

学校代码 10463

密 级 □

河南工业大学  
硕 士 学 位 论 文

基于 THz 技术的小麦品质无损检测研究

作者姓名	赵中原
指导教师	张元 教授
学科门类	工学
学科专业	计算机科学与技术
培养单位	信息科学与工程学院
完成时间	二〇一六年三月

## 关于学位论文的独创性声明

本人郑重声明：所呈交的论文是本人在指导教师指导下独立进行研究工作所取得的成果，论文中有关资料和数据是实事求是的。尽我所知，除文中已经加以标注和致谢外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含本人或他人为获得河南工业大学或其它教育机构的学位或学历证书而使用过的材料。与我一同工作的同志对研究所做的任何贡献均已在论文中做出了明确的说明。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：赵中原 日期：2016年6月14日

## 学位论文使用授权书

本人完全同意河南工业大学有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名：赵中原

日期：2016年6月14日

指导教师签名：张

日期：2016年6月14日

# **A Study on Nondestructive Testing for Wheat Quality Based on THz Technology**

A Dissertation Submitted for the Degree of Master

**Candidate: ZHAO Zhong-yuan**

**Supervisor: Prof. Zhang Yuan**

College of information science and engineering  
Henan University of Technology, Zhengzhou, China



## 摘 要

粮食安全关系经济发展和社会稳定。目前,在粮食质量安全检测领域,多数还在沿用人工检测的方法,虽然已有很多生化检测方法可以应用,但生化检测方法专业性强、操作程序复杂,很难在粮食检测领域推广。物理的特别是光谱的检测方法以其快速性和非破坏性受到人们的重视,成为一种很有前景的方法。但传统光谱学的方法仍然存在着设备昂贵、检测范围窄、精度差的缺点。考虑到太赫兹光谱是一种新型的光谱技术,且在生物化学领域取得了重大的应用成果,可以弥补现有光谱方法的很多不足,因此本文考虑将太赫兹光谱学的方法应用到粮食质量安全检测领域,试图做一些前瞻性的研究工作,以期有新的发现。

本文以四种不同品质小麦和 8 种不同种类小麦为检测对象,分别对它们的太赫兹光谱进行了研究,同时也对小麦储藏过程中常见的黄曲霉毒素的太赫兹光谱进行了研究,对提取的黄曲霉毒素的浓度进行了定量的分析。采集正常小麦、虫蚀小麦、发霉小麦和发芽小麦四种品质的小麦太赫兹吸收光谱,对光谱采用支持向量机结合主成分分析进行分类识别。

首先采用主成分分析法,对吸收光谱提取 4 个主成分,对折射率光谱提取 3 个主成分,然后把提取的主成分作为 SVM 的输入,利用 SVM 算法实现对四种小麦样品进行分类识别,结果表明,应用 PCA-SVM 模型,要比单独使用 SVM 模型,样品分类效果要好。同时,本文算法与其他太赫兹光谱分析常用方法的比较说明,本文的方法具有更高的分类识别精度。

对 8 种不同种类的小麦,采集太赫兹吸收谱和折射率谱,采用了偏最小二乘算法 PLS 构建回归模型,对 8 种小麦样品的种类进行了区分。通过实验,确定了小麦太赫兹光谱的最佳光谱区域,利用间隔偏最小二乘算法构建了 iPLS 模型,并与 PLS 模型进行了比较,证明了 iPLS 模型具有更好的区分性能。

对于小麦储藏过程中的黄曲霉毒素的检测,开发了一整套毒素提取方法,构建了毒素浓度与其太赫兹吸收光谱对应的线性回归模型。为了进一步提高毒素浓度预测值的可信度,基于多源信息融合的理论,提取多种太赫兹光谱对应频点的值作为证据,应用 D-S 证据推理算法计算预测浓度可信度,并把计算结果再次作为 D-S 证据推理的输入,提出了二重 D-S 证据推理算法,算法实验结果表明,本文的模型能够大大提高黄曲霉毒素浓度预测的可信度。

**关键词:** 小麦品质, 太赫兹光谱, 主成分分析, 支持向量机, 偏最小二乘, D-S 证据推理

## Abstract

Food Security concerns economic development and social stability. At present, in food quality safety detection, the main detection method is still manual method. Although there have been many biochemical detection methods, it is difficult to popularize in food domain due to complex operation programs. Spectrum methods greatly attract people's attention because of its speediness and non-destructive. But traditional spectrum have still many shortcomings such as expensive device, narrow detection range. Considering terahertz spectrum is a new spectrum technology, and has many applications in biochemistry, and can make up much insufficient for traditional spectrum, we tried to apply terahertz spectrum in food quality safety detection and find new rules.

In this paper, we took four different-quality wheat and eight sort of wheat as detection objects, and studied their terahertz spectrums respectively. At the same time, we also studied AFT B1's terahertz spectrum in wheat, and do a quantitative analysis for their concentrations. For normal wheat, worm-eaten wheat, mildew wheat and sprout wheat, the four types of wheat, we collected their absorption spectrum, and adopted SVM combining PCV model to identify their classifications.

At first, we used PCA to abstract four main components for absorption spectrum and three main components from refractive index spectrum as input variables for SVM, and use SVM to realize identification. The experimental results indicated that comparing with only using SVM, PCA-SVM has better performance for identification. In addition, compared with other methods, PCA-SVM has also better performance.

For eight sorts of wheat, we used iPLS method to construct regression model to identify them. Through doing experiments, we determined the best area in terahertz spectrum, and constructed a iPLS model, and by comparing with PLS model, we proved that the iPLS model has better performance than PLS.

For AFT in wheat, we developed a set of attraction method from wheat, and construct a regression model for forecasting their concentration. In order to improve forecasting performance, we improved D-S evidence inference algorithm. We attracted more frequency dots for fusion as evidence, and construct double evidence inference algorithm. The experimental results indicated that the model proposed in the paper can greatly improve the reliability for forecasting AFT's concentration in wheat.

**Key words:** Wheat quality, Terahertz Spectrum, PCA, SVM, D-S evidential reasoning

# 目 录

1 绪论.....	1
1.1 课题研究的背景和意义.....	1
1.2 国内外研究现状及存在问题.....	2
1.2.1 课题国内外研究现状.....	2
1.2.2 课题目前存在的问题.....	4
1.2.3 太赫兹技术及其应用现状.....	4
1.4 本文研究内容及章节安排.....	5
1.4.1 研究内容.....	5
1.4.2 章节安排.....	6
1.5 本章小结.....	6
2 THz 检测原理.....	7
2.1 THz 检测技术简介.....	7
2.1.1 THz 辐射及其特性.....	7
2.1.2 THz 辐射的产生与探测.....	8
2.1.3 THz 光谱技术.....	11
2.1.4 THz 光谱技术在农业领域中的应用.....	12
2.2 ZOMEGA-Z3 型 THz 时域光谱系统.....	13
2.3 光学参数的提取与计算.....	16
2.4 本章小结.....	18
3 小麦品质的 PCA-SVM 分类识别研究.....	19
3.1 小麦品质的定义及分类.....	19
3.1.1 小麦籽粒形态品质.....	19
3.1.2 营养品质.....	20
3.2 样品制备.....	20
3.3 样品的 THz 光谱分析.....	21
3.4 PCA-SVM 分类识别算法.....	23
3.4.1 PCA 算法.....	24
3.4.2 SVM 算法.....	24
3.4.3 PCA-SVM 算法.....	27
3.5 分类识别结果及分析.....	27

3.5.1 PCA 的分类识别结果.....	27
3.5.2 SVM 的分类识别结果.....	28
3.5.3 PCA-SVM 分析结果.....	29
3.5.4 PCA-SVM 与其他算法的识别结果比较.....	30
3.6 本章小结.....	31
4 小麦品质的 iPLS 分类识别研究.....	33
4.1 样品准备.....	33
4.2 PLS 算法和 iPLS 算法.....	33
4.3 样品的 THz 光谱分析.....	34
4.4 PLS 分析.....	36
4.5 iPLS 分析.....	37
4.6 模型与问题讨论.....	38
4.7 本章小结.....	38
5 储粮中黄曲霉毒素的 THz 定量分析.....	41
5.1 储粮真菌及其危害性.....	41
5.2 黄曲霉毒素太赫兹检测方法总体流程.....	41
5.3 黄曲霉毒素的提取和样品的制备.....	42
5.4 黄曲霉毒素的太赫兹光谱分析.....	44
5.5 多光学参量信息融合模型.....	47
5.5.1 信息融合模型框架.....	47
5.5.2 结合二重 D-S 推理算法的太赫兹光谱信息融合模型.....	48
5.6 实验结果与模型评价.....	50
5.7 本章小结.....	51
6 结论与展望.....	53
6.1 总结与结论.....	53
6.2 问题与展望.....	54
参考文献.....	55
致谢.....	61
个人简历.....	63

# 1 绪论

## 1.1 课题研究的背景和意义

安全储粮事关国计民生。确保储粮安全一直是国家与政府非常重视的问题。我国北方粮食主要以小麦为主，小麦在储藏期间，易受各种储粮害虫、霉菌损害，造成粮食产后减损严重，同时，储藏在大型粮库中的小麦发生霉变往往不易发现，并造成小麦品质的严重劣变，直接影响到人民群众的身体健康。因此，除了粮食监管部门及时有效地管理之外，研究开发可靠有效的粮食品质检测技术与系统 also 具有重要的意义。

关于粮食品质的检测，目前已经很多技术手段可用。从国内现状看，有人工检测、生化检测和物理检测方法。人工检测靠人的感官识别小麦品质，检测结果与检验员的经验有关，因而可靠性和稳定性较差。生化检测是用生物和化学的方法进行检测，存在着专业性强，检测速度慢，检测程序复杂的问题，难以在粮食行业推广应用。物理检测方法是指采用如光谱分析仪等物理检测设备对粮食品质进行检测。国内粮食品质检测技术与仪器的研发目前已取得了一定的进展，相继开发了小麦硬度指数测定仪、小麦粉加工精度测定仪、稻谷整精米率检测仪、高水分玉米水分快速检测仪、近红外粮食品质分析仪、谷物脂肪酸值测定仪、粮食感官检验图谱等一批重要的粮食检验仪器与设备，对粮食质量检验工作的进一步发展打下坚实的基础。

随着数字信号处理技术的空前发展，利用快速的检测技术客观评价粮食的最终用途，以促进粮食在市场上的快速流通，日益受到世界各国的广泛关注。快速、经济、无损的检测技术成为粮食品质研究的重要发展方向，近红外光谱、拉曼光谱、数字图像分析技术以及基于免疫技术的试剂盒等将越来越广泛用于粮食品质快速检验。

根据联合国粮农组织的统计，发达国家每年储粮损失达到 5%，而我国平均粮食产后损失约为 8%。造成粮食产后减损的主要原因之一是未能对储粮品质进行实时快速精准检测，导致储粮发生霉变、虫害和品质下降。同时，随着储粮时间的增加，粮食的品质随着发生改变。市场上时有出现以次充好的情况，一些不法商贩将陈化、霉变的粮食进行加工处理后重新流入市场。因此，粮食储藏和流通环节中的这些现状，迫切要求快速、精准、便携的检测设备的开发和推广，以保证粮食安全和正常的市场秩序。



## 1.2 国内外研究现状及存在问题

### 1.2.1 课题国内外研究现状

粮食在储藏过程中,随着粮食自呼吸作用的减弱、水分的散失,粮食自身理化特性发生变化,粮食逐渐陈化,另一方向,由于储藏条件不当,粮食也可能生虫、生霉,导致粮食陈化加剧,粮食品质发生劣变而无法食用。因此,及时对储粮品质进行检测,是保证粮食品质安全的一项重要措施。粮食品质检测的人工方法通常靠目测、牙咬等主观感受,可靠性难以保障。虽然可以通过复杂的生化检测方法检测粮食中含的蛋白质、脂肪、纤维素等,但均需要较高的专业知识和检测条件,一般储粮企业难以实施。事实上,粮食品质检测多集中在虫害、霉菌等这些储粮生物有害物的检测上,因为这些生物有害物是构成粮食品质劣变的主要因素。

储粮虫害的检测,多采用目测法、筛选法和诱集检测法<sup>[1]</sup>,可靠性较差。由于粮虫多集中在粮堆表面活动,便于观察和探测,近年来发展起来了及其视觉、图像处理、近红外光谱等非接触、非破坏性的物理检测方法,使得检测效果大大提高,已经能够满足实际应用的要求。对于储粮中真菌的检测,特别是厌氧类真菌(如酵母菌)的检测,目前还有诸多难题。粮食霉变是造成粮食产后损失的主要原因,且粮食霉变产生的毒素对人的健康影响很大,因此对储粮真菌的检测,特别是真菌早期活动的检测,是粮食品质检测的一个重要内容。

目前国内的真菌检测方法大多数仍沿用传统检测方法,但现代生物学方法的技术也已成熟并开始应用。

传统方法主要包括稀释平板计数法、显微镜计数法等<sup>[2-4]</sup>,目前仍在我国大部分地区使用。传统方法虽然有检测费用低的优点,但其检测周期长,准确性差,不能满足用户需求快速、精准检测的要求。

针对以上的问题,人们对传统方法进行了改进,先后开发出了纸片法、螺旋接种法、表面荧光滤膜法、等格滤膜法等<sup>[5-8]</sup>,这些方法虽然提供了检测的效果,但仍需较长的检测时间,且对于早期的真菌(活动初期的真菌)检测没有考虑。

真菌在生长过程中,会产生一些代谢物质或挥发物质,对这些物质(如麦角甾醇、细胞壁多糖、各种酶类等)进行检测<sup>[9-17]</sup>,也是检测真菌的一种有效的方法,国外研究机构对此报道较多,但多数成果处于研究阶段,实际应用较少。

物理检测方法检测速度快、安全可靠，具有自己的特色优势，是极受欢迎的一种检测方法。在真菌检测领域，这种方法极具发展前景。目前物理检测方法主要有传感器方法和光谱学方法。

### (1) 传感器法

针对真菌等微生物检测的传感器又称为电子鼻。Borjesson 等人用 MOSFET 传感器、SnO<sub>2</sub> 半导体和红外测量器对稻谷的霉变过程进行了检测<sup>[18-19]</sup>，发现传感器在区分 4 种霉变程度的分类时，正确率达到 75%。Keshri 等人运用传感器法区分出了以小麦为原料的食品中已发芽的孢子和已生长了 48 小时的真菌<sup>[20]</sup>。A. Jonsson 等人用传感器和人工神经网络(ANN)对小麦、黑麦和大麦进行了真菌检测和霉变程度的分类<sup>[21]</sup>，识别率达到 83%。Borjesson 和 Olsson 运用 10 个 MOSFET 传感器、6 个 MOS 传感器和一个 CO<sub>2</sub> 传感器对 40 个大麦样品中的真菌毒素进行了检测<sup>[19]</sup>，检出 86% 的样品受到真菌污染，高于靠人工气味辨别近 20 个百分点。

电子鼻的优势是适合在线实时检测。但电子鼻在检测真菌挥发物时，易受到环境等各种外界因素(如气流、通风情况等)影响，因而检测的可靠性大大降低，严重时失去应用价值。实验室的电子鼻检测，存在着样品代表性不足、无法实时在线的问题。总体来说，目前传感器法仍处于研究阶段，还难以实用，且检测结果不理想。

### (2) 光谱学的方法

应用光谱学的方法检测微生物及其代谢物始于 20 世纪 50 年代。目前在生物学方面的研究主要集中在药物与活细胞的相互作用上<sup>[22]</sup>，近年发展较快。柴阿丽、李金萍等利用红外光谱对真菌进行检测，发现各类群菌体的红外光谱具有特异性，可作为对真菌进行鉴别的依据<sup>[23]</sup>。在近红外(NIRS)的应用方面，Whitlow 等运用 NIRS 检测玉米青贮饲料，检测出 8% 的样品的黄曲霉素含量  $>10\mu\text{g} \cdot \text{kg}^{-1}$ <sup>[24]</sup>。Hans 对 NIRS 检测麦粒中的赤霉烯酮进行了研究，在波长 670-1100nm 范围内，发现具有较高的相关系数<sup>[25]</sup>。Jan, Giorgio 等人也对真菌的合成物—麦角甾醇进行了测定<sup>[26,27]</sup>。

红外光谱法由于其特征峰与生物分子的官能团有着对应关系，因而可以探测小麦的营养成分的组成结构，目前已得到广泛应用。但由于红外辐射穿透性较差，无法探测到小麦表层下的真菌初发情况，因而不能直接做小麦霉变早期的检测，而且往往需要将小麦粒破开进行检测，是一种破坏性检测技术，因而给检测带来了一定的麻烦。

### (3) 生物光子辐射检测法

光子辐射是生命的一个普遍现象，任何生命体都在向外辐射光子。近年来，生物光子学理论发展很快，相应地产生了针对有机体的生物光子分析技术(简称 BPAT)。目前，该技术主要针对人体的生物光子辐射<sup>[28]</sup>，以及人体癌变组织的生物光子辐射<sup>[29]</sup>和种子的发芽能力<sup>[30]</sup>。由于这种技术对检测环境要求极高，相应的信号处理难度很大(主要是目标发出的光子和非目标物体发出的光子的区分)，因而其应用受到很大限制，在粮食行业的应用还不具有可行性。

20 世纪 80 年代以来，随着基因技术的发展，真菌毒素的检测技术进入到分子水平，如出现了酶联免疫法、基因芯片法等<sup>[31-35]</sup>等现代生物学的方法。基因芯片法和酶联免疫吸附法具有最低的检测限和最优的识别率，但需要开发针对不同真菌的基因芯片和酶，具有极强的专业性，不适合粮食行业的日常检测。为适应储粮检测市场需求，还需寻找简便、精度高、周期短、费用低、能进行早期检测的新理论、新方法<sup>[36-37]</sup>，本课题正是为解决这一科学问题而提出的。

#### 1.2.2 课题目前存在的问题

小麦霉菌的早期检测，是小麦防霉、治霉的重要保证。目前的检测技术虽然很多，但适合粮食企业做大面积霉菌普查检测的技术并不多。一般说来，生化检测方法普遍存在着检测周期长、检测程序复杂、检测人员专业性要求高的问题，快速的生化检测方法检测精度往往较低，很多只能定性检测，对小麦防霉治霉指导作用有限。物理检测方法是一种很有发展前景的方法，且易于实现在线。但在目前的物理检测方法中，电子鼻的检测方法受环境影响较大，红外光谱法在很多情况下都是一种破坏性的方法，且现有物理检测方法检测项目单一，难以在深度和广度上全面反映检测对象的状况。因此，适用于粮食行业的检测技术，还需要进一步的开发和完善，太赫兹光谱作为一种新型的光谱技术，由于其优良的技术特点，可以弥补传统检测方法的诸多不足，因而本文将太赫兹技术作为研究的手段，对储粮小麦的品质检测提出了一整套新的检测方案。

#### 1.2.3 太赫兹技术及其应用现状

太赫兹波 (0.1-10THz) 是处于半导体电子器件和光子器件之间的波段，受到太赫兹的产生、采集等技术限制，直到上世纪 80 年代末得以推广应用。在上世纪 80 年度以前，由于缺乏有效的太赫兹波产生和检测的技术，太赫兹波段的应用开发一直处于空

白，形成所谓的太赫兹空隙(THz Gap)。近 30 年来，随着激光技术、超快技术、半导体技术的发展，太赫兹的光源和接收检测问题逐步得到解决，太赫兹辐射的应用也逐渐广泛开展起来。目前发达国家太赫兹技术已经开始广泛应用，而我国的太赫兹技术的应用还处于起步阶段。太赫兹波具有良好的穿透性和宽带性，且其频率刚好落在大多数生物分子集体振动和转动频率范围内，这就使太赫兹辐射针对生物分子的检测具有特殊的优势：可以携带更多关于生物分子的内部构成、构象的信息。目前，太赫兹光谱技术已逐步应用到生物制药、医学诊断、材料科学、反恐安检、食品质量安全检测等多个领域。

针对太赫兹技术在粮食中的应用，英国曼彻斯特大学、美国杜克大学、日本东京工业大学等单位分别针对特定食品中某些成分含量进行了检测和分析，得到了大量有用数据；在国内，首都师范大学、北京科技大学、中国科学院电子学研究所等科研院所也开展了相关研究，但都是在一些特定应用场合，针对某种特定物质的检测，目前还没有科研单位开展各类粮食中不同物质成分在太赫兹波段光谱特性的国家标准的研究和制定工作。

## 1.4 本文研究内容及章节安排

### 1.4.1 研究内容

课题以小麦为研究对象，通过采用太赫兹时域光谱技术，对正常小麦、发芽小麦、虫蚀小麦以及霉变小麦进行分类识别研究，对不同地域、不同品种小麦进行分类识别研究，同时对粮食中的黄曲霉毒素进行了定量研究。具体研究内容如下：

1. 研究太赫兹时域光谱系统(THz-TDS)的搭建和光学参数的提取以及样品的制备方法。
2. 研究正常小麦、发芽小麦、虫蚀小麦以及霉变小麦的太赫兹光谱分析。利用主成分分析对太赫兹光谱数据进行特征提取，利用支持向量机对提取的特征进行分类识别，同时与神经网络、主成分分析、支持向量机单独使用时分类结果进行比较研究。
3. 研究利用偏最小二乘(PLS)和间隔偏最小二乘(iPLS)对不同地域、不同种类的小麦进行识别的方法，建立基于 iPLS 的小麦种类识别线性回归模型。
4. 利用多源信息融合和证据理论对储粮中的黄曲霉毒素进行定量分析。

## 1.4.2 章节安排

第1章绪论。阐述了课题的研究背景和意义，重点介绍了应用物理检测方法进行储粮品质检测的技术现状，特别是储粮真菌的检测技术现状，总结了目前检测技术的不足。简单介绍了太赫兹技术及其应用现状。简要综述了本文的研究内容。

第2章太赫兹检测原理。本章主要讨论了太赫兹检测的机理。包括太赫兹的概念和特点，太赫兹的产生和探测原理，太赫兹波技术应用的进展，特别是太赫兹波在农业领域中的研究进展。本章还对课题所使用的实际的太赫兹时域光谱系统(ZOMEGA-Z3)进行了原理介绍，阐述了液态样品和固态样品的光学提取参数和提取方法。

第3章小麦品质的 PCA-SVM 分类识别研究。分别对四种小麦样品(正常小麦、发芽小麦、虫蚀小麦和霉变小麦)应用 ZOMEGA-Z3 系统进行太赫兹光谱测量。给出了四种样品的太赫兹时域谱、频域谱、吸收系数谱和折射率谱。对各样品光谱进行了分析。分别应用 PCA、SVM、PCA-SVM 对四种小麦样品进行识别，并将识别结果与 BP 算法进行比较，通过比较实验，支出 PCA-SVM 算法具有最好的识别结果。

第4章小麦品质的 iPLS 分类识别研究。本章主要采用 PLS 算法和 iPLS 算法对不同地域、不同品种的小麦进行识别研究。

第5章储粮中黄曲霉毒素的 THz 定量分析。本章主要对黄曲霉毒素的定量分析的太赫兹光谱的实现方法进行了研究。主要讨论了黄曲霉毒素在粮食中的存在现状，黄曲霉毒素的提取方法和样品制作方法。黄曲霉毒素的标准谱库的创建方法。利用信息融合技术和 D-S 证据理论对黄曲霉毒素浓度的识别。

第6章结论与展望。对论文的主要工作和取得的结果进行了总结，根据研究结果得出课题的主要结论。并对课题的研究预期进行了预测，对本领域的研究前景进行了展望。

## 1.5 本章小结

本章阐述了课题的研究背景和意义，综述了国内外粮食质量安全检测的研究现状，重点讨论了粮食品质物理学检测的技术现状和存在的问题，并对太赫兹技术的应用现状做了简明的介绍。最后，简要介绍了本文的研究工作和各章节的安排。

## 2 THz 检测原理

### 2.1 THz 检测技术简介

#### 2.1.1 THz 辐射及其特性

太赫兹(Terahertz)辐射是指对位于微波和红外辐射之间的一段电磁波辐射的统称。它的震荡频率在  $10^{12}$  Hz ( $1\text{THz}=10^{12}\text{Hz}$ ) 左右。该波段是由从电子学领域向光学领域过度的一段波段, 又被称为远红外辐射, 而我们一般所称的太赫兹波段, 是指频率在  $0.1\text{T}-10\text{T}$  范围的电磁波, 在有些特殊场合, 特指频率在  $0.3-3\text{T}$  范围内的电磁波, 本文所称的太赫兹波就是特指这个频率范围内的电磁波。图 1 给出了太赫兹波段在电磁波谱中的位置。

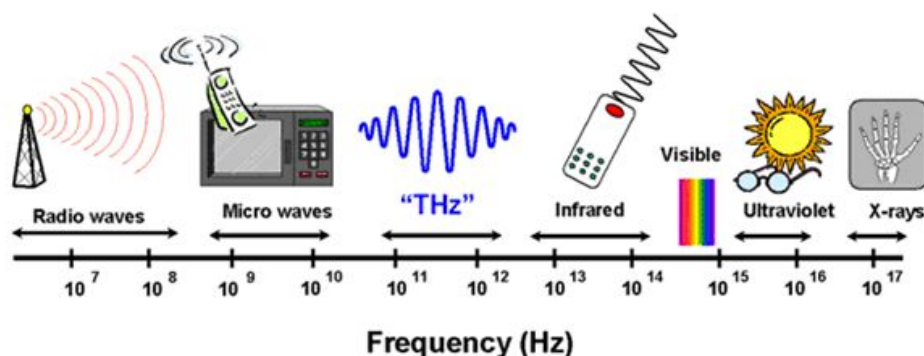


图 1 电磁波谱与太赫兹波段

尽管太赫兹辐射在自然界中普遍存在, 对太赫兹波段的利用却是近 30 年的事。在 20 世纪 80 年代以前, 由于缺少高效率的太赫兹发射源和灵敏的太赫兹探测器, 限制了太赫兹技术的研发应用, 而太赫兹两端的微波和红外波段已经得以广泛应用, 从而使太赫兹成为整个太赫兹波段中不为人们所熟悉的一段“空白”, 形成所谓“太赫兹间隙 (THz gap)”。近二十年, 随之超快光电子(optoelectronic)技术的发展, 太赫兹的发射源和探测器技术得以取得突破, 太赫兹技术的应用随之也快速发展起来。目前太赫兹技术已广泛应用于航天、安检、材料检测、医学诊断、生物制药等多个方面。

太赫兹目前形成广泛的应用得益于它具体很多独特的性质, 正是这些独特的性质赋予太赫兹辐射几个重要的特性, 下面简单介绍一下这些特性:

#### (1)高透性

太赫兹对很多介电材料和非极性液体有很好的穿透性, 这可以使太赫兹对不透明的



物体进行透射成像，如对信封中的物品进行成像。太赫兹能够穿透常见的包装物，如纸张、纸板、塑料、陶瓷、织物等，这使得太赫兹可以作为 X 射线或超声波成像的有力补充，应用在安全检查、医学影像诊断上。

### (2)安全性

太赫兹辐射的另一个显著的特点就是它的安全性。太赫兹辐射的光子能力非常低，只有毫电子伏，低于各种生物分子的化学键的键能，不会对生物分子的结构造成破坏。因而利用太赫兹对人体进行检查，不会对人体造成辐射损害。

### (3)光谱分辨特性

太赫兹波段处于大多数生物分子转动和振动(包括集体振动)能级范围内，使得生物分子在这一波段显现出强烈的吸收和色散特性。因而经过样品的太赫兹波包含了样品丰富的光谱信息，我们可以利用太赫兹波所包含的光谱信息，对样品的物化特性进行鉴别，甚至能够鉴别样品的组成成分或者种类。

## 2.1.2 THz 辐射的产生与探测

由于缺乏太赫兹波段的高效的发射源和灵敏的探测器，太赫兹波段一直未能得以很好的利用，形成所谓的太赫兹间隙(THz gap)。近二、三十年来，随着激光超快技术的发展，高功率、低成本的便携式太赫兹发射源被开发出来，同时太赫兹波的探测技术也得到解决。现在太赫兹辐射的产生主要有窄带连续波技术和宽带脉冲技术两种方式。窄带连续波技术包括：自由电子激光器、非线性光学混频技术、热辐射源泉技术等<sup>[38,39]</sup>；宽带脉冲技术主要有：光整流、等离子体震荡、光电导天线等，其中光整流、光电导天线应用得比较广泛<sup>[40,41]</sup>。宽带脉冲技术的特点是：原理简单、易于实现的产生装置等。宽带脉冲技术在太赫兹时域光谱系统(THz-TDS)中的应用较为广泛。

目前使用最为广泛的太赫兹发射源是光电导天线，该项技术由 Auston<sup>[42, 43]</sup>在 20 世纪 90 年代初提出。其基本原理是：在砷化镓(GaAs)晶体表面涂敷金子膜制成偶极子天线，天线经超快激光脉冲照射后，当入射光子能量大于晶体禁带宽度，光子被吸收，晶体产生大量载流子。当在外加偏置电场作用下时，载流子就会加速运动，形成瞬变光电流，从而激发出太赫兹脉冲并向外辐射。太赫兹波产生的原理如图 2 所示。

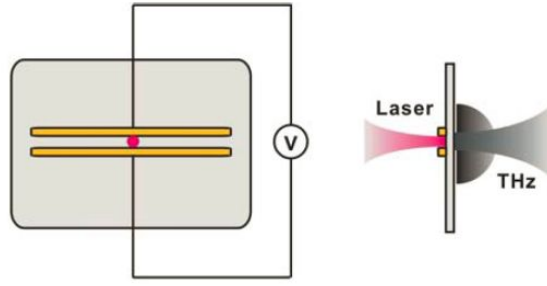


图 2 光电导天线以及脉冲 THz 辐射

20 世纪 90 年代，张希成等人<sup>[44]</sup>提出基于光整流效应的 THz 脉冲产生方法。当具有二阶非线性性质的材料与光相互作用时，由于差频效应，会产生一个与光强成正比的直流电场，称该现象为光整流效应。光整流效应的原理可描述如下：

一个二阶非线性过程可表示为：

$$P(0)=\chi^{(2)}(\omega,-\omega, 0) E(\omega) E^*(-\omega) \quad (2.1)$$

入射激光脉冲为：

$$E(t)=A(t) \cos(\omega t) \quad (2.2)$$

上式中， $P$  为电极化强度， $\chi^{(2)}$  为二阶非线性极化率， $\omega$  为基频， $E(\omega)$  是  $E(t)$  的傅立叶变换。如以光脉冲激发，产生的直流电场就是一个与光脉冲的包络相关的时间的函数。材料内部将产生一个交变的电场，该电场激发电磁波。其电场与该交变电场对时间的二阶微分成正比。如式(2.3)所示，其中  $t$  表示时间， $0$  表示频率为零。

$$E(t) \propto \frac{\partial^2 P(0,t)}{\partial t^2} \quad (2.3)$$

基于光整流效应产生太赫兹波的原理如图 3 所示：利用超短激光脉冲入射到非线性介质(如 ZnTe、LiTaO3、GaSe 等晶体)中，产生低频电极化场，从而可以辐射出脉冲 THz 电磁波。

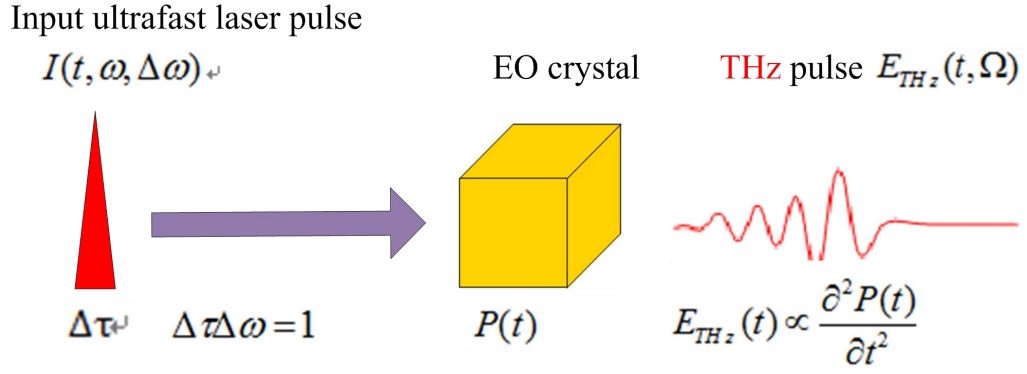


图 3 光整流效应产生 THz 波示意图

由于太赫兹辐射分为连续波辐射和脉冲辐射，因而其探测方法也分为连续波的探测和脉冲波的探测。在探测赫兹时域光谱系统(THz-TDS)中，常用的两种太赫兹波探测方法为：光导取样法(PC Sampling)和电光取样法(FS-EOS)，两种方法的原理如图 4 所示。

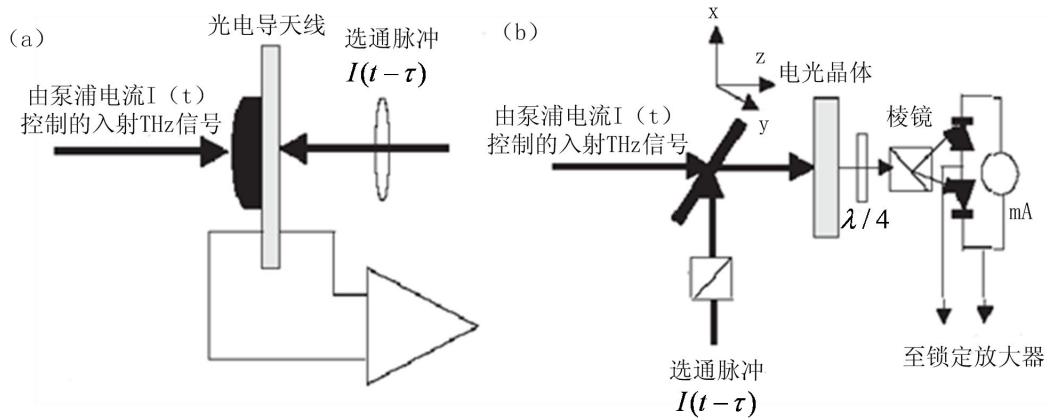


图 4 光电导取样和电光取样<sup>[38]</sup>

光导取样法采用光电导天线探测太赫兹波<sup>[45]</sup>。其基本原理是：将飞秒激光脉冲分路，产生一路探测光，照射到电光晶体上，产生载流子。当太赫兹脉冲也照射到电光晶体上时，相当于给载流子增加了偏置电压，使载流子加速运动，产生瞬时电流，从而激发出太赫兹电场。

电光取样法依据电光晶体内普克尔效应，由张希成等人<sup>[44]</sup>在 20 世纪 90 年代提出并研制出来，它实际上是张希成等人提出的光整流技术的逆过程，目前有逐渐替代光电导技术的趋势。电光取样法的基本原理是：将太赫兹脉冲和探测激光共线，同时穿过电光晶体(如锗化锌)，产生普克尔效应。即载有样品信息的太赫兹电场在电光晶体中引发电光效应，并对其折射率椭球进行调制，使探测光的偏振态随之改变。通过测量探测光

的偏振态，可间接得到太赫兹脉冲的波形。

### 2.1.3 THz 光谱技术

太赫兹科学技术有两大基本应用：太赫兹光谱和太赫兹成像。本文主要研究太赫兹光谱在粮食检测领域中的应用。目前太赫兹光谱技术有三种：时域光谱技术、发射光谱技术和时间分辨光谱技术。本文采用的是太赫兹时域光谱技术，英文缩写为 THz-TDS。

#### 1. 太赫兹时域光谱技术

太赫兹时域光谱系统由飞秒激光器、太赫兹发射模块、探测模块及时间延迟线组成。根据样品和测试要求的不同，还可以配置成反射式和透射式两种系统。太赫兹时域光谱系统结构如图 5 所示。

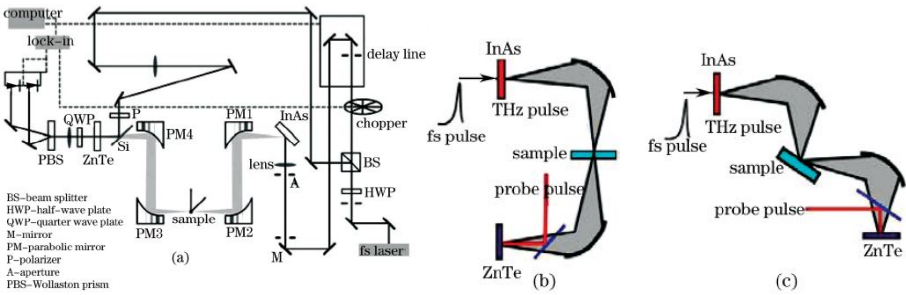


图 5 太赫兹时域光谱系统 (a)系统结构 (b)透射式 (c)反射式

THz-TDS 的工作原理是：飞秒激光经分束镜后，被分成功率较强的泵浦光和功率相对较弱的探测光。泵浦光经时间延迟线后，入射到太赫兹发射晶体上，产生太赫兹脉冲，太赫兹脉冲再经聚焦后入射到探测晶体上。探测光经多次反射后，与太赫兹脉冲共线入射到探测晶体上。延迟线的作用是保证探测光和太赫兹脉冲的相位一致。通过探测晶体的太赫兹脉冲电场改变了太测光的偏振态，经 1/4 波片后，再经偏振分束镜分束，最后经双眼光电探头引入至锁相放大器，最后经由计算机数据采集，间接得到太赫兹脉冲电场的变化波形。

#### 2. 太赫兹发射光谱技术

典型的太赫兹发射光谱系统原理图如图 6 所示。飞秒激光的绝大部分能量用于激励样品，使之产生光生电流，或产生光生极化，从而辐射出太赫兹脉冲。而飞秒激光的小部分能量用于产生探测光，探测光不是用来探测样品的太赫兹光学特性的，而是需通过垂直欺骗器，基于自由空间的电光取样技术探测由辐射出的太赫兹脉冲产生的自由空间电磁瞬变<sup>[46]</sup>。

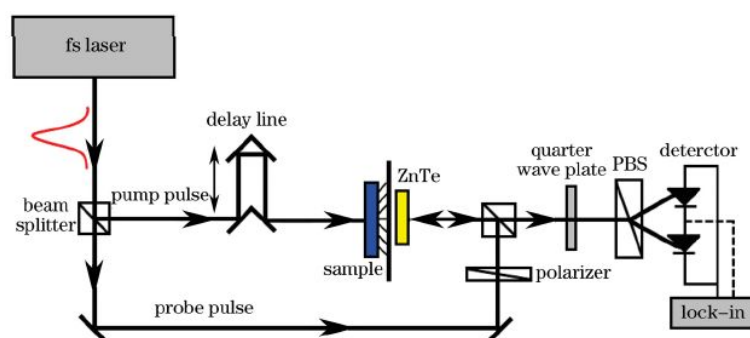


图 6 太赫兹发射光谱技术

### 3. 太赫兹时间分辨光谱技术

太赫兹时间分辨光谱技术是 THz-TDS 技术和光泵浦技术相结合的一种非接触电场探测技术。通过该项技术，我们可以观测到样品的动态变化信息，其时间分辨率最高可达 200fs。

以上三种光谱技术都是宽带线性光谱探测技术。相对于傅立叶变换红外光谱及时，太赫兹光谱技术具有如下优势：(1)传统的傅立叶变化红外光谱是非相干测量，只能得到功率谱，而太赫兹光谱属于相干测量，不仅可得到功率谱，还可以得到相位信息。尤其需要指出的是，通过测量参考信号和简单计算，我们还可以得到反映样品重要理化特性的折射率谱和吸收系数谱。(2)太赫兹光谱的信噪比和稳定性要好于红外光谱。(3)太赫兹光谱的动态范围高于红外光谱。

#### 2.1.4 THz 光谱技术在农业领域中的应用

##### 1. 水分检测

水分子是典型的极性分子，由于其集体振动频率处于太赫兹范围，因而对太赫兹光子能量有强烈的吸收。太赫兹波在水中的透射深度为  $100\mu m$ ，水在 1THz 的吸收系统达  $150cm^{-1}$ 。由于水对太赫兹的强烈吸收作用，可以利用这一特性对食品或农产品的微量水进行检测和分析。S.Hadjiloucas 等人<sup>[47,48]</sup>早在 1999 年就开展了这项工作，他们在 0.1-0.5T 的频率范围内分别研究了不同叶子的含水量。随后，利用太赫兹光谱技术检测水分含量的研究陆续报道出来。

##### 2 育种方面

利用太赫兹技术可以对种子类型和质量进行识别。由于种子是复杂的有机混合物，没有特征吸收峰出现，因而通常不能利用指纹谱进行识别，但一般不同类型和质量的种子的折射率都会发生变化，可以利用这一参数的差异对种子的真伪进行鉴别。这一方面

的研究可以参考孙金海等人<sup>[49,50]</sup>的研究。

### 3 农药残留的检测

农产品的农药残留问题是一直受到很大关注的食品安全问题。太赫兹辐射对很多农业分子具有特征吸收峰，通过密度泛函理论和量子化学软件对特征吸收峰进行确认，就可以判断出农产品中是否有某种类型的农药残留。

### 4 生物分子的识别

由于太赫兹光子能量只有毫电子伏特，不会破坏有机分子中的化学键(如氢键)，且大多数生物分子的集体振动和转动频率处于太赫兹波段范围内，因而太赫兹波段是研究生物分子特性的理想光源，可作为农产品品质检测的重要技术手段。

## 2.2 ZOMEGA-Z3 型 THz 时域光谱系统

太赫兹是 21 世纪最有发展潜能的高科技。太赫兹光谱仪是目前最具发展潜力的一个新兴产业，这也是继拉曼光谱仪、远红外光谱仪后的又一个在特殊波段研究物质与电磁波相互作用的光谱仪。太赫兹技术的研究起始于上个世纪的欧美高校研究所，2005 年开始的欧美高校研究所与企业联手开发太赫兹光谱仪，其中美国本土的科技企业占一半，Zomega 公司为代表的开发了宽带太赫兹光谱仪；应该 Teraview 公司的研发销售能力更好些，在医学、半导体材料等领域开拓了巨大的全球市场。除此之外，立陶宛、德国、日本、加拿大等国也相继成立公司进行太赫兹光谱仪的研发与销售。而中国在太赫兹技术的研究起始于上个世纪九十年代末，在太赫兹光谱和成像方面，首都师范大学的太赫兹教育部重点实验室团队的成果受到国内外的一致认可。2014 年年初，河南工业大学粮食信息处理与控制教育部重点实验室斥资 149 万元引进美国 ZOMEGA 公司 Z3 型实验室级太赫兹时域光谱和成像系统(THz-TDS)。该系统的引进使河南工业大学具备了独立研发太赫兹应用技术的实力，同时也为本课题的实施提供了硬件上的支撑平台。该系统原理性结构如图 7 所示。



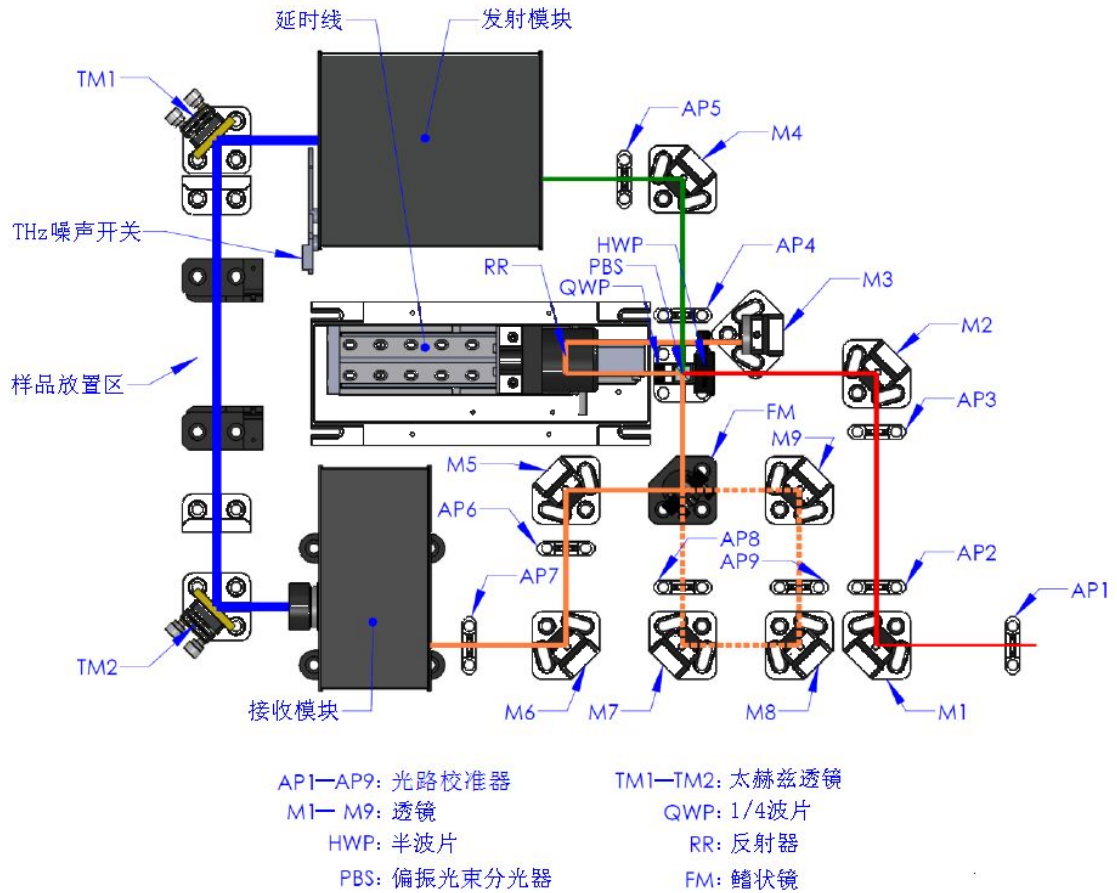


图 7 THz 时域光谱仪的基本结构

输入光束(飞秒激光束)用红色表示,而泵浦光,探测光和太赫兹光以绿色、橙色和蓝色线分别标示出来。使用光路校准工具可在选择点(AP1-AP9)上检查光路是否发生偏移。

飞秒激光束(红色)穿过机箱后板孔,由两个转向镜(M1 和 M2)引导进入半波片(HWP)和偏振分束器(PBS),将激光束分为泵浦光(绿色)和探测光(橙色)。泵浦光和探测光的功率比例由半波片的旋转角度控制,最优比例依输入功率调整。本仪器探测光的功率约为 15m W, 剩余功率用于泵浦。

探针光束经由 PBS 和四分之一波片(QWP, 以 45 度角固定安装以产生圆偏振光)传输后, 从安装于延迟线中的反射器(RR)反射至透镜 M3, 再由 M3 沿着同样的光路经由 RR 反射回 PBS, 再经由 PBS 到达鳍状镜 FM。

延时线中的 RR 的位置可以调整, 延时线的作用是: 保证探测光和由泵浦光产生的太赫兹光在到达接收模块时相位一致。

探测光经由鳍状镜 FM 后, 再通过 M5 和 M6 被反射到接收模块中。但需要注意的是, 当测量长度超过 350 毫米(1.2 ns)时, 就要折叠鳍状镜(FM), 使探测光经由额外的延

时(由 M7-M9 橙色虚线标示出来)后, 再由 M5 和 M6 导向到接收模块中。

经 PBS 分束的泵浦光通过 M4 被导入到发射模块。发射模块包含光电导天线,泵浦光束和太赫兹光束的聚焦器件。

发射模块的内部结构如图 2.8 所示, 包含了一个用于产生太赫兹的发射器(PCA).泵浦光(绿色)从右边进入, 被 M1 和 M2 反射后被引入透镜 L1,L1 把泵浦光聚焦到 THz 发射器(PCA)上, 发射器发射一个宽带的 THz 脉冲。这个 THz 脉冲被 L2 聚焦后, 从发射模块中射出, 经 TM1z(如图 1)反射, 经过实验区(该区域可放置样品), 然后被 TM2 反射, 最后进入接收模块。图 8 中的 Shutter 用于测量噪声时阻塞 THz 波。

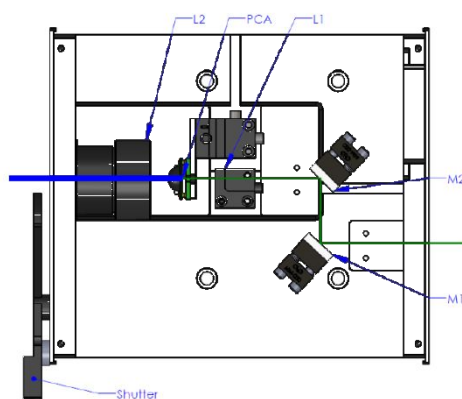


图 8 发射模块的内部结构

接收模块, 如图 9 所示。包含一个电光 (EO) 晶体 THz 检测器, 一个自平衡光电二极管检测器(图中没有画出),一个 650nm 碟用激光二极管(LD).THz 脉冲从模块的前面进入并被一个 50mm 厚的透镜 L1 聚焦。探测光(橙色)从接收模块的后面进入并被透镜 L2 聚焦, 经由 M1 被引导到 ITO(纳米铟锡,一种 N 型氧化物半导体), ITO 可以把太赫兹波束和探测光束整合起来使它们共线传输, 并在电光晶体(EO)的表面被聚焦, 一个四分之一波片(QWP),沃尔什棱镜(WP)和一个平衡检测器(BD)被用于探测 THz 波束的电场, 这个电场在电光晶体中产生了极化改变。

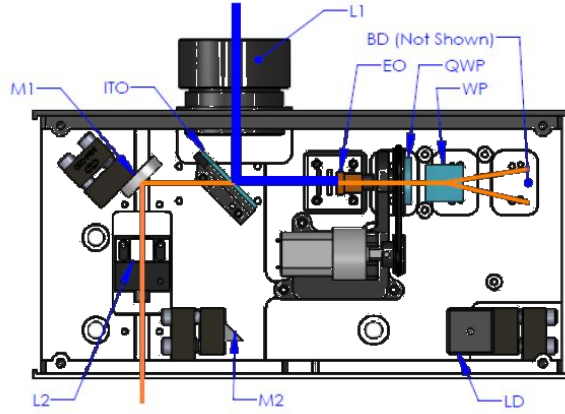


图 9 接收模块的内部结构

该产生了极化改变的电场被锁相放大后，引入计算机显示出太赫兹波形。该波形由于是经过了样品后的波形，因此携带了光与样品的丰富的物理化学信息。

## 2.3 光学参数的提取与计算

通过太赫兹光谱不仅能够得到样品的时域谱和频域谱，还可以通过简单的计算，得到样品的折射率谱和吸收系数谱，这是红外和其他光谱所不具有的。样品折射率可用如下关系表示：

$$N(\omega) = n(\omega) - jk(\omega) \quad (2.4)$$

其中， $N(\omega)$ 表示样品复折射率， $n(\omega)$ 为实折射率， $k(\omega)$ 为消光系数，表示了样品的吸收特性。

据文献[51-53]可得到复透射函数为：

$$H(\omega) = \frac{E_s(\omega)}{E_{ref}(\omega)} \quad (2.5)$$

其中 $E_{ref}(\omega)$ 为空载时光谱(参考光谱)， $E_s(\omega)$ 为样品光谱。如图 9(a)所示。

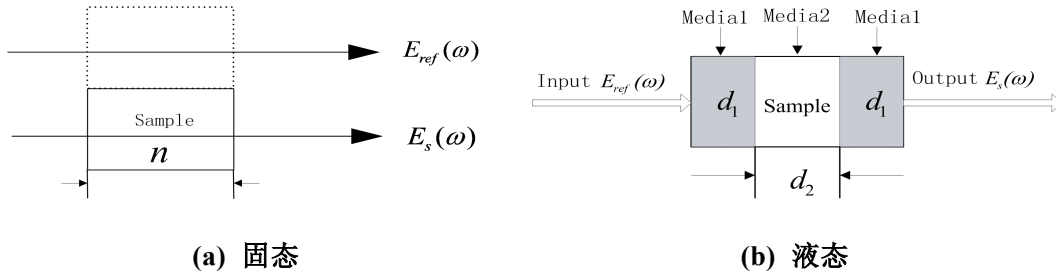


图 10 固态和液态样品的太赫兹传输模型

当样品厚度  $d$  较大且样品吸收较弱时,忽略 Fabry-Perot 效应,样品的折射率  $n(\omega)$  和吸收系数  $\alpha(\omega)$  可分别表示如下:

$$n(\omega) = \frac{\varphi(\omega)}{\omega d} c + 1 \quad (2.6)$$

$$a(\omega) = \frac{2}{d} \ln \left[ \frac{4n(\omega)}{\rho(\omega)(n(\omega)+1)^2} \right] \quad (2.7)$$

式中,  $c$  为光速,  $d$  为样品厚度,  $\omega$  为角频率。  $\varphi(\omega) = \varphi_s(\omega) - \varphi_{ref}(\omega)$ , 为样品与参考信号相位差,  $\rho(\omega)$  为样品与参考信号幅值比。

若样品为液体, 则其透射模型如图 9(b)所示。这时在计算  $n(\omega)$  和  $\alpha(\omega)$  时需要考虑容器影响。根据文献[54]计算方法如下:

$$n_2(\omega) = \frac{\arg[H_{Measure}(\omega)]c}{\omega d_2} + n_0 \quad (2.8)$$

$$a_2 = \frac{2}{d_2} \ln \left[ \frac{n_2(n_1 + n_0)^2}{|H_{Measure}(\omega)|(\tilde{n}_1 + n_2)^2 n_0} \right] \quad (2.9)$$

式中  $H_{Measure}(\omega)$  为系统传递函数。

上式中,  $\tilde{n}_1$  为容器的折射率, 计算方法如下:

$$n_1(\omega) = \frac{\arg[H_{Measure}(\omega)]c}{2\omega d_1} + n_0 \quad (2.10)$$

$$a_1 = \frac{1}{d_1} \ln \left[ \frac{16n_1^2 n_0^2}{|H_{Measure}(\omega)|(\tilde{n}_1 + n_0)^4} \right] \quad (2.11)$$

若被测样品为混合物, 可以通过公式(2-9)获得吸收度  $absorbance(\omega)$  [55]为:

$$absorbance(\omega) = -\log \left( \left| \frac{E_s(\omega)}{E_{ref}(\omega)} \right|^2 \right) \quad (2.12)$$

根据 Lambert-Beer 定律[56],  $absorbance(\omega)$  与样品含量成正比, 进而可以对混合物组分进行定量分析。

## 2.4 本章小结

本章概述了太赫兹辐射的特性，简单描述了太赫兹辐射的产生和探测原理。讨论了几种太赫兹光谱技术，重点讨论了太赫兹时域光谱系统(THz-TDS)在农产品质量安全检测域中的一些应用。本章还说明了论文所采用的一种实验室级的太赫兹时域光谱系统，即美国 ZOMEGA 公司的 Z3 系统的构成和原理。最后，本章讨论了固态样品和液态样品的光学参数(折射率和吸收系数)的获取方法。本章的内容为后续各章节的理论基础。

## 3 小麦品质的 PCA-SVM 分类识别研究

### 3.1 小麦品质的定义及分类

小麦品质是一个十分复杂的概念。由于小麦的用途不同，人们对品质的要求也不同。小麦的品质有多种指标体系，通常可以分为：小麦籽粒形态品质、营养品质、磨粉加工品质、小麦面粉理化特性、小麦面团流变学特性。本节主要介绍小麦籽粒的形态品质和营养品质。

#### 3.1.1 小麦籽粒形态品质

##### 1. 籽粒形态

小麦籽粒形态有卵圆形、长圆形、椭圆形和短圆形。籽粒形状越接近圆形，出粉率越高，也越容易磨粉。

##### 2. 籽粒整齐度

籽粒整齐度是指籽粒大小和形状的均匀性。籽粒整齐度高的小麦品种，出粉率高，皮损少。

##### 3. 籽粒饱满度

籽粒饱满度一般可分为4级。1级：饱满；2级：较饱满；3级：不饱满；4级：秕瘦。它是衡量形态品质的重要指标。

##### 4. 籽粒颜色

小麦籽粒颜色可分为白粒、红粒、以及黄色、红黄过渡颜色等。该形态对小麦品质影响不大。

##### 5. 胚乳质地

胚乳质地表现在硬度和角质率两个方面。根据籽粒横断面胚乳组织的密实程度，将小麦籽粒分为硬质、半硬质和粉质三种。

##### 6. 角质率

角质率是指角质胚乳在小麦籽粒中所占的比例。角质率有多种测定方法，人工直观判断的方法是常见的方法。

##### 7. 籽粒硬度

籽粒硬度反映了蛋白质与淀粉结合的紧密程度。小麦籽粒硬度的测定方法有直观法



(牙咬)、研磨法、压力法、近红外仪测试等方法。

### 3.1.2 营养品质

#### 1 蛋白质含量

小麦蛋白质含量对其品质有较大影响。普通小麦蛋白质含量在 7%—22%之间，主要种类是清蛋白、球蛋白、 谷蛋白和麦谷蛋白。蛋白质含量在 15%以上的适合做面包，10%以下的适合做糕点、饼干等。含量在 12.5%—13.5%的适合做馒头、面条等。

#### 2. 氨基酸

小麦所含蛋白质是由 20 多种氨基酸组成的。但小麦的蛋白质中，赖氨酸的含量较低，平均只有 0.36%左右，而赖氨酸又是人体最为需要的氨基酸之一。因此，提高赖氨酸的含量，是小麦育种面临的一个重要方面。

#### 3. 碳水化合物

碳水化合物约占麦粒重量的 70%，是小麦中含量最多的组分。主要包括淀粉、糊精、纤维素以及各种糖类。碳水化合物是人体主要的能量来源。

#### 4 脂质

小麦中脂质含量较低，一般为 2.9%，主要由不饱和脂肪酸组成。

#### 5 维生素

小麦中的维生素主要是维生素 B、维生素 E 和泛酸。维生素 A 的含量很低，几乎不舍维生素 C 和 D。

#### 6.矿物质

小麦中含有的各种矿质元素有：钙、钾、磷、铁、锌、锰、铜、镉等。在小麦和面粉中，矿物质用灰分来表示，含量(干基)约为 15%—2.2%。

## 3.2 样品制备

小麦在霉变、虫蛀后，由于其组分的含量和化学性质发生变化，势必产生品质上的变化。本文选择虫蛀、霉变、发芽和正常小麦作为实验对象。实验前，需制作适合太赫兹检测的样品。其过程如下：首先将样品去杂质，然后用磨粉机磨成粉状，用电子天平称重，使每种样品保证数量相同，实验中取各种样品 0.16g 可满足后续压片的需要。将 0.16g 粉状样品放入压片机模具内，用 10MPa 压力保持 5 分钟，使之被压制成厚度为 1mm

左右，直径为 1cm 的原片，用电子游标卡尺测其厚度并记录下来，以便后续的光学参数计算。注意要保持压片上下表面平行无裂痕。实验的小麦样品来源于中央储备粮郑州直属库，不同品质小麦样品的属性如表 1 所示，制成的小麦待测样品如图 11 所示。

表 1 不同小麦样品属性

类型	品种	蛋白质含量	脂质含量
正常	郑麦 366	15.25%	6.45%
霉变	郑麦 366	11.45%	4.25%
发芽	郑麦 366	13.25%	5.14%
虫蚀	郑麦 366	9.74%	2.97%

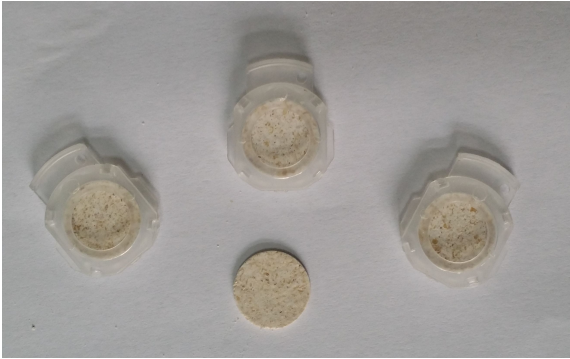


图 11 制成的小麦待测样品

### 3.3 样品的 THz 光谱分析

根据 3.2 节的样品制备方法，分别对虫蛀、霉变、发芽和正常小麦制备各 30 个样品，共 120 个样品。对每个样品采用 ZOMEGA-Z3 系统进行太赫兹辐射扫描，频率选择在 0.2~1.6 THz 作为有效扫描频段，每个样品扫描 3 次，计算平均光谱作为样品数据，共获得 120 个样品光谱数据。将所有样品数据分为 2 组，每组 60 个样品光谱数据，一组作为训练集，一组作为预测集，用于以后构建采用智能算法对未知样品进行识别的模型。四种小麦样品的典型光谱如图 12 所示。经过快速傅立叶变化，得到四种小麦频域光谱如图 13 所示。再根据公式 (2.6)和(2.7)，得到四种样品吸收光谱和折射率谱如图 14 和 15 所示。

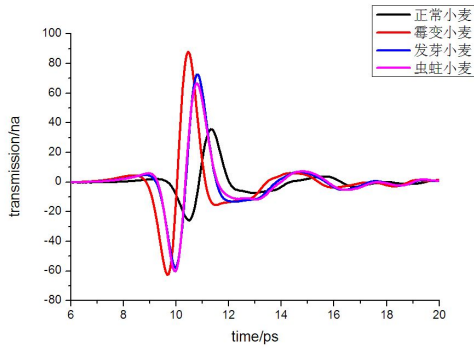


图 12 四种样品和参考的 THz 时域光谱

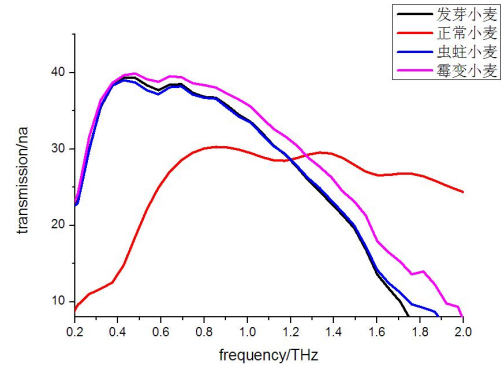


图 13 四种样品的 THz 频域谱

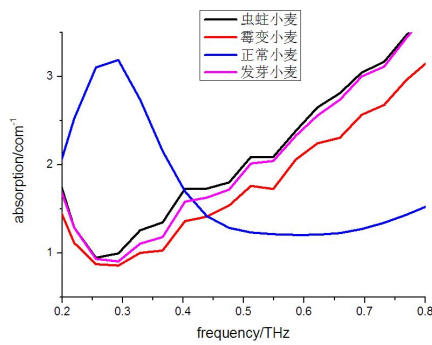


图 14 四种样品的吸收光谱

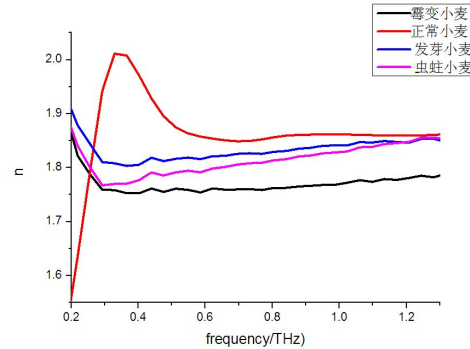


图 15 四种样品的折射率谱

由于虫蛀、霉变、发芽和正常小麦分子结构发生了变化，导致其分子间的振动频率和模式发生了变化，引起对太赫兹辐射的响应不尽相同。表现在光谱上，在 0.2-1.6THz 频段上的响应特性也不同。图 12 是四种样品的太赫兹时域谱，可以看出，四种样品谱线接近，说明系统是稳定的；同时，时间的延迟和脉冲幅度的衰减（相比参考信号）说明样品的折射率和吸收系数不同。在图 13 的频谱中，可以看到样品对太赫兹高频部分吸收较多且有较为明显的震荡，这主要是由于样品中的水分和密度的不同造成的。样品水分和密度越大，对太赫兹波的吸收越多。在 0.2-1.6THz 频段，四种小麦的频谱曲线存在明显区别，表明其生化和吸收特性不同，因此可以利用这一频段的太赫兹光谱对样品进行鉴别。

根据图 14 的吸收谱，四种样品在太赫兹波段均没有明显的吸收峰，但它们的吸收曲线并不相同，并且具有可区分性。从图中可以看出，正常小麦样品的吸收系数最大，而霉变、虫蚀和发芽小麦样品的吸收系数均较正常小麦低，表明这些小麦样品组分结构发生变化。在太赫兹有效波段的高频部分，样品的吸收系数均随频率的增加而增加，这可能是由于太赫兹波在样品内的散射造成的。在 1THz 处，各种样品的吸收系数如表 2

所示。

根据图 15 所示的折射率谱，可知四种小麦样品的折射率在 1.507-1.567 之间，并且有较大的可区分性。其中，正常小麦样品的折射率最大，且不同小麦样品的折射率有如下关系：

$$n_{\text{虫蚀小麦}} < n_{\text{发霉小麦}} < n_{\text{发芽小麦}} < n_{\text{正常小麦}} \quad (3.1)$$

四种小麦样品的折射率均随着频率的增大而下降，折射率的不同表明样品内部构成组分发生结构和含量上的变化，因此，不同状态的小麦可以对应到不同的折射率值上，应此可以通过测量小麦样品折射率值来鉴别小麦品质。四种小麦样品在 1THz 处的折射率值如表 2 所示。

表 3.2 四种小麦样品在 1THz 处的吸收系数和折射率

样品类别	1THz 处吸收系数	1THz 处折射率
正常小麦	278.863cm <sup>-1</sup>	1.543
发芽小麦	222.404 cm <sup>-1</sup>	1.532
发霉小麦	199.319cm <sup>-1</sup>	1.523
虫蚀小麦	196.035 cm <sup>-1</sup>	1.527

为了消除样品水分、年份、产地和品种对实验结构的影响，我们选择同一粮库同一年入仓小麦，并给予霉变和虫害培养。太赫兹测量的结果表明，虽然各样品均无特征吸收峰，但由于它们的折射率和吸收系数谱不同，因此是可以利用太赫兹光谱对小麦品质区分的。结合化学计量学的方法，我们可以设计出一个更好的分析模型对小麦品质进行鉴别。

### 3.4 PCA-SVM 分类识别算法

对于复杂物质的太赫兹检测，由于大多缺乏特征吸收峰，因此无法从吸收谱上直接这类物质进行判断分类。一般多采用化学计量学的方法，寻求光谱的统计规律，进而对物质进行定性和定量分析。在这类光谱分析中，使用最多的是一类非线性规划法，如主成分分析(PCA)，支持向量机(SVM)等，目前也有用神经网络算法和遗传算法进行分析的。本文主要讨论 PCA、SVM 算法在太赫兹光谱分析中的应用。

### 3.4.1 PCA 算法

主成分分析法(PCA)<sup>[57,58]</sup>是一种有效的研究多元统计分布问题的方法。它能够通过将原始高维数据降维,突出影响事物本质的主成分,从而将复杂问题简化,便于对复杂事物的研究。

主成分分析法的基本原理如下:选取数据量较少的新变量,使其线性组合能够表示原变量。新变量又称为主成分,各个主成分是互相正交的,并且能够最大限度地表征原变量所表达的数据信息。主成分可通过对协方差矩阵进行特征分解获得(此时又称为特征向量)并获得其权值(又称为特征值)。在各个主成分中,第一主成分拥有最大的信息量,第二主成分次之,依此类推。主成分分析的计算步骤如下<sup>[59]</sup>:

(1) 对数据矩阵  $X_{m \times n}$  ( $m$  为对象数量,  $n$  为对象维数) 进行标准化处理, 计算其标准化矩阵的协方差  $S_{n \times n}$ 。

$$X_i^* = \frac{X_i - \text{mean}(X_i)}{\text{std}(X_i)}, (i=1,2,\dots,n) \quad (3.2)$$

(2) 计算  $S_{n \times n}$  的前  $j$  个特征向量  $\mu_i$  及其对应的特征值  $\lambda_i$ 。

(3) 计算第  $j$  主成分  $Z_i$ , 即  $Z_i = \mu_i \times X^* (i=1,2,\dots,n)$ 。

可以证明,分解出的各个主成分所包含的原变量的信息量可用其对应的特征值去表征。事实上,往往前面几个主成分就包含了原变量所包含的绝大部分信息。因此,在满足一定误差范围内,可以只选取前面几个主成分来简化模型和算法。定义前  $k$  个主成分的累积方差贡献率为:

$$R = \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (3.3)$$

一般的,当前  $k$  个主成分的累积方差贡献率  $>80\%$  时,就可用这  $k$  个主成分表示原变量,此时,就把原高维空间降低到  $k$  维空间,便于我们对复杂问题的分析。

### 3.4.2 SVM 算法

支持向量机(Support Vector Machine, SVM)是一种常用的统计分类与回归分析方法,

其基本原理可参考文献<sup>[60,61]</sup>。本文采用 SVM 的“点对点”算法来对不同品质小麦进行分类，其基本思路是，将样本空间通过一个非线性映射函数，映射到一个高维特征空间中，使样本空间的非线性可分问题转化为特征空间的线性可分问题。

在 SVM 模型中，自变量  $x$  和因变量  $y$  的关系可表示为

$$y = f_{(x)} + N' \quad (3.4)$$

其中  $f_{(x)} = w^T \cdot \phi(x) + b$ ， $w$  和  $b$  为回归函数参数， $\phi(\cdot)$  为核函数。 $N'$  为噪声，由误差容忍度( $\varepsilon$ )定义。

函数  $f_{(x)}$  可通过训练集对 SVM 训练得到，其中训练集可表示为：

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\} \subset R^d \times R \quad (3.5)$$

式中， $y_k$  为输入变量  $x_k$  的实验值， $w$  和  $b$  分别是在最小误差条件下得到的。

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (3.6)$$

使得

$$\begin{aligned} w^T \cdot \phi(X_i) + b - y_i &\leq \varepsilon + \xi_i^*, \\ y_i - w^T \cdot \phi(X_i) - b &\leq \varepsilon + \xi_i, \\ \xi_i^*, \xi_i &\geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (3.7)$$

上式中， $C$  为常数，表示惩罚亏损的程度； $\xi_i^*$  和  $\xi_i$  分别是松弛变量； $n$  是样品数量。

通过采用一系列二重拉格朗日乘子  $\alpha_i$  和  $\alpha_i^*$  获得目标函数如下：

$$\sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \phi(x_i)^T \cdot \phi(x_j) \quad (3.8)$$

使得

$$\begin{aligned} \sum_{i=1}^n (\alpha_i - \alpha_i^*) &= 0 \\ 0 &\leq \alpha_i, \alpha_i^* \leq C \\ i &= 1, 2, \dots, n \end{aligned} \quad (3.9)$$

其中， $w$ 和 $b$ 可由 Karush–Kuhn–Tucker 的最优解计算得到。另外，SVM 回归函数可表示为：

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i)^T \phi(x) + b \quad (3.10)$$

若 $\bar{\alpha}_i = 0$ ，拉格朗日乘子 $(\alpha_i - \alpha_i^*)$ 可由 $\bar{\alpha}_i$ 表示；若 $\bar{\alpha}_i \neq 0$ ，这些数据集即为最终回归函数的支持向量。上式可改写为：

$$f(x) = \sum_{i=1}^n \bar{\alpha}_i \phi(x_i)^T \phi(x) + b \quad (3.11)$$

SVM 回归函数也可通过核函数 $k(x_i, y_i) = \phi(x_i)^T$ 来构建[62]。采用不同的核函数，可以生成不同的支持向量机。通常采用的核函数有：线性核函数、多项式核函数及 RBF 核函数。即：

- 1 线性核函数：  $k(x_i, y_i) = x_i \cdot y_i$
- 2 多项式核函数：  $k(x_i, y_i) = (x_i \cdot y_i + 1)^d$
- 3 RBF 核函数：  $k(x_i, y_i) = \exp(-\frac{\|x_i - y_i\|^2}{\gamma^2})$

核函数参数 $C, \gamma$ 的选取十分重要，选取过大或过小都会对分类结果造成很大影响，往往需要通过反复试验确定其最优值。本文中，我们采用五倍交叉验证的网格搜索算法来确定 $C, \gamma$ 的最优值[63]。

本文采用均方根误差(RMSE) [64] 评估回归模型的分类精度。RMSE 定义为：

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad (3.12)$$

式(3.12)中， $N$ 表示训练集中样品的数量； $Y_i$ 表示第 $i$ 个样品在数据集中的实际值， $\hat{Y}_i$ 表示第 $i$ 个样品在回归模型中的预测值。



### 3.4.3 PCA-SVM 算法

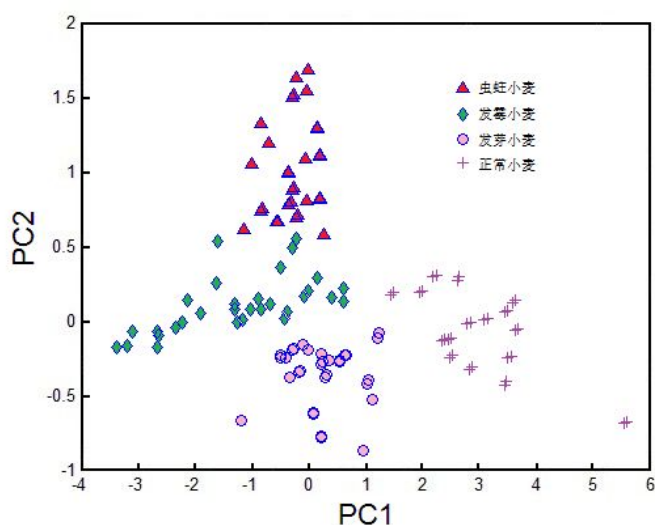
在本文对小麦品质的分类算法中，若直接使用 SVM，则由于特征过多而使计算量过大，因此本文提出了将 PCA 和 SVM 结合的方法，即所谓的 PCA-SVM 算法。其基本思路为：首先利用 PCA，在保证不遗漏主要信息的基础上，降低数据空间维数，提取出四个主成分作为 SVM 的输入，然后再采用 SVM “点对点” 算法对小麦的品质类型(发霉、虫蚀、发芽和正常小麦)进行预测识别。

## 3.5 分类识别结果及分析

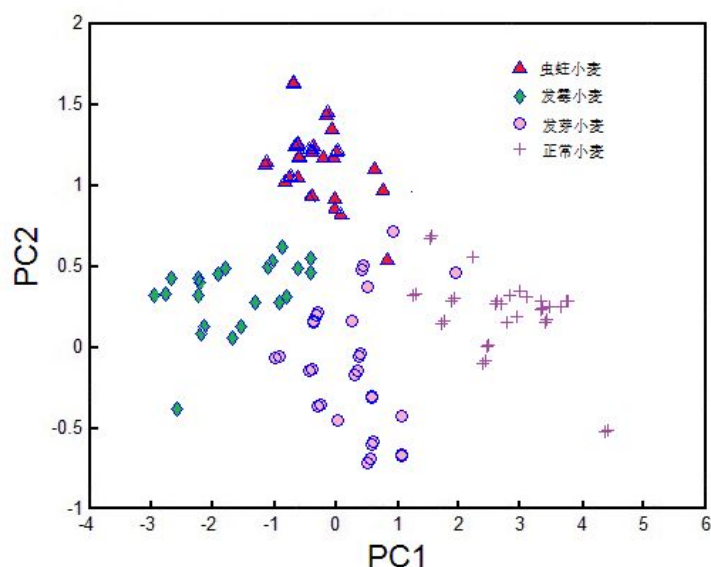
### 3.5.1 PCA 的分类识别结果

PCA 分析方法通过提取表征事物主要特征的主成分，将高维特征空间降为低维特征空间，本文将四种不同品质小麦进行太赫兹吸收谱和折射率谱测量，获取的光谱数据达 256 维，对于吸收谱，提取的前四个主成分的贡献率分别为 91.13%，5.34%，1.88% 和 0.19%，总贡献率达到 98.54%，使维数降低到了 4 维；对于折射率谱，提取的前 3 个主成分的贡献率分别为 85.13%，10.45%，2.16%，累计方差贡献率为 97.74%。

如图 16 为四种小麦样品第一主成分和第二主成分二维坐标系散点图。由图可以看出，120 个不同种类的小麦样品分别集中在相对独立的区域，说明采用 PCA 的方法理论上是可以将小麦样品的种类区分开来的。为了使用支持向量机对小麦样品进行分类，我们把样品吸收谱的前 4 个主成分提取出来作为支持向量机的输入变量，把样品折射率谱的前 3 个主成分提取出来作为支持向量的输入变量。



(a) 四种样品吸收光谱的主成分分析得分散点图



(b) 四种样品的折射率光谱主成分分析得分散点图

图 16 四种样品的主成分分析得分散点图

### 3.5.2 SVM 的分类识别结果

本文中，每种样品制作 30 个样本，四种样品共 120 个样本，从这 120 个样本中随机抽取 60 个样本作为训练集，用于构建 SVM 分类模型，剩下的 60 个样本用于构建的 SVM 模型进行分类。SVM 模型分类的效果与核函数的选择有很大关系，本文中采用三种常见的核函数：线性核函数、径向基函数和多项式核函数对样本进行分类。设置每种核函数不同的参数  $C$ ,  $\gamma$ ，可以得到不同的回归模型，微调这些参数，可以显著改善这些回归模型的预测精度。表 3 为使用不同核函数建立的回归模型的分类精度。

表 3 不同核函数分类精度

核函数类型	总分类精度(%)	每种样品分类精度 (%)			
		正常	发霉	发芽	虫蚀
线性	88.83%	100%	81%	93.32%	81%
多项式	83.33%	86.68%	73.32%	86.67%	86.66%
径向基	80%	93.32%	66.68%	86.64%	73.36%

由表 3 可以看出，三种核函数均能将正常小麦与其他非正常小麦区分开来，但对不同品质小麦样品，SVM 的分类结果有所不同，采用不同的核函数的 SVM，对同一种品质的小麦的分类结果亦有所不同。但 SVM 总的分类效果不错。就核函数在本文的分类效果来说，径向基核函数优于多项式核函数，而多项式核函数又优于线性核函数。

由于核函数参数对分类结果影响很大，因此需要优化，以提高模型分类效果。本文通过改变参数 $\gamma$ ，构建了 12 个 SVM 模型，并给出了每个模型的 RMSE(均方根误差)值，如图 17 所示。由于可以看出， $\gamma$ 的最优值为 4，由此得到的 C 的最优值为 2。

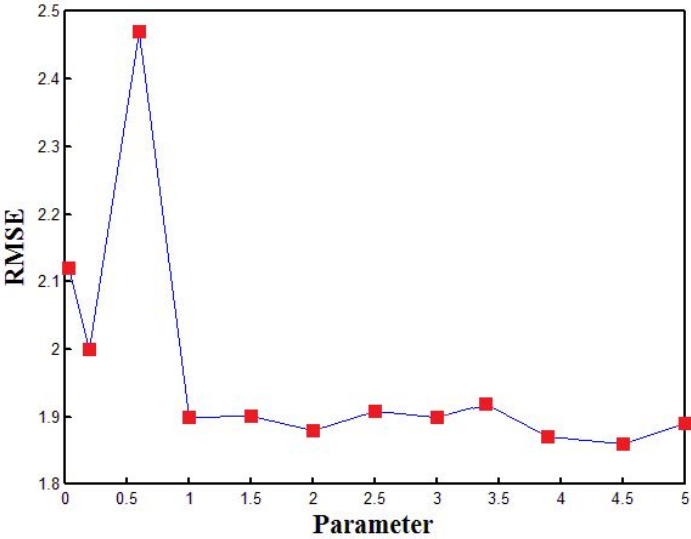


图 17 12 种 SVM 模的均方根误差

### 3.5.3 PCA-SVM 分析结果

SVM 对于样本数较少的分类是一种很好的方法。对于本文获取的小麦太赫兹光谱数据，若直接采用 SVM 分类识别，由于数据量太大而使分类算法耗时较长，本文先采用 PCA 方法对光谱数据进行降维，提取其中的 4 个主成分作为 SVM 的输入，再用 SVM 对小麦品质进行模式分类，大大简化了算法，降低了算法运时间。

仿照前面的 SVM 分析方法，构建 16 个模型，以优化核函数参数 $\gamma$ 。取不同 $\gamma$ 值，计算其 RMES，绘制图形如图 18 所示。由 3.8 可以看出，当 $\gamma=3.5$ 时，RMES 值最小，此时 $\gamma$ 最优，再根据网格探索优化算法得到 C 的最优值为 1.6。采用前面讨论的 3 种核函数，得到各自预测精度如表 4 所示。

表 4 PCA-SVM 算法的预测结果

核函数类型和总预测精度(%)		每种样品预测精度 (%)			
		正常	霉变	发芽	虫蚀
线性	96.67%	100%	93.35%	100%	93.31%
多项式	95%	100%	93.30%	100%	86.7%
径向基	93.33%	100%	86.65%	100%	86.69%

由表 4 可以看出, 在使用了 PCA 算法提取的主成分作为 SVM 输入的情况下, 三种核函数的预测精度均达到了 90% 以上, 显然高于单独使用 SVM 的预测精度, 其中, 区分正常和发芽小麦的准确率达到了 100%。实验结果表明, 采用 SVM 联合主成分分析, 可以大大提高本文小麦品质分类的准确率。

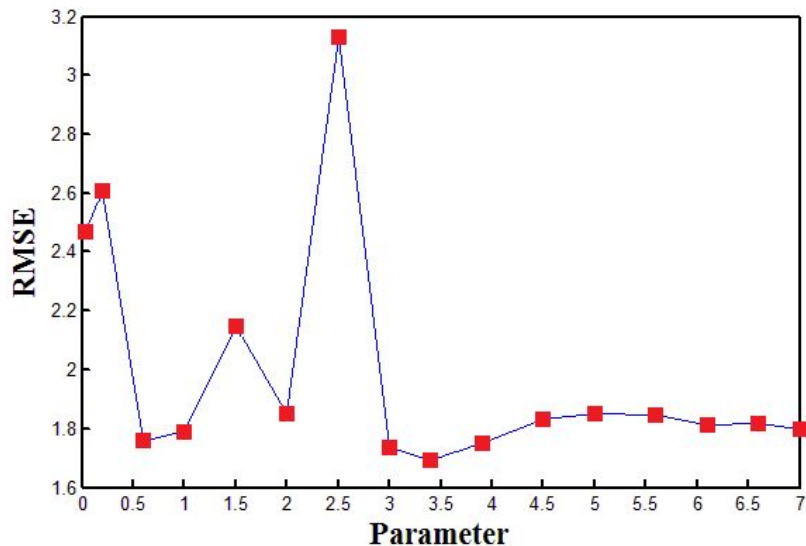


图 18 参数  $\gamma$  的均方根误差(PCA-SVM)

### 3.5.4 PCA-SVM 与其他算法的识别结果比较

在太赫兹光谱分析中, 常见的算法有 PLS、PCR 和 BP<sup>[65]</sup>, 本文将 PCA-SVM 算法与这些常见算法进行了比较, 以说明在小麦品质检测上, PCA-SVM 算法的优势。比较结果如表 5 所示。

表 5 PCA-SVM 与其他算法预测精度的比较

算法	预测精度 (%)	
	吸收系数	折射率
PCR	50%	46.68%
PLS	58.34%	63.35%
BP	83.35%	78.36%
PCA-SVM	93.34%	91.68%

由表 5 可见, 采用本文的 PCA-SVM 算法, 对小麦品质进行分类, 其精确度要明显高于其他算法。这是因为经过 PCA 算法, 提取出了吸收光谱和折射率光谱的主成分, 提取的主成分既包含了原始变量的大部分信息, 同时也使原始变量维数降低, 而 SVM

是适合处理小样本分类的算法，PCA 与 SVM 的结合，提高了 SVM 的预测精度。而对于其他算法，从表中可以看出，预测结果均不是很理想，特别是 PCR 的分类精度还不到 50%，PLS 的分类精度也只有 60%左右，分类性能明显低于本文的方法。因此表 3.5 再次说明了本文 PCA-SVM 模型的优越性。

综上所述，利用太赫兹时域光谱技术，结合化学计量学的方法，是可以对不同品质的小麦进行分类预测的。实验表明，各种小麦样品并没有特征的吸收峰，这是因为小麦是复杂的混合物，太赫兹光谱只是反映了所有成分的整体响应。但由于小麦品质发生劣变后，其组分构成发生变化，因而使其吸收谱和折射率谱的整体情况也发生了变化，从而使太赫兹光谱响应发生变化，正是基于这一点，利用太赫兹探测技术可以实现小麦样品的模式分类预测。当然，如果要提高预测精度，我们还可以综合吸收谱和折射率谱，共同作为 SVM 的输入，进一步提高 PCA-SVM 预测的精度。

### 3.6 本章小结

本章将太赫兹时域光谱技术与化学计量学的方法结合，用于实现对四种品质的小麦的分类识别。首先采用主成分分析法，对吸收光谱提取 4 个主成分，对折射率光谱提取 3 个主成分，然后把提取的主成分作为 SVM 的输入，利用 SVM 算法实现对小麦样品的分类识别，实验表明，应用 PCA-SVM 模型，要比单独使用 SVM 模型，样品分类效果要好。另外，本文也把 PCA-SVM 模型与其他太赫兹光谱分析常用模型进行了比较，说明了 PCA-SVM 模型的优势。本文的实验结果表明，采用太赫兹时域光谱技术，结合化学计量学的方法，对农产品进行品质检测和分析的方法是可行的。



## 4 小麦品质的 iPLS 分类识别研究

### 4.1 样品准备

本章中，我们采用太赫兹时域光谱系统分析8种不同种类的小麦。小麦样品来源于2013年储藏的小麦，尽量代表不同地域的粮食主产区。这些小麦营养组分构成、理化特性各不相同。8种小麦的营养成分和理化特性参数见表6所示。

表 6 8 种小麦的部分理化特性一览表

No.	Wheat Variety	Bulk density (g/L)	Crude Protein Content (%)	Water Content (%)	Imperfect Grain (%)	Gluten Content (%)
2	Zhouyuan 9369	790	14.9	12.5	3.4	33.0
3	Aobiao	845	14.0	11.5	1.0	26.0
4	DNS	840	14.5	11.9	1.6	38.0
5	Jiamai	830	13.8	12.2	1.8	39.0
6	Jinan17 wheat	773	15.6	11.5	3.0	34.0
7	Zhoumai 27	798	13.2	12.1	3.8	33.0
8	Yunong 416	787	14.3	12.5	4.0	32.5

小麦样品均从各地粮库中直接获取，在制作成样品前未进行过加工和处理。在制成样品前，首先对其去杂质，清洗晾干，去除灰尘并使其保持水分一致。然后用磨粉机磨2分钟，用200目筛子筛成均匀的粉末，然后取不同种类的小麦粉末样品各0.16克，放在压片机上压片，压片时保持5吨压力3分钟。取出压片，测出其精确的厚度，放入一次性密封袋内，贴上标签，标签上注明小麦种类和样品厚度。8种小麦分别制成样品，总共制成8个压片的样品，以备太赫兹光谱的检测。

### 4.2 PLS 算法和 iPLS 算法

本章用偏最小二乘(PLS)回归这一化学计量学的方法研究不同种类小麦的太赫兹光谱。PLS通过计算两个矩阵的最大的协方差矩阵，用于发现矩阵间的最佳相关性，设X为输入矩阵，Y为输出矩阵，PLS分析算法依据下列公式：

$$\begin{aligned} X &= TP^T = E \\ Y &= UQ^T = F \end{aligned} \quad (4.1)$$

上式中，T和U是得分矩阵。P和Q是载荷矩阵，P可以看作是X和Y之间的协方差矩

阵，U可以看作是Y和U之间的协方差矩阵。E和F是剩余矩阵。在本文中，把0.2-1.6T频段的小麦样品的吸收系数谱和折射率谱作为输入矩阵X，把小麦样品的种类作为输出矩阵Y，各种小麦样品的理化特性如表4.1所示。

间隔偏最小二乘法是一种选择优化技术，它可以从整个光谱区域识别出重要的光谱区域，从而消除来自其他区域的干扰。iPLS把整个光谱区域等间隔的划分出若干个子区域，对每个子区域建立相应的PLS模型。计算每个区域的交叉验证-均方根误差(RMSECV)，从而选出均方根误差最小的区域。可以利用预测值R和参考值的相关系数、校验集和预测集的几个误差评估校验模型，RMSECV和R的计算公式如下：

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^N (Y_r^i - Y_p^i)^2}$$

$$R = \frac{\sum_{i=1}^n (y_r^i - \bar{y}_r)(y_p^i - \bar{y}_p)}{\sqrt{\sum_{i=1}^n (y_r^i - \bar{y}_r)^2 \sum_{i=1}^n (y_p^i - \bar{y}_p)^2}} \quad (4.2)$$

上式中，n为校准样品集中样品的数量， $y_r^i$ 是第i个样品的参考值， $y_p^i$ 是第i个样品的预测值。 $\bar{y}_r$ 是样品参考值的平均值， $\bar{y}_p$ 是样品预测值的平均值。

### 4.3 样品的 THz 光谱分析

为了消除随机噪声，提升信噪比，每个样品的太赫兹光谱被测三次，因此样品的太赫兹光谱是在0.2-2.5T范围的三次平均光谱。参考信号的光谱在每个样品测量后测试一次。图19为8种小麦样品的太赫兹时域光谱图和经过快速傅里叶变换后的频域光谱图。图20为经过计算的吸收系数光谱图和折射率光谱图，计算公式见第2章式(2.6)和(2.7)。

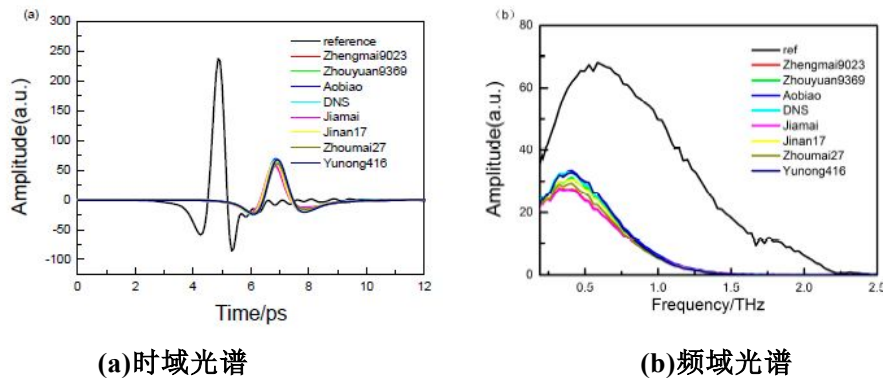
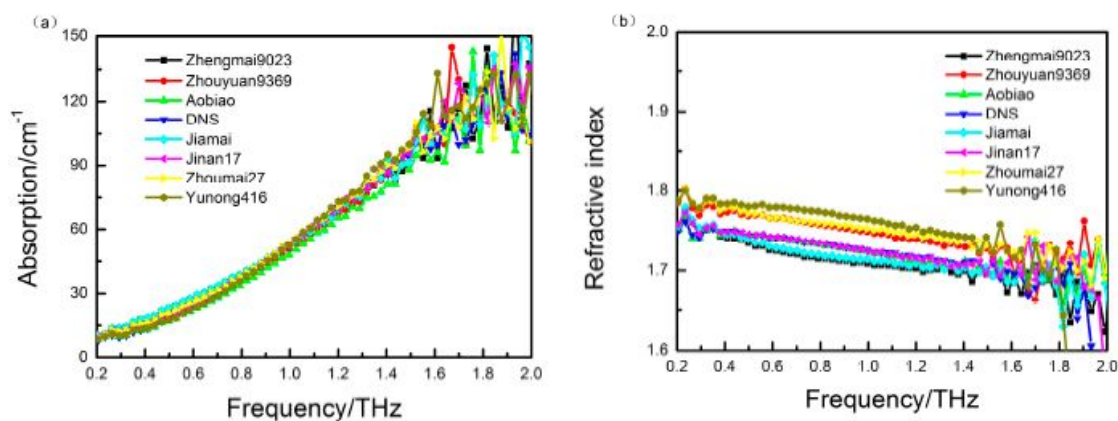


图 19 8 种样品的太赫兹时域光谱和频域光谱(频率范围:0.2-2.5THz)





(a)吸收系数光谱

(b)折射率光谱

图 20 8 种样品的太赫兹吸收系数谱和折射率谱(频率范围:0.2-2.0THz)

如图19所示，太赫兹频谱测量的有效频率范围是0.2-2.5THz。样品的光谱具有较窄的频带。尽管样品之间的频谱曲线区分不大，但样品和参考信号的光谱区分较大，说明样品对太赫兹辐射有较大的吸收。从频域谱来看，样品的幅度不同，说明各种样品对太赫兹辐射的吸收不同。

由于各种小麦的样品具有相似的成分和复杂的化学结构，因而测量获得的太赫兹光学参数也十分相似，这一点从图20可以看出来。为了能够显示出不同小麦品种光学参数的不同，我们将8种小麦制成20个样品，分别计算出每种小麦样品的吸收系数如图21所示。需要注意的是，由于系统动态范围的限制，样品的太赫兹谱信噪比将在高于1.5THz的范围内降低。由于8种样品均没有吸收峰，它们的吸收谱也没有明显的差别，我们采用了PLS回归模型来对这些具有小的差异的光谱进行分类识别。

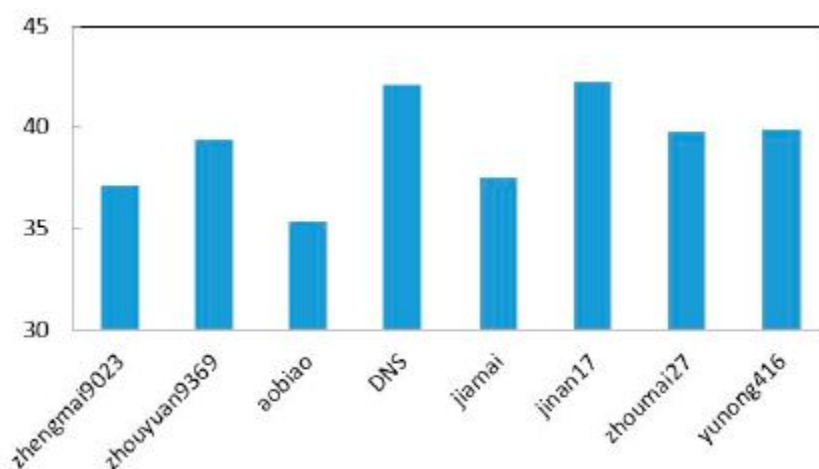


图 21 8 种小麦样品的平均吸收系数

#### 4.4 PLS 分析

我们采用MATLAB的PLS工具包执行PLS算法，对8种小麦的太赫兹光谱进行分析。MATLAB使用的版本是2012a，PLS工具包使用的是Version4.0(Eigenvector Research Inc., Wenatchee, USA)。利用PLS，在光谱数据和小麦种类之间建立相关，在光谱中总共使用了128个频率变量。8种小麦被编号为从1到8，如表6所示，作为PLS的输出。PLS的校正模型采用交叉验证算法。在本章中，0.2-1.5THz范围的吸收系数谱和折射率谱作为PLS的输入以获得最佳的交叉验证模型，并用此模型预测小麦样品的种类。8种样品集共160个样品被用于实验。所有的样品被随机地分为2个样品集：矫正集(含有96个样品)和预测集(含有64个样品)，本文的交叉验证模型的性能如表7所示。

表 7 太赫兹光谱的 PLS 模型的交叉和验证结果

Interval Variable	Frequency Range (THz)	Factors	Calibration		Cross Validation	
			R	RMSEC	R	RMSECV
吸收系数	0.2-1.5 THz	5	0.987	0.759	0.983	1.028
折射率	0.2-1-5 THz	5	0.982	1.472	0.979	1.684

由表4.2可以看出，对于本文的全太赫兹光谱PLS模型，吸收系数谱的模型比折射率谱有更高的R值、更低的RMSE值和RMSECV值，说明基于吸收系数谱建立的PLS模型预测结果优于折射率谱。

图4.4说明了基于吸收系数谱建立的PLS模型对小麦种类的验证结果。模型中参考线表明预测值和实际值吻合。由图22看出，所有预测值都未对参考线有较大偏离，说明预测值与实际值符合度良好，同时也说明了采用太赫兹光谱PLS模型是可以对小麦样品的种类进行识别的。

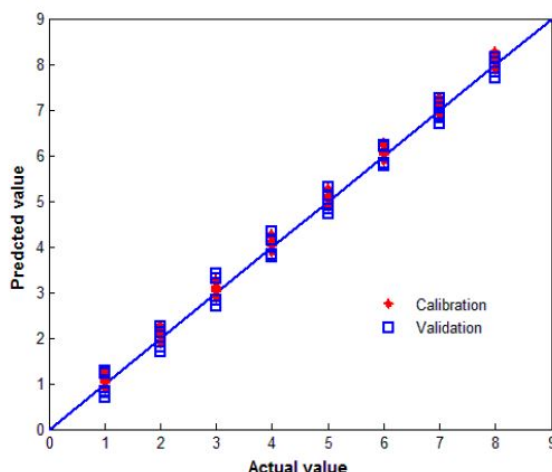


图 22 PLS 模型对小麦种类的校正和确认结果

#### 4.5 iPLS 分析

为了进一步提高PLS模型对小麦种类的预测精度，本文对样品吸收光谱数据设置间隔为8，进行等间隔分成16等分，在每个间隔上执行iPLS算法。小麦样品全吸收光谱分割示意图如图23所示。图23中的柱状图表示每个间隔的RMSECV值。

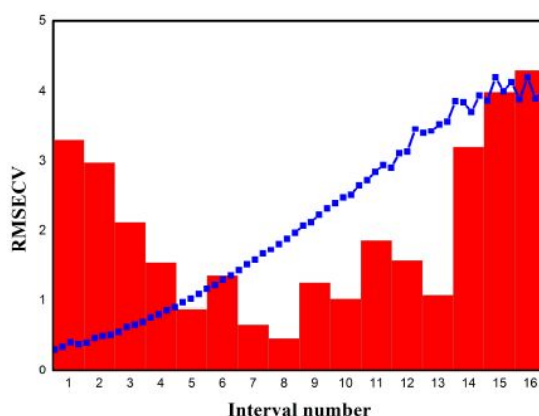


图 23 小麦样品的太赫兹光谱的 iPLS 分割结果

在图23中，每个柱状图的宽度是相同的，而每个柱状图的高度表示了经由iPLS模型计算的RMSECV值。其中第8个柱状图，间隔在57-64，相应的频率范围在787.5-900GHz，有着最小的RMSECV值，因而被选定为用于开发本文iPLS模型的区域。对于本文的小麦种类分类，本文的iPLS模型预测精度要优于PLS模型。

iPLS选择光谱中与小麦种类最相关的部分，可以有效地溢出噪声和其他区域的干扰，因而可以提高模型预测的性能。在本章的实验中，我们把整个光谱区域分成不同数目的间隔，用于发现最佳的光谱区域。我们设置间隔的宽度为4和16，从而把整个光谱

分成32个和8个间隔，与先前设置的间隔的宽度为8，把光谱等分成16个间隔的情况，进行了RMSEC值和RMSECV的比较，比较结果如表8所示。

表 8 优化后的 iPLS 回归模型预测结果

IntervalVariable	Frequency Range	$R_{cal}$	RMSEC	RMSECV
4	0.731-0.956 THz	0.991	0.768	1.260
8	0.787-0.900 THz	0.992	0.573	0.967
16	0.675-1 THz	0.984	0.837	1.237

如表8所示，宽度为8的iPLS模型具有最佳的RMSECY值(0.967)和最佳的RMSEC值(0.992)，并且具有最高的R值(0.992)。与PLS比较表明，iPLS模型的性能要优于PLS模型。

#### 4.6 模型与问题讨论

基于太赫兹吸收谱，我们针对8种小麦品种的分类识别问题分别构建了PLS回归模型和iPLS回归模型。由前面的讨论可知，iPLS回归模型的性能主要取决于所选太赫兹谱的范围和宽度，合适的谱范围和宽度有利于提供模型预测的精度。谱范围选择过宽，样品太赫兹光谱中的非相关信息或噪声就会加大，这无疑会影响甚至降低模型预测的精度。另一方面，谱的选择宽度过窄，一些具有特征性质的信息将会丢失，同样会降低模型预测的精度<sup>[66]</sup>。

构建的模型的预测精度同样也受到样品的太赫兹吸收谱本身的影响。被测的样品的太赫兹光谱会受到如实验环境、实验步骤、背景噪声、时域光谱系统的不稳定性的影响<sup>[67]</sup>。同样，样品本身也会对测得的光谱产生影响，如制作样品的过程，样品本身的化学成分、样品种植时的环境、气候、收获日期等等的影响。为了构建更为准确的PLS回归模型，需要考虑这些因素。精确的PLS回归模型不仅能够区分不同种类的小麦样品，而且能够区分小麦不同的收获地域和收获时间，这都是今后在构建模型需要考虑的。

#### 4.7 本章小结

本章采用太赫兹光谱和PLS算法对8种小麦样品的种类进行了确认和预测。使用太赫兹时域光谱系统测量了8种小麦样品在0.2-2.0THz内的吸收谱和折射率谱，使用偏最小二乘算法PLS构建回归模型，对8中小麦样品的种类进行了区分。模型的交叉验证相关系数

(R)和交叉验证均方根误差(RMSECV)分别是0.985和1.162。另外,通过选择光谱中的最佳区域,开发了间隔偏最小二乘算法(iPLS)用于模型的优化,优化后的模型的 $R=0.992$ ,  $RMSECV = 0.967$ 。实验表明,采用iPLS优化后的模型具有更好的预测精度。本文的研究表明,采用太赫兹光谱,结合采用PLS算法,对小麦品种进行非破坏性的检测和识别是可行和可靠的。



## 5 储粮中黄曲霉毒素的 THz 定量分析

### 5.1 储粮真菌及其危害性

我国粮库储藏的小麦一般储期可长达3-5年。小麦在储藏期间，如果温湿度调控不好，极易发生霉变。霉变易发生于南方省份湿热环境，或者是粮食入仓时本身水分过大引起的。引起小麦霉变的真菌有多种，常见的有曲霉、根霉、青霉、毛霉等。尤其以曲霉危害最严重，也最容易发生。在曲霉中，以黄曲霉、赭曲霉、杂色曲霉为最常见，其中以黄曲霉危害性最大。

黄曲霉毒素(AFT)是二氢呋喃香豆素的衍生物，是一类结构相似的化合物，主要类型有B1,B2,G1,G2，以及它们的代谢物M1和M2。主要存在于霉变的花生、玉米、小麦中，各种坚果、豆类作物中以及土壤中也有存在。黄曲霉毒素的主要危害是肝损害，当摄入量较大时，极易造成急性肝坏死，慢性中毒时，常诱发肝癌。因此，黄曲霉毒素的检测是各类食品安全检测项目中的必检项目。

赭曲霉毒素是另一种有着严重危害性且分布广泛的真菌毒素，其中以赭曲霉毒素A1为分布最广、危害最为严重。赭曲霉毒素A广泛分布与各种食品和饲料中，储藏小麦中赭曲霉毒素的分布不容忽视。赭曲霉毒素对人体的危害主要是肾脏和肝脏。肾脏是危害的主要器官，慢性中毒易导致肾功能衰竭。

杂色曲霉毒素主要有杂色曲霉菌和构巢曲霉菌产生，也是一类结构详细的化合物，主要污染、小麦、玉米、花生、大豆等作物，但毒素不如黄曲霉毒素强，主要是肝肾毒害。在杂色曲霉毒素污染较为严重的地区，肝癌发生率一般较高。

霉菌在生长的初期往往不易发现，这给霉菌的防治带来一定的困难。事实上，霉菌防治往往在其发展初期有效，而当霉菌生长中后期，粮食已经变质且不能食用。目前，关于霉菌的早期检测技术还有待开发，但目前这项技术仍是制约我国粮食安全的一个瓶颈<sup>[36,37]</sup>。本文的太赫兹检测方案，也正是为解决这一问题而提出来的。

### 5.2 黄曲霉毒素太赫兹检测方法总体流程

黄曲霉毒素太赫兹检测方法的总体流程如图24所示。

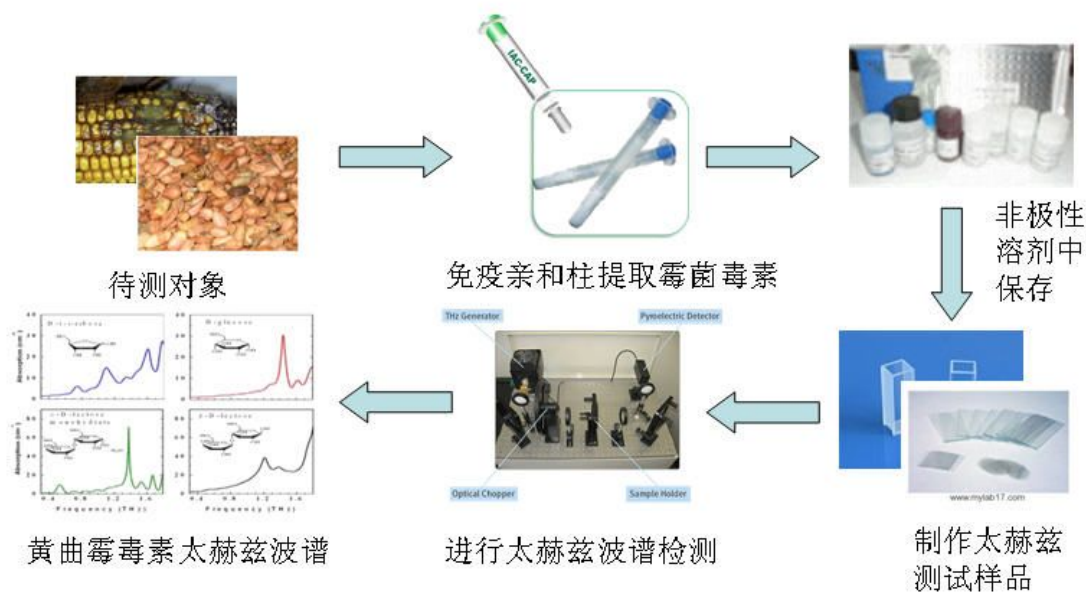


图 24 黄曲霉毒素检测的总体思路

1. 将怀疑有黄曲霉菌感染的粮食、食品和食用油作为待检对象，并制作待检物的样品。
2. 利用黄曲霉毒素免疫亲和柱对待检物中含有的黄曲霉毒素进行提取。
3. 提取的黄曲霉毒素保存在非极性、易挥发的有机溶剂(如甲醇、氯仿)中，并低温、密封保存。
4. 检测时首先制作用于太赫兹波谱检测的样品。考虑到黄曲霉毒素检测浓度过低带来的不利影响，本技术方案采用了痕量检测方法。该方法的基本过程为：将含有黄曲霉毒素的溶剂滴到载玻片上，并涂抹均匀，然后让有机溶剂挥发，有机溶剂挥发完后就会在载玻片上留下一层黄曲霉毒素的薄层，考虑到薄层不均匀或过薄影响检测效果，可多滴几次，并使溶剂完全挥发，以增加薄层厚度，增强检测效果。
5. 载有黄曲霉毒素薄层的载玻片放在太赫兹时域波谱检测系统(THz-TDS)上采用透射法进行波谱测量，得到黄曲霉毒素的波谱。
6. 对黄曲霉毒素样品的太赫兹光谱进行分析，采用最小二乘支持向量机建立回归模型；采用证据推理算法对黄曲霉毒素的浓度进行定量预测识别。

### 5.3 黄曲霉毒素的提取和样品的制备

1. 样品均质



将生霉小麦去杂质后粉碎、过筛，使成粉状，置于锥形瓶中，加水后在摇床上进行均质。

## 2. 粗滤和稀释

用槽纹纸过滤均质液15ml，再加30ml水稀释，用玻璃注射器提取15ml稀释样品。

## 3. 过柱

连接免疫亲和柱、玻璃注射器和空气压力泵，开启空气压力泵并调节其压力，使样品液体以6ml/min流速缓慢通过免疫亲和柱，待液体全部过柱后，再通过2-3ml空气。

## 4. 黄曲霉毒素的洗脱

用色谱级甲醇1.0ml洗脱免疫亲和柱，注意甲醇洗脱要慢、流速要均匀，并控制在1-2ml/min。用玻璃试管搜集洗脱液。至此，黄曲霉毒素已溶解在洗脱液中。

## 5. 免疫亲和柱的清洗

用清水清洗免疫亲和柱2-3次，按使用说明书上的保存方法进行处理后，将免疫亲和柱放入冰箱中保存以备下次使用，保存温度为2-8℃。免疫亲和柱可多次反复使用。

## 6. 待测样品的保存

如果要保存样品，可将洗脱液放入不透光玻璃瓶中密闭，放入冰箱中。注意冰箱温度控制在2-8度，玻璃瓶应密闭良好，以免甲醇挥发。

## 7. 制作太赫兹痕量样品

由于用此法收集黄曲霉毒素浓度较低，直接用液体池或比色皿支撑检测样品，太赫兹响应会十分微弱甚至没有响应，为此，本文采用痕量检测的方法，以有效解决因黄曲霉毒素浓度过低造成太赫兹响应不敏感的问题。具体过程如下：

- (1) 用注射器扎透保存有样品的玻璃瓶上的橡皮塞，提取部分溶液到注射器中
- (2) 将注射器中的溶液滴在载玻片上，用另一载玻片将溶液摊平
- (3) 放在通风处使溶剂挥发，此时载玻片上只留下黄曲霉毒素的薄层
- (4) 重复上述过程，直至载玻片上留有的黄曲霉毒素的薄层有足够的厚度

### [注意事项]

- (1) 在载玻片上摊平溶液时应小心、缓慢，尽量均匀不要溢出载玻片外
- (2) 操作过程要做好防护工作，戴口罩、手套、穿工作服，以免有毒溶剂对皮肤和呼吸道造成伤害。

### 5.4 黄曲霉毒素的太赫兹光谱分析

利用太赫兹光谱对黄曲霉毒素进行定量分析，首先按一定的浓度梯度制备黄曲霉毒素B1标样，将黄曲霉毒素溶于乙腈中，分别制成5ppb、10ppb、25ppb、50ppb、100ppb、500ppb、1ppm和10ppm的标样，利用THz-TDS测得其太赫兹时域谱如图25所示。由图25可以看出，参考信号和样品信号的峰值差别较大，且延时差别也较大；但样品之间的峰值和延时差别不大，将部分信号放大，见图25，可以看到，随着样品中黄曲霉毒素B1浓度的增大，其吸收幅值逐渐下降，说明采用太赫兹光谱分析区分黄曲霉毒素的浓度具有一定的可行性。

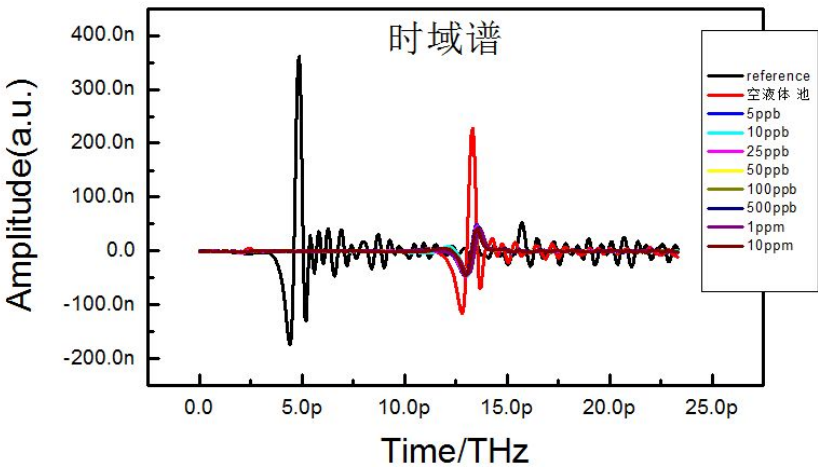


图 25 黄曲霉毒素 B1 不同浓度样品的太赫兹时域光谱图

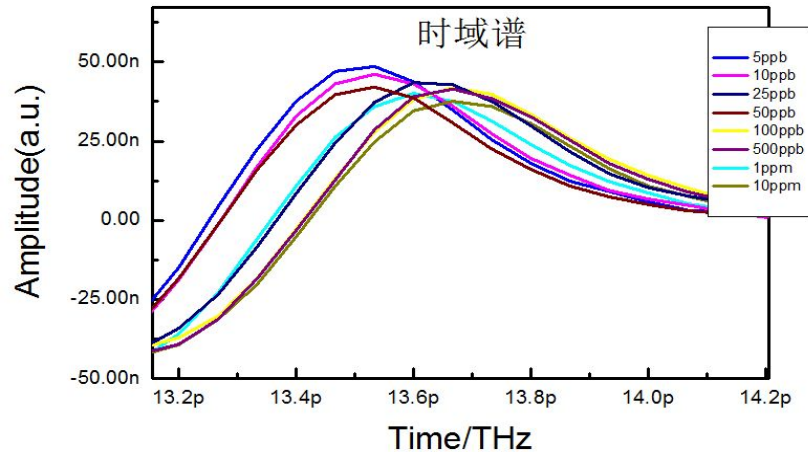


图 26 黄曲霉毒素 B1 不同浓度样品太赫兹局部时域波形放大图

对如图26所示的太赫兹时域谱进行快速傅立叶变换，可以得到如图27所示的不同浓度样品的太赫兹频域谱。由频域谱图可以看到，各浓度曲线比较接近，相互之间有交叉

现象。在小于0.5T的低频部分，频谱表现的规律不明显，这可能是THz-TDS系统在开始测量时信号不稳定造成的。从0.5T以后，各浓度的频谱幅度基本上是随着黄曲霉毒素浓度的升高而下降的，说明黄曲霉毒素对太赫兹辐射有着较强的吸收作用。

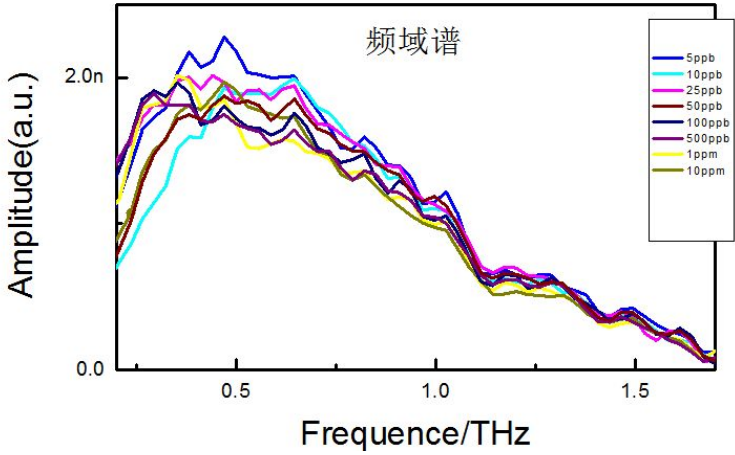


图 27 黄曲霉毒素 B1 不同浓度样品太赫兹频域光谱图

按照前面第3章给出的计算折射率和吸收系数的公式，对获取的太赫兹时域光谱数据进行计算，可以得到不同黄曲霉毒素浓度的太赫兹折射率谱和吸收谱如图28和图29所示。折射率反映了样品对太赫兹波的色散情况。从样品的折射率谱看，折射率随浓度的不同而不同。从吸收谱上看，不同浓度样品吸收谱较为接近，且在高频段有较大幅度的震荡，考虑震荡由噪声引起，因此本文把吸收谱限定在0.5-1.5T范围内进行考察。

从样品的吸收谱可以看出，在0.5T-1.5T的范围内，吸收系数总体上随频率增加而上升，且随样品中黄曲霉毒素浓度的增大而增大，虽然有个别谱有交叉现象，这可能是由于噪声、浓度配比不准确、光源不稳定造成的；但总体上是有一定规律的。虽然各样品没有特征吸收峰，但考虑到吸收系数可信性强，能够较好反映不同样品浓度的比例关系，因此我们把吸收谱作为黄曲霉毒素定量分析的对象。

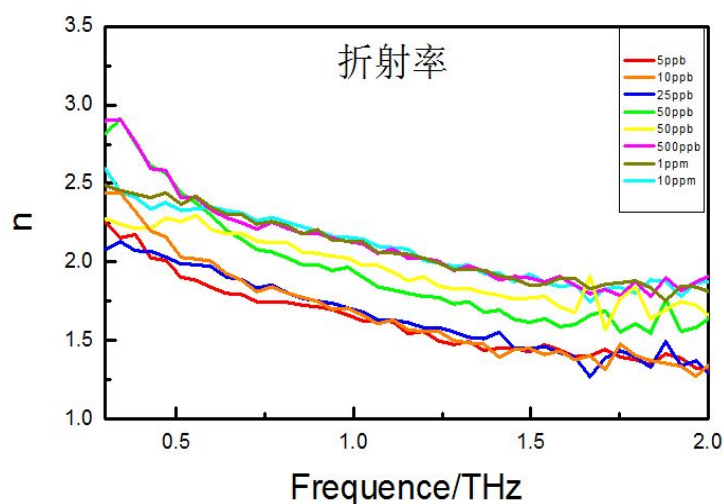


图 28 黄曲霉毒素 B1 不同浓度样品太赫兹折射率谱图

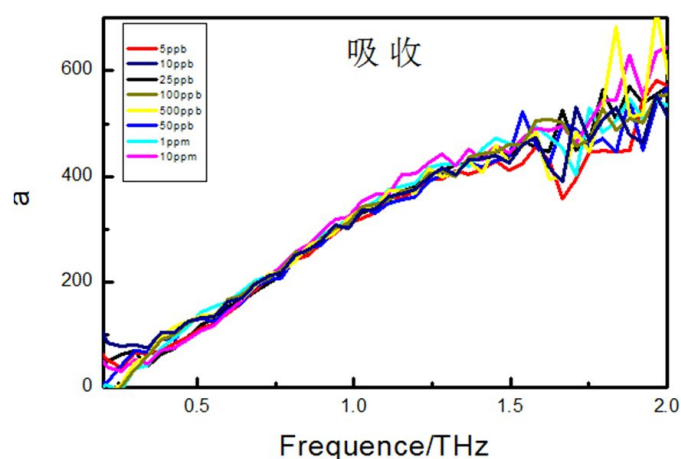


图 29 不同浓度黄曲霉毒素 B1 样品太赫兹吸收谱

鉴于吸收系数与样品黄曲霉毒素浓度具有近似线性关系，本文采用最小二乘支持向量机构建样品黄曲霉毒素的浓度和其吸收系数的线性回归方程。选择0.5-1.5THz范围内的吸收谱，经过平滑处理后，计算平均吸收系数。平滑处理是为了消除噪声，其方法是在平滑点附件取若干点进行平均或拟合。计算出的平均吸收系数经过归一化后，作为模型的输入值，得到样品黄曲霉毒素浓度与吸收系数关系图如图30所示，由图30，得到线性回归方程为：

$$y=0.14886x-0.07352$$

由图30可知，样品中的黄曲霉度B1浓度与其太赫兹吸收系数近似成线性关系。

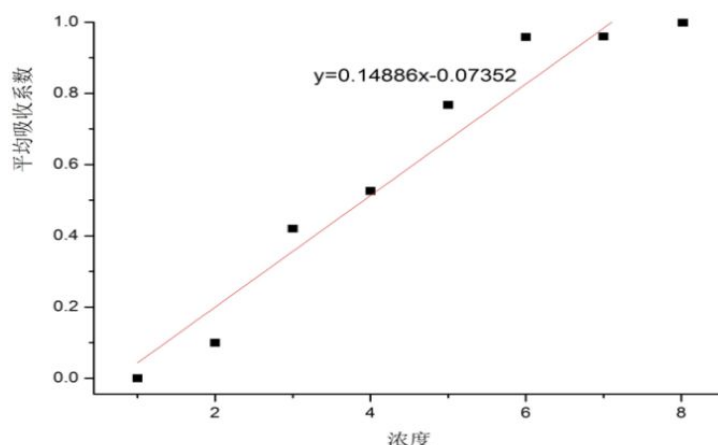


图 30 黄曲霉毒素 B1 的浓度与其太赫兹平均吸收系数关系图

制备新的样品作为校正集样品，测得其太赫兹光谱并计算其平均吸收系数，代入到上述的模型中，得到对应样品的预测值，在利用预测值计算出校正集样品的R值和RMSEC值，利用留一交叉校验法得到校正集样品的预测浓度和参考浓度，再进一步计算出相应的R值和RMSEC值，得到模型评价参数表如表9所示。

表 9 本文回归模型评价参数表

校正集		交叉验证集	
R	RMSEC	R	RMSECV
0.992	0.533	0.992	0.629

对预测样品集进行太赫兹吸收谱测量，将测量结果代入到上述模型中，得到预测样品的黄曲霉毒素预测浓度，再计算预测浓度与实际浓度的误差(实际浓度可通过建立标样的方法预知)，误差计算的结果表明，本模型的预测误差在20%以内，表明该模型具有一定的浓度预测能力，可以用来对黄曲霉毒素的浓度进行定量分析，但本文采用的方法比较适合小样本，模型还不是最优的，模型的预测精度还有很大的提升空间，下面我们给出另一种模型，可能具有更高的预测精度。

## 5.5 多光学参量信息融合模型

### 5.5.1 信息融合模型框架

前面的最小二乘支持向量机模型黄曲霉毒素的预测精度在80%以上，还不能完全满足实际应用的要求。对于不同黄曲霉毒素浓度的太赫兹吸收光谱，由于受到光源稳定性、噪声等的影响，只作为预测模型唯一的信息来源，难免会存在可靠性的问题。因此，就

这需要综合利用我们所获得的所有光学信息，互相取长补短，才能进一步提高模型的预测精度，在这方面，多源信息融合的理论和方法为我们提供了好的思路。

从文献<sup>[68-70]</sup>来看，信息融合常用的算法有很多，其中证据理论(又称为D-S证据推理)和人工神经网络算法应用较多。本文对D-S证据推理算法进行改进，把证据推理的结果再次作为证据，提出了一种二重证据推理算法来提高证据推理的精度，从而提高黄曲霉毒素B1的浓度预测精度。

本文采用的证据推理基本模型如图21所示，该模型把样品的太赫兹时域谱、频域谱、折射率谱和吸收谱都作为推理的证据，构成多光学参数的融合模型，以克服单一光学参数的不可靠性。

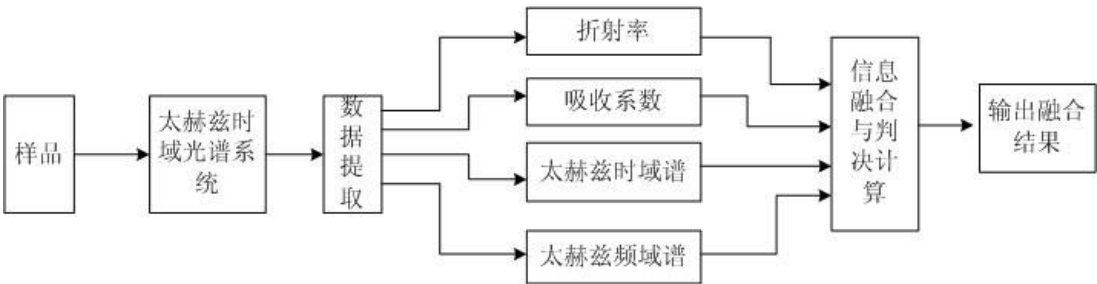


图 31 黄曲霉毒素 B1 浓度预测太赫兹多光学参数融合模型

上述融合模型是数据集上的融合，需融合的多源数据采用的是有限数目的离散值，而我们得到的样品太赫兹光谱的幅值是连续值，因此需要对光谱的连续幅值进行抽样，获得固定频点的光谱数据，作为融合模型的输入数据进行建模。

### 5.5.2 结合二重 D-S 推理算法的太赫兹光谱信息融合模型

利用D-S证据理论进行推理时，理论上多个证据的联合可信度要高于单个证据的可信度，但具体应用到本例中，由于获取的太赫兹光谱数据难以区分，甚至有交迭现象，只采用一个频点的多个数据的联合可信度，有时候仍然会有一定的偏差，为了进一步提高D-S证据推理的精度，我们在多个频点上提取多个光谱的数据，对每一频点应用D-S证据推理算法进行联合可信度计算，然后再把计算结果作为应用D-S证据推理计算联合可信度的新的证据，在多个频点上计算二次联合可信度，依此来提高D-S证据推理的精确性。

基于上述对D-S证据理论的改进思想，一个改进后的二重D-S证据推理的模型如图32所示。该模型将多个频点上的联合可信度再次作为D-S证据推理的证据，获得更为精确

的推理效果。

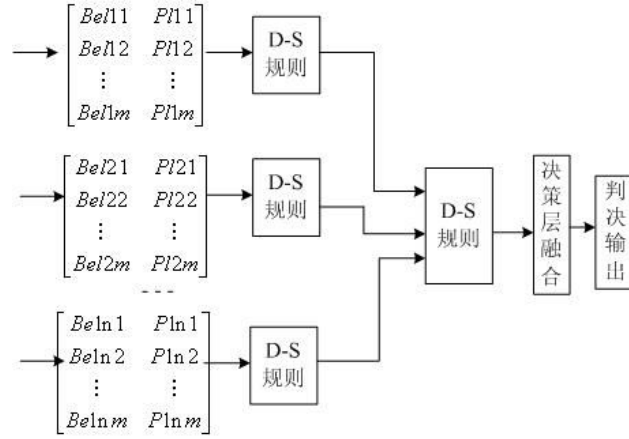


图 32 二重 D-S 证据推理模型

下面对本二重D-S证据推理模型的推理过程和精度举例说明。

假定在频点  $f_1$  处，对应的吸收系数和折射率对命题“黄曲霉毒素的浓度为10ppm”的支持度分别为：

$$m_1(a) = 0.6 \quad m_1(\Theta) = 0.4$$

$$m_2(a) = 0.5 \quad m_2(\Theta) = 0.5$$

而在频率点  $f_2$  上，吸收系数和折射率的值对命题“黄曲霉毒素的浓度为10ppm”的支持度分别为：

$$m_1(a) = 0.7 \quad m_1(\Theta) = 0.3$$

$$m_2(a) = 0.4 \quad m_2(\Theta) = 0.6$$

在频点  $f_1$  处，依据D-S证据规则，得到的联合可信度为：

$$Bel_1(a) = \begin{bmatrix} m(a) = 0.80 \\ m(\Theta) = 0.20 \end{bmatrix}$$

$$Bel_2(a) = \begin{bmatrix} m(a) = 0.82 \\ m(\Theta) = 0.18 \end{bmatrix}$$

而在频点  $f_2$  处，得到的联合可信度为：

根据图5.9的二重D-S证据推理模型，把以上两组联合可信度  $Bel_1(a)$  和  $Bel_2(a)$  再次作为证据，计算得到对命题“黄曲霉毒素的浓度为10ppm”的支持度为：

$$m_1(a) + m_2(a) = \begin{bmatrix} m_1(a) \cap m_2(a) = 0.656 \\ m_1(a) \cap m_2(\Theta) = 0.144 \\ m_1(\Theta) \cap m_2(a) = 0.164 \\ m_1(\Theta) \cap m_2(\Theta) = 0.036 \end{bmatrix}$$

这样就得到2个频点下联合可信度为：

$$Bel(a) = \begin{bmatrix} m(a) = 0.964 \\ m(\Theta) = 0.036 \end{bmatrix}$$

由此可见，在把某一浓度作为预测值时，由两个频点计算得到的联合可信度高于1个频点得到的联合可信度。由此可以推断，如果我们提取的作为证据的频点越多，计算的对浓度预测值的联合可信度就越精确。这是因为通过对多重的联合可信度的计算，减弱了某些频点由于干扰或测量的问题造成的本身可信度下降的问题，从而抵消了某些频点的推理误差。

## 5.6 实验结果与模型评价

取四种不同来源、发霉程度不同的小麦作为实际样品，然后按照5.3节的方法提取黄曲霉毒素。此时黄曲霉毒素的类型和浓度是未知的。用厚度为0.1mm的比色皿或液体池制成样品，在太赫兹时域光谱系统中测得样品时域谱、经快速傅里叶变换后得到频域谱、折射率谱和吸收系数谱。为了消除噪声和光源不稳定的影响，在光谱中间的0.5T-1.5T范围内，等间隔地取10个频点下的频谱值、折射率谱值和吸收谱值。然后根据已知浓度对应的各谱值的经验值为每个谱值设置初始支持度，经由5.5.2节描述的计算方法，计算多个抽样频点的联合可信度。根据计算结果，将最大联合可信度对应的浓度作为测量浓度。表10列出了部分实验的结果。

表 10 模型计算的四种样品的联合可信度

	5ppm	10ppm	25ppm	50ppm
样品 1	0.784	0.655	0.330	0.288
样品 2	0.654	0.816	0.704	0.425
样品 3	0.696	0.755	0.532	0.246
样品 4	0.322	0.224	0.016	0.128

例如在上表中，对于样品1，由于黄曲霉毒素浓度为5ppm具有最高的可信度，因而样品1中的黄曲霉毒素B1浓度的最佳预测值为5ppm，样品4由于各浓度可信度都比较低，



可以有三种可能：(1)浓度不在测定的范围内；(2)不是黄曲霉毒素B1，(3)未含有黄曲霉毒素B1。

如图33所示为采用不同数目的频点得到的浓度预测值的可信度。由图33可知，所选取的频点数越多，也即融合的次数越多，对黄曲霉毒素预测的浓度值的可信度越高，从而判定的准确性越高。

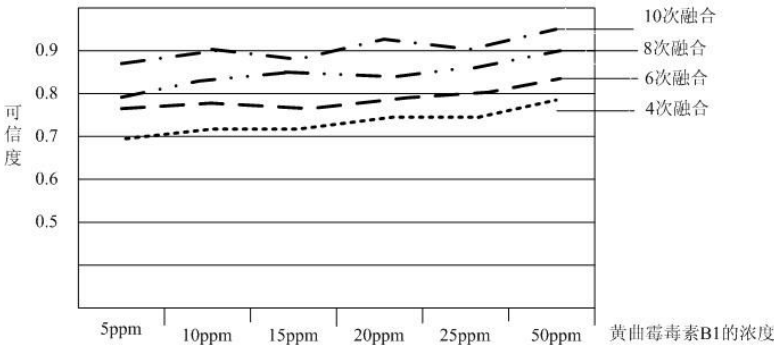


图 33 黄曲霉毒素浓度预测值的可信息与其数据融合次数的关系

### 5.7 本章小结

在本章中，我们综述了储粮真菌现有检测技术的研究现状，特别是物理检测方法的研究现状，给出了黄曲霉毒素的太赫兹检测方法的整体流程，详细描述了黄曲霉毒素太赫兹检测样品的制备过程。采用太赫兹吸收光谱，构建了基于最小二乘支持向量机的太赫兹吸收光谱与其浓度的线性回归模型。针对太赫兹吸收光谱受外界因素干扰和影响，光谱之间难以区分的问题，基于多源信息融合的理论，对传统的证据推理方法进行改进，提出了基于光谱离散频点的二重D-S证据推理模型，模型的应用表明，采用本文的二重证据推理方法，预测浓度值的可信度大大提高了。



## 6 结论与展望

### 6.1 总结与结论

太赫兹波段是电磁波谱最后被开发利用的波段，太赫兹辐射具有一些特殊的性质，使它在对物质特别是有机分子的探测上具有其他波段无法比拟的优势，本文首先对太赫兹波的产生和探测进行了初步的原理性的描述，综述了太赫兹波段在各领域特别是农业领域的应用。针对粮食品质的检测技术，本文结合化学计量学的方法，分别应用太赫兹频域谱、吸收谱和折射率谱，对不同品质小麦样品、不同种类小麦样品的区分以及小麦储藏过程中容易出现的黄曲霉毒素的检测等问题开展了一系列实验性的研究工作，本文的主要研究工作包括以下方面：

1. 将太赫兹时域光谱技术与化学计量学的方法结合，用于实现对四种品质的小麦的分类识别。首先采用主成分分析法，对吸收光谱提取4个主成分，对折射率光谱提取3个主成分，然后把提取的主成分作为SVM的输入，利用SVM算法实现对小麦样品的分类识别。实验表明，应用PCA-SVM模型，要比单独使用SVM模型，样品分类效果要好。另外，本文也把PCA-SVM模型与其他太赫兹光谱分析常用模型进行了比较，说明了PCA-SVM模型的优势。

2. 采用太赫兹光谱和PLS算法对8种小麦样品的种类进行了确认和预测。使用太赫兹时域光谱系统测量了8种小麦样品在0.2-2.0THz内的吸收谱和折射率谱，使用偏最小二乘算法PLS构建回归模型，对8种小麦样品的种类进行了区分。通过选择光谱中的最佳区域，开发了间隔偏最小二乘算法(iPLS)用于模型的优化。实验表明，采用iPLS优化后的模型具有更好的预测精度。

3. 给出了黄曲霉毒素的太赫兹检测方法的整体流程，详细描述了黄曲霉毒素太赫兹检测样品的制备过程。采用太赫兹吸收光谱，构建了基于最小二乘支持向量机的太赫兹吸收光谱与其浓度的线性回归模型。针对太赫兹吸收光谱受外界因素干扰和影响，光谱之间难以区分的问题，基于多源信息融合的理论，对传统的证据推理方法进行改进，提出了基于光谱离散频点的二重D-S证据推理模型。实验结果表明，通过采用本文的二重证据推理方法，大大提高了预测浓度值的可信度。

通过总结本文以上的工作，我们得到以下结论：采用太赫兹光谱技术，再结合化学

计量学的方法，可以对小麦的品质、品种和小麦中黄曲霉毒素的含量进行定性和定量的鉴别和分析。从中也可以看出，作为红外光谱和拉曼光谱的一种有益的补充，太赫兹光谱技术在农产品质量安全检测领域具有瞩目的应用前景。

## 6.2 问题与展望

太赫兹技术作为一种新型的光谱技术，近年来的应用得到很大的重视，我们把它作为粮食微观信息的一种探测手段，在基础理论研究和应用研究方面取得了一定的成绩，也为后续的研究打下了一定的基础，但在整个的研究过程中，也存在着一些问题，主要包含以下方面：

(1) 样品的种类还不够。在研究中，我们只涉及了四种品质的小麦和8个种类的小麦样品，没有考虑小麦更多的种类以及同种小麦不同地域、不同气候条件的差别，使得模型的适用性还不强。

(2) 在真菌毒素的太赫兹检测方面，只进行了黄曲霉毒素的检测，没有进行其他种类毒素的检测，也没有考虑多种毒素混合在一起的情况。而且，毒素的检测限不高，精度还难以与现有的生化方法相比。

解决上面的问题，除了采用更多丰富的样品以外，为了提高检测的精度，还需进行样品制备方法、化学计量学算法的研究，提高样品中待测物质的检测限。同时，通过设备改造，增加仪器的太赫兹频谱宽度，争取能够探测到更多有价值的光谱信息。

## 参考文献

- [1] 薛传平,吴梅松,李元明.高度重视小麦储藏损耗—确保国家小麦安全[J].粮油仓储科技通讯. 2009(2):12-13.
- [2] 赵丽瑾.国内微生物检测产品的现状和分析[J].中国卫生检验杂志, 2000, 10(6):667-668
- [3] 陈维斌,宋玉清.农产品微生物快速检测方法的研究[J].杂粮作物, 2001,21(4):4-46
- [4] 尹在宁,王火根.粮情测控系统在使用中存在的问题及分析[J]. 粮油仓储科技通讯. 2003(2):50-51.
- [5] 李华.平板菌落计数的改进方法[J].生物学通报. 2006,41(1):51-52
- [6]王秀芹,张敏,颜萍. 食品微生物检测技术研究进展[J]. 微生物学杂志,2012,06:77-79.
- [7]徐笑. 食品微生物检测技术现状与展望[J]. 环境保护与循环经济,2012,12:25-27.
- [8] 王苗. ATP 荧光微生物检测法在食品卫生监控领域中的应用与展望[J]. 中国食品杂志. 2004,16(3):266-267
- [9] Tuo Wang, Andre Salazar, Olga A. et al. Structure and Dynamics of Brachypodium Primary Cell Wall Polysaccharides from Two-Dimensional  $^{13}\text{C}$  Solid-State Nuclear Magnetic Resonance Spectroscopy[J]. Biochemistry, 2014, 53 (17), pp 2840–2854
- [10] Francesca Cicogna, Calogero Pinzino, Sabrina Castellano, et al. Interaction of Azole Compounds with DOPC and DOPC/Ergosterol Bilayers by Spin Probe EPR Spectroscopy: Implications for Antifungal Activity[J]. J. Phys. Chem. B, 2013, 117 (40), pp 11978–11987
- [11] Wen-Sen He, Ji Yin, Han-Shan Xu, et al. Efficient Synthesis and Characterization of Ergosterol Laurate in a Solvent-Free System[J]. J. Agric. Food Chem., 2014, 62 (48), pp 11748–11755
- [12] 蔡静平.小麦微生物活性及其应用的研究[J]. 中国粮油学报. 2004, 19( 4): 76-79
- [13] Yong Kyoung Yoo, Myung-Sic Chae, Ji Yoon Kang, et al. Multifunctionalized Cantilever Systems for Electronic Nose Applications[J]. Anal. Chem., 2012, 84 (19), pp 8240–8245
- [14] Hanjiang Zhu, Xueqi Li, Charles F, et al. Ultrahigh Performance Liquid Chromatography Analysis of Volatile Carbonyl Compounds in Virgin Olive Oils[J]. J. Agric. Food Chem., 2013, 61 (50), pp 12253–12259.
- [15] Olesya I. Zhuravleva, Maria P. Sobolevskaya, Elena V, et al. Meroterpenoids from the Alga-Derived Fungi *Penicillium thomii* Maire and *Penicillium lividum* Westling[J]. J. Nat. Prod., 2014, 77 (6), pp

1390–1395.

- [16] Ron G. Buttery and Gary R. Takeoka. Cooked Carrot Volatiles. AEDA and Odor Activity Comparisons. Identification of Linden Ether as an Important Aroma Component[J]. J. Agric. Food Chem., 2013, 61 (38), pp 9063–9066.
- [17] Stefanie Wroblewitz, Liane Hüther, Remy Manderscheid. Effect of Rising Atmospheric Carbon Dioxide Concentration on the Protein Composition of Cereal Grain[J]. J. Agric. Food Chem., 2014, 62 (28), pp 6616–6625.
- [18] Borjesson, T., Eklov, T., Jonsson, A., et al. Electronic nose for odor classification of grains. Cereal Chemistry, 1996, (73), pp 457–461
- [19] Borjesson, T., Olsson, J. Electronic nose for quality control in cereals: relations to ergosterol and mycotoxin content. In: Proceedings of Symposium on Electronic Noses in the Food Industry, 16–17th November, 1998, Stockholm, Sweden, pp. 24–28
- [20] Keshri, G., Magan, N., Voysey, P., Use of an electronic nose for the early detection and differentiation between spoilage fungi[J]. Letters in Applied Microbiology, 1998, (27), pp 261–264.
- [21] Jonsson, A., Winqvist, F., Schnurer, J., Sundgren, H., Lundstrom, I., Electronic nose for microbial quality classification of grains. International Journal of Food Microbiology, 1997, (35), 187–193.
- [22] K. L. Andrew Chan and Pedro L. V. Fale. Label-Free in Situ Quantification of Drug in Living Cells at Micromolar Levels Using Infrared Spectroscopy[J]. Anal. Chem., 2014, 86 (23), pp 11673–11679.
- [23] 柴阿丽, 李金萍, 石延霞, 谢学文, 李宝聚. 基于傅里叶变换红外光谱和聚类分析的真菌鉴别[J]. 光谱学与光谱分析, 2010, 11: 2941–2944.
- [24] Whitlow L W, W. M. Hagler Jr. W M. Mycotoxins in feeds[J]. Feedstuffs, 2004, 76(38), pp 66–67.
- [25] Hans Pettersson, Lena Aberg. Near infrared spectroscopy for determination of mycotoxins in cereals. 2003, 14(4), pp 229–232.
- [26] Jan U. Porep, Martin E. Erdmann, Adrian Korzendorfer, et al. Rapid determination of ergosterol in grape mashes for grape rot indication and further quality assessment by means of an industrial near infrared/visible (NIR/VIS) spectrometer – A feasibility study[J]. Food Control, 2014, (43), pp 142–149
- [27] Giorgio Mancinelli, Maria Letizia Costantini, Loreto Rossi. Predicting ergosterol in leaf litter by near-infrared spectroradiometry: A preliminary assessment[J]. European Journal of Soil Biology, 2014,

(63), pp49-54

- [28] Roeland Van Wijk, Eduard P.A. Van Wijk, Herman A. van Wietmarschen, et al. Towards whole-body ultra-weak photon counting and imaging with a focus on human beings: A review[J]. Journal of Photochemistry and Photobiology B: Biology, 2014, (139), pp39-46
- [29] Mortaza Iranifam. Analytical applications of chemiluminescence methods for cancer detection and therapy[J]. TrAC Trends in Analytical Chemistry, 2014, (59), pp156-183
- [30] Yan, Y., Popp, F.A., Rothe, G.M. Correlation between germination capacity and biophoton emission of barley seeds[J]. Seed Science and Technology, 2003, 31(2), pp 249-258.
- [31] 张宁,李敏,李培武,张奇,张文,李冉. 酶联免疫吸附法(ELISA)测定黄曲霉毒素 B<sub>1</sub> 的样品基质效应研究[J]. 中国油料作物学报, 2014, 03:404-408.
- [32] 刘光明.用 PCR-ELISA 技术检测转基因食品的研究[J].食品科学.2003 (1):101-105.
- [33] S Baumgartner, S I Steiner, S Kloiber, et al. Towards the development of a dipstick immunoassay for the detection of trace amounts of egg proteins in food[J]. Euro Food Res Techno, 2002, 214:168-170.
- [34] 朱慧莉,黎锡流.酶联免疫吸附法及其在食品分析中的应用[J].食品工业科技.2001, 22(2):80-82
- [35] Karl Heinz Trebesius, Lorenz Lititz, Kristin Adler, et al. Culture independent and rapid identification of bacterial pathogens necrotizing fasciitis and streptococcal toxic shock syndrome by Quiescence in situ hybridisation[J]. Med Microbial Immunol, 2000, (188), pp169-175.
- [36] 程树峰,唐芳,伍松陵.小麦真菌危害早期检测方法的研究[J].中国粮油学报.2011, 26(4):85-88..
- [37] 冯永建,何学超,刘云花等.小麦真菌危害早期检测技术初步探索[J].小麦储藏.2010, 39(6):44-48.
- [38] Dragoman D, Dragoman M. Terahertz fields and applications [J]. Progress in Quantum Electronics, 2004, 28(1): 1-66.
- [39] Sherwin M. Applied physics - Terahertz power [J]. Nature, 2002, 420(6912): 131-132.
- [40] Jepsen P U, Jacobsen R H, Keiding S R. Generation and detection of terahertz pulses from biased semiconductor antennas [J]. Journal of the Optical Society of America B-Optical Physics, 1996, 13(11): 2424-2436.
- [41] Winnemisser C, Jepsen P U, Schall M, et al. Electro-optic detection of THz radiation in LiTaO<sub>3</sub>, LiNbO<sub>3</sub>, and ZnTe [J]. Applied Physics Letters, 1997, 70(23): 3069-3071.
- [42] Auston D H, Cheung K P. Coherent Time-Domain Far-Infrared Spectroscopy [J]. Journal of the

- Optical Society of America B-Optical Physics, 1985, 2(4): 606-612.
- [43] Cai Y, Brener I, Lopata J, et al. Design and performance of singular electric field terahertz photoconducting antennas [J]. Applied Physics Letters, 1997, 71(15): 2076-2078.
- [44] Xu L, Zhang X C, Auston D H. Terahertz Beam Generation by Femtosecond Optical Pulses in Electrooptic Materials [J]. Applied Physics Letters, 1992, 61(15): 1784-1786.
- [45] Zhang J, Hong Y, Braunstein S L, et al. Terahertz pulse generation and detection with LT-GaAs photoconductive antenna [J]. Iee Proceedings-Optoelectronics, 2004, 151(2): 98-101.
- [46] 张红梅.基于气体传感器阵列的几种农产品品质检测研究[D].浙江大学,2007
- [47] Hadjiloucas S, Karatzas L S, Bowen J W. Measurements of leaf water content using terahertz radiation [J]. Ieee Transactions on Microwave Theory and Techniques, 1999, 47(2): 142-149.
- [48] Hadjiloucas S, Galvao R K H, Bowen J W. Analysis of spectroscopic measurements of leaf water content at terahertz frequencies using linear transforms [J]. Journal of the Optical Society of America a-Optics Image Science and Vision, 2002, 19(12): 2495-2509.
- [49] 孙金海.用太赫兹时域光谱技术对玉米种子和系列毒品的研究[D].首都师范大学, 2005.
- [50] 孙金海,沈京玲,郭景伦等.太赫兹时域光谱技术在玉米种子鉴定中的实验研究[J].光学技术.2008,34(4):541-546.
- [51] Dorney T D, Baraniuk R G, Mittleman D M. Material parameter estimation with terahertz time-domain spectroscopy [J]. Journal of the Optical Society of America a-Optics Image Science and Vision, 2001, 18(7): 1562-1571.
- [52] Duvillaret L, Garet F, Coutaz J L. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy [J]. Ieee Journal of Selected Topics in Quantum Electronics, 1996, 2(3): 739-746.
- [53] Duvillaret L, Garet F, Coutaz J L. Highiy Precise determination of optical constraints and sample thickness in terahertz time-domain spectroscopy [J]. Applied Optics, 1999, 38(2): 409-415.
- [54] 刘畅,王新柯,孙文峰等.非极性有机溶剂光学参数的太赫兹波精确测量[J].光谱与光谱分析, 2011, 31(11): 2886-2890.
- [55] Liu H. Terahertz spectroscopy for chemical and biological sensing applications[D]. NY, USA: Rensselaer Polytechnic Institute, Center for Terahertz Research, 2006.



- [56] 陈敏伯, “计算化学-从理论化学到分子模拟”, 科学出版社, 2009.
- [57] Katrin S, Philippe C C, Reinald B, et al. Automatic selection of a representative trial from multiple measurements using Principle Component Analysis [J]. Journal of Biomechanics, 2012, 45(13): 2306-2309.
- [58] Noori R, Sabahi M S, Karbassi A R, et al. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set [J]. Desalination, 2010, 260(1-3): 129-136.
- [59] Parkkinen J P S, Hallikainen J, Jaaskelainen T. Characteristic Spectra of Munsell Colors [J]. Journal of the Optical Society of America a-Optics Image Science and Vision, 1989, 6(2): 318-322.
- [60] He M, Yang G L, Xie H Y. A hybrid method to recognize 3D object [J]. Optics Express, 2013, 21(5): 6346-6352.
- [61] Cristianini N, Shawe-Taylor J, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods [M]. Cambridge University Press, Cambridge, 2000.
- [62] Cortes C, Vapnik V. Support-Vector Networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [63] Maali Y, Al-Jumaily A. Self-advising support vector machine [J]. Knowledge-Based Systems, 2013, 52: 214-222.
- [64] Zhang Y, Peng X H, Chen Y, et al. A first principle study of terahertz (THz) spectra of acephate [J]. Chemical Physics Letters, 2008, 452(1-3): 59-66.
- [65] Marengo E, Bobba M, Robotti E, et al. Hydroxyl and acid number prediction in polyester resins by near infrared spectroscopy and artificial neural networks [J]. Analytica Chimica Acta, 2004, 511(2): 313-322.
- [66] 许俐, 钱敏娟, 方如明, 等. 大米加工精度的图像识别方法 [J]. 农业工程学报, 1996, 12(3): 172-175.
- [67] 许俐. 机器视觉测量大米留胚率的研究 [J]. 江苏理工大学学报, 1997, 18(6): 8-11.
- [68] 赵宗贵. 信息融合技术现状、概念与结构模型[J]. 中国电子科学研究院学报, 2006, 04: 305-312.
- [69] 张明路, 戈新良, 唐智强, 刘兴荣. 多传感器信息融合技术研究现状和发展趋势[J]. 河北工业大学学报, 2003, 02: 30-35.
- [70] 周前祥, 敬忠良, 姜世忠. 多源遥感影像信息融合研究现状与展望[J]. 宇航学报, 2002, 05: 89-94.



## 致谢

值此即将毕业之际，回首三年人学习和生活，感触颇多，要感谢的人也很多。本论文是在尊敬的张元教授的指导下完成的。我首先要衷心的感谢我的导师张元老师。自研究生入学以来，张老师不论从学习上和生活上都给予了我极大的帮助，更是传授了我许多做人的道理。张老师渊博的学识、严谨的治学作风、明锐的学术眼光和勤奋的工作精神使我受益匪浅；并且为人坦诚，对待同学更是悉心教导，在广大师生中口碑极佳，他对待科学的严谨态度和锐意进取的精神更是我今后学习和工作的楷模。在这次毕业论文课题的选择、实验与写作的整个过程都一直指引我不断地去探索、实践，因此我的论文能顺利地完成，要非常感谢张老师的教导、帮助与支持。

特别感谢廉飞宇老师和葛宏义师兄及蒋玉英师姐在实验及数据分析上给予我的理论指导与帮助。

感谢我们宿舍的成员卢佳明，高星，张辰在研究生期间的一直陪伴；感谢孙彪瑞、苏庭奕、剧新刚、张李健与李鹏鹏在学习与生活上的帮助；感谢研究生所有的同学一直以来的互相帮助。

深深地感谢我的父母、爱人、孩子和亲朋好友，在漫长的求学过程中，他们始终如一的为我的学业给予了最大的支持和鼓励，他们对我的关爱和教诲是我前进的动力，是他们让我明白了我身上的责任。

再一次向所有关心、帮助、支持和鼓励过我的老师、同学、亲人和朋友致以最诚挚的谢意！



## 个人简历

### 个人概况:

姓 名:	赵中原	性 别:	男
民 族:	汉	出生年月:	1981.7
学 历:	硕士	专 业:	计算机科学与技术
研究方向:	太赫兹技术与应用研究	导 师:	张元
论文题目:	基于 THz 技术的小麦品质无损检测研究		

### 教育经历:

2013 年 9 月~2016 年 6 月 河南工业大学信息科学与工程学院 计算机科学与技术  
硕士;

2007 年 3 月~2009 年 7 月 周口师范学院计算机科学系 计算机科学与技术 本科。

2003 年 9 月~2006 年 6 月 周口师范学院计算机科学系 计算机应用与维护 专科

### 科研成果:

1. ZHAO Zhong-yuan, LIAN Fei-yu, Yuan Zhang. Aflatoxin B1 Detected by Terahertz time-domain Spectroscopy , 第八届生物医学工程与信息学国际学术会议(BMEI 2015)
2. 赵中原, 廉飞宇, 张元. 一种粮堆中无线传感器网络的络覆盖保持路由算法 , 周口师范学院学报 (录用)