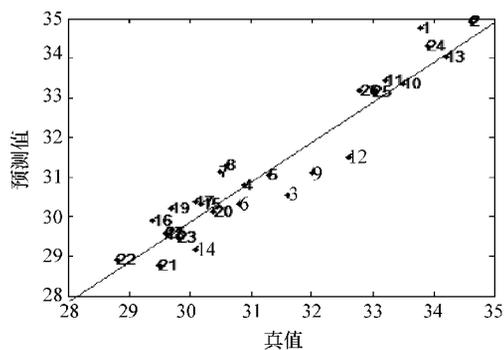


图 7-29 面筋的 SVM 预测结果

7.5.4 小麦粉模型比较分析

3 种建模方法的建模结果见表 7-8，由表 7-8 可知，ANN BP 建模的 RMSECV 和 RMSEP 相对 PLS 方法有所降低，但是在训练过程中发现 ANN BP 很不稳定，会出现每次输出的结果不相同的情况，而 SVM 法建模 RMSECV 和 RMSEP 相对前两种方法都会减小，而且输出结果稳定。

表 7-8 建模结果比较

小麦粉成分	建模方法	内部交叉 检验 R^2	内部交叉检验 标准差 RMSECV	预测 R^2	预测标准差 RMSEP
灰分	PLS	0.9469	0.0357	0.9258	0.0379
	BP	0.9551	0.0329	0.9530	0.0302
	SVM	0.9759	0.0241	0.9974	0.0070
水分	PLS	0.8799	0.2739	0.8616	0.2892
	BP	0.9136	0.2323	0.8585	0.2930
	SVM	0.9241	0.2205	0.8659	0.2858
面筋	PLS	0.8295	0.7558	0.8867	0.5916
	BP	0.8623	0.6791	0.8669	0.6210
	SVM	0.8714	0.6563	0.8972	0.5457

分析以上结论的原因，统计学理论表明学习机器的实际风险由经验风险（即训练误差）和置信范围两部分组成。BP 网络算法是遵循经验风险最小化的准则，该准则不能在顾及经验风险的同时兼顾置信范围，在样品有限的条件下，训练误差小，但并不是总能得到较好的预测效果。SVM 学习算法是遵循结构风险最小化的准则，该准则在顾及经验风险的情况下能够兼顾置信范围，其原理是通过把函数集构造成为一个函数子集序列，并在每个子集中找到最小的经验风险，在每个子集间综合考虑经验风险和置信范围，使取得实际的风险最小。SVM 模型是基于结构风险最小化的原则建立的模型，它同时兼顾了置信范围和经验风险这两个最小化风险的泛函问题，而训练得到的结果通常为全局最优解。所以，SVM 模型好于 PLS、ANN BP 两种模型。

7.6 小结

为了解决传统小麦粉检测时间长、操作复杂等难题,研究提出基于近红外光谱检测小麦粉品质的研究思路。首先利用近红外光谱分析技术结合 PLS 法定量分析小麦粉中水分、灰分及面筋 3 种有效成分。然后,为了排除异常样品的干扰,利用马氏距离法和蒙特卡洛采样法剔除异常样品来提高模型进度。之后采用 K-S 法对样品集进行划分,为了提高模型的精确度,尝试利用 iPLS 法、SiPLS 法、BiPLS 法和 CARS 法进行特征谱区筛选,从而对小麦粉有效成分的定量分析模型进行优化。最后,引入非线性回归方法(SVM 和 ANN BP)对模型的稳定性及预测能力进行了比较分析。

参考文献

- [1] 姚惠源. 我国面粉加工业的发展方向 [J]. 粮食加工, 2007, 32 (6): 7-9.
- [2] 王玉庭. 中国小麦消费现状及趋势分析 [J]. 中国食物与营养, 2010, 5: 47-50.
- [3] 刘元彪. 小麦粉品质关键指标及相关评价 [J]. 西部粮油科技, 1998, 23 (1): 39.
- [4] 孙辉, 尹成华, 赵仁勇, 等. 我国小麦品质评价与检验技术的发展现状 [J]. 粮油工程, 2010, 17 (5): 16-17.
- [5] 白剑侠, 赵欣. 制粉厂的生产检验 [J]. 技术粮食工程, 2010, 52 (3): 53-54.
- [6] 刘景沿, 彭黔荣, 鲍峰伟, 等. 主成分分析-马氏距离法优选近红外模型定标样品 [C]. 北京: 中国石化出版社, 2006, 10 (1): 322-326.
- [7] 陈斌, 邹贤勇, 朱文静. PCA 结合马氏距离法剔除近红外异常样品 [J]. 江苏大学学报, 2008, 29 (4): 277-279.
- [8] 邹小波, 赵杰文. 农产品无损检测技术与数据分析 [M]. 北京: 中国轻工业出版社, 2007.
- [9] Zou Xiaobo, Zhao Jiewen, Huang Xingyi, et al. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji', apple based on different PLS models [J]. Chemometrics and Intelligent Laboratory Systems, 2007, (87): 43-51.
- [10] Zou Xiaobo, Zhao Jiewen, Li Yanxiao. Selection of the efficient wavelength regions in FY-NIR spectroscopy for determination of SSC of 'Fuji' apple based on BiPLS and FiPLS models [J]. Vibrational Spectroscopy, 2007, 44: 220-227.
- [11] 连承波, 赵永军, 李汉林, 等. 基于支持向量机回归的煤层含气量预测 [J]. 西安科技大学学报, 2008, 28 (4): 707-710.
- [12] 张灵帅, 王卫东, 谷运红, 等. 近红外光谱的主成分分析-马氏距离聚类判别用于卷烟的真伪鉴别 [J]. 光谱学与光谱分析, 2011, 31 (5): 1254-1257.
- [13] 罗竞佳. 基于 PCA/ICA 的多数据流关联及模式发现 [D]. 大连: 大连理工大学, 2009.
- [14] 张玉荣, 付玲, 周显青. 基于 BP 神经网络小麦含水量的近红外检测方法 [J]. 河南工业大学学报 (自然科学版), 2013, 34 (1): 17-20.
- [15] Chen L, Huang J F, Wang F M, et al. Comparison between back propagation neural network and regression models for the estimation of pigment content in rice leaves and panicles using hyperspectral data [J]. International Journal Of Remote Sensing, 2007, 28 (16): 3457-3478.

基于近红外光谱的淀粉品质检测方法研究

8.1 简介

淀粉是以谷类、薯类、豆类为原料,不经过任何化学方法处理,也不改变淀粉内在的物理和化学特性加工而成的。它是日常生活中必不可少的作料之一,如煎炸烹炒,做汤勾芡都少不了要用到淀粉。随着食用淀粉在现代食品加工工业中的广泛应用,淀粉生产和加工贸易取得了较大的发展。常见的产品主要有玉米淀粉、马铃薯淀粉、红薯淀粉和绿豆淀粉等,不同种类的淀粉价格差别较大,有的相差高达 10 倍以上,但是不同种类淀粉颗粒的宏观外观和普通物化指标差别不明显,无法辨认。由于缺乏相应的食用淀粉鉴别检验技术标准,国内淀粉市场严格监管很难执行。一些不法商家为追求利润,常把廉价的、大量的玉米淀粉包装成马铃薯淀粉或绿豆淀粉来销售,从而提高销售价格,从中牟取暴利。为保护合法生产经营者和消费者的利益,有必要进行淀粉种类及淀粉掺杂的鉴别。传统的感官评定方法一方面需要经验,另一方面其检测结果受主观因素影响较大,精确度难以保证。而常规的理化分析方法不仅费时费力,还破坏样品。因此有必要研究一种简单、快速、无损的淀粉品质鉴别技术。

近红外光谱分析技术是 20 世纪 90 年代以来发展最快、最引人注目的光谱分析技术,它以其快速、高效、取样简单以及无污染等独特的分析优点,已广泛应用于农业、食品、医药、石油、化工、烟草、化妆品等行业。在食品分析方面,近红外光谱技术已被应用于苹果、水稻、酸奶和黄酒等的品种鉴别,而对于淀粉品质的快速检测还少见报导。

8.2 基于聚类分析的淀粉种类定性识别方法研究

8.2.1 样品制备

实验用玉米淀粉和土豆淀粉均是从市场购买的不同品牌或同一品牌不同批次的淀粉,将其编号,1~26 号为土豆淀粉样品,27~58 号为玉米淀粉样品。

8.2.2 光谱采集

采用德国 Bruker 公司傅里叶变换近红外光谱仪,波数范围 $12500 \sim 4000\text{cm}^{-1}$,波长间隔 8cm^{-1} ,扫描 64 次后取平均,环境温度 $23 \sim 25^\circ\text{C}$ 。漫反射样品杯附件。将上述淀粉样品放置在旋转样品台的样品杯中,进行近红外光谱采集。

26 个土豆淀粉样品和 32 个玉米淀粉样品的近红外漫反射光谱如图 8-1 所示。由图可以看出,在 $12500 \sim 4000\text{cm}^{-1}$ 范围内较为相似,具有许多淀粉之间的相似信息,峰形、峰位差别很小,无法直接鉴别。利用化学计量学方法将原光谱进行数学处理,采用聚类算法进行鉴别,能突出样品之间化学成分的细小差异,从而达到鉴别的目的。

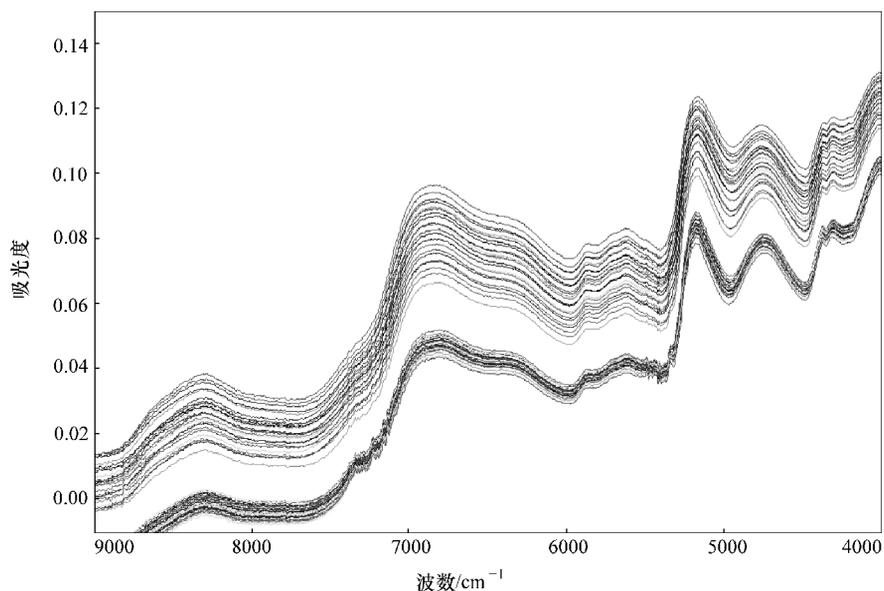


图 8-1 58 个样品的近红外漫反射光谱图

8.2.3 结果与分析

运用 OPUS 中的聚类分析功能选项, 从土豆淀粉中选择 22 个样品的原始光谱用作建模, 土豆淀粉中的 6 号、11 号、17 号和 23 号用作预测。从玉米淀粉中选择 28 个样品的原始光谱用作建模, 玉米淀粉中的 30 号、36 号、41 号和 50 号用作预测, 进行聚类分析。光谱预处理方法选择矢量归一法, 光谱范围选择 $9000 \sim 4000 \text{cm}^{-1}$ 。聚类后结果如图 8-2 所示, 从图中可以看出, 聚类结果与实际样品分类情况完全一样, 精确度 100%。

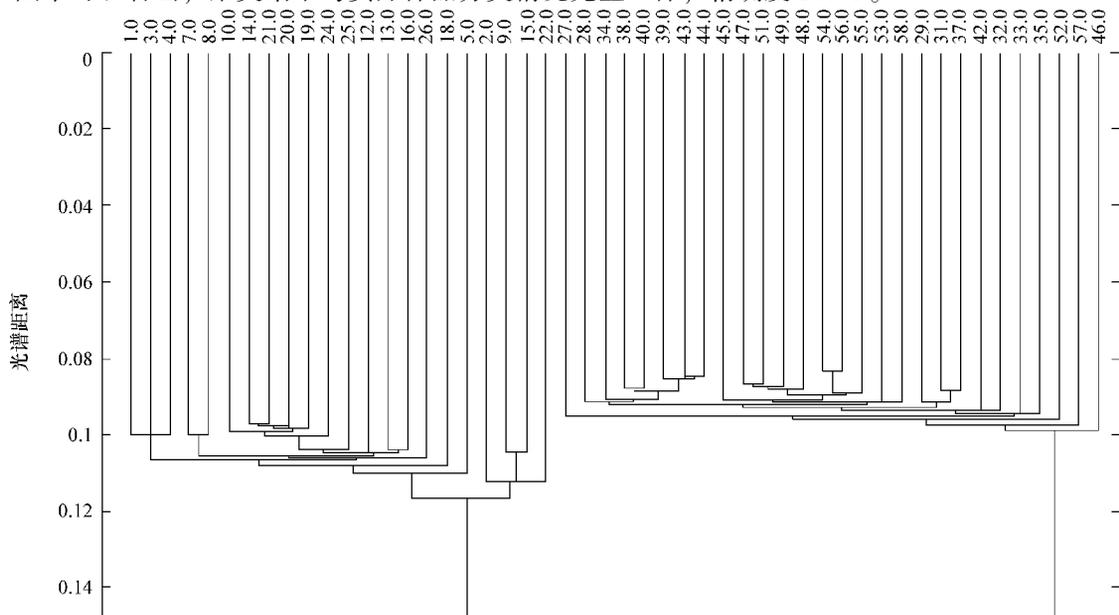


图 8-2 聚类分析结果

近红外光谱技术在食品品质检测方法中的研究

为检验聚类分析模型的预测能力，考察模型对预测集样品的精确度，用该模型分别对 6 号、11 号、17 号、23 号土豆淀粉样品和 30 号、36 号、41 号和 50 号玉米淀粉进行验证，验证结果如图 8-3 所示。从预测的结果可以看到预测精确度达到 100%。

研究采用近红外光谱结合聚类分析法进行了土豆淀粉和玉米淀粉的鉴别，鉴别结果精确度达到 100%，此方法比常规鉴别法简便、准确，且更具有科学性，从而为淀粉种类的鉴定提供了一种新的方法和手段。由于本实验研究的淀粉种类、品牌、批次以及产地有限，无法完全代表淀粉种类的多样性，因此需要进一步收集扩充淀粉样品的种类、品牌、批次以及产地，考察淀粉的多样性和鉴别的精确度，使该方法成为一种淀粉种类鉴别的可靠手段。应用近红外光谱技术快速鉴别淀粉种类，这是一项符合我国食品市场管理的极有应用前景的技术。

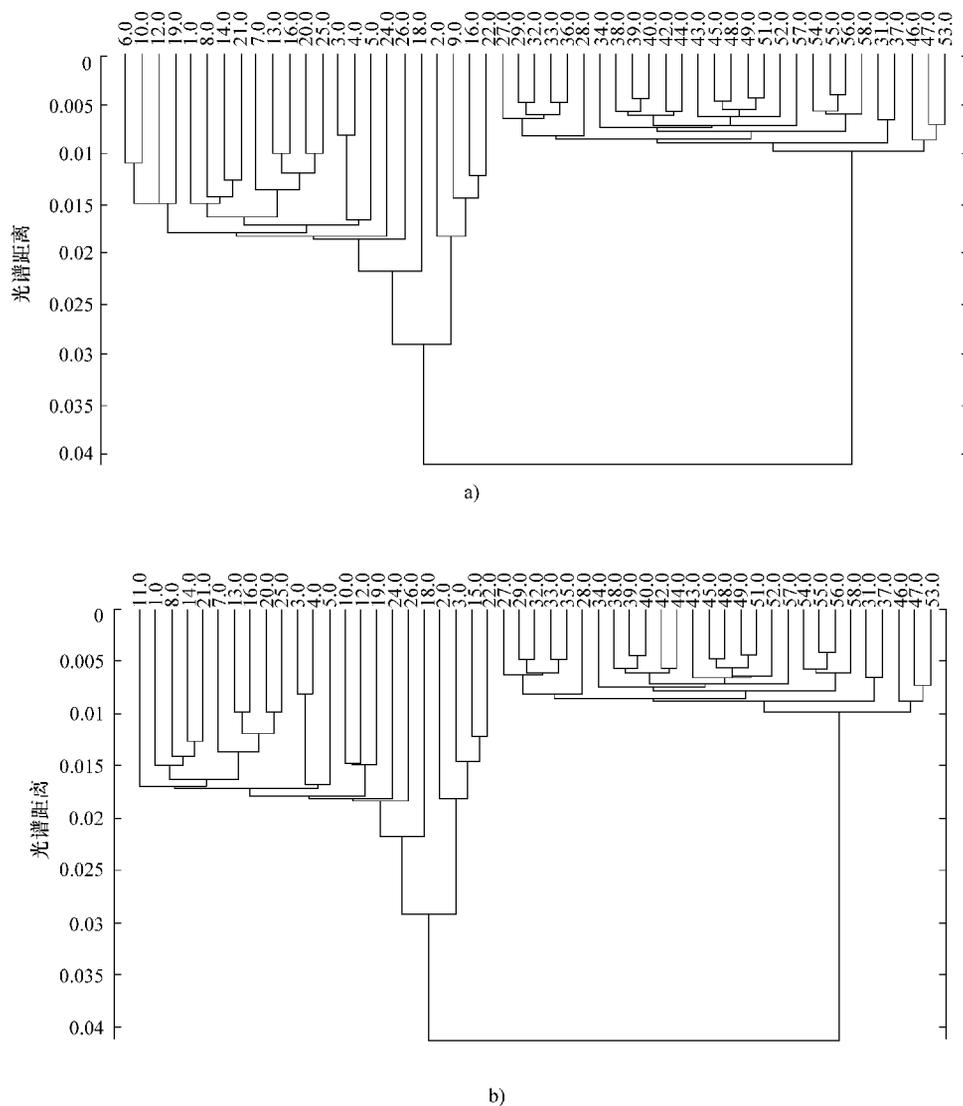


图 8-3 淀粉样品预测结果

a) 6 号土豆淀粉预测结果 b) 11 号土豆淀粉预测结果

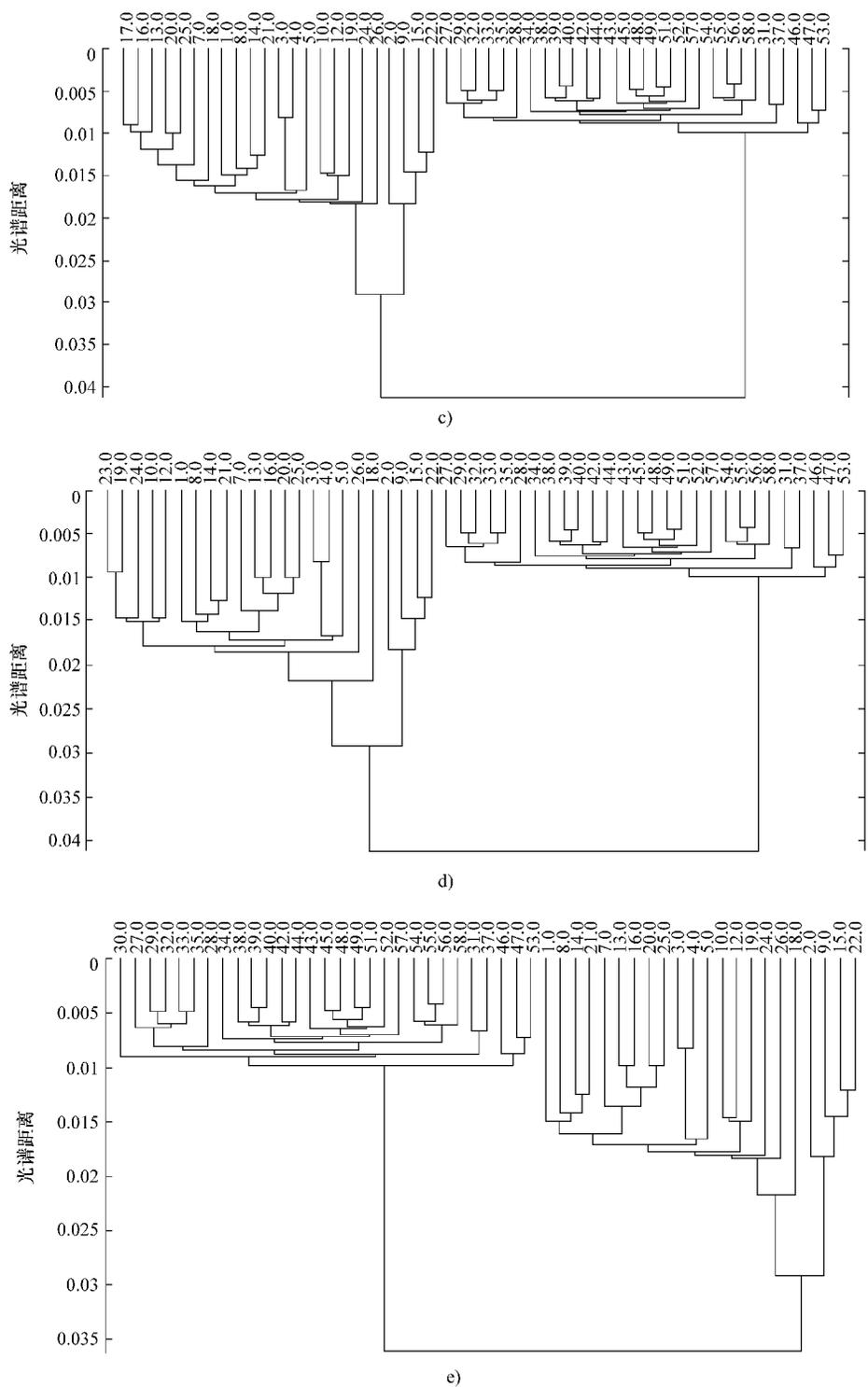
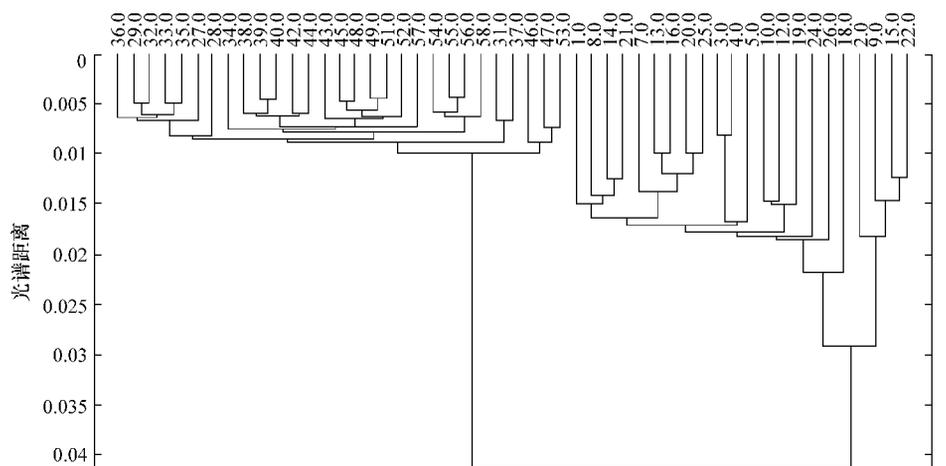
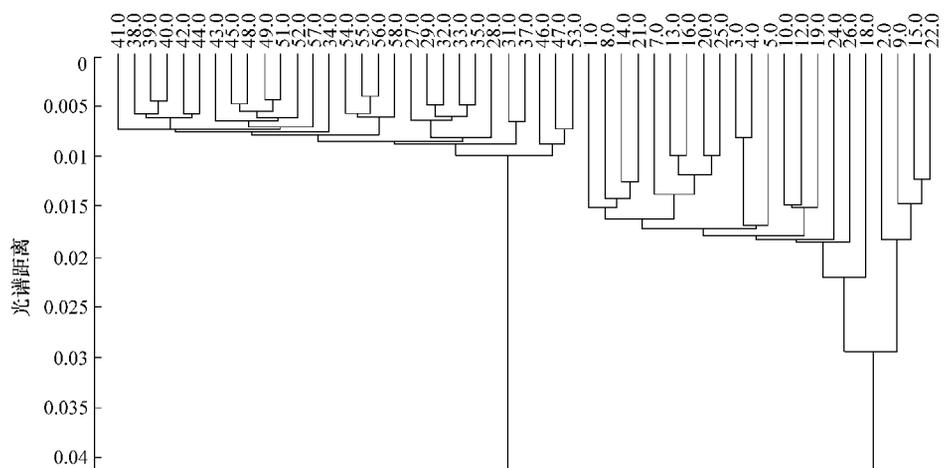


图 8-3 淀粉样品预测结果 (续)

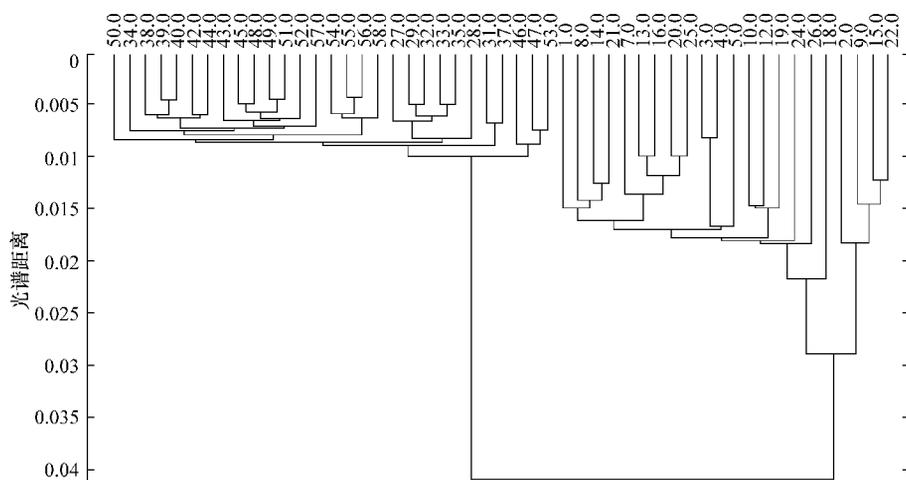
c) 17 号土豆淀粉预测结果 d) 23 号土豆淀粉预测结果 e) 30 号玉米淀粉预测结果



f)



g)



h)

图 8-3 淀粉样品预测结果 (续)

f) 36 号玉米淀粉预测结果 g) 41 号玉米淀粉预测结果 h) 50 号玉米淀粉预测结果

8.3 基于 SVM 的淀粉种类定性识别方法研究

8.3.1 样品制备

在各大超市中购买不同批次不同厂家生产的马铃薯、玉米两种淀粉，按名称各取 20g，其中马铃薯淀粉 28 个、玉米淀粉 21 个。将其编号，1~28 号为马铃薯淀粉，29~49 号为玉米淀粉。然后将样品分为校正集和预测集。1~14 号、29~39 号为校正集样本，另外收集 15~28 号、40~49 号作为各自的预测集样品。

8.3.2 光谱采集

采用德国 Bruker 公司生产的 VERTEX70 近红外光谱分析仪对原始样品进行全谱测定。检测器为 Bruker 公司的专利数字检测器。20g 分别装入近红外光谱分析仪的固体测量池进行淀粉样品近红外光谱采集，样品杯选用减少光谱散射的高性能样品杯，采用透反射采样模式，对 $12500 \sim 4000 \text{cm}^{-1}$ 谱区扫描，分辨率为 8cm^{-1} ，扫描 64 次。

49 个淀粉样品未经任何化学处理，逐一扫描样品。样品采集光谱前，将样品逐一摇动，使得样品均匀，避免漏光现象的存在，从而得到准确的光谱图。测得的样品近红外光谱如图 8-4 所示。

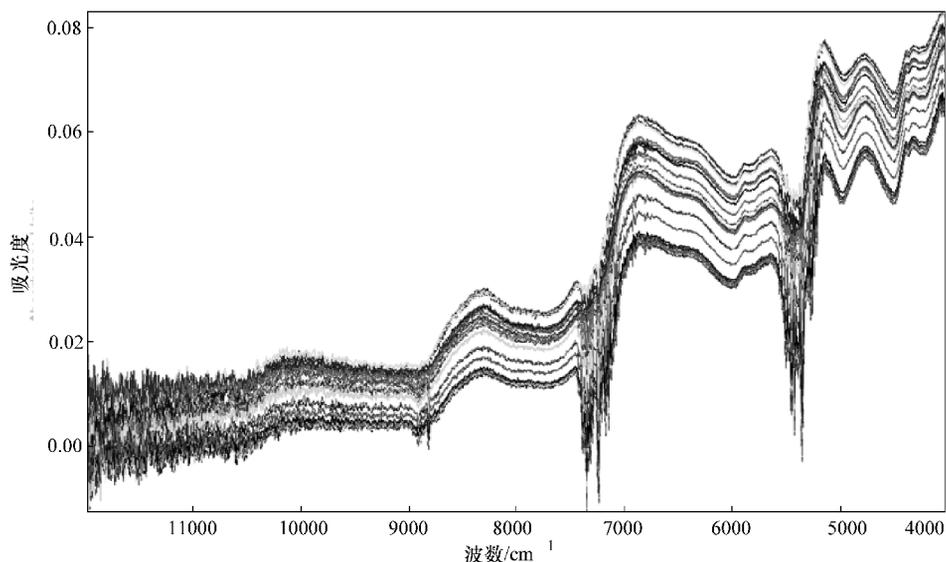


图 8-4 样品近红外光谱图

由图可以看出淀粉谱图的峰形，峰位有一定的差别，利用化学计量学的方法对光谱进行预处理，采用 SVM 的算法进行鉴别，可以突出样品之间化学组成含量上的微小差别，从而达到分类的目的。

8.3.3 结果与分析

不同种类的淀粉，其吸光度会出现较大的差异。可以根据样品光谱特性的差别，探索

SVM 在淀粉种类检测中的可行性，具体流程如图 8-5 所示。

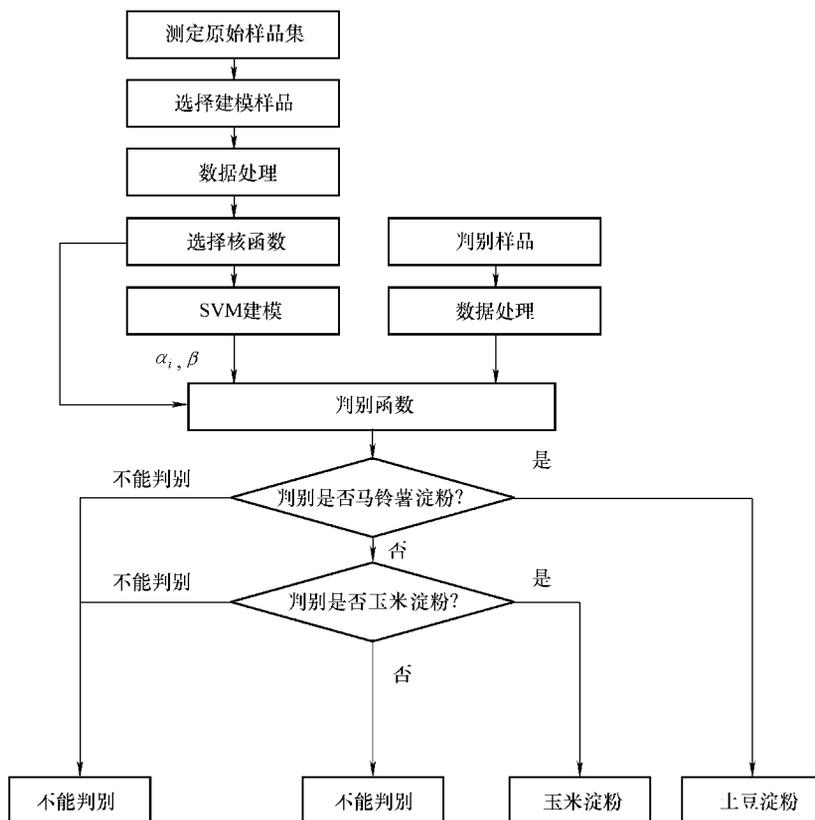


图 8-5 判别淀粉种类流程

1. 选择训练集和数据处理

在利用 SVM 进行分类判别时，训练集选取的好坏直接影响判别的精确度。在实际的应用过程中，选择训练集的原则多种多样。在本实验中，通过对原始样品集的光谱数据进行分析，选择要分类的两类样品光谱图中，吸光度最接近的几组样品组成训练集，剩下的样品作为校验集。利用这个原则选择训练集，使训练样品尽可能真实地反映该类样品的光谱特性，即使再加入新的待预测样品数据，同样可以最大限度提高判别的精确度。选取 17 个样品组成训练集，剩下的 24 个样品作为校验集。

由图 8-6 可知，在整个波谱段，样品吸光度变化的范围较大，而且吸光度和波数之间的非线性关系明显，在整个波谱段判别淀粉样品种类难度较大。为了使得判别过程简单，提高样品检测的精确度，并且使算法的鲁棒性较好，即使样品数据某些波数受到噪声的干扰，依然具有较好的检测效果，在整个波数段的高、中和低三部分选取 5 个有代表意义的波数段。波数段的选择原则是使训练集中的支持矢量个数少。因为 Vapnik 指出测试未知样品的最大出错概率和支持矢量的个数如式 (8-1) 所示：

$$E(\text{Pr}(\text{error})) \leq \frac{E(\text{支持矢量个数})}{\text{训练矢量的个数} - 1} \quad (8-1)$$

即支持矢量的个数越少，SVM 的期望风险的上界越小，其泛化能力则越强。一般来说，在利用 SVM 进行样品数据分类时，为了最大限度地提高支持矢量的训练和预测能力，需要

对样品数据进行尺度变换,把样品数据 x 变换到 $[-1, 1]$ 或者 $[0, 1]$ 范围内。

2. 选择适当的核函数和参数

在利用 SVM 进行样品分类判别中,核函数 $\ker(x^T, x)$ 和惩罚因子 C 的选择对判别的结果影响较大,目前国际上对核函数的选择还没有形成统一的模式,只有凭借经验和实验的对比进行寻优。惩罚因子 C 的选取需要在平衡最大分类间隔和最小分类误差中。在本次实验中,通过分析了不同种类的核函数和惩罚因子对判别结果影响,对核函数和惩罚因子在 SVM 中有了一个直观的了解。

3. 利用 SVM 对训练集进行建模

通过选取适当的核函数 $\ker(x^T, x)$ 和惩罚因子 C ,即可求解优化问题,利用 SVM 对训练集样品进行训练学习,求解 Lagrange 乘子 α_i 和偏置项 b 。利用求解得到的 α_i 和 b ,再利用分类判别函数 (8-25),即可对待测样品进行分类。在本次实验中,在 MATLAB 6.5 仿真环境下,采用 SVM 工具包来实现对样品数据的训练和预测。例如,选取高斯径向基函数 (RBF) 作为核函数,并且取 $\sigma = 1$;选择惩罚因子 $C = 2000$,则对训练集样品进行训练学习过程如图 8-6 和图 8-7 所示。

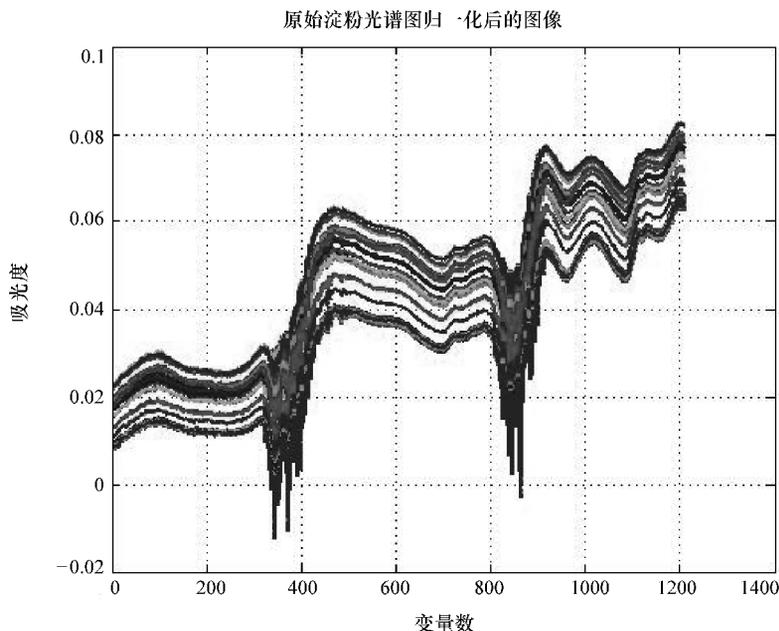


图 8-6 训练集样品

4. 预测样品的判别

利用 3. 中计算得到的 Lagrange 乘子 α_i 和偏置项 b ,通过分类决策函数式 (8-1) 对预测样品进行分类判别。为了提高判别的精确度,利用式 (8-2) 对预测样品 x 进行分类:

$$\begin{cases} f(x) > 1 \Rightarrow \text{类别 1} \\ f(x) < -1 \Rightarrow \text{类别 2} \\ -1 \leq f(x) \leq 1 \Rightarrow \text{不可判别} \end{cases} \quad (8-2)$$

给定某个预测样品,利用分类判别函数对预测样品进行分类,其分类过程如图 8-8 ~ 图 8-11 所示。

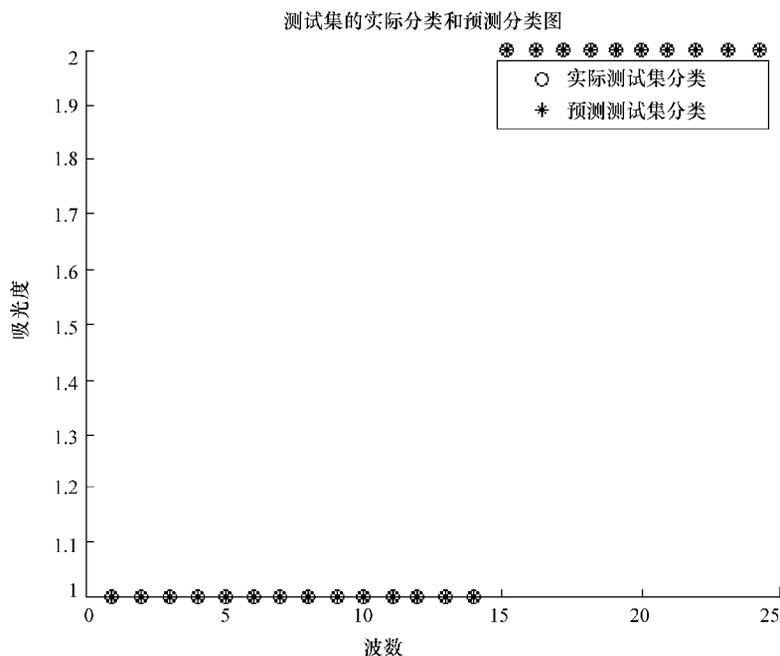


图 8-7 SVM 训练结果

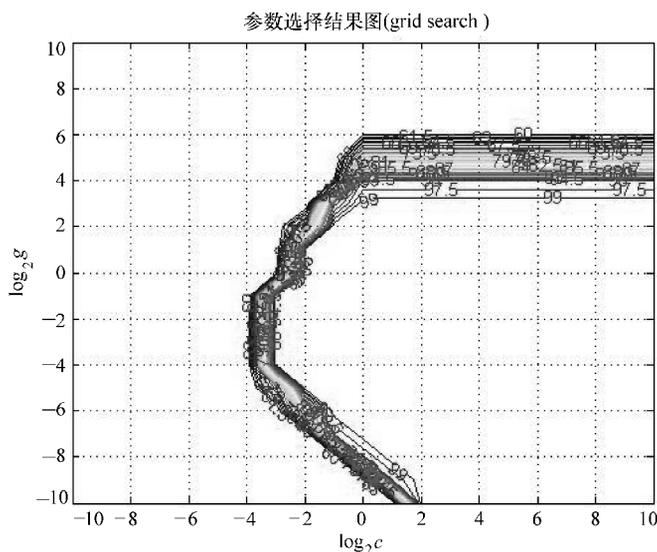


图 8-8 粗略选取参数

在 SVM 工具包中，设置程序使分类函数 $f(x) > 1$ 的样品点用红色的“+”表示；分类函数 $f(x) < -1$ 的样品点用绿色的“*”表示；分类函数 $-1 \leq f(x) \leq 1$ 的样品点用黑色“○”表示，即可根据分类结果图判别出预测样品的分类结果，当然也可以根据判别函数的值来对预测样品进行分类。因为预测样品中数据较多，可能出现判别结果不一致的情形。规定如果对预测样品中各点的判断精确度在 85% 以上，则可以判定该预测样品属于某一类。

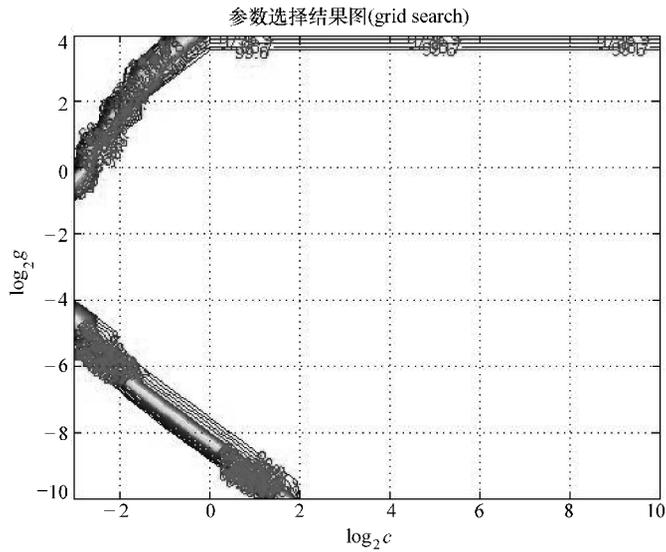


图 8-9 精细选取参数

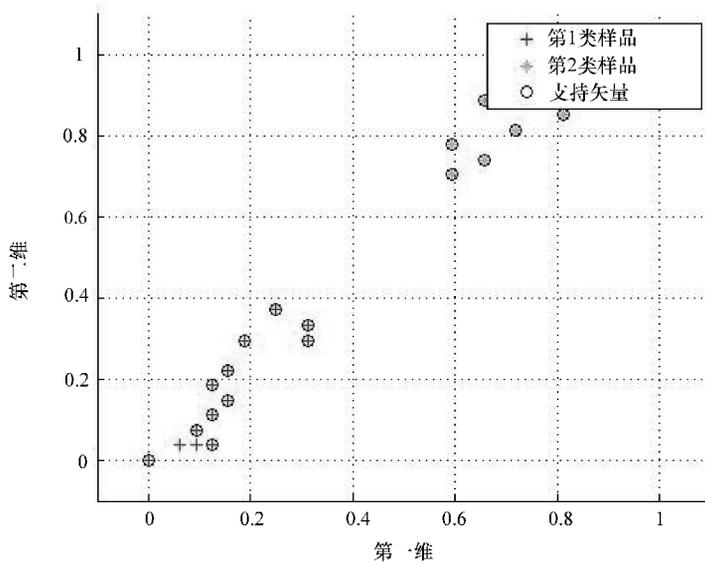


图 8-10 分类精确度

5. 结果分析

利用 SVM，根据之前所述步骤对原始样品集进行判别。选取不同的核函数和惩罚因子 C ，可以得到表 8-1 ~ 表 8-3 的判别结果。

```

打印粗略选择结果
Best Cross Validation Accuracy = 100% Best c = 0.125 Best g = 0.0625
打印精细选择结果
Best Cross Validation Accuracy = 100% Best c = 0.125 Best g = 0.0625
Accuracy = 100% (24/24) (classification)
打印测试集分类精确度
Accuracy = 100% (24/24)
bestCVaccuracy =
    100
bestc =
    0.1250
bestg =
    0.0625
Accuracy = 100% (25/25) (classification)
Accuracy = 100% (24/24) (classification)
Mean squared error = 0.903155 (regression)
Squared correlation coefficient = -1.#IND (regression)
\\
    
```

图 8-11 预测样品分类结果

表 8-1 SVM 分类结果

惩罚因子 C	线性		$Poly (p=1)$		$RBF (\sigma=1)$	
	支持矢量数	预测率	支持矢量数	预测率	支持矢量数	预测率
10	264	37.5% (9/24)	274	100% (24/24)	269	100% (24/24)
50	70	100% (24/24)	74	100% (24/24)	70	100% (24/24)
100	41	100% (24/24)	44	100% (24/24)	41	100% (24/24)
500	23	100% (24/24)	23	100% (24/24)	23	100% (24/24)
1000	22	100% (24/24)	23	100% (24/24)	22	100% (24/24)
2000	22	100% (24/24)	23	100% (24/24)	22	100% (24/24)
5000	22	100% (24/24)	23	100% (24/24)	22	100% (24/24)

表 8-2 多项式核函数 $Poly$ 的 SVM 分类结果 ($C=100$)

p	支持矢量数	预测率
1	44	100% (24/24)
2	25	100% (24/24)
3	23	100% (24/24)
4	22	100% (24/24)
5	22	100% (24/24)

表 8-3 高斯径向基核函数 RBF 的 SVM 分类结果 ($C=100$)

σ	支持矢量数	预测率
0.5	25	100% (24/24)
1	41	100% (24/24)
2	119	100% (24/24)
3	239	100% (24/24)
4	395	100% (24/24)
5	518	95.8% (23/24)

表 8-1 ~ 表 8-3 中的支持矢量数表示在 5 个波谱段和两次训练学习中的支持矢量的总数。由实验结果可以看出，利用 SVM 对淀粉类别进行判别是可行的，实验中几乎能够完全对预测样品进行判别。表 8-1 显示了惩罚因子 C 对判别结果的影响，随着惩罚因子 C 的值变大，表示不容忍判别错误的程度变大，支持矢量的个数越来越小，这也说明对未知样品的

判别错误概率变小。但是当支持矢量大到一定程度时,则其对判别结果的影响越来越小。表 8-2 显示多项式核函数中 p 对判别结果的影响,随着多项式次数 p 的变大, SVM 的个数变小,则表示对未知样品判别的精确度的上界变小。表 8-3 显示高斯径向基核函数中参数 σ 对判别结果的影响,随着 σ 的变大,支持矢量的个数变大,则表示对未知样品判别的精确度的上界变小,判别出错的概率变大。

本节建立了基于聚类分析的淀粉种类定性分析模型,该模型预测精确度达 100%;同时研究了基于 SVM 法判别淀粉种类的校正模型,该模型能够准确地判别淀粉种类。分析了核函数、SVM 数对于模型预测能力的影响。从定性分析的角度探索性研究了 SVM 对于淀粉品质检测的应用。

8.4 基于近红外光谱的淀粉含水量快速分析方法研究

含水量是淀粉品质的重要指标,其含量高低直接影响着淀粉的品质,目前国标法测淀粉水分含量是凭借称量淀粉样品干燥后损失的质量,用样品损失质量占样品原质量的百分比表示。其测试原理是将样品放在 130 ~ 133℃ 的电热烘箱内干燥 90min,得到样品的损失质量。

国标法测淀粉含水量耗时、费力,而且浪费样品,因此需要研究一种快速无损的检测方法。而淀粉中分子在近红外光谱区的吸收主要由分子中含氢基团(C-H、N-H、O-H、S-H 等)的组合频和倍频吸收组成,淀粉水分结构中的 O-H 对近红外光谱中的合频信息贡献较大,其数量反映了淀粉中水分的含量。

因此本研究结合近红外光谱的优点,基于 PLS 法建立了淀粉中淀粉含水量的定量分析模型,并分析比较了不同数据预处理方法和不同波长对模型预测能力的影响。

8.4.1 样品制备

实验收集了由 56 个来自不同品牌、不同厂家、不同批次的淀粉样品(包括马铃薯淀粉、玉米淀粉、红薯淀粉、绿豆淀粉等)组成的样品集。

本实验的样品水实际含量均用国标法测得。即采用烘箱法对 56 个样品的水含量进行测定,得到的结果作为校正模型建立时水的真实含量。

8.4.2 光谱采集

采用德国 Bruker 公司生产的 VERTEX70 近红外光谱分析仪对原始样品进行全谱测定。将 56 个原始样品各 20g 分别装入近红外光谱分析仪的固体测量池进行淀粉样品近红外光谱采集,样品杯选用减少光谱散射的高性能样品杯,采用透反射采样模式,对 12500 ~ 4000 cm^{-1} 谱区扫描,分辨率为 8 cm^{-1} ,扫描 64 次。淀粉样品光谱如图 8-12 所示。

8.4.3 淀粉含水量 PLS 模型的建立与分析

1. 样品集数据的分析

为了使得所建立的校正模型对未知样品有良好的预测能力,校正集样品的淀粉含水量变化范围应大于未知样品的含量变化范围,并且淀粉的含量在整个变化范围内是均匀的。基于此,在 56 个原始的淀粉样品中按 3:1 的比例划分校正集样品和预测集样品。其中,按浓度

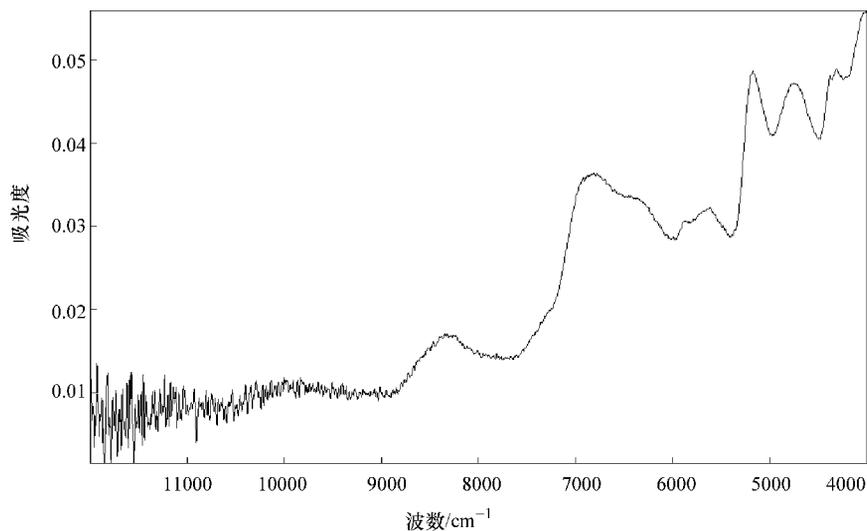


图 8-12 淀粉样品光谱图

梯度法在不同的含水量范围内选取 42 个淀粉样品用于建立校正模型，剩下的 14 个样品用于预测该模型。

2. 测量淀粉的近红外光谱

经全波段近红外光谱扫描得到光谱信息，如图 8-13 所示。

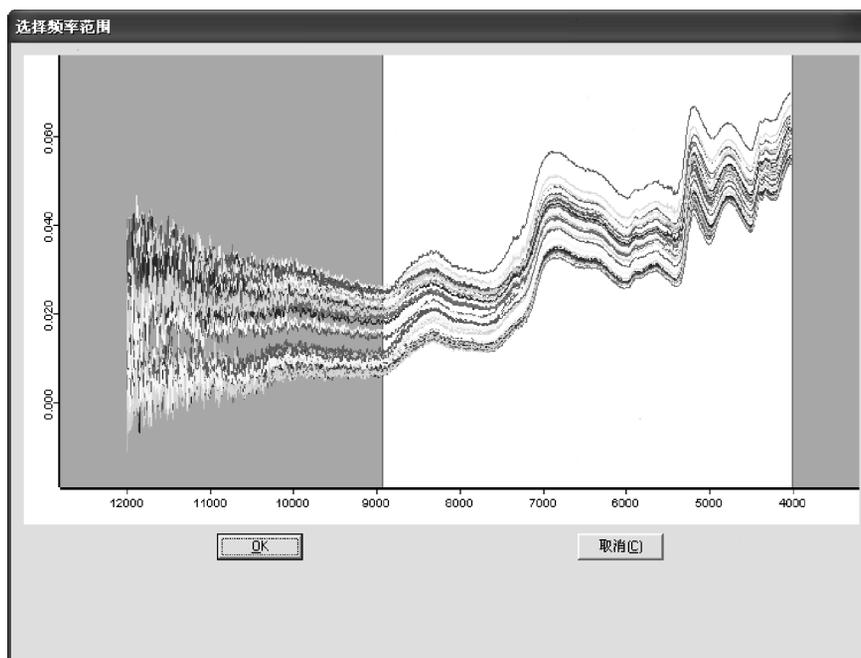


图 8-13 样品原始光谱

3. 光谱数据的预处理及模型的建立

用 OPUS6.5 光谱分析软件进行数据预处理分析, 经过波长最优化后选择波长范围为 $8439.4 \sim 7941.8\text{cm}^{-1}$ 、 $6464.5 \sim 5970.8\text{cm}^{-1}$ 和 $5481 \sim 3999.8\text{cm}^{-1}$ 的波段进行分析建模, 校正集样品经预处理和波长最优化后的光谱如图 8-14 所示。

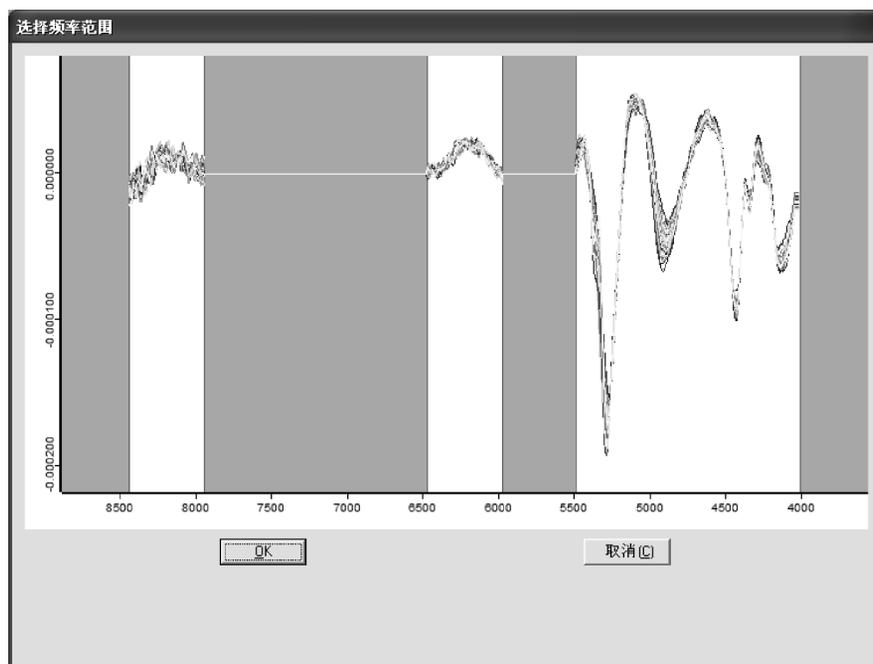


图 8-14 经预处理后的样品光谱图

在对样品含水量建立校正模型时, 不同的预处理方法, 得到了不同的最佳主成分数, 其主成分数与 RMSECV 的关系如图 8-15 ~ 图 8-19 所示。

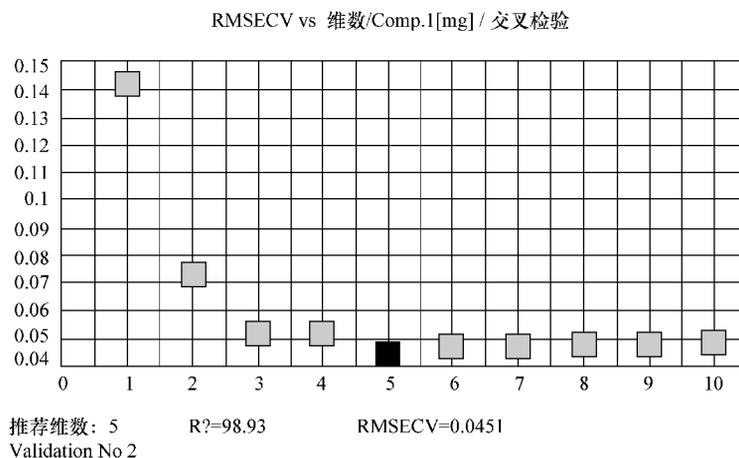


图 8-15 多元散射校正

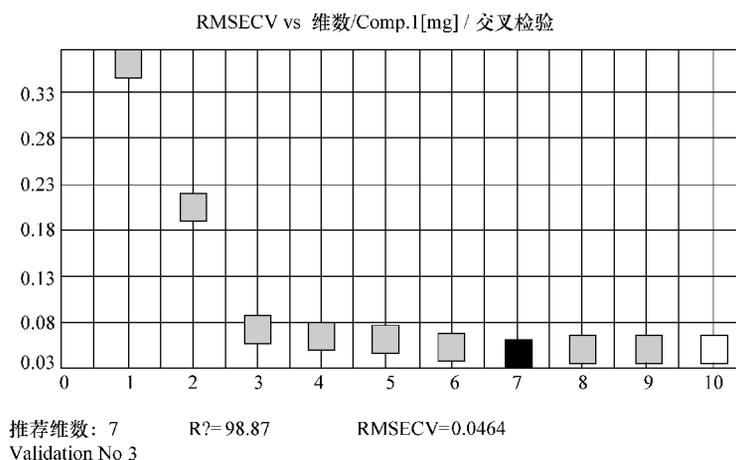


图 8-16 消除常量偏移量

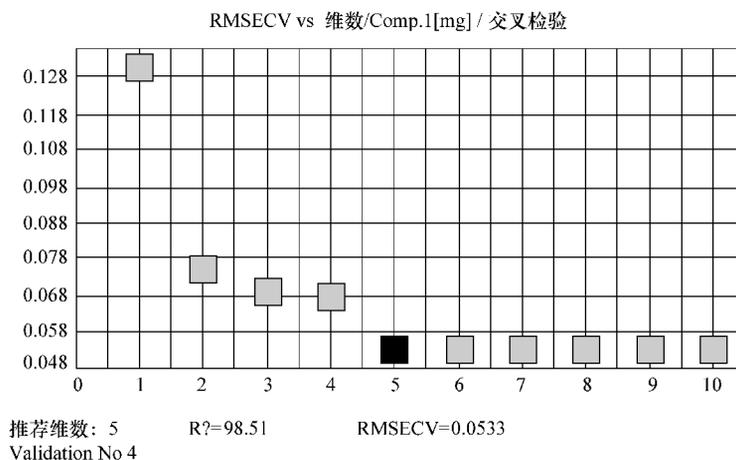


图 8-17 矢量归一

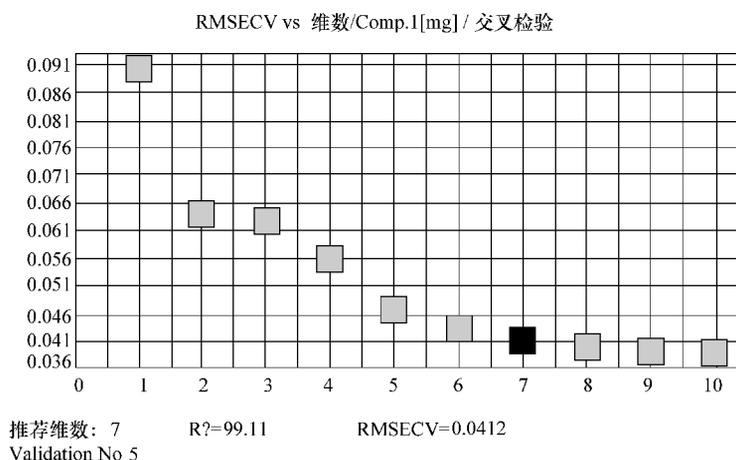


图 8-18 一阶导数和矢量归一

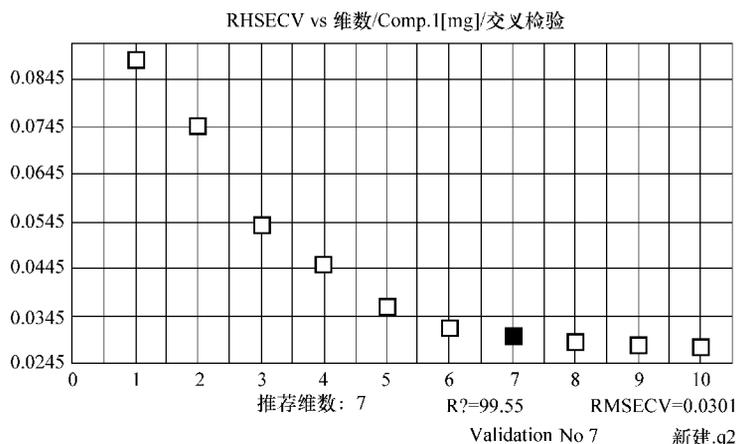


图 8-19 一二阶导数、矢量归一、多元散射校正与消除常数偏移量结合

经光谱最优化分析，一阶导数、二阶导数、多元散射校正法和矢量归一化等预处理方法得到的校正模型，其决定系数和正交标准差均未达到较高的预测精度，而采用一二阶导数、矢量归一、多元散射校正与消除常数偏移量结合的方法却得到较高精确度的校正模型。处理光谱后，建立的模型具有相对其他预处理方式更好的模型精确度， R^2 为 99.55，RMSECV 为 0.0301，其训练结果较为精确。

4. 淀粉近红外光谱模型的预测

将利用浓度梯度法选取的 14 个预测集样品代入模型中进行淀粉含量的预测，得到预测决定 R^2 为 99.97，正交标准差 RMSEP 为 0.0104，其预测值与标准值如图 8-20 所示，预测曲线与标准曲线夹角很小，可见预测值趋近于真实值，并且样品散点密集的分布在曲线两侧，预测样品均在校正模型的可预测范围内，无超出预测范围的样品，该模型具有较好的适用性。

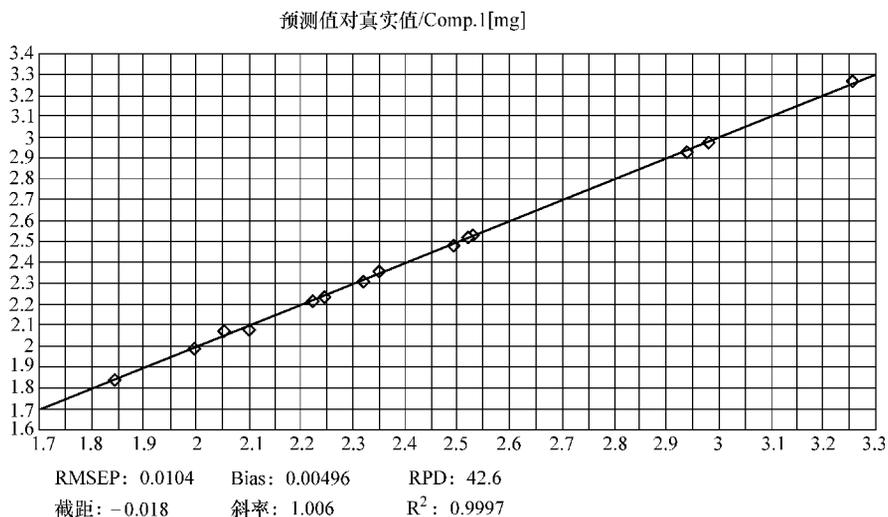


图 8-20 预测集样品真实值与预测值的相关曲线

样品淀粉水分含量的真实值与预测值见表 8-4。

表 8-4 样品淀粉水分含量的真实值与预测值

样品编号	真实值	预测值	样品编号	真实值	预测值
4	2.244	2.2337	32	2.319	2.3074
8	2.1	2.0803	36	1.844	1.8335
12	1.995	1.9903	40	2.35	2.3552
16	2.529	2.5284	44	2.226	2.2173
20	3.256	3.2668	48	2.521	2.5203
24	2.492	2.4792	52	2.939	2.925
28	2.981	2.9765	56	2.053	2.0657

8.5 马铃薯淀粉掺杂玉米淀粉的近红外光谱检测方法研究

马铃薯淀粉，是自然淀粉中的一种，在世界整体淀粉生产中，马铃薯淀粉占有很大的份额。其产量仅次于玉米，居第二位。马铃薯淀粉的颗粒，在所有淀粉产品中是最大的，且其形状也最接近圆形。在许多理化指标上也优于其他淀粉，其中以初始糊化度低、成糊黏度高和成糊透明度高最为显著。

玉米又称玉蜀黍。玉米最初一直是作为人类主要的粮食作物而进入人类社会的，玉米的淀粉含量高，能够为人类的生存发展提供必需的能量。随着社会的发展、科学技术的进步，在满足了人类基本生存的基础上，玉米逐渐被用作饮料以及其他深加工产品的原料，因而大大促进了世界玉米生产的发展。目前，它不但是世界三大粮食作物之一，而且其产量已跃居世界粮食品种的首位。

玉米淀粉是淀粉中最主要的品种，占世界淀粉总量的 80%。玉米淀粉是玉米加工的初级产品，不但可直接使用，更是重要的工业原料，广泛应用于食品、制糖、发酵、医药、造纸、化工等行业。全球利用玉米进行深加工而生产的产品超过 2000 种，我国玉米深加工产品有 500 多种。目前世界上大约 2/3 的玉米淀粉用于食品、医疗、饮料等方面，1/3 用于造纸、包装、纺织、石油等方面。

1. 用于食品加工

玉米淀粉作为原料可以直接用于粉丝、粉条方便面、火腿肠、肉制品、冰激凌等方面的加工，需求稳定。

2. 用于生产淀粉糖

淀粉糖是玉米淀粉深加工产量最大的一类产物。淀粉糖的种类繁多，附加值较大，主要包括葡萄糖、麦芽糖、麦芽糖浆、果糖等产品，以麦芽糖浆、高麦芽糖浆和葡萄糖为主，这类品种占淀粉糖比重的 80% 以上。淀粉糖可直接用于食品添加，也是下游产品的原料。例如，葡萄糖、麦芽糖、果糖，可直接用于食品，加氢处理后，分别得到山梨醇、麦芽糖醇和甘露糖醇，它们统称多元糖醇，应用于医学、食品工业、化妆品工业、农业等领域。

3. 用于生产赖氨酸、味精、氨基酸、柠檬酸等

赖氨酸是动物体内不能合成必须从外界摄取的必需氨基酸，它的缺乏将影响其他氨基酸

的吸收,因此随着饲料业、饲养业的发展,赖氨酸的需求量也越来越大。我国现在是世界最大的赖氨酸生产国,每年消耗大量的玉米淀粉。

味精(谷氨酸钠)作为鲜味剂广泛地应用于饮食业。亚洲人对味精情有独钟,有的国家人均年消费已达1kg多。我国是味精生产大国,产量居世界第一,产量突破100万t,消耗玉米淀粉150万t。

以淀粉为原料生产的氨基酸还有异亮氨酸、精氨酸、结氨酸等。近年来氨基酸的世界需求量增长迅速。用玉米淀粉生产的有机酸有柠檬酸、乳酸、苹果酸、衣康酸、草酸等。其中柠檬酸产量较大,柠檬酸在食品工业、医药业、洗涤业、纺织工业、化妆品业上都有广泛的应用。

4. 生产变性淀粉

变性淀粉是利用物理、化学和酶等手段改变天然淀粉的性质,使其符合各行各业应用的需要,变性淀粉的使用量和范围非常广,目前我国已经生产出了预糊化淀粉、酸化淀粉、氧化淀粉、醚化淀粉、酯化淀粉、交联淀粉、接枝淀粉等应用到纺织、造纸、食品、石油、医用、建筑、农业饲料、日用化工等行业共70多个品种。其中以造纸行业的应用量最大,占总量的50%~60%。

变性淀粉在食品工业的应用中可以增强食品的韧性和黏度,改善食品的口感,延长保存期,由于通过变性提高了淀粉糊的透明度,因此可以改善食品的外观,提高其光泽度。所以在生产方便食品、肉制品、调味料、酸奶、果冻、糖果、松脆点心、豆沙馅、休闲食品等都可以添加变性淀粉,以提高产品质量。

在造纸业,作为添加剂,能够生产出各种高级纸,同时还可减轻对环境的污染。变性淀粉在纺织工业用量较大,主要用在丝纱上浆、印花糊料。在石油工业中变性淀粉主要是用于石油钻井液、压裂液和油气生产的多种场合。总之变性淀粉的应用范围广、专用性强、品种多,是一个市场潜力大、不断发展中的产品。

5. 用于医药行业

就医药工业而言,淀粉是抗生素工业最重要的原料,因为几乎所有抗生素都采用淀粉发酵法生产,如销量极大的青霉素、头孢菌素、四环素、土霉素、金霉素、链霉素与各种氨基糖苷类抗生素等,无一不是用淀粉为底物经工业微生物发酵、提取而成。另外,淀粉的另一重要用途是作为药物赋型剂,早期各国药厂生产的片剂绝大多数使用玉米淀粉为填充剂及粘合剂。

马铃薯淀粉由于自身所具有的优良特性成为食品加工中的重要组成部分,广泛应用于方便食品、果子羹、浓缩调料汁、香肠、火腿肠、膨化食品的生产中。马铃薯淀粉的高白度、高透明度、高黏度、低糊化温度等特殊性能是其他淀粉无法替代的。但是一些生产企业为了谋取非法利润,常把价格低1/2的玉米淀粉掺杂到马铃薯淀粉中来销售,最多的添加到80%。掺杂后的马铃薯淀粉由于其黏度低、糊化温度高,会影响食品或菜肴的成色和品质,严重地损害了消费者的利益。可是由于马铃薯淀粉和玉米淀粉颗粒的宏观外观差别不明显,肉眼无法辨认,因此目前多靠传统的感官评定方法来鉴别,该方法一方面需要经验,另一方面其检测结果受主观因素影响较大,精确度难以保证。而常规的理化分析方法不仅费时费力,还破坏样品。因此有必要研究一种简单、快速、无损的淀粉品质检测技术。

8.5.1 样品制备

实验用玉米淀粉和马铃薯淀粉均是从超市购买不同品牌或同一品牌不同批次的淀粉，在马铃薯淀粉中随机掺入浓度为0%~100%的玉米淀粉，共制备43个掺杂样品，并以掺杂后马铃薯淀粉的浓度作为样品的真值，43个掺杂样品中37个样品用于定量建模，6个用于模型验证。

8.5.2 光谱采集

傅里叶变换近红外光谱仪 VERTEX 70，配有漫反射样品杯和 OPUS6.5 光谱采集及分析软件。将上述淀粉样品放置在漫反射样品台的样品杯中，然后进行近红外光谱采集。波数范围 $12000 \sim 4000 \text{cm}^{-1}$ ，波数间隔 8cm^{-1} ，扫描64次取平均，环境温度 $23 \sim 25^\circ\text{C}$ 。43个掺杂样品的近红外漫反射光谱如图8-21所示。

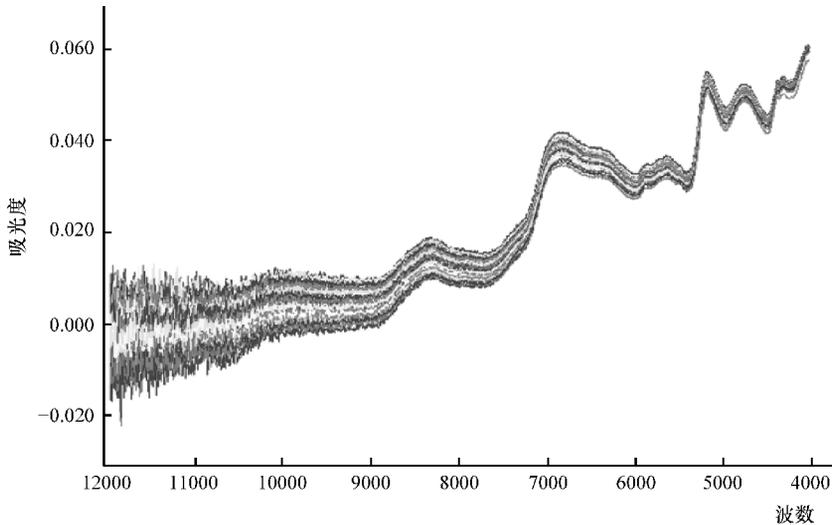


图8-21 43个掺杂淀粉样品的近红外漫反射光谱图

8.5.3 定量分析模型的建立

将43个样品应用于近红外定量分析，确定校正集37个样品，6个样品用于模型验证。通过 OPUS6.5 软件的分析 and 优化，选择最优处理算法，通过比较建立掺杂淀粉的 PLS 模型的优劣，寻找淀粉的吸收光谱较丰富的波段，分析表明淀粉对光谱信息贡献量最大的谱区范围是 $7502.1 \sim 6800.1 \text{cm}^{-1}$ 和 $5052.8 \sim 4597.7 \text{cm}^{-1}$ ，维数为5。图8-22所示为掺杂淀粉近红外光谱图交叉验证后马铃薯淀粉含量的预测值与真值。

交互验证结果表明预测值和化学值之间具有显著的相关性，校正样品均匀地分布在回归线的两侧，且交互验证得到的校正 R^2 为 0.9891，RMSECV 为 0.397，偏差为 0.00793。

8.5.4 定量分析模型的验证

为了验证定量模型的预测精确度，实验用检验集的6个样品进行预测，预测结果如

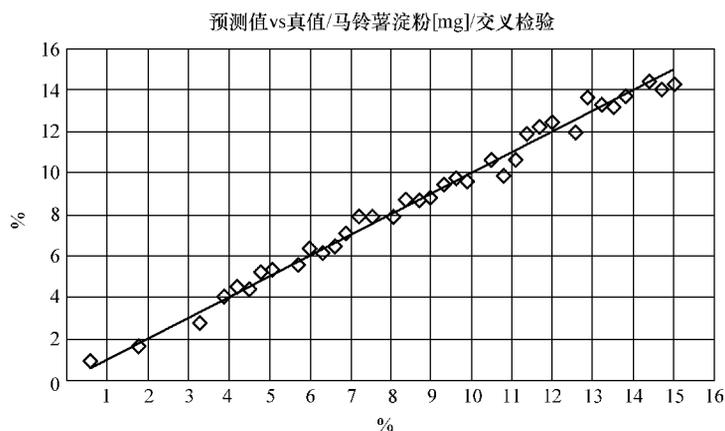


图 8-22 近红外光谱交叉验证预测值与真值

图 8-23 所示。预测 RMSEP 为 0.265，偏差为 0.17， R^2 为 0.999，详细分析结果值见表 8-5。

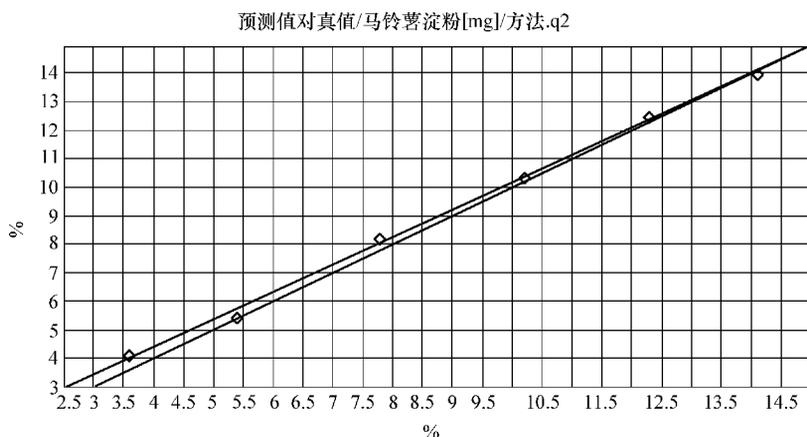


图 8-23 验证集样品的近红外光谱交叉验证近红外光谱预测值与化学值

表 8-5 验证集样品的近红外预测值

文件名	真值	预测值	偏差	单位
5	14.1	13.971	0.129	mg
11	12.3	12.435	-0.135	mg
18	10.2	10.309	-0.109	mg
26	7.8	8.1788	-0.3788	mg
34	5.4	5.4462	-0.0462	mg
40	3.6	4.0786	-0.4786	mg

初步研究结果表明，利用近红外漫反射光谱分析技术，能够准确、快速地检测马铃薯淀粉的品质。由于受条件所限，本研究仍然存在一些不足之处，如采用的样品数量还不够大。目前，正向定标样品扩充更多的淀粉品种，提高定标预测的稳定性。利用近红外光谱分析技术快速检测淀粉掺杂只是建立快速检测机制的一个开始，今后还将陆续开展测量淀粉中的含

水量等指标检测工作，为今后快速、无损检测淀粉品质提供最新、最全的数据支持。

8.6 淀粉掺杂滑石粉的近红外光谱检测方法研究

滑石主要成分是滑石含水的矽酸镁，分子式为 $Mg_3Si_4O_{10}(OH)_2$ 。滑石属单斜晶系。晶体呈假六方或菱形的片状。通常成致密的块状、叶片状、放射状、纤维状集合体。无色透明或白色，但因含少量的杂质而呈现浅绿、浅黄、浅棕甚至浅红色，解理面上呈珍珠光泽。硬度 1，比重 2.7~2.8。

滑石具有润滑性、耐火性、抗酸性、绝缘性、熔点高、化学性不活泼、遮盖力良好、柔软、光泽好、吸附力强等优良物理、化学特性，由于滑石的结晶构造是呈层状的，所以具有易分裂成鳞片的趋向和特殊的润滑性，如果 Fe_2O_3 的含量很高则会减低它的绝缘性。

滑石粉的主要用途如下：

- 1) 化妆品级 (HZ)：用于各种润肤粉、美容粉、爽身粉等。
- 2) 医药—食品级 (YS)：医药片剂、糖衣、痲子粉和中药方剂、食品添加剂、隔离剂等。
- 3) 涂料级 (TL)：用于白色体质颜料和各类水基、油基、树脂工业涂料、底漆、保护漆等。
- 4) 造纸级 (ZZ)：用于各类纸张和纸板的填料、木沥青控制剂。
- 5) 塑料级 (SL)：用于聚丙烯、尼龙、聚氯乙烯、聚乙烯、聚苯乙烯和聚酯类等塑料的填料。
- 6) 橡胶级 (AJ)：用于橡胶填料和橡胶制品防粘剂。
- 7) 电缆级 (DL)：用于电缆橡胶增剂、电缆隔离剂。
- 8) 陶瓷级 (TC)：用于制造电瓷、无线电瓷、各种工业陶瓷、建筑陶瓷、日用陶瓷和瓷釉等。
- 9) 防水材料级 (FS)：用于防水卷材、防水涂料、防水油膏等。
- 10) 微细滑石粉：用于高级油漆涂料、塑料、电缆橡胶、化妆品、铜版纸涂料、纺织润滑剂等。

近年来，部分淀粉企业为了谋取利益，向淀粉中添加大量滑石粉，以增加淀粉的重量。食用滑石粉是医药、食品行业的添加剂，具有无毒、无味、口味柔软、光滑度强等特点，因此可以食用，但食用过量或长期食用有致癌性。

正常淀粉中矿物质（以灰分计）的含量一般在 0.5% 以下，如淀粉中掺入了石膏、滑石粉等，皆能使小淀粉中的灰分增加。在灰分中测出钙离子、硫酸根、二氧化硅，就能定性掺入的物质。

1) 灰分的测定方法：称取样品 2g 放入预先灼烧的 550℃ 恒重的坩埚中，在电炉上加热至炭化，再放入 550℃ 的马费炉中，灼烧 2h，取出冷却降温。如果灰化不完全，再加水或硝酸使灰分湿润，微温至干，然后再放在马费炉中灰化 2h，取出冷却至 200℃，移至干燥器中，30min 后称重，计算灰分。

2) 二氧化硅定性方法：向测定完灰分含量后的灰分中，加入 2 倍量以上的研成细末的氢氧化钾，混合均匀，于 600℃ 熔融，冷后加水溶解，向水溶液中滴加 (1:1) 盐酸，使之

呈酸性，如果有胶状物析（ H_3SiO_3 ），说明检出了二氧化硅，同时作空白对照。

正常的小淀粉，一般用此法检不出二氧化硅，但掺入滑石粉等无机物则可检出。

3) 钙离子和硫酸根检验方法：取样品灰分，加（1:1）盐酸溶液 10mL，加热溶解、过滤，滤液分成两份，一份溶液中加入 1% 氧化钡溶液 1mL，如果产生大量沉淀，说明检出了硫酸根，同时作空白对照。再在另一份滤液中加入饱和草酸铵溶液 1mL，滴加（1:1）氨水呈弱碱性，产生大量沉淀，则为阳性，同时作空白对照。

灰分中如果仅检出钙离子、硫酸根，可认为是掺入石膏，如果同时检出二氧化硅及上述两种离子，可认为是检出了滑石粉或大白粉。当前市场上出售的大白粉，是将滑石粉精制加工而成，其成分与滑石粉相同。

上述方法虽能测定滑石粉的含量，但是试验所需试剂、材料、仪器等多而繁琐，此外也会破坏被测样品。因此本章研究应用近红外漫反射光谱分析技术，探索研究了检测淀粉中掺杂滑石粉的可行性，建立了相应的定量分析模型，并用独立的检验集对校正模型的可靠性进行了验证。

8.6.1 样品制备

实验用的淀粉均是从市场购买不同品牌或同一品牌不同批次的淀粉，用电子分析天平准确称量，在淀粉中随机掺入浓度为 0% ~ 25% 的滑石粉，共制备 32 个掺杂样品，并以掺杂后滑石粉的浓度作为样品的真值。

8.6.2 光谱采集

实验仪器用德国 Bruker 公司傅里叶变换近红外光谱仪、漫反射样品杯、OPUS6.5 光谱采集及分析软件。

将上述淀粉样品放置在旋转样品台的样品杯中，然后进行近红外光谱采集。波数范围 $12500 \sim 4000\text{cm}^{-1}$ ，波长间隔 8cm^{-1} ，扫描 64 次取平均，环境温度 $23 \sim 25^\circ\text{C}$ 。

纯滑石粉近红外光谱如图 8-24 所示。

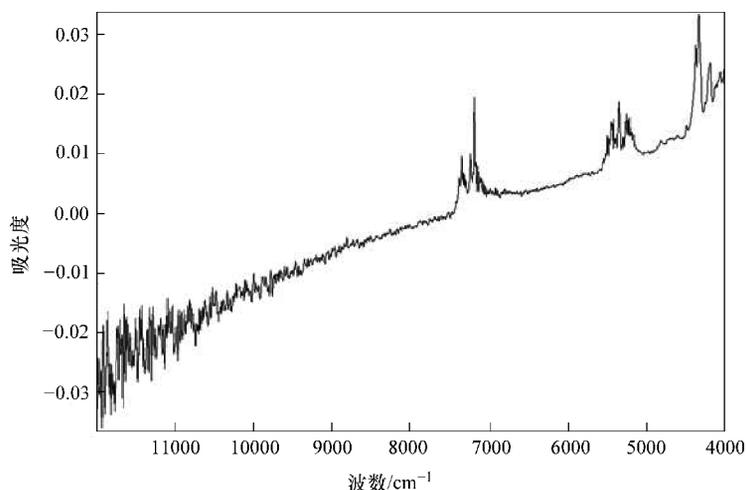


图 8-24 滑石粉样品光谱图

32 个混合后的淀粉样品的近红外漫反射光谱如图 8-25 所示。由图可以看出，在 $9000 \sim 4000\text{cm}^{-1}$ 范围内较为相似，带有许多淀粉之间的相似信息，峰形、峰位差别很小，无法直接鉴别。

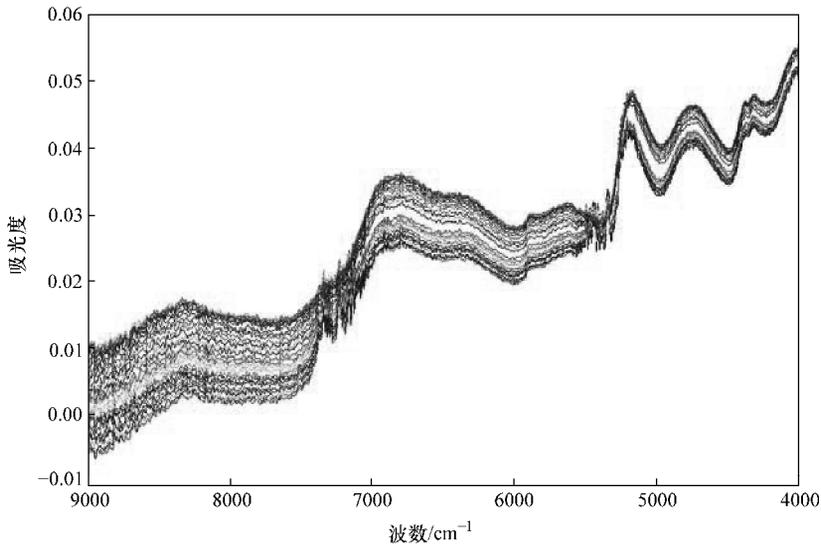


图 8-25 32 个掺杂淀粉样品的近红外光谱图

8.6.3 实验方法设计

利用化学计量学方法将原光谱进行数学预处理，采用 OPUS6.5 软件分析光谱数据。通过选择信息较丰富的光谱波段及对数据进行预处理，实现对原始光谱数据进行优化，建立精确度较高的预测模型。回归统计方法采用 PLS 法，PLS 法利用主成分分析将吸光度矩阵和浓度矩阵先分别分解为特征矢量和载荷矢量，然后用 PLS 法在这些稳变量之间建立相互关系，从而得到吸光度矩阵与浓度矩阵之间的数学校正模型。PLS 法的优点是尽可能在自变量中提取出与因变量相关性最大的组成部分，这对于分析微含量的物质有益。

8.6.4 定量分析模型的建立

将 32 个样品应用于近红外光谱定量分析，确定校正集 26 个样品，检验集 6 个样品。通过 OPUS6.5 软件的分析 and 优化，选择最优处理算法，通过比较建立混合淀粉的 PLS 模型的优劣，寻找淀粉的吸收光谱较丰富的波段，分析表明淀粉对光谱信息贡献量最大的谱区范围是 $7502.1 \sim 5446.3\text{cm}^{-1}$ ，维数为 4。图 8-26 所示为淀粉中滑石粉含量近红外光谱图交叉验证近红外光谱预测值与化学值。

交互验证结果表明，近红外光谱预测值和化学值之间具有显著的线性相关性，校正样品均匀地分布在回归线的两侧，且交互验证得到的校正 R^2 为 0.9939，RMSECV 为 0.118，偏移为 0.00102。

8.6.5 定量分析模型的验证

为了验证定量模型的预测精确度，实验用检验集的 6 个样品进行预测，预测结果如图 8-27 所示。

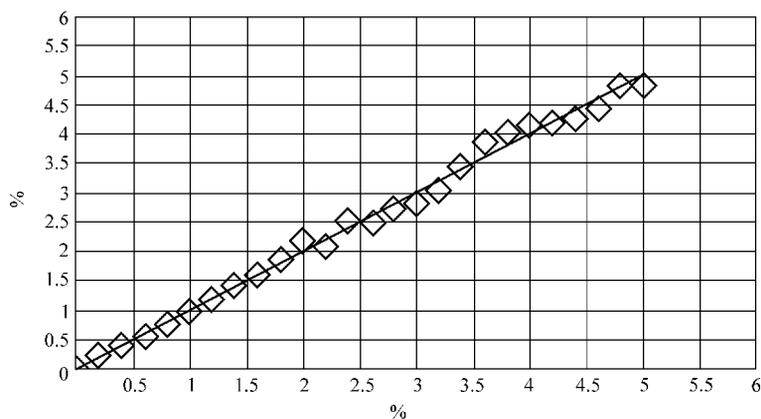


图 8-26 近红外光谱交叉验证近红外光谱预测值与化学值

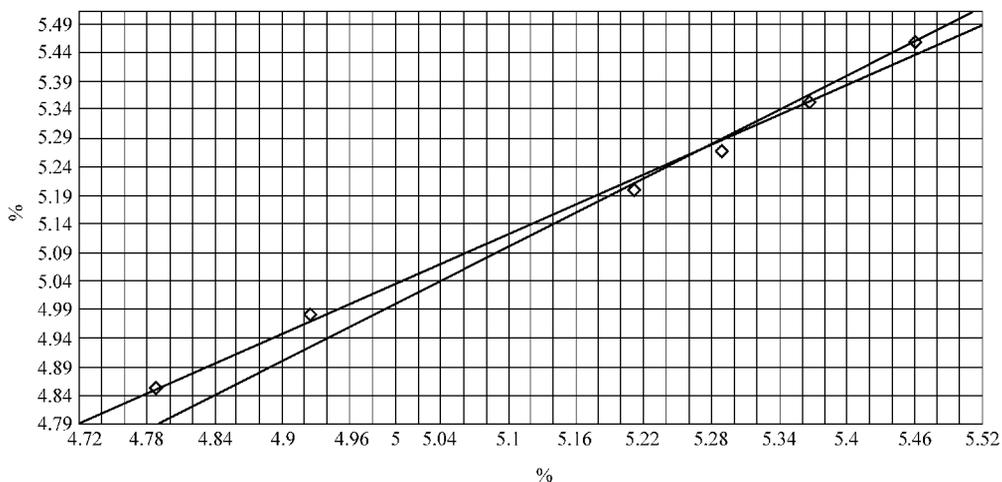


图 8-27 检验集样品的近红外光谱交叉验证近红外光谱预测值与化学值

详细分析结果见表 8-6。

表 8-6 样品的真实值与预测值

样品编号	真实值	预测值	样品编号	真实值	预测值
27	4.789	4.8512	30	5.212	5.196
28	4.925	4.9799	31	5.289	5.267
29	5.367	5.3535	32	5.46	5.4579

利用近红外光谱技术建立了淀粉中掺杂滑石粉的检测模型，初步实验结果表明，利用近红外反射光谱分析技术，能够准确、快速地检测出淀粉中掺杂滑石粉的目的。在实际应用前，有必要扩大样品集的范围，以进一步检验模型和提高模型的适应性。

8.7 小结

本章运用近红外光谱技术快速、无损的鉴别淀粉品质，采用聚类分析法和 SVM 的定性

近红外光谱技术在食品品质检测方法中的研究

建模方法,对玉米淀粉和马铃薯淀粉进行了分类。实验结果表明,利用近红外光谱法可以准确地进行分类,同时对淀粉中水分的含量进行了定量分析,结果显示 R^2 为 99.55, RMSECV 为 0.0301, 其训练结果较为精确。同时本章利用近红外光谱对马铃薯淀粉中掺杂玉米淀粉及掺杂滑石粉进行了定量分析,实验结果表明,运用近红外光谱可以快速、准确地对淀粉品质进行定性和定量检测。

参考文献

- [1] 尹华涛,徐建华.我国淀粉种类鉴别技术研究获突破 [N].中国质量报,2019-04-06.
- [2] 李燕萍,钱莹,段钢.采用近红外光谱测定木薯乙醇发酵液中乙醇、甘油和葡萄糖含量 [J].食品与发酵工业,2009,35(8):117-121.
- [3] <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [4] 史月华,徐光明,沈晓燕,等.近红外光谱法定量分析维生素 E [J].分析化学,2000,28(5):587-589.
- [5] 王东丹,李天飞,吴玉萍,等.近红外光谱分析技术在烟草化学分析上的应用研究 [J].云南大学学报,2001,2(3):135-137.
- [6] Scholkopf B, Smola A J, Williamson R C, et al. New support vector algorithms [J]. Neural Computation, 2000, 12(5): 1207-1245.
- [7] 芦永军,陈华才,吕进,等.茶多酚中总儿茶素的近红外光谱分析 [J].分析化学,2005,33(6):835-837.
- [8] 张洪江,吴金红,梅捍卫,等. FT-NIRS 技术应用于稻米直链淀粉含量分析研究 [J].植物遗传资源学报,2005,1(6):91-95.
- [9] 于丽燕,方如明,陈斌.近红外光谱透射法测定酱油的主要成分 [J].食品科技,2001,1:61-62.
- [10] 陈文杰,谭小力,王竹云,等.用傅里叶变换近红外光谱仪测定油菜种子品质指标的研究 [J],陕西农业科学,2002,(8):6-9.
- [11] 熊成,董庆利,曾静,等.近红外光谱分析技术在肉品品质检测中的应用 [J].食品与发酵工业,2010,36(12):141-145.
- [12] 甘莉,孙秀丽,金良,等. NIRS 定量分析油菜种子含油量、蛋白质含量数学模型的创建 [J].中国农业科学,2003,12(36):1609-1613.
- [13] 刘燕德,应义斌.基于 MATLAB 语言的苹果糖度近红外光谱定量分析 [J].浙江大学学报,2004,38(10):1371-1374.

总结与展望

9.1 本书的主要内容总结

本书主要介绍了本课题组基于近红外光谱分析技术在果蔬类农药残留量的检测、食用植物油品质、小麦粉、淀粉的品质检测中的应用研究。本书的主要研究结论：

1. 首先较系统地综述了近红外光谱法的分析特点以及发展应用过程

主要介绍了近红外光谱具有样品无需预处理、多组分同时测定、分析速度快、无污染、低成本、重现性好，效率高、对测试人员的要求不高，易培训、易推广等优点。同时介绍了近红外光谱的发展历程，近年来，随着近红外光谱仪器的不断更新，光谱预处理、定量校正模式及模型传递等近红外光谱分析技术得到了进一步的发展。为了更好地了解近红外光谱的应用，本章还重点介绍了近红外光谱技术在食品安全检测领域的研究进展，如近红外光谱在农残检测、食用油品质检测、面粉品质检测领域的研究进展。主要介绍了近红外光谱在这三大领域的应用现状以及产生的问题，并简单介绍了在这三大领域检测中所采用的预处理以及建模方法。其次介绍了基于近红外光谱的数据处理分析方法，主要包括光谱数据预处理方法以及多元校正方法。介绍了本课题的研究基础及主要课题。

2. 总结阐述了近红外光谱分析的理论基础和基本流程

在基本的理论基础介绍中提到了近红外光谱分析检测技术的两大类，即透射技术和反射技术。近红外光谱分析的基本流程，主要包括样品的预处理、模型的建立和预测。然后针对流程中所提到的预处理程序和模型建立，重点介绍了几种常用的预处理方法（如数据规范化处理、光程校正、基线校正、导数法、平滑处理、主成分分析法、马氏距离、杠杆值法、蒙特卡洛交叉验证法、傅里叶变换、正交信号校正、标准正态变量变换等）和模型建立方法（如多元线性回归、主成分回归法、PLS 法、ANN BP 法、聚类分析法、SVM 法等）。为了更好地进行预处理和建模，详细介绍了农药残留检测的前处理技术，包括超临界流体萃取、固相微萃取、微波辅助萃取技术、凝胶渗透色谱、基质固相分散萃取、固相萃取等。并提出了近红外光谱定标模型的评价指标，主要有 R^2 、RMSPCV、SEP。最后介绍了近红外光谱仪器的发展和应用。并且介绍了本研究的主要平台：VERTEX 70 傅里叶红外光谱仪、MPA 多功能傅里叶变换近红外光谱仪以及紫外分光光度计的性能指标和优缺点。

3. 主要介绍了农药残留量快速检测方法的研究

首先针对蔬菜中有机磷农药残留量的现状，提出了国内、国外关于食品检测方面的几项标准。然后重点介绍了国内外蔬菜中农药残留检测技术，主要包括色谱法、生物传感器法、生物化学测定法、发光菌法及免疫分析法等。对这几种方法的分类和检测原理以及应用现状进行了详细的介绍。最后提出了本书进行农药残留量检测方法时所采用的预处理和建模方法，以及使用的仪器。

近红外光谱技术在食品品质检测方法中的研究

主要探索运用近红外光谱定量分析模拟液体中的微量有机磷农药毒死蜱，通过设计不同的实验背景，逐步探究近红外光谱在测定溶液中微量毒死蜱的应用情况。实验设计按照毒死蜱在溶液中的浓度由高到低，依次为 3.13 ~ 10mg/kg、3.03 ~ 10mg/kg、0.005 ~ 0.1mg/kg。背景由简单到复杂，即由甲醇、水、毒死蜱的溶液到甲醇、水、维生素 C、蔗糖和毒死蜱的溶液。针对农药残留痕量分析研究，分别从近红外光谱技术方向进行了综合分析研究，以毒死蜱农药标准物质和炔螨特农药标准物质作为制作样品的研究对象，从农药水溶液的初步研究开始，基于近红外光谱的微量农药残留检测分析，通过对毒死蜱和炔螨特光谱数据进行不同预处理方法的优化后建立了 PLS 法数学模型和农药溶液的 ANN BP 模型，通过对比不经任何优化方法处理的光谱数据模型和其他常用优化方法，确定了光谱数据优化对建立模型的重要贡献：降低样品数据噪声和其他干扰。将其应用于萝卜农残检测试验中，各模型都得到了良好的评价参数，且具有很好的相关性，证明了近红外光谱技术可用于微量农药残留量的初步分析。

4. 研究了基于近红外光谱分析技术的蔬菜中农药残留量检测方法并建立模型

采用 PLS 法完成了模型建立及分析，阐述了应用近红外光谱分析法测定大白菜和菠菜的毒死蜱含量。配制后混合溶液样品，经过采样分辨率、采样次数及采样模式的选择，使用德国 Bruker 仪器公司的 VERTEX 70 型傅里叶变换红外光谱仪采集混合溶液样品的近红外光谱；其次采用 10 种不同的光谱预处理方法对样品的近红外光谱进行处理；最后经过波长优化、主成分数目的确定，建立了基于 PLS 法的近红外校正模型。另外本章还探索了基于近红外光谱的微量农药残留检测分析技术应用于萝卜农残检测试验中，各模型都得到了良好的评价参数，且具有很好的相关性。

通过以上实验表明，近红外光谱分析技术对蔬菜中农药残留检测具有巨大的潜力和应用价值，对实际农药残留快速检测具有一定的借鉴意义。

5. 采用近红外光谱技术结合模式识别方法建立食用油种类、花生油和橄榄油掺伪定性识别模型的过程

实验结果表明，采用近红外光谱与 SVM 技术建立了性能较好的花生油和橄榄油掺伪定性鉴别模型，对实验中预测集和测试集的预测率均能达到 100%；而采用近红外光谱与系统聚类法结合能有效地识别食用油种类。实验表明将近红外光谱 - 模式识别技术相结合作为一种快速无损鉴别食用油掺伪方法具有较强的可行性和实用性，这里工作的探索有力地推动了近红外技术应用于食用油品质定性快速识别的进程。

6. 采用近红外光谱技术结合化学计量学方法建立了食用油脂肪酸含量、过氧化值和酸价的定量分析模型

探讨了光谱分辨率的设定对食用油近红外模型性能的影响。实验结果表明，特征谱峰的选取结合适当的光谱预处理可以有效地提高近红外模型的预测性能，使食用油定量分析模型的实用性得到进一步提高。本章工作的探索有力地推动了近红外技术应用于食用油品质快速监管的进程。

7. 研究提出基于近红外光谱检测小麦粉品质的研究思路

首先利用近红外光谱分析技术结合 PLS 法定量分析小麦粉中水分、灰分及面筋三种有效成分。然后，为了排除异常样品的干扰，利用马氏距离法和蒙特卡洛采样法剔除异常样品来提高模型精确度。之后采用 K-S 法对样品集进行划分，为了提高模型的精度，尝试利用

常规区间偏最小二乘法、联合区间偏最小二乘法、向后区间偏最小二乘法和 CARS 算法进行特征谱区筛选,从而对小麦粉有效成分的定量分析模型进行优化。最后,引入非线性回归方法(SVM 和 ANN BP)对模型的稳定性及预测能力进行比较分析。解决了传统小麦粉检测时间长、操作复杂等难题。

8. 运用近红外光谱技术快速、无损的鉴别淀粉品质

采用聚类分析法和 SVM 的定性建模方法,对玉米淀粉和马铃薯淀粉进行了分类,实验结果表明,利用近红外光谱法可以准确地进行分类。同时对淀粉中水分的含量进行了定量分析,结果显示 R^2 为 99.55, RMSECV 为 0.0301,其训练结果较为精确。同时本章利用近红外光谱对马铃薯淀粉中掺杂玉米淀粉及淀粉中掺杂滑石粉进行了定量分析,实验结果表明,运用近红外光谱可以快速、准确地对淀粉品质进行定性和定量检测。

9.2 展望

实践证明,近些年来近红外光谱分析技术作为一种快速、无损、高效的分析技术受到越来越广泛的关注,该技术已在农业、食品、环境科学药物、医学、石化等多种行业中得到了广泛的应用。近红外光谱分析技术在农产品品质检测方面具有很大潜力。近红外光谱分析技术是一种“多快好省”的较为理想的现代分析技术。随着近红外光谱软硬件系统的不断改进和提高,近红外光谱的应用领域不断扩大,在很多领域得到了应用。

在今后的研究中,将开展①对多光谱信息融合检测方法的研究;②对多农药检测方法的研究;③对面粉、食用植物油的近红外模型传递的研究等。

参 考 文 献

- [1] 徐广通,袁洪福,陆婉珍.现代近红外光谱技术及应用进展[J].光谱学与光谱分析,2000,20(2):134-142.
- [2] 黄艳,王锡昌.近红外光谱分析在食品检测中的最新进展[J].食品研究与开发,2007,28(7):137-140.
- [3] 陆婉珍,袁洪福,褚小立,等.当代近红外光谱分析技术[M].北京:中国石化出版社,2000.
- [4] 徐广通,袁洪福,陆婉珍.现代近红外光谱技术及应用进展[J].光谱学与光谱分析,2000,20(02):134-142.
- [5] 隋淑霞.基于光谱技术的农药残留量智能检测方法研究[D].北京:北京工商大学,2008.
- [6] 郑光.基于近红外光谱技术的蔬菜中农药残留量检测方法研究[D].北京:北京工商大学,2010.
- [7] 索少增.农药残留量的多光谱检测方法研究[D].北京:北京工商大学,2012.
- [8] 李慧.基于近红外光谱的食用油品质检测技术研究[D].北京:北京工商大学,2011.
- [9] 吴胜男.小麦粉品质的近红外光谱检测方法研究[D].北京:北京工商大学,2013.
- [10] 董秀丽.基于红外光谱的小麦粉添加物检测方法研究[D].北京:北京工商大学,2013.

地址：北京市百万庄大街22号
邮政编码：100037

电话服务

服务咨询热线：010-88361066

读者购书热线：010-68326294

010-88379203

网络服务

机工官网：www.cmpbook.com

机工官博：weibo.com/cmp1952

金书网：www.golden-book.com

教育服务网：www.cmpedu.com

封面无防伪标均为盗版



机械工业出版社E视界 机械工业出版社微信公众号

上架指导 工业技术 / 食品工业

ISBN 978-7-111-52026-9

策划编辑◎顾谦 / 封面设计◎陈沛

ISBN 978-7-111-52026-9



9 787111 520269 >

定价：39.00元