

分类号 O657.33

UDC 543.4

密级 公开

编号 10299S0509092



江蘇大學

學位論文

应用近红外光谱分析技术在线检测牛奶 成分的研究

Application of Near Infrared Spectroscopy Analytical Technology
to on-line Determination of Constituents in Cow Milk

邹贤勇

指导教师 陈斌 教授

江苏大学食品与生物工程学院

申请学位级别 硕士 专业名称 食品科学

论文提交日期 2008年5月 论文答辩日期 2008年6月

学位授予单位和日期 江苏大学 2008年6月

答辩委员会主席 _____

评阅人 _____

2008年6月

学位论文版权使用授权书


本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权江苏大学可以将本学位论文的全部内容或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

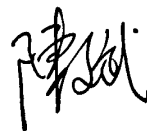
保 密 ， 在 3 年解密后适用本授权书。

不保密 。

学位论文作者签名：

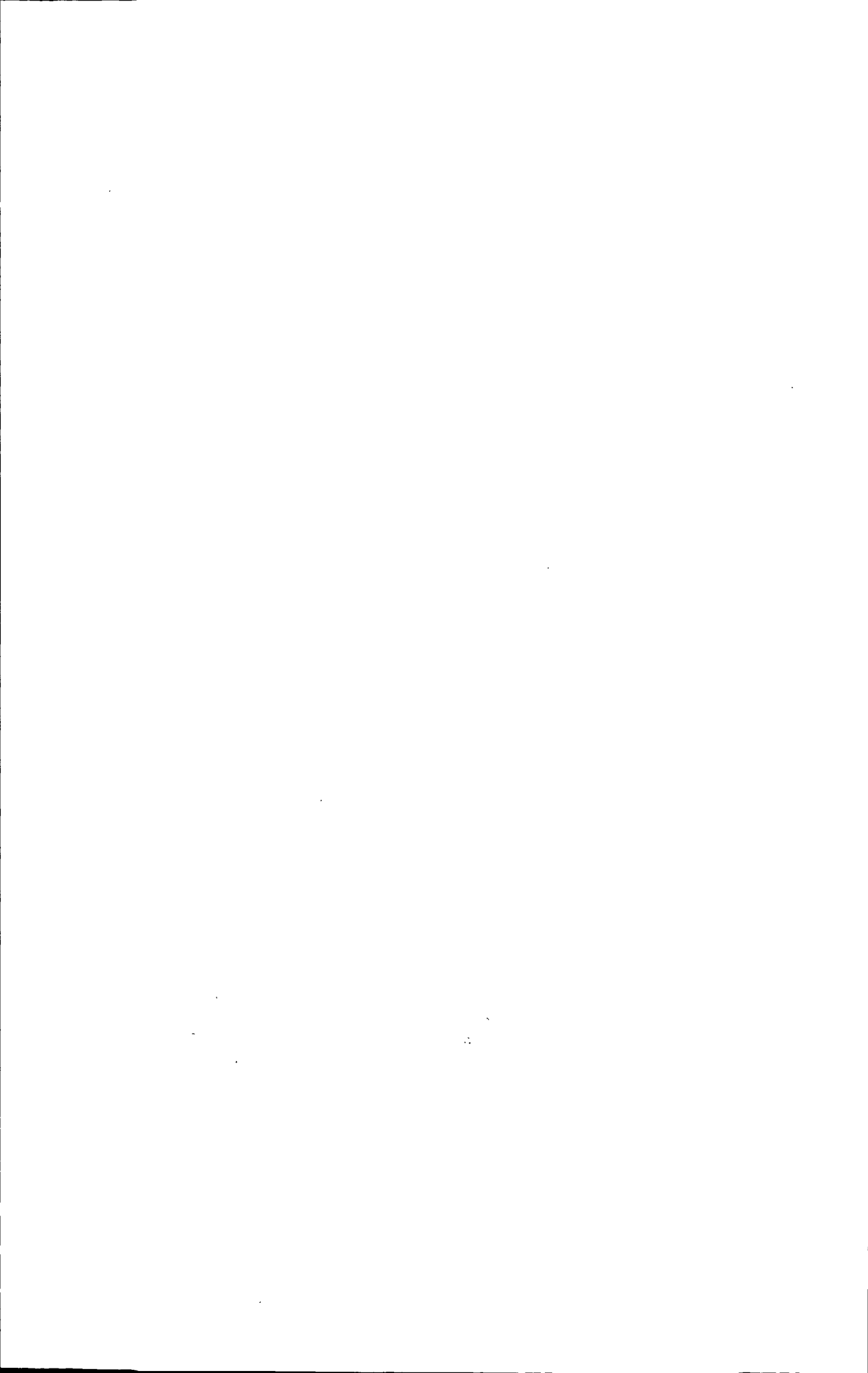


指导教师签名：



2008 年 6 月 9 日

2008 年 6 月 9 日



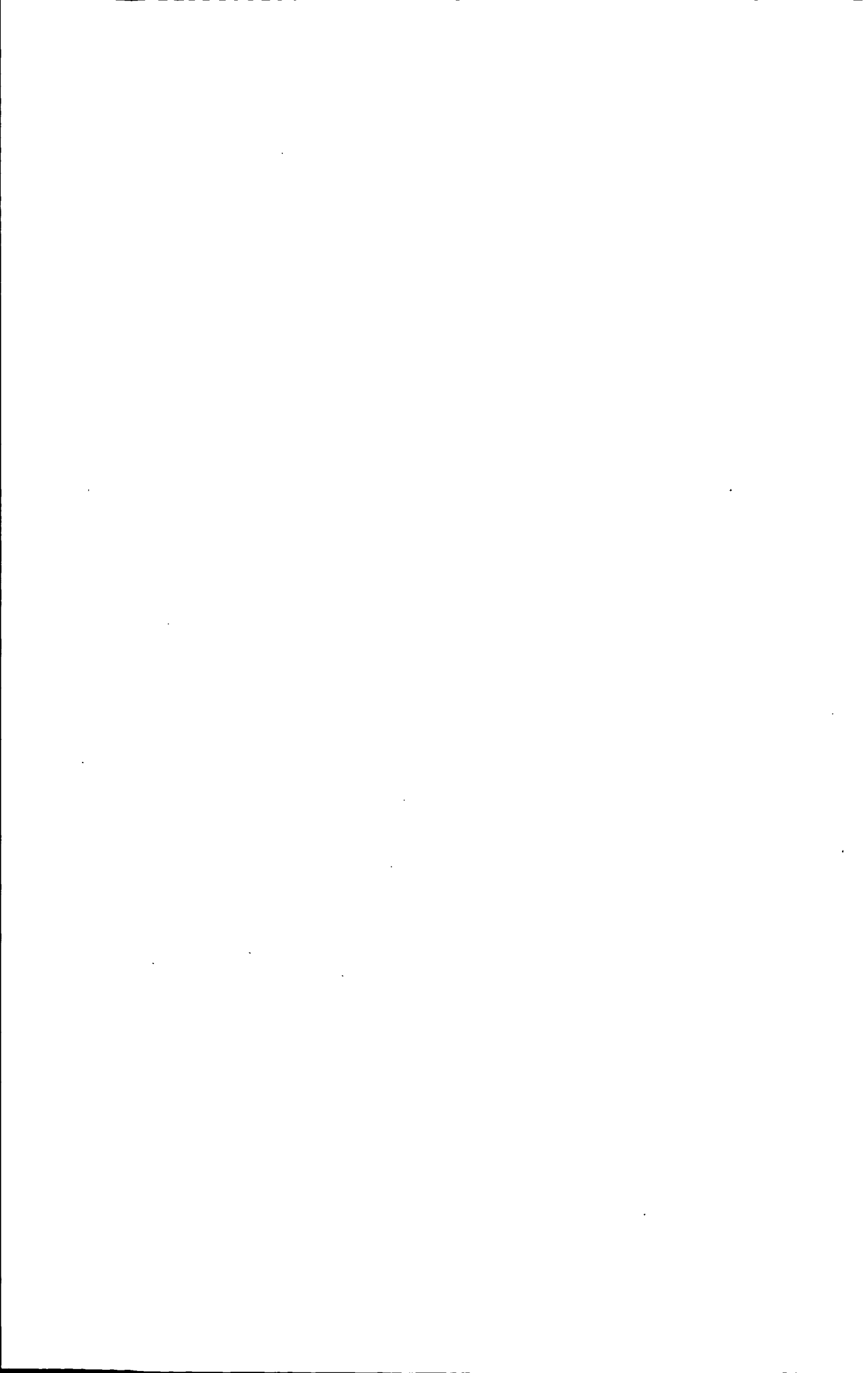


独创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已注明引用的内容以外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：郭俊勇

日期：2008年6月9日



应用近红外光谱分析技术在线检测牛奶 成分的研究

**Aplication of Near Infrared Spectroscopy Analytical Technology
to on-line Determination of Constituents in Cow Milk**

专业名称 食品科学

指导教师 陈 斌 教授

姓 名 邹 贤 勇

2008 年 6 月



摘 要

本论文基于本课题组自行研制的近红外液态在线检测试验台,以牛奶为研究对象,以实现牛奶主要成分蛋白质和脂肪含量在线检测为目标,综合多学科的知识对近红外光谱数据中的信息提取和模型建立中遇到的关键问题给出了解决方案,论证了该技术的可行性。该研究对实现牛奶成分含量快速、实时、准确的检测具有重要的指导意义和参考价值。主要有以下研究内容:

1. 以实现牛奶光谱在线检测的可行性和稳定性为原则,基于本近红外在线检测设备,对试验的各项参数(本底、扫描次数、牛奶温度、流速、光纤探头光程)进行了优化,这些参数的优化为牛奶样品光谱采集提供了参考;

2. 比较了多种预处理方法的效果,找出了蛋白质和脂肪的最佳预处理方法,同时探讨了 MSC 方法在本试验中的可行性并分析了原因;比较了各预处理方法对不同温度下预测模型的预测效果,表明了仅使用光谱预处理方法不能消除温度对模型的影响,无法建立温度稳健模型;选择了蛋白质和脂肪最佳建模波段,然后利用偏最小二乘法分别建立了牛奶样品的蛋白质和脂肪校正模型并对模型的预测性能进行了评价,蛋白质预测模型相关系数、预测均方差和平均偏差分别为:0.906 7、0.308 7 和 0.201 7;脂肪预测模型相关系数、预测均方差和平均偏差分别为:0.929 9、0.399 6 和 0.246 1。结果表明:本试验建立的牛奶蛋白质和脂肪的模型预测性能较好,基本上可以满足牛奶成分的在线检测精度要求;

3. 模型的稳健性是评价在线近红外分析技术的一个重要指标。在 26℃~41℃ 范围内,研究了温度对预测牛奶成分的影响。把 36℃ 下建立的校正模型用来预测不同温度下的验证集发现,36℃ 下建立的牛奶成分模型只对相同温度下的预测模型有效,这说明了样品温度对预测结果的重要性;

4. 采集了不同温度下牛奶的光谱组成温度补偿校正集,通过补偿校正方法的引入,建立了温度全局模型,实现了模型对温度因素的稳健性。补偿校正方法的引入虽然在一定程度上降低了预测模型精度,但该方法简单可行,还可用于建立其他影响因素(如一些难于控制的干扰因素:气泡、压力、样品状态等)的稳健预测模型。该方法为进一步研究在线近红外分析模型稳健性提供了重要的参考。

关键词:近红外光谱,牛奶成分,在线检测,稳健模型,偏最小二乘法



ABSTRACT

In this paper, the objectives possibility of using near infrared spectroscopy(NIRS) analysis technique is to investigate the spectral for on-line detection of the major constituents (protein and fat content) of milk, which was based on NIR on-line liquid detecting equipment developed by the study group. The study was carried out by several subjects methods to come up with solution on the abstraction of weak spectral signal and establishing model and the feasibility of the technology was demonstrated. The study can provide guiding significance and reference value for rapid, real-time and correct non-detection of the major constituents of milk.

The main contents and conclusions are as follows:

1. Based on NIR on-line liquid detecting equipment, in order to achieve the feasibility and stability of on-line detecting the spectrum of milk, the various parameters of the experience (reference, scanning frequency, temperature and flow rate of milk, optical length of the optical-fiber probe) were optimized, which provided a reference for further sampling the spectrum of milk.
2. In order to find the best method of pretreatment, various methods were used to pretreat the original spectrums, and the feasibility of multiple scatter correction(MSC) also was investigated in the study and the reason was analyzed; The effect of different pretreatment on the validation model under different temperature was researched. The results showed that the robust calibration model on temperature was unable to established only by the methods of spectrum pretreatment. The best spectrums band was Chosen to predict protein and fat content, and then the calibration Model was established by the method of partial least squares(PLS) and the forecast capacity of the model was verified. In the protein validation model, correlation coefficient, root mean square error of prediction and mean bias are 0.990 2、0.155 8 and 0.124 6, respectively; In the fat validation model, correlation coefficient, root mean square error of prediction and mean bias are 0.980 5、0.207 2 and 0.149 6, respectively. The results showed that: the model established to forecast milk protein and fat content had a nice

forecast capacity, which can meet the needs of on-line detection accuracy for milk constituents.

3. the robustness of calibration model is a important index in on-line NIR analysis technology. In the scope of 26 ~ 41°C, the effect of temperuatre on measurement of milk constituents by NIR spectroscopy was studied. The calibration model built in 36°C was used to predict the valibration in different temperature. The result showed that the calibration model built in 36°C can only predict the valibration under the same temperature, which proved that the sample temperature had importance on forecasting results.
4. In order to compensate the effect of temperature on measurement by NIR spectroscopy , calibrations of milk constituents on different temperature were developed. To a certain extent, the method of compensation for temperature correction could reduce the accuracy of the prediction model ,and it is simple and feasible, can also be used to establish rebust varibration model on other factors(such as bubble, pressure, state of samples, which were difficult to be controlled). The method provided an important reference for further study of the robustness of calibration model in the on-line near infrared analysis technology.

KEY WORDS: near-infrared spectrum, milk, on-line determination, Robust calibration model, partial least square

目 录

第一章 绪论	1
1.1 近红外光谱分析技术简介	1
1.1.1 近红外光谱分析技术的发展	1
1.1.2 近红外光谱分析技术的应用	2
1.1.3 近红外光谱分析技术的基本原理	3
1.1.4 近红外光谱分析技术特点	4
1.2 近红外光谱分析中的化学计量学理论	5
1.3 本课题研究的目的、意义及内容	6
1.3.1 牛奶近红外光谱分析国内外研究现状	6
1.3.2 课题研究的日和意义	8
1.3.3 本文的主要内容	9
第二章 在线近红外光谱分析牛奶成分的基础	10
2.1 在线近红外光谱分析系统	10
2.1.1 硬件技术	10
2.1.2 软件技术	13
2.2 牛奶成分近红外在线检测的可行性	14
2.2.1 牛奶体系的构成及光学特性	14
2.2.2 近红外光谱分析技术在线检测牛奶成分的可行性	14
2.3 本章小结	15
第三章 试验过程和方法	16
3.1 试验材料	16
3.1.1 试验设备与仪器	16
3.1.2 牛奶样品的收集	17
3.2 检测系统参数的选择	17
3.2.1 本底的选择	17
3.2.2 扫描次数的选取	18
3.2.3 样品温度的选择	19
3.2.4 最佳光程和流速的选择	20
3.3 样品光谱的采集与数据处理	23
3.3.1 样品光谱的采集	23

3.3.2 数据处理与模型评价指标	23
3.4 本章小结	24
第四章 牛奶成分校正模型的建立及验证	26
4.1 试验方案	26
4.2 最佳数据预处理方法的选择	26
4.2.1 几种预处理方法的原理	26
4.2.2 预处理方法结果与分析	28
4.2.3 多元散射校正 (MSC) 适用性探讨	29
4.3 牛奶成分模型的建立及验证	30
4.3.1 最佳建模波段的选择	30
4.3.2 牛奶蛋白质和脂肪模型的建立及验证	31
4.4 本章小结	33
第五章 温度对牛奶近红外在线检测的影响	35
5.1 试验方案	35
5.2 样品温度对牛奶成分预测结果的影响	35
5.2.1 温度对牛奶光谱的影响	35
5.2.2 温度对牛奶成分预测结果的影响	36
5.2.3 光谱预处理方法对模型温度稳健性的影响	37
5.3 温度补偿校正方法的效果	38
5.4 本章小结	41
第六章 总结及展望	42
6.1 研究总结	42
6.2 本文的创新点	43
6.3 展望	43
致 谢	44
参考文献	45
硕士期间发表论文	48
附录 A 60 个牛奶样品的化学检测值	49
附录 B 50 个奶粉样品的化学检测	50

第一章 绪论

牛奶作为一种营养全面的理想乳品,越来越受到人们的青睐。实时、快速、准确地检测牛奶成分对实现乳业生产过程的自动化及提高乳品质量具有重要意义。目前牛奶成分检测仍以化学分析方法为主导,虽然其检验结果准确,但过程耗时费力、破坏样品、更无法进行实时在线检测。

近年来,随着计算机技术和化学计量学的飞速发展,现代近红外(Near Infrared, NIR)光谱分析技术以其高效、快速、无损、适合在线检测等优点在分析测试领域中(特别是在工业在线分析中^[1,2])得到越来越广泛的应用。

1.1 近红外光谱分析技术简介

近红外光是指波长介于可见区(VIS)和中红外区(MIR 或 IR)之间的电磁波,其波长范围约为 780~2 500nm,波数范围约为 12 500~4 000 cm^{-1} 。习惯上又将近红外光谱划分为短波近红外(780nm~1 100nm)和长波近红外(1 100nm~2 500nm)两个区域。近红外谱区是在 1800 年由 Herschel 发现的,是人类认识最早的非可见光谱区^[3]。

近红外光谱信息主要反映有机物分子中 C-H、N-H、O-H、S-H 等含氢基团的倍频与合频吸收,可以用来对物质进行定性分析和定量分析^[4]。近红外光谱分析技术是一种间接测量技术,即需要结合化学计量学方法建立校正模型,将校正集样品的光谱与样品已知参数关联,然后通过建立的校正模型和测定的未知样品的光谱信息进行分析,最终得到未知样品的组成信息。

1.1.1 近红外光谱分析技术的发展

近红外光谱分析技术的发展经历了以下几个主要阶段。1800 年 Herschel 第一次发现了近红外区域^[5]。19 世纪末,Abney 和 Festing 在近红外短波区域首次记录了有机化合物的近红外光谱^[6]。1928 年 Brackett 测得第一张高分辨率的 NIR 图,并对有关基团的光谱特征进行了解释^[7]。由于缺乏可靠的仪器基础,20 世纪 50 年代以前,近红外光谱的研究只限于为数不多的几个实验室中,而且没有得到实际应用。20 世纪 50 年代中期,Kave^[8]首先研制出能准确得到 NIR 光谱的仪器,一些公司也相继开发了商业化的仪器。Norris 等人在近红外光谱漫反射技术上所做的大量工作^[9],掀起了近红外光谱应用的一个小高潮,近红外光谱分析技术在测定农副产品的品质方面得到了广泛

使用。进入 20 世纪 60 年代中后期, 由于经典近红外光谱分析技术暴露出灵敏度低、抗干扰性差的弱点, 加之各种新的分析技术的出现, 使人们淡漠了近红外光谱分析技术。20 世纪 80 年代中后期, 随着计算机技术的发展和化学计量学研究的深入, 加之近红外光谱仪器制造技术的日趋完善, 促进了现代近红外光谱分析技术的快速发展。20 世纪 90 年代至今, 通过一些厂家和科研单位的积极努力, 在近红外光谱仪器的研制、软件研发等方面取得了一定的成绩。尤其是光纤技术以及化学计量学等软件技术在近红外光谱分析仪器上的应用, 带动了现场、实时、在线检测技术的发展。同时, 借助强有力的 Internet 网络资源, 带动了近红外光谱分析技术的网络化, 部分科研院所已成功研制出了近红外网络远程分析软件^[10,11]; 但此技术只是处于研究阶段, 成熟的产品还有待发展和提高。

1.1.2 近红外光谱分析技术的应用

现代近红外光谱分析技术经过半个多世纪的发展, 已经在食品、精细农业、环境、化学、高聚物、药物、纺织、石油化学、生物医学、生命科学、化妆品以及医学等十几个领域中得到成功应用。

农产品品质分析是近红外分析技术应用的第一个领域, 1974 年加拿大谷物中心使用近红外方法代替传统的凯氏定氮法测定了小麦中的蛋白质^[12]。现今近红外分析技术在农产品分析及作物育种方面已得到长足发展, 也是近红外技术应用最成熟的一个领域。近红外光谱技术应用于这一领域主要有谷物类组分分析(例如水分、蛋白、淀粉^[13,14])、菜籽油中芥酸含量测定^[15]等。1996 年近红外透射光谱(NIRT)用于测量整粒谷物, 现在 NIRT 已成为测量大豆蛋白质和脂肪含量及小麦中蛋白质含量的美国官方方法^[16]。

在食品分析中, 近红外光谱分析技术已经成功地用于谷物中水分、蛋白质、脂肪、淀粉等化学成分的测定。此外, 在食品工业中的应用还涉及到肉类、奶制品以及各种液体饮料品质指标的测定等, 这些指标包括水分、淀粉、蛋白质、油脂、糖分、纤维以及灰分等的含量测定^[17]。江苏大学陈斌(2001 年)使用近红外光谱分析技术快速检测食醋中总酸、挥发酸、还原糖含量^[18]。

在石油化工工业分析上, 1989 年 Jeffrey J.K. 等建立了使用短波近红外光谱分析技术测定无铅汽油的辛烷值方法^[19], 这项开创性的工作将近红外光谱分析技术应用引入到石化领域; 徐广通、袁洪福等人将 CCD 近红外光谱仪用于济南炼油厂柴油调和控

制分析,发现近红外光谱方法非常适合测定十六烷值^[20];2006年鲍峰伟、刘景艳介绍了近红外光谱分析技术定性、定量方法在石油化工领域的应用^[21]。

随着近红外光谱分析及计算机软件技术的发展,近红外光谱分析技术在制药工业中的应用日趋广泛,显示出了巨大的潜力。包括药物中活性成分的测定、固体药剂的非破坏性表征^[22]、药物生产过程(包括合成、混合、加工、制剂、压片及包装过程)的在线监控及原料或产品的鉴定等。

在纺织工业,李晓薇等用近红外光谱定量分析技术建立了一套混纺毛织品中羊毛的质量分数测定方法^[23],改变了凭感官指标测定需要经验且因人而异、通过理化分析耗时、操作困难的弊端。Wayne 和 Tincher 等人通过 3~4 个波长进行回归准确地测定了聚酯纤维和棉花混合物中两者的比例^[24];江苏大学近红外光谱分析组与中国计量学院合作,也成功的将近红外光谱分析技术应用到了纺织品含棉量的定量分析中^[25]。

1.1.3 近红外光谱分析技术的基本原理

近红外光谱是分子振动光谱的倍频和合频吸收光谱^[26,27],主要是 X-H 键(X 为 C, O, N, S 等)的吸收,不同基团产生的光谱在吸收峰位置和强度上有所不同,根据朗伯-比耳吸收定律,随着样品成分含量的变化,其光谱特征也将发生变化。不同的官能团在近红外区域具有不同的吸收位置,表 1-1 列出了各种含氢官能团在红外区域的谱带归属,各类化合物及官能团的近红外光谱特征见文献^[28]。

表 1-1 各种含氢基团在红外区的倍频吸收谱带归属 (nm)

Tab1-1 Range of different hydrogenous radical's frequency multiplication in NIR

基团	C-H	N-H	O-H
伸缩振动基频	3 300	2 940	2 740
弯曲振动基频	6 900	6 250	7 700
合频	2 300	2 200	2 000
一级倍频	1 745	1 540	1 450
二级倍频	1 210	1 040	960
三级倍频	934	785	730
四级倍频	762		

近红外光谱定量分析的基本公式是朗伯-比尔 (Lambert-Beer) 定律,简称比尔定律,即:

$$A = \lg \frac{I_0}{I} = KCL \quad (1-1)$$

式中： I_0 为入射光强； I 为出射光强； A 是透射率倒数的对数，称为光密度或吸光度，它是物体组成和波长的函数； K 为比例系数，它与物质的吸光系数等因素有关； L 为透过液体的光程长度； C 为液体的浓度。

式(1-1)反映了液体浓度与吸光度之间的数学关系，称为朗伯-比尔定律。对于多组分(混合溶液)而言，分析认为由于分子的红外光谱受周围分子的影响甚小，多组分的光谱是其各组分光谱的简单算术和，就是把多组分看成每种组分有独立的吸光度。因而朗伯-比尔定律亦适用于多组分体系的定量分析^[29]。

1.1.4 近红外光谱分析技术特点

近红外光谱技术之所以成为一种快速、高效、适合过程在线分析的有利工具，是由其技术特点决定的，近红外光谱分析的主要技术特点如下：

(1) 无前期处理，无污染，快速 近红外光在检测样品时，不需要进行任何前处理，也不需要任何化学试剂。和常规分析方法相比，既不会对环境造成污染，又可以节约大量的试剂费用，是一种绿色分析技术。近红外仪器的测定时间短，几分钟甚至几秒钟就可以完成测试，并打印出结果；

(2) 无损检测 由于光谱分析无需对被测样品进行预处理，所以样品不会被破坏，不会对被测样品产生影响；

(3) 多组分同时测定 多组分同时测定是近红外技术得以大力推广的主要原因。通过一次全光谱扫描，即可获得样品中各种化学成分的光谱信息，再由相应的数学模型计算，就可得到样品多种化学成分的含量；

(4) 测定重现性好 光谱测定有很好的重现性，测试结果受人为干扰少，与常规的化学方法相比，近红外光谱分析一般可显示出更好的重现性和精确性；

(5) 在线检测 由于近红外技术能够及时快捷的对样品进行检测，在生产中，可以在生产流水线上配置近红外装置，对原料和成品及半成品进行连续在线检测，有利于及时地发现原料及产品品质的变化，便于及时调控，维持产品质量的稳定。光纤导管和光纤探头的开发应用使远距离检测成为现实。并且远距离检测技术特别适用于污染严重、高压、高温等对人体和仪器有损害的环境应用，为近红外网络技术的发展奠定了基础。

(6) 现代近红外光谱分析也有其固有的弱点

一是测试灵敏度相对较低，这主要是因为近红外光谱作为分子振动的非谐振吸收跃迁几率较低，一般近红外倍频和合频的谱带强度是其基频吸收的 10 到 10 000 分之一，就对组分的分析而言，其含量一般应大于 0.1%；

二是一种间接分析技术，近红外光谱分析方法所依赖的模型必须先利用标准方法或参考方法对一定范围内的样品测定出组成信息或性质的数据。同时，模型的建立需要投入大量的人力、物力和财力，另外分析结果的准确性与模型建立的质量和模型的合理使用有很大的关系；

三是近红外光谱容易受到测量条件^[30]（如：温度、仪器、环境、装样人）以及样品状态等因素的影响，通过仪器所获得的光谱会受到测量条件的影响，从而会引起光谱的不确定性。

1.2 近红外光谱分析中的化学计量学理论

现代近红外光谱分析技术是集计算机技术、分析化学方法、计算数学等多个学科的发展而不断发展壮大的，而化学计量学^[31,32]的发展为现代近红外光谱分析技术的发展奠定了基础。近红外光谱中化学计量学方法的研究主要涉及 3 个方面的内容：一是光谱预处理方法和波长优选方法的研究，目的是针对特定的样品体系，通过对光谱的适当处理，减弱以至于消除各种非目标因素对光谱的影响，净化谱图信息，为校正模型的建立和未知样品组成或性质的预测奠定基础；二是近红外光谱定性和定量校正方法的研究，目的在于建立稳定、可靠的定性或定量分析模型；三是校正模型传递技术的研究，也称近红外光谱仪器的标准化，目的是将在一台仪器上建立的定性或定量校正模型可靠地移植到其他相同或类似的仪器上使用，从而减少建模所需的时间和费用。

(1) 预处理方法和波长优选方法的研究

光谱的预处理^[33]主要包括噪音消除、归一化/正规化处理、光谱范围的优化选择、中心化及标准化处理等。预处理掉的噪音主要来自高频随机噪音、基线漂移、信号本底、样品不均匀、光的散射等。常用的预处理方法有：Savitsky 与 Colay 卷积平滑方法^[34]；峰谷点扯平、偏置扣减、微分处理和基线倾斜等基线校正方法；光谱的标准归一化处理 (SNV)；多元散射校正技术 (MSC) 以及最近发展起来的正交信号处理技术 (OSC)^[35]等。

近红外光谱分析技术中，有效的特征提取能够很大程度提高模型的预测能力和简

化模型,这个也是最近才发展起来的光谱处理技术。目前波长优选方法主要有相关系数法,逐步回归法,无信息变量消除法(Uninformative Variables Elimination, UVE),连续投影法(Successive Projections Algorithm, SPA),遗传算法(Genetic Algorithm, GA)等。

(2) 近红外光谱定性分析方法的研究

样品的近红外光谱包含了组成与结构信息,而样品的品质参数也与其组成、结构相关。因此,在样品的近红外光谱和其品质参数间存在着内在联系。使用数学方法对两者进行关联,可确定这两者间的定性关系。建立定性分析模型常用的判别分析方法主要有主成分分析(Principal Component Analysis, PCA)^[34],马氏距离(Mahalanobis)^[36]、欧式距离(Normalized)^[36]等,模式识别(Pattern Recognition)^[37]和人工神经网络(Artificial Neural Network, ANN)^[38]等也得到越来越多的应用。而 Mahalanobis 方法同定量分析中的 PLS 方法结合的模式识别算法也有较多的应用^[39]。

(3) 近红外光谱定量分析方法的研究

建立光谱定量分析模型的多元校正方法主要包括多元线性回归分析(Multivariate Linear Regression, MLR)、逐步回归分析(Stepwise Regression Analysis, SRA)、主成分回归分析(Principle Component Regression, PCR)和偏最小二乘回归分析(Partial Least Square, PLS)等。

(4) 校正模型传递技术的研究

模型转移方法(MT)包括无标量模型转移方法——有限脉冲响应算法(FIR),有标量模型转移方法——直接校正算法(DS)、分段直接校正算法(PDS)、和专利算法(Shenk's)^[39,40],通过这些算法可以将一台仪器上建立的模型可靠的移植到其他仪器上使用,或者将某一条件下建立的模型适用于同台仪器的其他条件下的光谱数据,同时通过此算法也可以在仪器上采集光谱数据的同时,对光谱数据不断修正,来消除仪器变动、环境等因素引起的仪器光谱数据的漂移,然后再对修正后的光谱数据建模,从而提高模型的预测精度。

1.3 本课题研究的目的是、意义及内容

1.3.1 牛奶近红外光谱分析国内外研究现状

1. 国内研究现状

天津大学徐可欣研究组,主要研究有:李庆波等^[41](2002年)利用他们自行研制的

以声光可调谐滤光器为核心器件的牛奶成分自动检测仪器,测定了牛奶中脂肪、蛋白质、乳糖等成分含量,测量值与参考值具有良好的相关性,测量的重复性变异系数优于 0.02,表明近红外光谱测量方法可以满足牛奶主要成分的实际测量要求;常敏^[42] (2004 年)在牛奶近红外光谱分析研究中采用了散射和透射两种方式,结果表明脂肪在反向接收散射的方式下测量结果较好,而蛋白质的测量用透射更好。王丽杰等^[43] (2004 年)利用近红外漫反射光谱法,在 1 100~1 700 nm 区域采用偏最小二乘法实现了牛奶中蛋白质、脂肪和乳糖的快速测量;王云等^[44] (2006 年)研究了近红外光谱法在牛奶主要成分分析中的应用,对比了不同近红外区域的检测结果。结果表明在长波段 (1 700~2 500nm) 检测牛奶中脂肪及蛋白质含量的准确性最高。

中国农业大学韩东海研究组,主要研究有:通过对液态乳的不同加工工艺、不同成分的近红外光谱检测鉴别的探讨,发现近红外不仅可以对牛奶常规成分进行检测,也可以对加入还原乳的液态乳和不同的加工工艺处理的牛乳进行检测。其中在还原乳的定量鉴别方面 $R=0.971$, $RMSEP=7.76\%$, $RPD=5.13$ 。定性鉴别方面对于还原乳浓度达到 20% 以上时可以达到 100% 鉴别^[45];应用电荷耦合器件 (CCD) 近红外光谱仪 (NIRS) 在短波区域 (700~1 100nm),测定牛奶中蛋白质、脂肪、乳糖等几种常规成分的含量。采用散射校正 (Multiplicative Scatter Correction, MSC) 和偏最小二乘相结合的方法建立了定量模型,并对建立模型的稳定性进行了预测^[46];比较了利用短波近红外积分球漫反射和长波近红外透射定量测量的效果,结果表明两种方法都满足要求^[47]。

燕山大学毕卫红等人^[48] (2005 年)通过对牛奶近红外透射光谱的分析,确立最佳回归波长位置,采用多元线性回归方法建立了数学模型,预测效果比较理想;杨小丽^[49] (2005 年)利用近红外透射光谱 (800~1 700 nm) 快速测量牛奶中脂肪、蛋白质和乳糖的含量,采用偏最小二乘法回归,建立了测量光谱与牛奶主要成分的相关性,并比较了引入非线性建模方法的效果。

2. 国外研究现状

Roumiana T.^[50]等 (2000 年)采用透射方式对 258 个未经均质的原奶样品进行测量,结果表明脂肪和蛋白模型交互验证的相关系数 (R) 分别为 0.999 和 0.857,交互验证的标准偏差 ($SECV$) 分别为 0.115 和 0.086。

Roumiana T.^[51] (2001 年)介绍了几种非均脂牛奶分析和牛奶成分测量的不同方法。他们建立了检测牛奶的体细胞个数 (SCC) 和牛奶尿素氮 (MUN) 的近红外模型。

Pravdova V.^[52]等 (2001 年)应用近红外光谱对非均脂牛奶中体细胞数进行了预测,

对原乳样品在 400~1 100nm 范围进行近红外光谱测量,应用荧光电子法作为参考方法。

Young-Ah Woo.^[53]等(2001年)在近红外短波区域(700~1 100nm)采用透射方式对 112 个未经处理的原奶样品进行测量,结果表明脂肪的预测精 *SEP* 达到 0.06%,蛋白的预测精度 *SEP* 达到 0.10%。

Desmond B.^[54]等(2003年)建立了一个近红外光谱系统用来在线检测牛奶处理过程中的脂肪含量。该系统被用于在线牛奶处理过程中,在 800~1 100nm 波长范围内检测牛奶脂肪含量的变动,系统响应敏感。

Shuso K.^[55]等(2003年)在实验基础上构建了在线近红外光谱检测系统。该系统在 600~1 050nm 波长范围内,在挤奶过程中获得非均脂牛奶的近红外光谱。建立了牛奶的 3 个主要成分(脂肪、蛋白质和乳糖)、体细胞个数和牛奶尿素氮的标定模型。

Romdhane K.^[56]等(2006年)介绍了近红外光谱技术及其他光谱分析技术结合化学计量学方法(CCSWA 等)用于鉴别奶制品(奶酪等)的真假及地区来源一致性,并认为该技术用于在线检测是未来发展的方向。

1.3.2 课题研究的目的是和意义

随着社会的进步,人们生活水平的提高,牛奶已成为人类重要的营养食品之一。牛奶质量直接关系到牛奶产品的质量和销售,进而影响奶牛场、乳品厂的经济效益。目前牛奶成分检测仍以化学分析方法为主导,其检验过程耗时费力、破坏样品、更无法进行实时在线检测。把具有高效、快速、无损、适合在线检测等优点的现代近红外光谱分析技术引入到牛奶成分的在线检测对实现乳业生产过程的自动化及提高乳品质量具有重要意义。

近年来国内外对牛奶的近红外研究已经取得了一定的研究成果,但仍然存在一些不足:在国内,应用近红外光谱技术对牛奶的主要成分进行检测是可行的,但目前还未在工业在线检测方面得到实际应用,利用该技术进行牛奶成分在线检测有待进一步的研究;在国外,近红外在线检测牛奶成分的报道主要集中在短波近红外区域,在长波近红外方面研究较少,而且对于检测过程中的一些重要参数(如温度、流速、光程)的选择、各种干扰因素(如样品温度)的影响和稳健模型的建立等研究还不够完善。

本研究提出把近红外光谱分析技术应用于牛奶成分在线检测,期望能够在该领域得到突破,这对实现乳品工业生产自动化、提高乳品质量以及促进人民健康具有重要

意义。

1.3.3 本文的主要内容

本研究基于本课题组自行开发的近红外液体在线检测试验台，对牛奶的蛋白质、脂肪含量进行在线检测，本论文的主要研究内容有几方面：

(1) 以实现牛奶光谱在线检测的可行性和稳定性为原则，基于本近红外在线检测试验台，对检测系统的各项参数（本底、扫描次数、牛奶温度、流速、光纤探头光程）进行了优化，这些参数的优化为牛奶样品光谱采集提供了参考。

(2) 比较了各种预处理方法的效果，找出蛋白质和脂肪的最佳预处理方法，同时探讨了 MSC 方法在本试验中的可行性并分析了原因；选择了蛋白质和最佳建模波段，然后利用 PLS 分别建立了牛奶样品的蛋白质和脂肪校正模型并对模型的预测性能进行了验证。

(3) 在 26℃~41℃ 范围内，研究了温度对牛奶成分预测的影响，比较了各种预处理方法对不同温度下预测模型的预测效果。

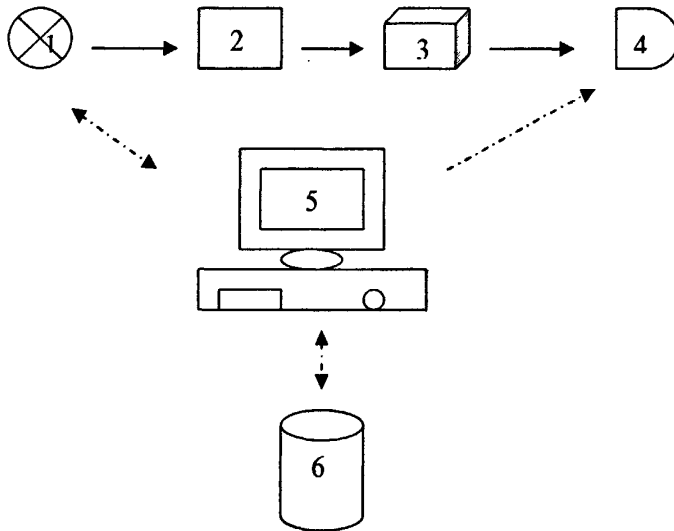
(4) 模型的稳健性是评价在线近红外分析技术的一个重要指标。采集了不同温度下牛奶的光谱组成温度补偿校正集，通过补偿校正方法的引入，建立了温度全局模型，部分消除了温度对模型稳健性的不利影响。

第二章 在线近红外光谱分析牛奶成分的基础

近红外光谱分析是从复杂、重复、变动的背景中提取弱信息，由两个要素组成：一是准确、稳定地测定样品的吸收或漫反射光谱谱图的硬件技术，这一硬件技术的主要要求是必须保持长时间的稳定性；另一个是利用多元校正方法计算测定结果的软件技术。

2.1 在线近红外光谱分析系统

与实验室型的近红外光谱分析仪相同，工业现场所用在线近红外光谱分析系统也由硬件、软件和分析模型 3 部分组成，但对其所包含的实际内容及各项性能指标却有着显著的差异。



1: 光源 2: 分光系统 3: 测样附件 4: 探测器
5: 仪器控制、数据处理系统 6: 数据库存储系统

图 2-1 近红外光谱仪器基本组成

Fig.2-1 the basic composition of NIR instrument

2.1.1 硬件技术

在线近红外光谱分析系统的硬件主要包括光谱仪、自动取样系统、测样装置、样品预处理系统部分^[57,58]。

光谱仪 光谱仪是整个在线分析系统的核心。近红外光谱分析仪器主要由光源、分光系统、测样附件、探测器、仪器控制和数据处理、数据库存储系统等组成^[17,59]，其工作流程如图2-1所示。目前，几乎所有类型的近红外光谱仪器，如固定波长滤光片、

扫描光栅色散、固定光路阵列检测器(CCD, PDA)、傅立叶变换和声光可调滤光器(AOFT)等,都有用于在线分析的报道^[60]。许多生产过程需每天24小时连续运行,在线分析仪也应无间断连续运行。在选择在线光谱仪时,抗环境干扰以保持自身稳定性的性能指标是首要考虑的问题。对于固定光路阵列检测器和AOFT两类仪器,由于内部没有移动光学部件,在长期稳定性方面具有较强的优势,其性能价格比也较为合理。

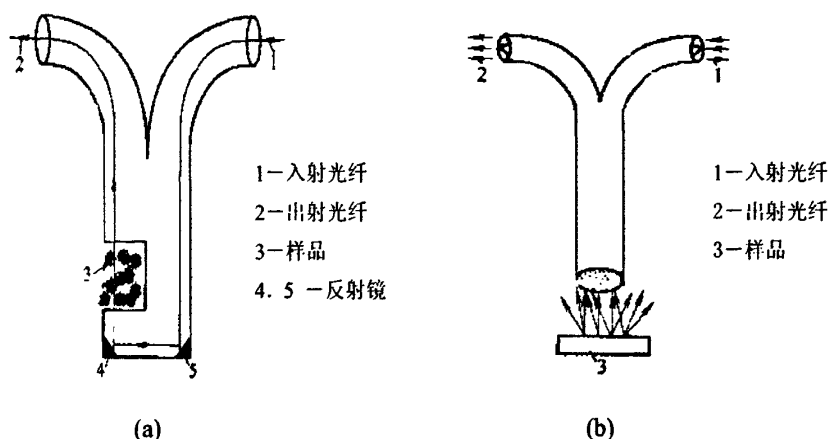


图 2-2 光纤探头附件示意图

(a) 透射光纤探头(b) 漫反射式光纤探头

Fig2-2. Plot of optical fibre probe

近红外区域的光可以用光纤进行有效传输,大多数在线NIR分析仪都采用光纤方式远距离传输光。对液体而言,在线近红外光谱分析中实现多路物料检测的最常用方式是采用光纤多路转换器。光纤探头主要有透射式和漫反射式两种(如图2-2所示)。

光纤具有以下优点:

(1) 直径细,易弯曲,可直接插入生产装置中,进行原位分析和实时跟踪检测,还可用于遥测分析;

(2) 光纤的化学和热稳定性、对电磁干扰不敏感等特性,使其可在困难条件或危险环境,以及复杂的工业生产现场中工作;

(3) 光纤具有价廉、轻巧、使用寿命长、安装和维护方便。更重要的是,采用光纤和光纤多路转换器,在线近红外光谱分析仪很易实现一台光谱仪检测多路物料,提高仪器的利用效率。

取样系统 取样系统设计是否合理将直接影响分析结果的可靠性,有时甚至决定着整个分析系统的成败^[61]。对于液体分析体系,取样系统主要有3种方式,即:泵抽采样、压差引样和定位实施测量。泵抽采样是通过在旁路上附加泵来实现的,多用于

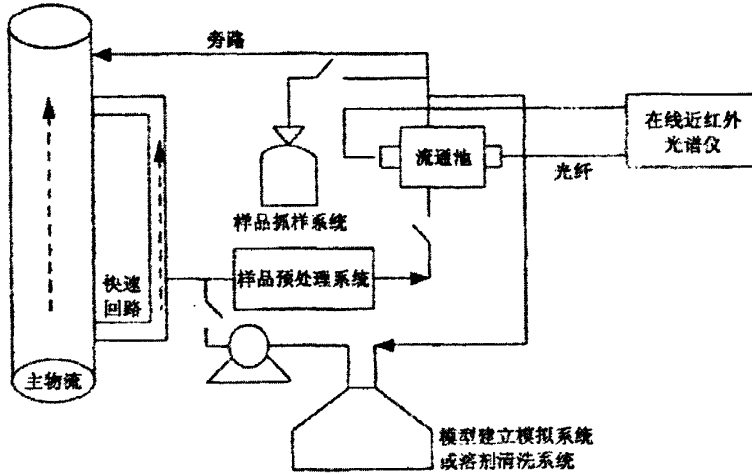


图 2-3 液体样品旁路分流取样系统

Fig2-3. sampling system for liquid sample by bypass distributary

取样点与测样装置(如流通池)之间无压力差的过程;压差引样则要求取样点与测样装置之间存在压力差,靠该压力差将主管线或装置中的样品经旁路引入测样装置。在旁路取样系统中(如图2-3),较长的取样管线以及较慢的输送,在以上两类取样系统与测样装置之间通常加有一个样品预处理系统,对样品的流速、压力进行控制,使其保持稳定,这是旁路取样方式的一个突出优点。而定位实施测量是将测样装置直接安装到装置流程或主管线的一种取样方式,如将光纤探头直接插入反应装置中或将流通池直接安装到装置管线中。与旁路取样方式相比,这种方式反馈速度快,实现了真正意义上的实时分析,但取样点处的压力、温度及流速的波动会对分析结果产生影响,苛刻的取样条件如高温、高压、易污染光学窗片等,也对光纤探头或流通池的设计和加工带来困难。

测样装置 近红外光谱在过程分析中所测量的样品较复杂,样品状态(固体粉末、颗粒、液体、浆状物等)、样品性质(样品形状、流动性、透光性等)以及温度等外界因素都有较大差别。因此,在进行光谱采集时,要根据所测样品的实际情况选择合适的测样方式和测样装置。对于透明和半透明液体,一般主要采取透射或透反射方式采集光谱,而对不透明液体一般则采用漫反射方式。用于液体测量的测样装置主要有流通池和插入式光纤探头两种形式。液体流通池主要用于流动性好以透射方式测量的样品,

可以通过光纤将流通池安装到采样点附近,也可通过管线将物料引入到置于光谱仪内部的流通池中。光纤探头可分为透(反)射式和漫反射式探头两类。将光纤探头直接插入主流路中可用来实时测量液体和固体样品,但这种方式会受到较多因素如压力、温度、气泡等对光谱测量的影响,探头还容易被污染,需定期清洗。

样品预处理系统 样品预处理系统的主要功能是控制样品的温度、压力和流速,以及脱除样品中的气泡、水分和机械杂质等影响因素,确保分析结果有效准确^[62,63]。对不同的测量体系,预处理系统的组成也不尽相同,一般由快速回路(减小滞后时间)、过滤(除尘、除机械杂质)、压力调节(减压或抽引)、温度调节(降温或加温)、有害或干扰成分处理(除水分等)、流量调节等组成。

样品温度变化将导致化学键振动能量和能级跃迁几率的变化,对光谱吸光度和吸收峰位置都有一定的影响,为获得准确的分析结果,保证样品测试时温度的恒定是非常重要的。因此,在流通池周围加有恒温装置。此外,对于流动性较差的液体样品(如高分子材料、植物黄油、润滑油等),在过程分析时必须具有一定的温度,保持其良好的流动性。对于无法进行恒温的情况,如将探头直接插入主管线,则需要建立温度校正曲线,修正分析结果。也可以将温度作为一个变量,与光谱数据一起参与建立模型。此外,将不同样品温度下采集的光谱组成混合样品集,建立温度全局稳健模型(Global Calibration),也是一种常用的方法。

样品的流动状况与流速变化等因素将影响光的散射,从而使光谱数据发生变化,成为影响最终分析结果的一个因素。同样,样品的压力变化也会对结果产生影响。因此,非常有必要对样品进行恒速和恒压控制,以获取准确可靠的分析结果。

2.1.2 软件技术

如果说光谱仪是整个在线分析系统的核心的话,那么NIR分析软件和模型可以说是整个系统的灵魂。在线近红外光谱分析系统的软件除具备必需的光谱实时采集和化学计量学光谱分析功能外,还应包括其他一些功能:如数据与信息显示功能;数据管理功能;通讯功能;故障诊断与安全功能;监控功能等。

分析模型在近红外光谱分析中处于核心地位。与实验室相比,建立一个适用范围广、稳健性好的在线近红外分析模型将更为复杂。在建立模型时,应注意模型预测精度与模型稳健性之间的相互影响问题。一般来讲,若校正集中光谱采集条件完全相同(如光谱仪的环境温度、样品温度、压力和流速),则所建模型对相同条件下采集光谱

数据的预测准确性较高,但若采集条件发生了波动,其预测结果将会产生较大的偏差。因此,在建立模型时,往往人为地在一定范围内变动某些测试条件如样品的温度或流速,以提高模型的稳健性和预测能力。但这样做在一定程度上会降低模型的预测精度,在具体的实施过程中,应对其进行取折处理。也可将实验室建立的分析模型通过模型传递方法转换后用于在线分析。模型传递的方式有多种,可以将建立好的分析模型直接传递;可以将光谱在不同仪器间传递,重新建立模型;也可以将分析结果进行校正。

2.2 牛奶成分近红外在线检测的可行性

2.2.1 牛奶体系的构成及光学特性

牛奶是多种物质组成的混合物,各种物质相互组成复杂的分散体系^[64],其中包含真溶液、高分子溶液、胶体悬浮液、乳浊液及其过渡状态的分散体系(见表 2-1)。

表 2-1 牛奶体系的构成及特性
Table 2-2 the structure and characteristic of milk

构成	物质状态	组成成分	微粒直径 (nm)	散射类型
真溶液	分子或离子态	乳糖等	1	
高分子溶液	大分子态	乳蛋白	15-50	
胶体悬浮液	复合体胶粒	酪蛋白	30-800	瑞利散射
乳浊液	脂肪球	脂肪	100-10 000	米氏散射

光在牛奶中的传播特性非常复杂,其中吸收和散射作用是牛奶测量中需要考虑的重要因素。光通过媒质时,使透射光强减弱的因素有两个,一个是光的吸收,二是光的散射。

一般按散射粒子的大小把散射分成两类:一类是散射粒子的线度在 $1/5 \sim 1/10$ 波长以下的散射,称为分子散射或瑞利散射;另一类是散射粒子的线度与光波波长同数量级的散射,称为大粒子散射或米氏散射。需要指出的是,脂肪球经过均质后,其微粒直径下降到 $1\ 000\text{nm}$ 以下,一定程度上减少了颗粒的散射效应。

2.2.2 近红外光谱分析技术在线检测牛奶成分的可行性

近红外光谱是分子振动光谱的倍频和合频吸收光谱,主要是 X-H 键(X 为 C,O,N,S

等)的吸收,不同基团产生的光谱在吸收峰位置和强度上有所不同,根据朗伯-比尔吸收定律,随着样品成分含量的变化,其光谱特征也将发生变化。这是近红外光谱分析方法的理论基础。牛奶成分中脂肪、蛋白质分子等成分对近红外光具有吸收作用,因此从理论上讲,近红外光谱技术可应用于牛奶成分测定。近红外光谱技术同光纤技术相结合,使得近红外光谱分析技术应用与牛奶在线检测中成为可能。

2.3 本章小结

本章从近红外分析技术的基础入手,介绍了在线近红外分析系统的组成:硬件技术和软件技术。其中硬件技术包括光谱仪、自动取样系统、测样装置、样品预处理系统等部分;软件技术除具备必需的光谱实时采集和化学计量学光谱分析功能外,还应包括其他一些功能:如数据与信息显示功能;数据管理功能;通讯功能;故障诊断与安全功能;监控功能。最后根据牛奶的体系构成及其光学特性分析了近红外光谱分析技术在线检测牛奶的可行性。

第三章 试验过程和方法

3.1 试验材料

3.1.1 试验设备与仪器

本试验采用的设备是课题组自行研究开发的近红外在线检测试验台（如图3-1所示），它主要由包括：蠕动泵、流量计、水浴加热装置、传感器、储料罐、实验平台等。实物拍摄图见图3-2。

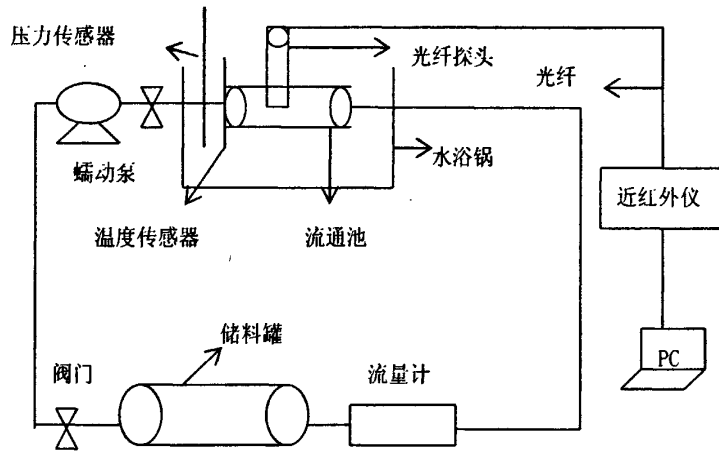


图 3-1 近红外在线检测试验台示意图

Fig.3-1 plot of NIR on-line detecting equipment

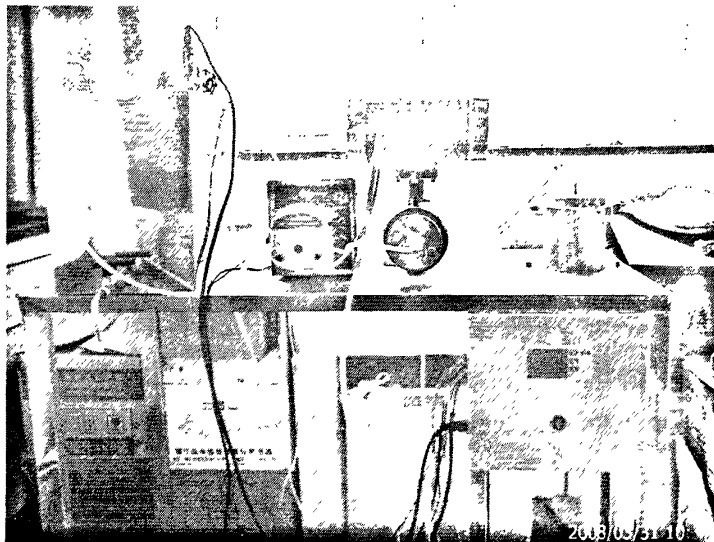


图 3-2 近红外在线检测试验台

Fig.3-2 NIR on-line detecting equipment

试验所用光谱仪器是美国 CONTROL DEVELOPMENT 公司生产的 NIR256-2.2T2

近红外光谱仪器（如图 3-3 所示）。

其技术指标如下：

波长线性范围：1 100~2 200 nm；

探测器：铟镓砷（InGaAs）；

波长准确性：优于 0.5nm；

光度准确性：小于 0.000 05AU；

电源：6V；

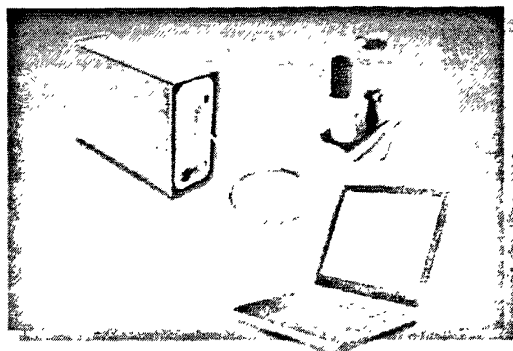


图 3-3 试验中使用的近红外光谱仪器

Fig3-3 NIR256-2.2T2 NIRS
instrument used in experiment

3.1.2 牛奶样品的收集

本试验收集了包括伊利、雀巢、完达山、南山等品牌的奶粉，牛奶样品都是由经过超高温瞬时加热和均质处理的奶粉调

制而成。利用 50 个奶粉样品加入一定容积的蒸馏水（30℃，密度 0.995 7g/ml）配置成 50 个牛奶基准样品约 350ml。由于试验中需要一定浓度范围的样品，其余 10 个样品由现有的基准样品通过混合配置得到，共计样品 60 个。

奶粉的化学值测量由课题组严格按照国标测定，蛋白质的测定采用凯氏定氮法（GB/T 5009.5-2003）^[65]，脂肪的测定采用哥特里—罗兹法（GB/T5413.3-1997）^[65]。牛奶样品的化学值由调制奶粉的质量和加水的质量计算得到，蛋白质和脂肪的化学值（质量百分比浓度）见表 3-1。

表 3-1 牛奶样品主要质量指标的化学值测定结果

Table 3-1 Summary of variation ranges, means, standard deviations of protein, fat of milk

检测成分	最大值	最小值	平均值	标准差
蛋白质 (%)	5.99	1.57	3.33	0.96
脂肪 (%)	6.75	1.59	3.96	1.39

3.2 检测系统参数的选择

3.2.1 本底的选择

在测定溶液的近红外光谱时，要求尽量减少来自溶剂的吸收。牛奶是以水作为溶剂的一种混合液体，含有大约 87%的水，水分子具有很强的极性，在近红外谱区有一些很强的吸收谱带。而蛋白质和脂肪的光谱吸收与水相比非常弱，所以在检测中要尽可能的消除水的干扰。本试验采用蒸馏水作本底，以减少牛奶中水溶剂对待测成分吸收光谱的影响。

3.2.2 扫描次数的选取

平均技术是提高光谱信噪比的有效方法, 假定测量过程中近红外信号与背景噪声是相互独立的, 且是加性关系, 即 $x_i(k)=s(k)+n_i(k)$ $i=1, 2, \dots, N$, 其中 $s(k)$ 是样品在近红外区的真正信息, $n_i(k)$ 表示第 i 次扫描时的背景噪声, $x_i(k)$ 表示第 i 次测量的表观数据, 公式如下所示:

$$\bar{x}(k) = s(k) + \frac{1}{N} \cdot \sum_{i=1}^N n_i(k) \quad (3.1)$$

则经过 N 次叠加以后, 与信号无关的噪声会大量的减少, 而信号以代数和的方式逐渐增大, 除以叠加次数 N 以后, 信号则保持不变, 所以, 通过增加扫描次数可以提高信噪比, 从而改善近红外的分析结果。

用某一牛奶样品在扫描次数分别为 8、16、32、48、64 次的情况下各测量 7 次光谱, 然后对光谱经过一阶微分和平滑处理后, 计算其标准差(如图 3-4),

$$\sigma_i = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \quad (3.2)$$

式中: x —样本平均值;

n —样本大小。

由于近红外光谱仪器受到探测器检测范围的限制, 在 1890nm 以上有很大的仪器噪声, 所以在此只考虑 1100~1889nm 范围内的标准差。将此范围内的每个扫描次数

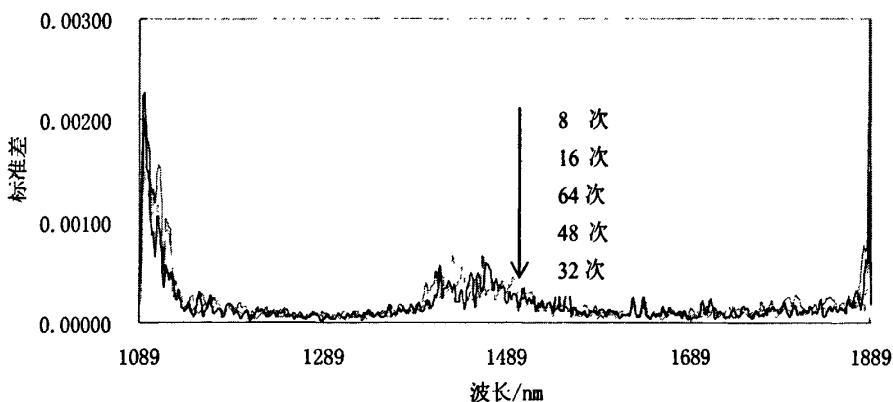


图 3-4 不同扫描次数的标准差示意图

Fig.3-4 The sketch map of standard deviation on different scans

下的标准差分别取平均值(如图 3-5):

$$\bar{\sigma} = \sum_{i=1}^m \frac{\sigma_i}{m} \quad (3.3)$$

式中： m —每条光谱的数据点数。

在图 3-4 中，当扫描次数为 8 时，对应的标准差曲线在最上层，随着扫描次数的逐渐增加，其标准差值就相应的减少，当扫描次数是 64 时，其对应的标准差曲线在最下层，扫描次数为 32 次的标准差与 48 和 64 次的几乎十分接近。

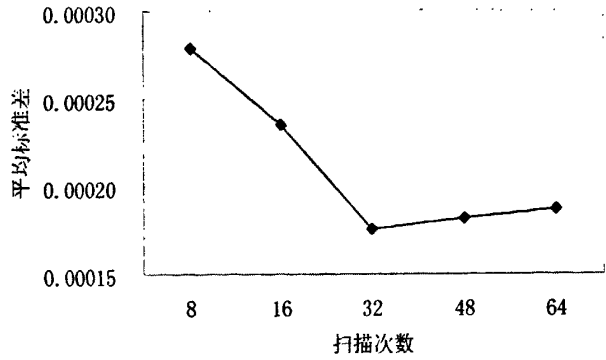


图 3-5 不同扫描次数下的平均标准差比较
Fig.3-5 The comparison of mean standard deviation on different scans

从图 3-5 中可以看出，扫描次数从 8 次增加到 32 次的过程中，标准差有明显减少，32 次达到最小，到了 48 和 64 次的时候，标准差略有上升；同时为了节省样品采集时间，因此选择 32 次作为试验的扫描次数。

3.2.3 样品温度的选择

在影响 NIR 分析准确性的诸多因素中，样品温度影响较严重。温度不仅影响样品的吸光度，而且也影响波长位移，导致谱带变形^[66]。而一定的温度不仅可以增强牛奶的流动性，还可以使游离的脂肪球得以溶解，减少散射的影响。

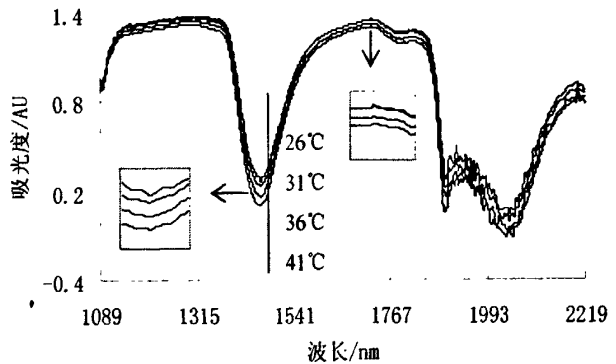


图 3-6 温度对牛奶光谱的影响
Fig3-6 Influence of temperature on NIRS

为了选择最佳的样品温度，选择温度为 26℃、31℃、36℃、41℃，待温度恒定后，在各个温度下连续采集某一牛奶样品 7 次，所有光谱数据经一阶求导和 11 点移动平均滤波（Moving Average Filter, MAF）平滑处理

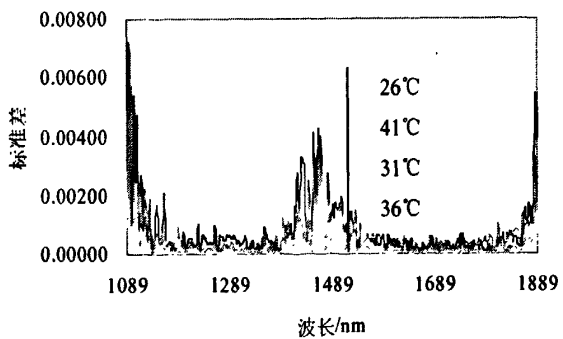


图3-7 各温度下光谱标准差比较
Fig.3-7 The standard error of the spectrums in different temperature.

后，对每个温度下的 7 条光谱求标准差。

表 3-2 不同温度下所得光谱的平均标准差

Table 3-2 The average standard error of the spectrums in different temperature.

样品温度/°C	26	31	36	41
平均标准差	0.001 562	0.001 204	0.000 945	0.001 484

观察各个温度下采集的光谱（见图 3-6）及比较各温度下光谱标准差及平均标准差（见图 3-7 和表 3-2）。从图 3-6 中可以看出，各个温度下的光谱谱型基本一致，但温度的变化导致了吸光度的变化，变化规律比较复杂。从表 3-2 中可以看出，各个温度下光谱的平均标准差都在 0.2% 以下，光谱稳定性都较好。相比较而言，36°C 平均标准偏差最低。这是由于乳脂肪球低于其熔点（约 35°C）的温度下呈乳糜化状态，即以脂肪以相互凝结的极小脂肪球形式存在^[67]，这不但会增加光谱的散射，而且降低了牛奶的流动性，增加了光谱的不稳定性。而温度过高也会导致牛奶化学成分的变化，从而改变光谱的吸光度。综合考虑选择样品温度为 36°C。

同时为了研究温度对牛奶成分预测模型的影响，也采集了一部分样品在各个温度下的光谱，详见第五章。

3.2.4 最佳光程和流速的选择

在牛奶成分在线透射测量中，一个很重要的问题是对光纤探头光程进行优化，使光探测器检测的信号中尽可能多的包含被测量成分的信息并减少其他成分信息的影

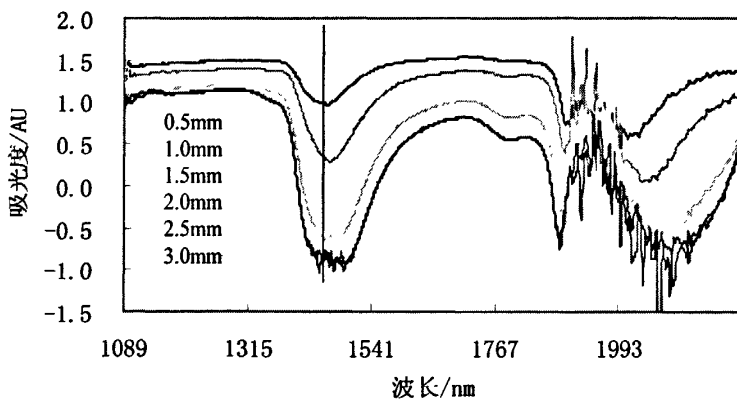


图3-8 某牛奶不同光程下光谱图的比较

Fig.3-8 The spectrums in different optical length

响。光程的选择与光在牛奶中的穿透深度有关，由吸收和散射共同决定的，吸收和散射越强，穿透深度就越小，反之越大。当在穿透深度的范围内接收光电能量时，能保

证较高的信噪比^[68]。

图 3-8 是某一样品在不同光程下测定的光谱,通过比较发现,当光程达到 3mm 时,牛奶光谱出现明显噪音,特别是在水的特征峰 1450nm 和 1900nm 附近噪音更明显。这是由于受到牛奶中大分子蛋白质的吸收和脂肪的散射影响,当光程达到 3mm 时,已超过光在牛奶中的有效穿透深度,导致检测器接收到的光能量减弱,信噪比较差。

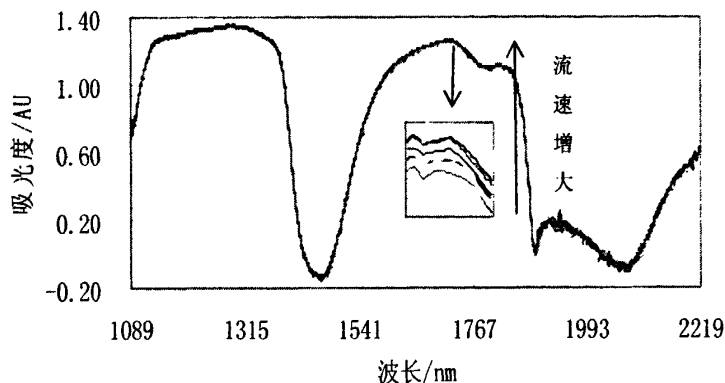


图 3-9 不同流速下某样品光谱的变化

Fig.3-9 comparison of milk spectrums at different velocity of flow

为了找出流速的改变对牛奶光谱的影响,选择光程为 2mm,通过蠕动泵(从 50r/min 到 300r/min, 间隔 50r/min) 改变某一牛奶样品流速从 0.115cm/s 到 0.287cm/s (间隔 0.034 cm/s), 比较不同流速下测定的光谱如图 3-9 所示。图中放大部分(1730nm 附近)为牛奶脂肪 C-H 键基频振动的一级倍频吸收,从图中发现,随着流速的增大牛奶光谱吸光度略有增大,流速从 0.115cm/s 增大到 0.287cm/s, 1730nm 处吸光度增大了约 0.20AU,说明了流速的变化对牛奶光谱产生了一定的影响。

以上结果说明,光纤探头光程长和样品流速的合理选择对本试验牛奶光谱的在线测定至关重要。为了确保近红外仪能稳定地获取牛奶样品的光谱而使牛奶平稳地透过光纤探头,则需要一定的光程长和一定的流速以避免光纤探头出现死容积^[69]。因此,为了

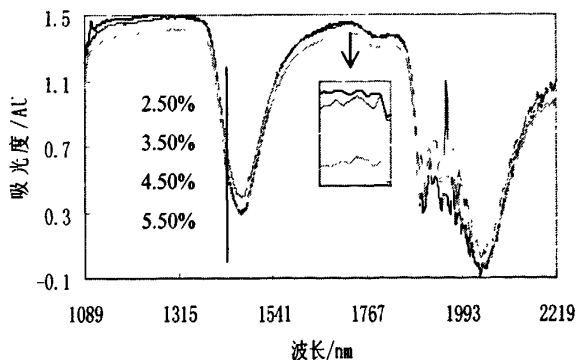


图3-10 改变牛奶浓度的光谱变化图

Fig.3-10 The spectrums in different content

找出可行的试验方案,选择光程范围为 0.5~2.5mm(间隔 0.5mm),流速范围为 0.115~0.287cm/s(间隔 0.034 cm/s)。在不同的光程和流速的水平下,保持光纤探头不动,通过连续添加奶粉依次改变样品蛋白质浓度为 2.50%、3.50%、4.50%、5.50%,观察采集的光谱是否随样品浓度的变化而变化。图 3-10 是在某一光程某一流速下改变样品浓度引起的光谱变化图,从图中可见,吸光度随样品浓度出现有规律的变化即浓度增大吸光度减少,这表明在该光程和流速组合下牛奶样品可以平稳透过光纤探头,透过情况结果见表 3-3。

从表 3-3 中可见,光程在 0.5mm 时牛奶都不能透过;光程为 1mm 时,只有当流速到了 0.287cm/s 才能透过;光程为 1.5mm 时,流速大于 0.252cm/s 可以透过;而光程为 2mm 和 2.5mm 时,流速大于 0.184cm/s 都可以透过。

表 3-3 不同光程、流速下的样品透过情况

Table.3-3 penetrating results of milk in different optical length & velocity of flow

流速 cm/s \ 光程 mm	0.115	0.149	0.184	0.218	0.252	0.287
0.5	×	×	×	×	×	×
1.0	×	×	×	×	×	√-1
1.5	×	×	×	×	√-2	√-3
2.0	×	×	√-4	√-5	√-6	√-7
2.5	×	×	√-8	√-9	√-10	√-11

注: √表示透过; ×表示未透过; √-1 表示方案 1

流速的变化会引起牛奶光谱吸光度的变化,流速过大也可能会由于形成层流等不利因素造成光谱不稳定,因此有必要选择一个合适的流速并加以恒定。为此,在可以透过探头的可行光程和流速组合下连续测定同一样品 7 次光谱,选择 1100~1889nm 光谱区间并求其平均标准差(见图 3-11)。从图中可见,采取方案 5 时平均标准偏差最小,即采用光程为 2mm、流速为 0.218cm/s 时,光谱稳定性最佳。

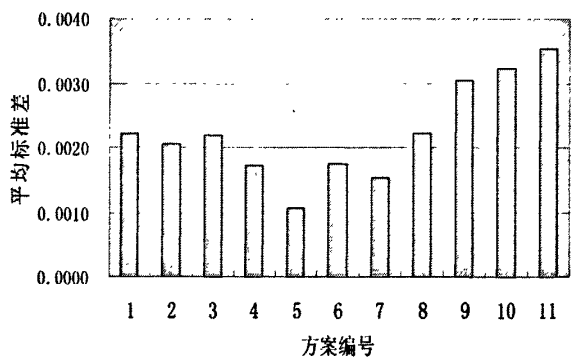


图3-11 不同透过方案的光谱平均标准差
Fig.3-11 The average standard error of the spectrums in different condition.

3.3 样品光谱的采集与数据处理

3.3.1 样品光谱的采集

在样品光谱采集前,将仪器预热 1h 以上,以保证仪器能稳定工作;同时确保在线检测设备能正常工作(如管道是否干净、通畅;流通池是否密封;温控设备是否正常)。打开温控设备,将水浴锅加热至预定温度后,将牛奶样品 350ml 放入储料罐,同时打开蠕动泵设定流速为 0.218cm/s,使样品循环流动以实现恒温、

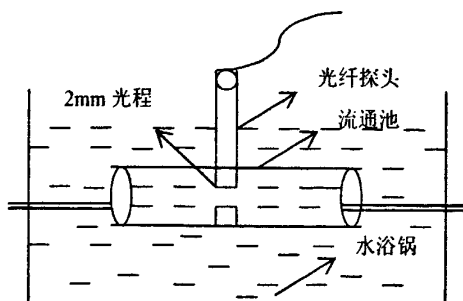


图 3-12 牛奶近红外光谱在线检测示意图

Fig.3-12 Plot of NIR on-line detection

恒速测量。待样品温度恒定后将光程为 2mm 的光纤探头放入流通池中,使探头与流通池底部垂直,并使探头光程间隙开口方向与牛奶流动方向一致,以确保牛奶能平稳地流过探头而不受气泡等因素的影响。此时开始采集牛奶样品的光谱,以蒸馏水为本底,仪器采集次数选择 32 次,积分时间为 0.022 700s,每个样品采集 3 次光谱取平均值。

3.3.2 数据处理与模型评价指标

对原始光谱采取移动窗口滤波平滑、一阶导数、二阶导数、归一化和多元散射校正等方法进行预处理,找出最佳预处理方法;利用相关系数法优选最佳建模波段;最后采用 PLS 建立牛奶主要成分蛋白质和脂肪的数学模型。文中涉及到的所有预处理方法、波长优选及建模算法均用 MATLAB7.0 编写完成。

模型评价指标包括校正集均方差(RMSEC)、相关系数(R)、预测集预测标准差(RMSEP)、残差(e)^[70]。

(1)校正集均方差(RMSEC) 校正集是指校正集样品的近红外预测值与参考值的标准偏差。公式表示为:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-1}} \quad (3.3)$$

式中, y_i 为校正集中第 i 个样本的实测值; \hat{y}_i 为该模型对第 i 个样本预测值, n 为校正集

样本数。

(2) 相关系数 (R) 相关系数表征的是预测值和参考值之间的相关度, 公式表示为:

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.4)$$

式中, y_i 和 \hat{y}_i 分别为样本集 (包括训练集和预测集) 中第 i 个样本的实测值和预测值, \bar{y} 为样本集中所有样本的实测值的平均值。

(3) 预测标准差 ($RMSEP$): 主要用于评价所建模型对外部样本的预测能力, $RMSEP$ 越小表明模型效果越好。公式表示为:

$$RMSEP = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.5)$$

式中: n 为预测集样品数, \hat{y}_i 为预测集预测过程中第 i 样品的预测值, y_i 为第 i 样品参考方法的测定值;

(4) 平均偏差 ($Bias$): 平均偏差物理意义明确, 简单易懂, 是评价模型的重要指标。公式表示为:

$$Bias = \frac{|y_i - \hat{y}_i|}{n} \quad (3.6)$$

式中: n 为预测集样品数, y_i 为第 i 样品参考方法的测定值, \hat{y}_i 为预测集预测过程中第 i 样品的预测值;

3.4 本章小结

本章从试验方法入手, 以实现牛奶光谱在线检测的可行性和稳定性为原则, 基于本近红外在线检测试验台, 对以下检测系统参数进行了优化:

由于试验中需要一定浓度范围的样品, 利用收集的伊利、雀巢、完达山、南山等品牌的 50 个奶粉样品配置了 50 个牛奶基准样品, 其余 10 个样品由基准样品通过混合配置的方式获得, 共 60 个样品; 为了减少水对牛奶吸收光谱的影响, 选择蒸馏水作为本底; 为了提高光谱信噪比并节约扫描时间, 选择扫描次数为 32; 为了增强牛奶的流

动性及减少牛奶脂肪球散射影响，选择最佳建模温度为 36℃；为了确保近红外仪能稳定地获取牛奶样品的光谱而使牛奶样品能够平稳的流过光纤探头以避免出现死容积，选择了最佳光程和流速分别为 2mm 和 0.218cm/s。这些参数的优化为牛奶样品光谱采集提供了参考。

第四章 牛奶成分校正模型的建立及验证

4.1 试验方案

在 36℃ 温度下, 以蒸馏水为本底, 选择光程为 2mm, 流速为 0.218cm/s, 通过近红外在线检测试验台测定 60 个牛奶样品的光谱。分别对原始光谱进行 MAF 平滑、基线校正、归一化、正规化等预处理, 找出最佳预处理方法, 同时探讨多元散射校正 (MSC) 预处理方法在本试验中的可行性; 最后选择最佳建模波段, 利用偏最小二乘法建立牛奶样品的蛋白质和脂肪校正模型并对模型的预测性能进行验证。

4.2 最佳数据预处理方法的选择

光谱实际测量过程中不仅包括了牛奶组分浓度信息, 还包括仪器噪声, 外界干扰等, 需要对原始光谱进行适当的预处理, 最大程度的提取牛奶的有效信息。本文中应用的预处理方法有: 平滑、微分处理、归一化、多元散射校正等^[7]。

4.2.1 几种预处理方法的原理

平滑处理是一种较为常用的预处理方法。由于受各种因素的影响, 近红外光谱仪

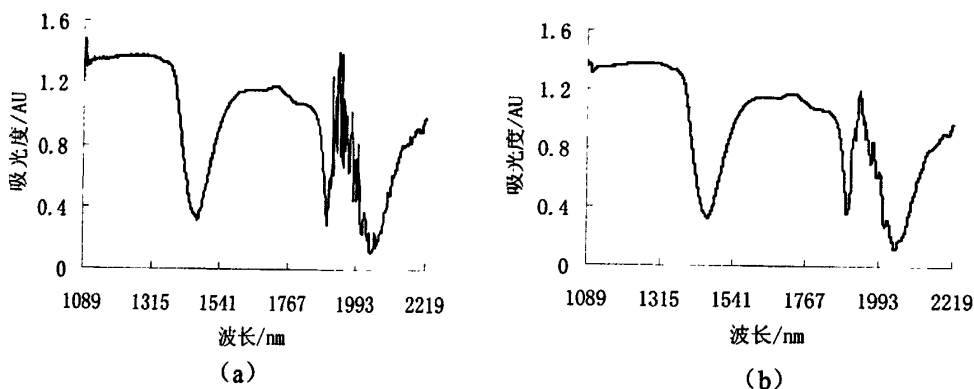


图 4-1 某一样品光谱平滑前后对比

(a) 原始光谱 (b) MAF 平滑光谱

Fig.4-1 comparison of origin and smooth spectra

所测得的光谱数据不可避免地伴随着随机误差和噪声, 得到的往往是一条不光滑的曲线, 平滑可以降低随机误差和噪声。图 4-1 是某一牛奶光谱经过 11 点 MAF 平滑处理前后的效果对比。

归一化处理 (normalize) 是一种较为简单和常用的预处理方法, 其主要目的是消除光程的变化或样品稀释浓度等变化对光谱产生的影响。本文采用平均值归一化即每个数据除以光谱的平均值后得到的光谱数据。图 4-2 是某一牛奶光谱经过平滑后进行归一化处理后的光谱, 处理后的光谱可以一定程度上消除牛奶样品稀释浓度的变化对光谱产生的影响。

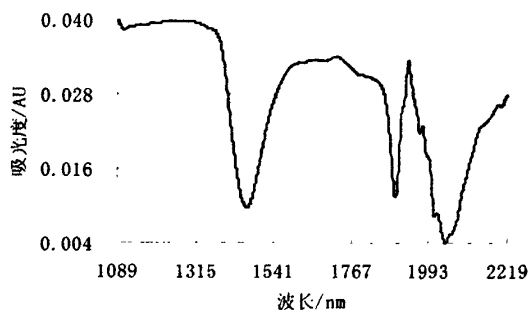


图 4-2 某一样品归一化处理后的光谱

Fig4.3 The spectra after normalize

微分处理 (derivative) 包括一阶导数(1st Der)、二阶导数(2nd Der)预处理, 微分处理方法是 NIR 光谱分析中常用的基线校正预处理方法。在对样品采集光谱时, 由于制样条件和仪器参数等对光谱的影响, 谱图可能发生平移、旋转以及图谱的其它变形。图 4-3 和图 4-4 分别是某一牛奶光谱经过平滑处理后进行一阶求导 (结果放大 1 000 倍) 和二阶求导 (结果放大 100 倍) 后的光谱。对光谱进行一阶微分处理能够明显消除光谱的基线漂移, 消除图谱平移误差, 克服谱带重叠, 强化谱带特征; 而对光谱进行二阶微分处理则可以消除光谱的旋转误差, 但须注意它同时会增加随机噪声, 降低信噪比。

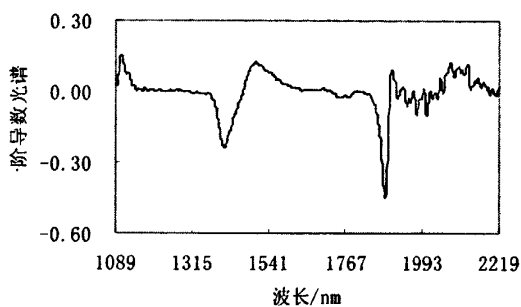


图 4-3 某一牛奶一阶导数光谱

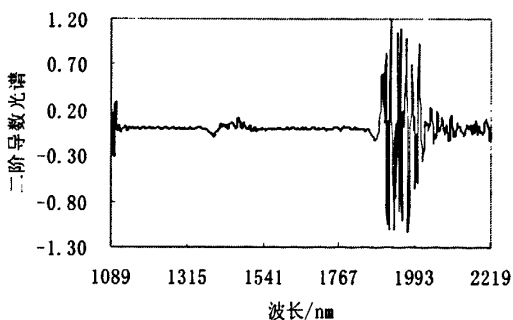
Fig4.3 The spectra after 1st derivation

图 4-4 某一牛奶二阶导数光谱

Fig4.3 The spectra after 2st derivation

多元散射校正的目的是校正每个光谱的散射并获得较“理想”的光谱。多元散射校正法假定与波长有关的散射对光谱的贡献和成分的贡献是不同的, 理论上说, 通过光谱上许多点的数据分析, 可以把这两部分分开。MSC 方法认为每一条光谱都应该与“理想”光谱成线性关系, 而真正的“理想”光谱无法得到, 可以用校正集的平均光

谱来近似。因此每个样品的任意波长点下反射吸光度值与其平均光谱的相应吸光度的光谱是近似关系,直线的截距和斜率可由光谱集线性回归获得,并用于校正每条光谱,截距大小反映样品独特反射作用,而斜率大小则反映样品的均匀度。具体公式如下所示:

$$\text{平均光谱: } \bar{A}_i = \sum_{i=1}^n A_i / n \quad (4-1)$$

$$\text{线性回归: } A_i = m_i \cdot \bar{A}_i + b_i \quad (4-2)$$

$$\text{MSC 校正: } A_{i(MSC)} = (A_i - b_i) / m_i \quad (4-3)$$

上述公式中, A 是校正集的光谱矩阵, A_i 为第 i 个样品的光谱, m_i 和 b_i 是第 i 个光谱 A_i 与平均光谱 \bar{A} 的线性回归的斜率与截距,都是列向量。通过调整 m_i 和 b_i 的使得在减少光谱差异的同时尽量保留原有的与化学成分相关的信息。图 4-5 是某一牛奶光谱经过平滑后进行 MSC 处理后的效果图。

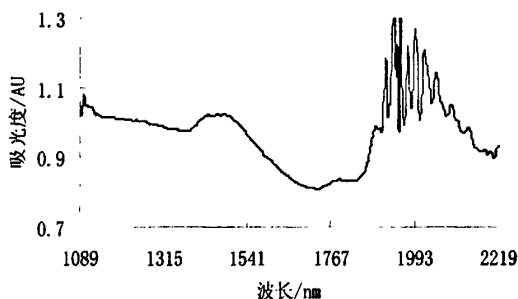


图 4-5 某一牛奶 MSC 光谱
Fig4.5 The spectra after MSC

4.2.2 预处理方法结果与分析

把 60 个牛奶样品按照浓度均匀的原则分成两部分,其中 45 个样品组成校正集用于校正模型的建立,剩下 15 个样品组成验证集用于验证校正模型的预测性能,校正集和验证集化学组成见表 4-1。

表 4-1 牛奶样品校正集和预测集的化学组成

Table.4-1 the chemical components of the calibration and valibration sets of milk

成分	校正集			验证集		
	浓度范围	平均值	标准差	浓度范围	平均值	标准差
蛋白质 (%)	1.57~5.99	3.37	1.03	2.05~4.66	3.33	0.82
脂肪 (%)	1.59~6.75	3.99	1.44	1.88~6.25	4.06	1.10

为了找出最佳预处理方法,对原始光谱分别使用不同的预处理方法或预处理方法的组合进行处理,然后利用 PLS 建立牛奶蛋白质和脂肪的预测模型,比较预测模型的评价指标(见表 4-2)。在 PLS 建模中,校正模型的主成分数由交互验证方法得到的预测残差平方和(PRESS)^[17]来决定, PRESS 值越小表明模型的预测性能越好,预测集相关系数(R)就越高,而预测集标准差(RMSEP)就越小。

从下表中可以看出：在蛋白质预测模型中，原始光谱经过 MAF 平滑、一阶导数处理和归一化处理后， R 和 $RMSEP$ 分别是 0.895 4 和 0.312 5，预测效果最佳；在脂肪预测模型中，原始光谱经过 MAF 平滑、一阶导数处理和归一化处理后， R 和 $RMSEP$ 分别是 0.918 0 和 0.408 5，预测效果最佳。

因此，对蛋白质和脂肪模型，选择的最佳预处理方法都为：MAF 平滑、一阶导数处理和归一化处理。

表 4-2 各种预处理方法 PLS 模型预测结果

Table 4-2 PLS model predicted results with combination pretreatment

牛奶成分	光谱预处理方法	主成分数	$RMSEC$	R	$RMSEP$
蛋 白 质	None	7	0.341 9	0.803 8	0.429 1
	MAF+D1	7	0.295 1	0.886 9	0.359 2
	MAF+D1+NOR	6	0.265 7	0.895 4	0.312 5
	MAF+D2	5	0.302 6	0.876 7	0.406 4
	MAF+D2+NOR	5	0.282 6	0.886 7	0.376 4
脂 肪	None	5	0.316 5	0.854 7	0.471 8
	MAF+D1	7	0.249 7	0.850 0	0.511 2
	MAF+D1+NOR	7	0.173 6	0.918 0	0.408 5
	MAF+D2	6	0.188 5	0.908 1	0.441 9
	MAF+D2+NOR	7	0.178 3	0.910 4	0.422 2

注：D1：一阶微分；NOR：归一化。

4.2.3 多元散射校正 (MSC) 适用性探讨

表 4-3 多元散射校正处理的结果比较

Table 4-3 Comparison of results multiplication scatter correction

牛奶成分	光谱预处理方法	主成分数	$RMSEC$	R	$RMSEP$
蛋 白 质	MAF+MSC	8	0.293 2	0.845 1	0.389 3
	MAF+D1+NOR	6	0.265 7	0.895 4	0.312 5
	MAF+D1+NOR+MSC	4	0.502 8	0.788 9	0.614 3
脂 肪	MAF+MSC	7	0.589 1	0.725 9	0.824 7
	MAF+D1+NOR	7	0.173 6	0.918 0	0.408 5
	MAF+D1+NOR+MSC	6	0.576 0	0.720 2	0.630 7

MSC 作为一消除多元散射的预处理方法，在乳制品成分近红外光谱检测方面的应用有较好的效果^[41,72]。为了考察 MSC 是否在本文中也有效果，选用 MSC 光谱和上一节预处理光谱的预测效果进行比较，列表 4-3。从表中可以看出，无论蛋白质模型还是脂肪模型，经过 MSC 预处理的光谱都达不到理想的预测效果，最好的光谱预处理组合仍然是上一节选择的最佳组合。分析其中原因：

(1) 本文采用的是透反式光纤探头，检测到的近红外光谱主要是牛奶样品的透射

光谱。而 MSC 一般用于去除近红外漫反射光谱中样品的镜面反射及不均匀性造成的噪声，消除漫反射光谱的基线及光谱的不重复性。

(2)本试验的牛奶都是经过瞬时高温处理的均质奶，脂肪球被均一化且粒度变小。同时本试验中的牛奶是在温度为 36℃ 的水浴锅中加热下进行光谱采集的，牛奶中的脂肪球由于其熔点低于 36℃ 得以溶解，因而大大减少了其对光谱散射的影响。

4.3 牛奶成分模型的建立及验证

PLS^[73]被认为近红外定量分析中最有效、使用最广泛的多元校正方法，它将因子分析和回归分析相结合，很好地解决了许多以往用普通多元线性回归难以解决的问题。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，去掉代表干扰组分和干扰因素主成分，选用有用的主成分参与回归。

本章采用 PLS 作为建立蛋白质和脂肪预测模型的数学方法，其主成分数的选择是采用交互验证（Cross Validation）方法来选取，根据预测残差平方和（PRESS）来选择最佳的主成分数。PRESS 越小，说明模型的预测能力就越好。由于 PLS 原理已经是一种相对成熟的理论体系，这里不再累述。

4.3.1 最佳建模波段的选择

选择含信息量多的最佳光谱区间建模可以使校正模型具有最好的预测能力，并简化数据运算，提高校正效率。因此，在建立 PLS 回归模型时首先要找出含信息量多的最佳光谱区间。

相关系数法^[74]是一种选择最佳光谱区间

较常用和简单有效的方法。相关系数法是将样品的光谱吸光度数据和样品成分含量数据相关联，得到相关系数图，以确定信息最强的光谱范围。图 4-6 和图 4-7 分别为牛奶原始光谱经过最佳预处理后（见 4.2 节）各点近红外吸光度与蛋白质和脂肪含量对

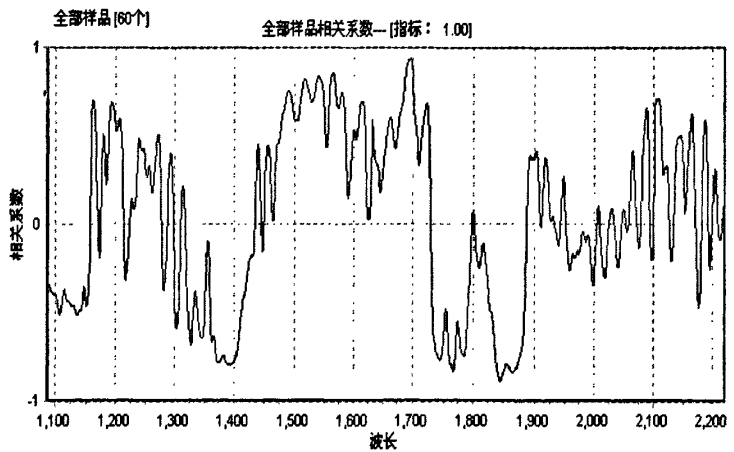


图 4-6 牛奶光谱与蛋白质相关系数

Fig4-6 Correlative coefficient of the spectra and protein

应于每个波长点的相关系数分布情况。当光谱与样品（组分）之间的相关系数较大时，其对应波长处的样品光谱含有最丰富的被测信息。因此，选择与样品（组分）之间的相关系数较大的光谱区间建模能够提高模型的预测能力。

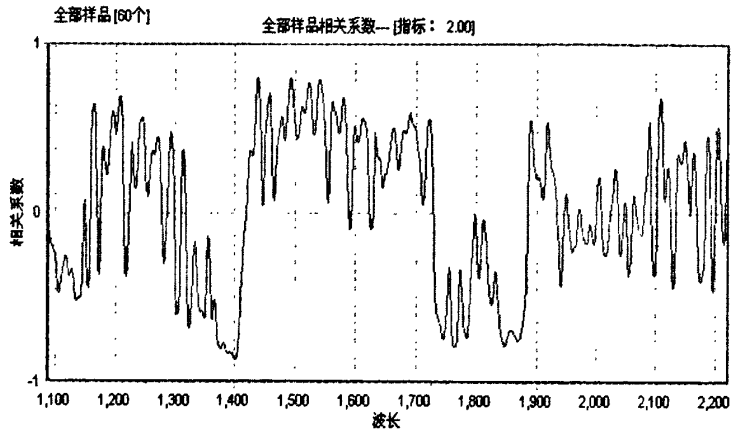


图 4-7 牛奶光谱与脂肪相关系数

Fig4-7 Correlative coefficient of the spectra and fat content

为了找出最佳的光谱区

间，选择对蛋白质和脂肪相关系数都较高的几个波段进行建模，比较各个光谱区间或区间组合下建立的预测模型的相关评价指标(见表 4-4)。

从表 4-4 中可见，在 1 370~1 410nm 和 1 680~1 860nm 中建立的蛋白质预测模型相关系数 (R) 和预测标准差 ($RMSEP$) 分别为 0.906 7 和 0.308 7，效果最好；而在 1 150~1 410nm 和 1 430~1 880nm 中建立的脂肪预测模型相关系数 (R) 和预测标准差 ($RMSEP$) 分别为 0.929 9、0.399 6，效果最好。选择与牛奶成分含量相关系数较高的光谱区间作为建模波段不仅降低了主成分数，简化了数据运算，而且提高了模型预测精度。

表 4-4 不同光谱区间 PLS 模型预测结果

Table 4-4 PLS model predicted results with combination spectra area

牛奶成分	光谱区间 (nm)	主成分数	$RMSEC$	R	$RMSEP$
蛋 白 质	All	6	0.265 7	0.895 4	0.312 5
	1 300~1 430、1 500~1 880	5	0.280 3	0.889 5	0.328 5
	1 150~1 220、1 350~1 900	6	0.278 0	0.887 3	0.337 0
	1 370~1 410、1 680~1 860	5	0.249 7	0.906 7	0.308 7
脂 肪	All	7	0.173 6	0.918 0	0.408 5
	1 300~1 430、1 500~1 880	6	0.212 0	0.882 8	0.416 5
	1 400~1 880	7	0.174 5	0.907 7	0.409 8
	1 150~1 410、1 430~1 880	6	0.165 7	0.929 9	0.399 6

4.3.2 牛奶蛋白质和脂肪模型的建立及验证

采用 PLS 方法分别建立由 45 个样品组成的校正集蛋白质和脂肪校正模型，然后

利用剩下的 15 个样品验证模型的预测性能（校正集与验证集化学组成见 4.2.2 节表 4-1），建模相关参数见表 4-5。其中校正模型优化指标有：相关系数（ R ）、校正集均方差（ $RMSEC$ ）；预测模型评价指标有：相关系数（ R ）、预测均方差（ $RMSEP$ ）和平均偏差（ $BIAS$ ）。

表4-5 牛奶蛋白质和脂肪模型的建立与验证结果

Table.4-5 the results of calibration and validation of milk protein and fat

模型	预处理方法	光谱区间	主成分数	校正集		验证集		
				R	$RMSEC$	R	$RMSEP$	$BIAS$
蛋白质	MAF+NOR+D1	1370~1410nm 和 1680~1860nm	5	0.9563	0.2497	0.9067	0.3087	0.2017
脂肪	MAF+NOR+D1	1150~1410nm 和 1430~1880nm	6	0.9916	0.1657	0.9299	0.3996	0.2461

从表4-5中可见，在蛋白质预测模型中，相关系数、预测均方差和平均偏差分别为 0.9067、0.3087和0.2017；在脂肪预测模型中，相关系数、预测均方差和平均偏差分别为0.9299、0.3996和0.2461。表4-6和表4-7分别为15个验证集样品蛋白质和脂肪的近红外光谱预测结果统计表。

对15个样品的实测值和预测值的相关性作图（见图4-8），从图中可以看出；牛奶光

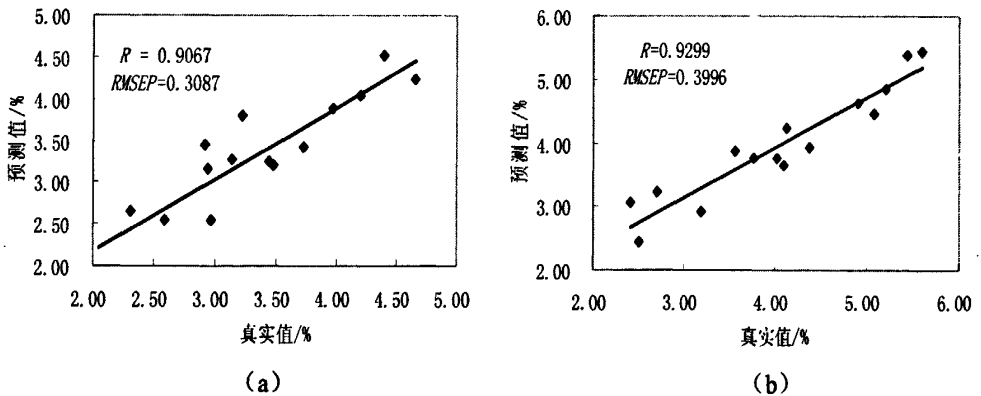


图4-8 牛奶成分近红外在线检测预测模型相关性图
(a)蛋白质 (b)脂肪

Fig.4-8 Correlations between prediction result and reference contents of milk using near-infrared spectroscopy

谱经过MAF平滑、一阶导数和归一化处理后，选择最佳建模波段，利用PLS建立的预测模型实测值与预测值具有较好的相关性，基本上可以满足牛奶工业在线检测的精度要求。

表4-6 15个牛奶样品蛋白质预测结果

Tabel.4-6 The protein results of 15 milk samples by NIRS

编号	实测值 (%)	预测值 (%)	相对误差 (%)	绝对误差
4	2.05	1.89	7.94	0.16
6	3.13	3.27	4.56	0.14
8	2.59	2.54	1.83	0.05
10	2.30	2.45	6.67	0.15
14	2.92	3.25	11.20	0.33
18	2.94	3.15	7.13	0.21
25	3.44	3.26	5.36	0.18
32	4.19	4.04	3.52	0.15
35	4.66	4.22	9.35	0.44
38	4.39	4.52	2.85	0.13
40	3.22	3.49	8.40	0.27
48	3.96	3.88	1.93	0.08
49	3.48	3.21	7.84	0.27
50	3.72	3.42	7.99	0.30
56	2.96	2.79	5.84	0.17

表4-7 15个牛奶样品脂肪预测结果

Tabel.4-7 The fat results of 15 milk samples by NIRS

编号	实测值 (%)	预测值 (%)	相对误差 (%)	绝对误差
5	4.38	4.00	8.73	0.38
13	4.12	4.23	2.66	0.11
18	5.60	5.44	2.81	0.16
25	2.50	2.45	1.80	0.05
30	3.76	3.77	0.15	0.01
34	5.44	5.37	1.22	0.07
36	5.21	4.87	6.59	0.34
40	2.71	2.49	8.28	0.22
42	3.57	3.88	8.69	0.31
43	3.18	2.91	8.39	0.27
45	2.40	2.67	11.20	0.27
48	5.08	4.46	12.27	0.62
50	4.91	4.62	5.89	0.29
53	4.01	3.75	6.47	0.26
57	4.09	3.75	8.30	0.34

4.4 本章小结

本章通过比较各种预处理方法的效果，找出蛋白质和脂肪的最佳预处理方法为：MAF平滑、归一化处理和一阶导数处理，同时探讨了MSC方法在本试验中的可行性并分

析了原因；选择了蛋白质最佳建模波段为：1 370~1 410nm和1 680~1 860nm，脂肪为：1 150~1 410nm和1 430~1 880nm，然后利用偏最小二乘法分别建立了牛奶样品的蛋白质和脂肪校正模型并对模型的预测性能进行了验证，蛋白质预测模型相关系数、预测均方差和平均偏差分别为：0.906 7、0.308 7和0.201 7；脂肪预测模型相关系数、预测均方差和平均偏差分别为：0.929 9、0.399 6和0.246 1。结果表明：本试验建立的牛奶蛋白质和脂肪的模型预测性能较好，基本上可以满足牛奶成分的在线检测精度要求。

第五章 温度对牛奶近红外在线检测的影响

在线近红外光谱分析结果的准确性受多种因素影响,如温度、压力、流速等。为保证在线近红外分析系统的准确性,需要对检测系统中的流通池进行恒温、恒压等处理。但同时为了建立一个适用范围广、稳健性好的在线近红外分析模型,往往人为地在一定范围内变动某些测试条件如样品的温度,以提高模型的稳健性和预测能力^[75]。

模型的稳健性是在线近红外分析技术中的一个重要指标。本章为进一步研究样品温度对近红外预测模型的影响以及建立一个在线近红外稳健分析模型,设计如下试验。

5.1 试验方案

在第四章采集的 60 个牛奶样品光谱的基础上,根据浓度均匀原则选择其中有代表性的样品作为校正集和验证集,作如下划分:

(1) 36℃校正集 选取 40 个样品,样品在 36℃下采集光谱,用来建立 36℃校正模型,称为 36℃校正模型。

(2) 温度补偿校正集 从(1)剩余的样品中,挑选出 10 个样品,样品温度在 26℃采集光谱;以此类推,在 31℃、41℃时各 10 个分别采集光谱。这样由 30 个样品组成 26℃、31℃、41℃温度补偿校正集。

(3) 验证集 由最后剩下的 10 个样品组成,分别在 26℃、31℃、36℃、41℃采集光谱,称为 26℃验证集(M1)、31℃验证集(M2)、36℃验证集(M3)、41℃验证集(M4)。

5.2 样品温度对牛奶成分预测结果的影响

5.2.1 温度对牛奶光谱的影响

图 5-1 为一验证集牛奶样品分别在 26℃、31℃、36℃和 41℃下采集的光谱图,从图中可以看出,在 26℃~41℃范围内,随着温度的升高,吸光度增大,但牛奶光谱的主要特征吸收峰的位置并未发生明显变动。说明在该温度范围内,温度的变化只影响

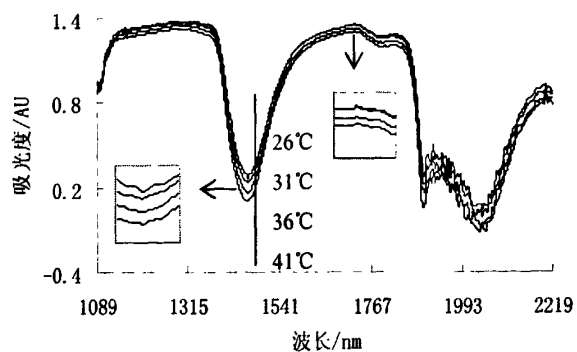


图 5-1 温度对牛奶光谱的影响

Fig5-1 Influence of temperature on NIRS

牛奶光谱的吸光度，而未对其特征吸收波长位置产生明显影响。图 5-2 为在不同温度下测定的能消除光谱的基线平移，消除图谱平移误差的一阶导数光谱。从图中光谱放大大部分可以看出，不同温度下的一阶导数光谱并未完全重合，这说明一阶导数预处理并不能完全消除不同温度光谱的差异，即温度不仅仅是造成光谱的平移。预处理方法对消除温度影响的有效性将在 5.2.3 节中展开讨论。

5.2.2 温度对牛奶成分预测结果的影响

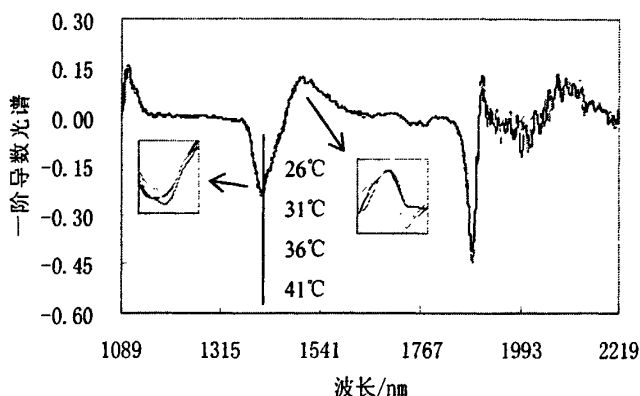


图 5-2 各温度下的一阶导数光谱比较

Fig5-2 The spectras after 1st derivation on different temperature

样品温度变化将导致化学键振动能量和能级跃迁几率的变化，对光谱吸光度有较大的影响。为了进一步说明温度变化对本研究中牛奶成分分析的影响，对由 36°C 下采集的 40 条光谱组成的 36°C 校正集建立 PLS 校正模型，然后对不同温度验证集进行预测，比较各个预测模型的预测均方差（见表 5-1）。从表中可以看出，在蛋白质预测模型中，26°C、31°C、36°C、41°C 验证集 *RMSEP* 分别为：0.570 2、0.593 6、0.308 7 和 0.640 2，36°C 验证集效果最好；在脂肪预测模型中，26°C、31°C、36°C、41°C 验证集 *RMSEP* 分别为：0.704 4、0.668 3、0.399 6 和 0.735 6，36°C 验证集效果最好。

表 5-1 36°C 校正模型预测不同温度下的样品比较

Table.5-1 Comparison of prediction of different temprature milk by 36°C calibration model

牛奶成分	主成分数	<i>RMSEC</i>		<i>RMSEP</i>		
		36°C 校正集	26°C 验证集	31°C 验证集	36°C 验证集	41°C 验证集
蛋白质	5	0.249 7	0.570 2	0.593 6	0.308 7	0.640 2
脂肪	6	0.165 7	0.704 4	0.668 3	0.399 6	0.735 6

以上结果说明：36°C 下建立的模型只对相同温度下的验证集有效，这进一步证明了

在线近红外光谱分析系统中，对样品温度进行恒定的重要性。

为了进一步找出样品温度对牛奶蛋白质和脂肪预测结果的规律性，作 36℃校正集对各个温度验证集牛奶蛋白质和脂肪平均偏差和温度的变化的关系图（如图 5-3）。从图中可以看出：在 26℃~41℃范围内，36℃校正模型对不同温度下蛋白质和脂肪的预测平均偏差与 36℃下的相比都较大，但并非呈明显的线性变化，其变化趋势有待进一步研究。这是由牛奶这种含复杂化学

成分的液体的特殊性质决定的，同时在线近红外分析系统受各种干扰因素的影响，比如气泡是液体样品进行在线分析需要考虑的干扰因素，液体可以从大气中溶解相当数量的气体，如果升高温度，气体便以气泡的形式释放。气泡的存在会引起光的散射和反射，产生假分析信号，影响最终分析结果。

5.2.3 光谱预处理方法对模型温度稳健性的影响

对样品进行定量分析所得结果的好坏除了取决于近红外光谱对样品待测组分、结构的表征外，在很大程度上还取决于光谱数据本身的质量，也就是各种因素对光谱的影响。

对 36℃校正集分别使用不同的预处理方法或预处理方法的组合进行处理，然后利用 PLS 建立牛奶蛋白质和脂肪不同温度下的预测模型，比较预测模型的评价指标（见表 5-2）。从下表中可见，经过 MAF 平滑、一阶导数、二阶导数、归一化等预处理的

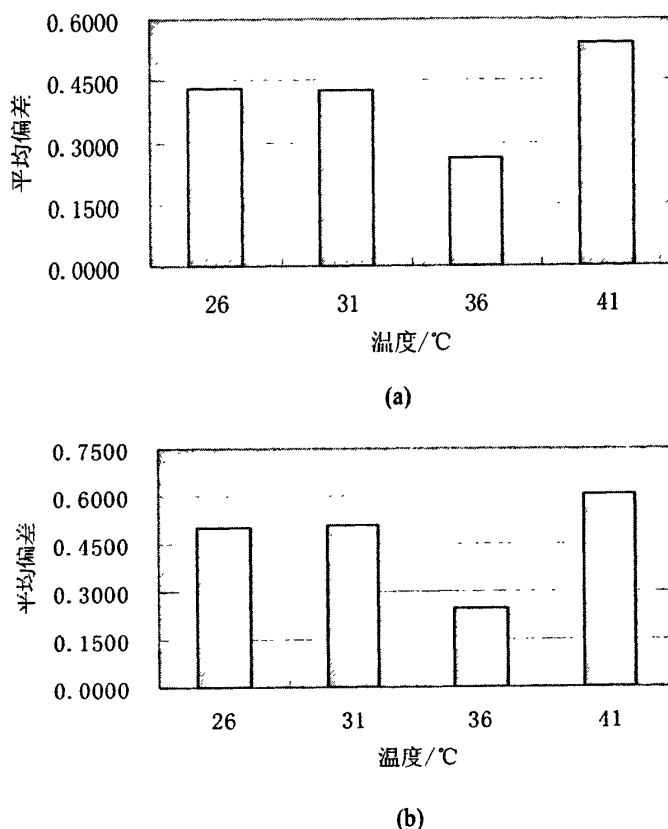


图 5-3 不同温度预测模型的平均偏差比较
(a)蛋白质(b)脂肪

Fig5-3 Bias of different temperature in valiration model

36℃校正集光谱并不能改善对其他温度下预测模型的精度。这说明仅使用光谱预处理方法不能消除温度对模型的影响,无法建立温度稳健模型。

表 5-2 几种预处理方法对不同温度光谱的预测效果比较

Table.5-2 Comparison of prediction results after different pretreatments

牛奶成分	预处理方法	主成分数	RMSEP			
			M1	M2	M3	M4
蛋 白 质	MAF+D1	7	0.567 1	0.565 7	0.369 3	0.451 2
	MAF+D1+NOR	5	0.570 2	0.593 6	0.308 7	0.640 2
	MAF+D2	5	0.621 4	0.655 5	0.416 4	0.589 5
	MAF+D2+NOR	5	0.524 6	0.635 4	0.386 5	0.475 1
脂 肪	MAF+D1	8	0.534 6	0.435 4	0.521 2	0.614 7
	MAF+D1+NOR	6	0.704 4	0.668 3	0.399 6	0.735 6
	MAF+D2	7	0.586 8	0.568 4	0.465 9	0.597 5
	MAF+D2+NOR	7	0.496 4	0.580 7	0.434 2	0.599 8

5.3 温度补偿校正方法的效果

为了研究温度补偿校正方法的效果,以及消除样品温度对牛奶成分预测结果的影响而建立一个能抗温度干扰的牛奶成分稳健模型,建立由 36℃校正集和 26℃、31℃、41℃温度补偿校正集组成的温度混合校正集。参照第四章的方法,选择最佳预处理方法和建模波段。原始光谱经 MAF 平滑、归一化处理和一阶导数预处理后,利用相关系数法进行建模波段优选,然后用 PLS 分别建立蛋白质和脂肪的温度混合校正模型,最佳主成分数由预测残差平方和确定。所建校正模型分别对 26℃、31℃、36℃、41℃验证集进行预测,校正模型优化参数和预测模型评价指标分别见表 5-3 和表 5-4。

表 5-3 校正模型优化参数效果比较

Table.5-3 the results of calibration set

校正模型比较	蛋白质		脂肪	
	R	RMSEC	R	RMSEC
36℃模型	0.956 3	0.249 7	0.991 6	0.165 7
温度混合校正模型	0.961 6	0.265 7	0.991 1	0.166 2

表 5-4 温度混合模型预测效果评价指标比较

Table.5-4 the results of valibration set

牛奶预测模型	26℃验证集		31℃验证集		36℃验证集		41℃验证集	
	R	RMSEP	R	RMSEP	R	RMSEP	R	RMSEP
蛋白质	0.899 0	0.330 1	0.901 1	0.329 7	0.900 4	0.324 7	0.897 2	0.403 9
脂肪	0.915 8	0.440 3	0.914 0	0.443 0	0.921 4	0.408 8	0.903 1	0.479 5

从表 5-3 中可知：蛋白质校正模型中，36℃校正模型和温度混合校正模型校正集

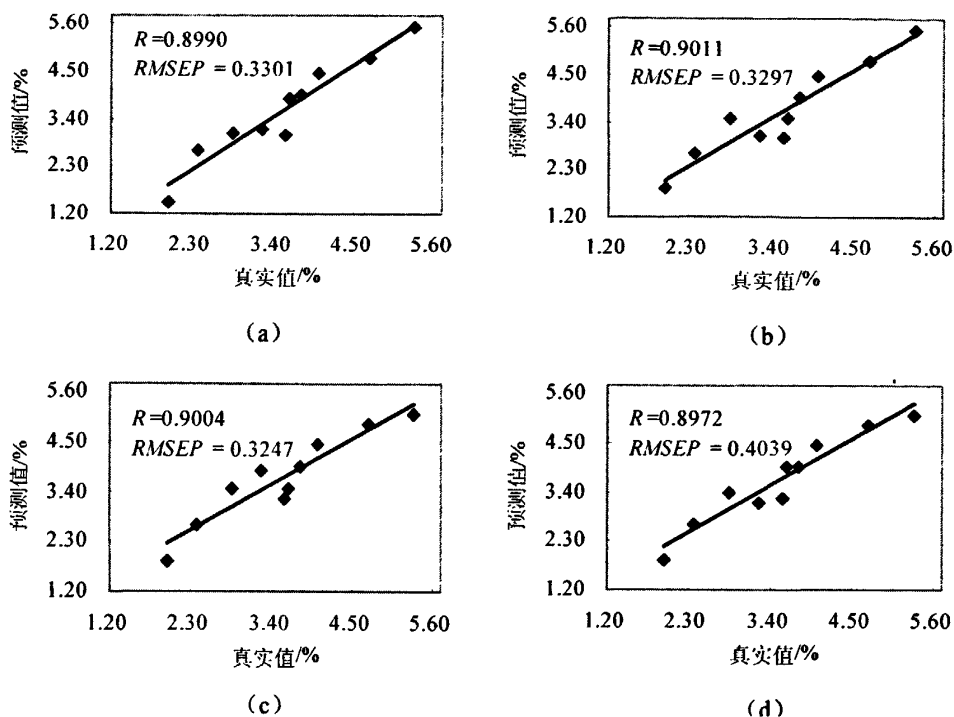


图 5-4 温度混合模型对不同温度验证集的蛋白质预测相关图

(a) 26℃验证集(b) 31℃验证集(c) 36℃验证集(d) 41℃验证集

Fig.5-4 plot of protein prediction results of different validation set by global calibration

相关系数 R 分别为 0.956 3 和 0.961 6，回判均方差 $RMSEC$ 分别为 0.249 7 和 0.265 7；脂肪校正模型中，36℃校正模型和温度混合校正模型校正集相关系数 R 分别为 0.991 6 和 0.991 1，回判均方差 $RMSEC$ 分别为 0.165 7 和 0.166 2，结果都非常接近，说明温度补偿校正集的加入对校正模型的影响不大。这是因为加入的补偿校正集样品数（10 个）相对与原来 36℃校正集相比较少，且分布均匀，不足以对校正模型造成较大影响。

从表 5-4 中可见：蛋白质预测模型中，温度混合模型对 26℃、31℃、36℃、41℃ 验证集相关系数 R 和预测均方差 $RMSEP$ 分别为 0.899 0、0.901 1、0.900 4、0.897 2 和 0.330 1、0.329 7、0.324 7、0.403 9，效果都较好；脂肪预测模型中，温度混合模型对 26℃、31℃、36℃、41℃ 验证集相关系数 R 和预测均方差 $RMSEP$ 分别为 0.915 8、0.914 0、0.921 4、0.903 1 和 0.440 3、0.443 0、0.408 8、0.479 5，效果都较好。图 5-4 和图 5-5 分别为温度混合模型对不同温度验证集蛋白质和脂肪含量真实值和预测值的相关图。

以上结果说明加入了温度补偿校正集的温度混合校正模型提高了校正模型对不同温度验证集的预测能力，能够消除样品温度对模型的影响，提高了模型的稳健性^[76]。

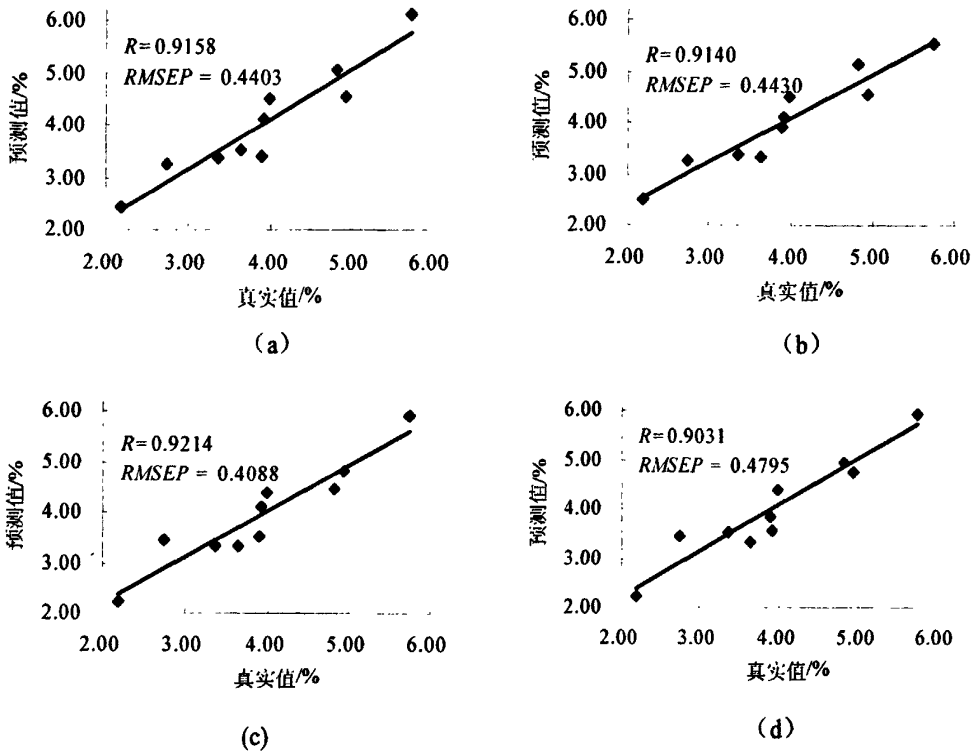


图 5-5 温度混合模型对不同温度验证集的脂肪预测相关图

(a) 26°C 验证集 (b) 31°C 验证集 (c) 36°C 验证集 (d) 41°C 验证集

Fig.5-5 plot of fat prediction results of different validation set by global calibration

但这样做同时降低了预测模型的精度,如温度混合校正集和 36°C 校正集分别对 36°C 验证集进行预测,蛋白质预测模型中, R 下降了 0.63 % 而 $RMSEP$ 增大了 0.016 0; 脂肪预测模型中, R 下降了 0.85 % 而 $RMSEP$ 增大了 0.009 2。在具体的实施过程中,如果有条件就要严格控制温度,比如可以在流通池周围加有恒温装置。此外,对于流动性较差的液体样品,在过程分析时必须具有一定的温度,保持其良好的流动性。对于无法进行恒温的情况,如将探头直接插入主管线,就要进行温度补偿^[69]。

NIR 分析技术对不同条件下样品测定均需要建立相应的分析模型,而建立分析模型需要消耗相当大的人力、物力和财力,加上在线 NIR 技术受到多种干扰因素的影响,因此稳健模型的建立至关重要;同时这也是实现模型传递、资源共享的前提条件^[77]。本文温度补充校正方法的引入为建立其他(如一些难于控制的干扰因素:气泡、压力、样品状态等)在线近红外稳健模型提供了参考。

5.4 本章小结

本章研究了不同温度对牛奶成分预测的影响,并将不同温度下在线采集的牛奶样品光谱组成温度补偿校正集,建立了温度全局模型,可实现模型对温度因素的稳健性。

在 26~41℃ 范围内,牛奶光谱随着温度的升高,吸光度增大,但其主要特征吸收峰的位置并未发生明显变动;把 36℃ 下建立的校正模型用来预测不同温度下的验证集发现,36℃ 下建立的牛奶成分模型只对相同温度下的预测模型有效,这说明了样品温度对预测结果的重要性;比较了各预处理方法对不同温度下预测模型的预测效果,表明了仅使用光谱预处理方法不能消除温度对模型的影响,无法建立温度稳健模型;采集了不同温度下牛奶的光谱组成温度补偿校正集,通过补偿校正方法的引入,建立了温度全局模型,实现了模型对温度因素的稳健性。

补偿校正方法的引入在一定程度上降低了预测模型精度,该方法简单可行,还可用于建立其他影响因素(如一些难于控制的干扰因素:气泡、压力、样品状态等)的稳健分析模型。该方法为进一步研究在线近红外分析模型稳健性提供了重要的参考。

第六章 总结及展望

6.1 研究总结

本论文基于本课题组自行研制的近红外液态在线检测试验台,以牛奶为研究对象,以实现牛奶主要成分蛋白质和脂肪含量在线检测为目标,得到了以下研究成果:

1. 为了减少水对牛奶吸收光谱的影响,选择蒸馏水作为本底;为了提高光谱信噪比并节约扫描时间,选择扫描次数为 32;为了增强牛奶的流动性及减少牛奶脂肪球散射影响,选择最佳建模温度为 36℃;为了确保近红外仪能稳定地获取牛奶样品的光谱而使牛奶样品能够平稳的流过光纤探头以避免出现死容积,选择了最佳光程和流速分别为 2mm 和 0.218cm/s。这些参数的优化为牛奶样品光谱采集提供了参考;

2. 比较了多种预处理方法的效果,找出蛋白质和脂肪的最佳预处理方法为:MAF 平滑、归一化处理和一阶导数处理,同时探讨了 MSC 方法在本试验中的可行性并分析了原因;比较了多种预处理方法对不同温度下预测模型的预测效果,表明了仅使用光谱预处理方法不能消除温度对模型的影响,无法建立温度稳健模型;选择了蛋白质最佳建模波段为:1370~1410nm 和 1680~1860nm,脂肪为:1150~1410nm 和 1430~1880nm,然后利用偏最小二乘法分别建立了牛奶样品的蛋白质和脂肪校正模型并对模型的预测性能进行了验证,蛋白质预测模型相关系数、预测均方差和平均偏差分别为:0.9067、0.3087 和 0.2017;脂肪预测模型相关系数、预测均方差和平均偏差分别为:0.9299、0.3996 和 0.2461。结果表明:本试验建立的牛奶蛋白质和脂肪的模型预测性能较好,可以满足牛奶成分的在线检测精度要求;

3. 在 26~41℃ 范围内,牛奶光谱随着温度的升高,吸光度增大,但其主要特征吸收峰的位置并未发生明显变动;把 36℃ 下建立的校正模型用来预测不同温度下的验证集发现,36℃ 下建立的牛奶成分模型只对相同温度下的预测模型有效,这说明了样品温度对预测结果的重要性;

4. 采集了不同温度下牛奶的光谱组成温度补偿校正集,通过补偿校正方法的引入,建立了温度全局模型,实现了模型对温度因素的稳健性。补偿校正方法的引入虽然在一定程度上降低了预测模型精度,但该方法简单可行,还可用于建立其他影响因素(如一些难于控制的干扰因素:气泡、压力、样品状态等)的稳健分析模型。该方法为进一步研究在线近红外分析模型稳健性提供了重要的参考。

6.2 本文的创新点

1. 针对国内近红外光谱分析在线检测牛奶成分目前还未在工业方面得到实际应用, 本文为该技术在乳品生产工业中牛奶成分的在线监测提供了参考;
2. 针对国外报道近红外光谱分析在线检测牛奶成分大多集中在短波近红外区域, 本文使用 1 100~2 200nm 长波近红外光谱得到了较满意的效果;
3. 本文对于检测过程中的一些重要参数(如温度、流速、光程)进行了优选, 这些参数的优化为牛奶样品光谱采集提供了参考。

6.3 展望

由于各种条件的限制, 本研究还存在以下有待改进的问题:

1. 由于试验条件的限制, 本在线检测系统还不够完善, 如流通池的设计未考虑到样品清洗的方便性, 光纤探头的固定有待进一步的完善;
2. 本研究选择奶粉加水配置样品, 与实际牛奶生产加工有一定的差异, 建立的模型也具有局限性, 如果建立一个更多样品适用范围更广的模型, 对于应用和指导生产实践会更适用。
3. 模型的稳健性是评价在线近红外分析技术的一个重要指标, 由于条件限制本文只考察了温度对模型稳健性的影响。其他因素(如流速、压力、样品状态)对模型的稳健性的影响, 有待进一些研究;
4. 把在线近红外光谱分析技术推广到牛奶成分检测的实际应用中还有一定差距, 如需降低仪器的成本、加大在线检测结果反馈的速度、增强模型对各种因素的抗干扰能力。

致 谢

攻读硕士学位的3年里,得到了导师陈斌教授的精心培养和悉心指导。尤其在本文的选题、研究方案的确定方面给予的大力支持,并且特别感谢他在我的论文撰写期间提出的宝贵建议,使我能够顺利地完成论文的各项工作。他严谨求实的治学态度、博大精深的专业知识、平易近人的学者风范和高度的责任感使我身受感染。在此,学生表示衷心的感谢和崇高的敬意!

同时感谢陆道礼老师在研究过程中给我的宝贵意见和无私的协助,他一丝不苟的实验态度和严谨的工作作风使我受益匪浅,给我留下了深深的印象。也要感谢朱芸、肖颖、李玉平、叶静、朱文静、蔡斌和颜辉以及学弟学妹们在本文研究和撰写过程中给予的帮助。衷心感谢所有曾帮助过我的老师、同学和朋友。

最后, 特别感谢一直默默支持我的家人!

参考文献

- [1] Galls, J.B., Lllman, D.L., Kowalski, B.R. *Anal. Chem.*, 1987, 59(9): 624A-637.
- [2] Workman J., *Appl. Spectros. Rev.*, 1999, 34(1&2): 1-89.
- [3] 徐广通, 袁洪福, 陆婉珍等. 近红外光谱技术快速测定柴油物理性质[J]. *石油学报*, 1999, 15(5): 63-65.
- [4] 祝诗平. 近红外光谱品质检测方法研究.[博士论文]. 中国农业大学, 2003.
- [5] Herschel W., *Investigation of the Powers of the prismatic Colours to heat and illuminate Objects; with Remarks that prove the different Refrangibility of radiant Heat. To which is added an Inquiry into the Method of viewing the Sun advantageously with Telescopes of large Apertures and high magnifying Powers*, *Philosophical Transactions*, 1800, 90: 255-326.
- [6] Abney W., Festing, E. R., *On the Influence of the Atomic Grouping in the Molecules of Organic Bodies on their Absorption in the Infra-Red Region of the Spectrum*, *Philosophical Transactions*, 1881, 172: 887.
- [7] Brackett F.S., *Proc. Natl. Acad. Sci.* 1928, 14: 857.
- [8] Kaye W. *Theory and principles of near infrared spectroscopy*[J]. *Spectrochim Acta*, 1955, 7: 181.
- [9] Ben-Gera I., K.H. Norris, *Direct spectrophotometric determination of fat and moisture in meat products*[J]. *Journal Food Science*, 1968, 33: 64.
- [10] 李军会. 中国农业近红外分析技术软件及网络系统的研制与开发[J]. *现代仪器*, 2000(5).
- [11] 朱锐. 智能型近红外光谱分析仪的研制.[硕士论文]. 江苏大学, 2006.
- [12] MBlanco, Villarroya I. *NIR spectroscopy: a rapid - response analytical tool*[J]. *trends in analytical chemistry*, 2002, 21(4): 240-250.
- [13] 彭玉魁, 李菊英. NIRS 同时测定小麦种子水份、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.
- [14] 刘建学, 吴守一, 方如明. 大米直链淀粉含量的近红外光谱分析[J]. *农业工程学报*, 2000, 16(3).
- [15] 陈蛋, 陈斌, 陆道礼等. 近红外光谱分析法测定菜籽油中芥酸的含量[J]. *农业工程学报*, 2007, 23(1): 234-237.
- [16] 张军, 陈瑾仙. 近红外光谱分析仪的发展及在农产品、食品中的应用[J]. *Spectral analysis*, 2003, (5): 24-291.
- [17] 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.
- [18] 陈斌. 用近红外光谱分析液态食品的品质.[博士论文]. 江苏理工大学, 2001.
- [19] Kelly JJ, Barlow CH, Jinguji TM, et al. *Anal Chem*, 1989, 61(4): 313-320.
- [20] 徐广通, 袁洪福, 陆婉珍. 近红外光谱法测定柴油十六烷值[J]. *石油学报*, 1999, 15(4): 62.
- [21] 鲍峰伟, 刘景艳. 近红外光谱分析技术在石油化工中的应用[J]. *贵州化工*, 2006, 31(6): 34-36.
- [22] 田丽蓉, 汤真, 刘福旋等. 近红外光谱-人工神经网络法用于粉末药品阿司匹林非破坏定量分析[J]. *分析试验室*, 2001, 20(1): 80-81.
- [23] 李晓薇等. 用近红外光谱定量分析混纺毛织品中羊毛的质量分数[J]. *中国纺织大学学报*, 2000, 26(6): 75.
- [24] Wayne C, Tincher and Anderaluk. *NIRS Analysis of Cotton /Polyester Yarns*[J]. *Spectroscopic*

- Analysis,1985,17(10):200~204.
- [25] 王小犬,陈斌,吴继明.相关分析技术用于消除样品背景对预测模型的影响[J].纺织学报,2007,3.
- [26] Ingle J D Jr and Crouch S R.Spectrochemical Analysis. New York: Pretice-Hall Englewood Cliffs,NJ,1988.
- [27] Osborne B G and Fearn T.Near Infrared Spectroscopy in Food Analysis. New York: John Wiley and Sons Inc.1986.
- [28] 丁东.近红外光谱技术及其在生物医学中的应用研究.[博士论文].吉林大学,2004.
- [29] 张宏勋.过程分析仪器[M].北京:冶金工业出版社.1984,11.
- [30] Thygesen L G.J. Near Infrared Spectrosc. ,2000,8:183.
- [31] 许祿,邵学广.化学计量学方法[M].北京:科学出版社.2004.
- [32] 朱尔一,杨梵原.化学计量学技术及应用[M].北京:科学出版社.2001.
- [33] 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,2004,07 (16).
- [34] 袁洪福,陆婉珍.现代光谱分析中常用的化学计量学方法[J].现代科学仪器,1998,(5):6~9.
- [35] 翁诗甫.傅里叶变换红外光谱仪[M].北京:化学工业出版社.2005.
- [36] 程永华,闰永胜.化学计量学在光谱分析中应用的研究进展[J].平顶山师专学报(自然科学版),1998,13(4):33~40.
- [37] 陆瑞才.模式识别—化学计量学的一个重要分支[J].湖州师范学院学报,2000,22(3):57~60.
- [38] 耿新青.化学计量学中人工神经网络的进展[J].鞍钢技术,2001,(3):47~49.
- [39] 陈斌,王豪.专利算法在白酒酒精度近红外光谱分析模型转移中的应用[J].红外技术,2006,28(4) :245~248.
- [40] 陈斌,吴继明.用 Shenk's 方法提高芥酸 NIR 模型预测精度[J].农业机械学报,38 (8) :101~104
- [41] 李庆波,汪曦,徐可欣等.牛奶主要成分含量近红外光谱快速测量法 [J].食品科学,2002,23(6) :121~124.
- [42] 常敏.应用红外光谱技术进行牛奶成分检测的研究.[硕士论文].天津大学.2004.
- [43] 王丽杰,徐可欣,郭建英.采用近红外光谱技术检测牛奶中脂肪、蛋白质及乳糖含量[J].光电子激光,2004,15(4):468~471.
- [44] 王云,徐可欣,常敏等.近红外光谱技术检测牛奶中脂肪及蛋白质含量校正模型的建立[J].光学仪器.2006,28(3):3~7.
- [45] 韩东海,鲁超,刘毅等.生鲜乳中还原乳的近红外光谱法鉴别[J].光谱学与光谱分析,2007,27 (3): 465~468.
- [46] 皮付伟,韩东海.基于短波近红外的牛奶检测研究[C].全国第一届近红外光谱学术会议论文集,2006,10(1): 506~510.
- [47] 韩东海,鲁超,皮付伟.近红光谱分析技术在液态乳检测上的应用[C].全国第一届近红外光谱学术会议论文集,2006,10(1):106~111.
- [48] 毕卫红,李超等.利用近红外光谱技术对牛奶中蛋白质含量进行定量分析[J].计量技术,2005,08: 34~36.
- [49] 杨小丽.牛奶中营养物质含量的近红外检测技术的研究[硕士论文].燕山大学,2005.
- [50] Tsenkova R.Ozaki Y. et al.Near-infrared spectroscopy for biomonitoring: Cow milk composition measurement in a spectral region From 1100 to 2400 namometers[J].J.Anim.Sci,2000,78:515~5 22.
- [51] Senkova R.T et al.somatic cell count determination in cow milk based on Near infrared

- spectroscopy a new diagnostic tool.[J].Anim.Sci,2001,79:2550~2557.
- [52] Pravdova V. et al. Calibration of somatic cell count in milk based on Near infrared spectroscopy[J].Analytica Chimica Acta ,2001,450:131~141.
- [53] Woo Y, Terzawa Y, Chen J Y, et al. Development of a New Measurement Unit(Milkspec-1) for Rapid Determination of Fat, Lactose, and Protein in Raw Milk Using Near-Infrared Transmittance Spectroscopy[J].Applied Spectroscopy,2002,56(5).
- [54] Desmond B, John A Brennan D. et al. Issues in development of NIR micro spectrometer system for on-line process monitoring of milk product[J]. Measurement,2003,33:67~74.
- [55] Kawamura S. et al. on-line near infrared spectroscopic sensing technique for Assessing milk quality milking[C]. ASAE.2003:27~30.
- [56] Karoui R, Baerdemaeker J.D. A review of the analytical methods coupled with chemometric tools, for the determination of the quality and identity of dairy products[J].Food Chemistry, 2007 (12): 621~640,
- [57] vandenBerg F.W.J., Hoefsloot H.C.J., Smilde A.K., Anal. Chem.,2002,74(13):3105~3111.
- [58] Smilde, A K vandenBerg F.W.J., Hoefsloot H.C.J., Anal. Chem.,2002,74(13):368A~373A.
- [59] 陆婉珍,袁洪福,褚小立,王艳斌.当代中国近红外光谱技术[C].全国第一届近红外光谱学术会议论文集.北京:中国石化出版社,2006.
- [60] 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.
- [61] Gary D N. Anal. Chem.,1981,53(3):489A~500A.
- [62] 金义忠,王育良,封沛 等.过程分析仪器成套系统工程应用研究[J].分析仪器,1997,(4):56~58.
- [63] 任军.浅谈在线分析系统技术[J].化工自动化及仪表,1996,23(4):61~62.
- [64] 郭本恒,吴 昊,陈 剑 等.乳品化学[M].北京:中国轻工业出版社,2001.8~9.
- [65] 技术监督行业工人技术考核培训教材编委会组编.乳及乳制品检验技术[M].北京:中国计量出版社,1997.
- [66] Thygesen L G. J. Near Infrared Spectrosc.,2000,8:183.
- [67] 骆承庠.乳与乳品工艺学[M].北京:农业出版社,1989:43~44.
- [68] 韩永辉.牛奶成分近红外测量的光学基础理论与实验研究.[博士论文]天津大学,2005.
- [69] 孙振东,万峰,赵玉春,范世福等.光纤光度分析仪设计的有关问题[J].分析仪器,2000,01.
- [70] 陆婉珍.现代近红外光谱分析技术(第二版)[M].北京:中国石化出版社,2006:174~203.
- [71] 陆婉珍,袁洪福.现代光谱分析中常用的化学计量方法[J].现代科学仪器,1998,5.
- [72] 褚鹏蛟.奶粉蛋白质含量光学快速检测方法的研究[硕士论文].天津大学,2005.
- [73] 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.
- [74] 孟祥龙.“镇江香醋”发酵液酒精度近红外光谱分析研究[硕士论文].江苏大学,2006.
- [75] 褚小立,袁洪福,陆婉珍等.在线近红外光谱过程分析技术及其应用[J].现代科学仪器,2004,2:3~17.
- [76] 褚小立,袁洪福,王艳斌等.近红外稳健分析校正模型的建立(I)一样品温度的影响[J].光谱学与光谱分析,24,6:666~671.
- [77] 李 勇.近红外分析模型稳健性研究[硕士论文].西北农林科技大学,2005.

硕士期间发表论文

- [1] 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. 江苏大学学报 (自然科学版). (已录用)
- [2] 朱文静, 陈斌, 邹贤勇. 基于滤光片型 NIR 光谱仪快速测定完整大豆蛋白含量的研究. 现代仪器.

附录 A 60 个牛奶样品的化学检测值

样品编号	蛋白质	脂肪	样品编号	蛋白质	脂肪
1	2.32	2.71	31	1.57	3.76
2	2.34	3.26	32	3.72	4.79
3	2.53	4.03	33	4.19	5.02
4	2.05	2.58	34	3.35	3.43
5	2.94	4.38	35	3.86	5.44
6	3.13	2.61	36	4.66	6.31
7	3.08	3.68	37	3.70	5.21
8	2.59	2.83	38	1.69	6.01
9	2.61	2.70	39	4.39	6.54
10	2.30	2.66	40	3.30	5.98
11	2.86	2.64	41	3.22	2.71
12	3.02	3.44	42	3.26	3.69
13	3.27	4.12	43	3.68	3.57
14	2.92	4.19	44	5.99	3.18
15	4.77	1.59	45	4.84	3.81
16	2.83	3.92	46	4.30	2.40
17	2.34	3.03	47	2.65	3.55
18	2.94	5.60	48	3.48	2.97
19	3.06	6.17	49	3.96	5.08
20	2.69	2.72	50	3.48	4.73
21	3.11	5.90	51	3.72	4.91
22	3.30	2.95	52	3.72	5.46
23	3.75	5.37	53	3.30	2.56
24	4.70	5.25	54	3.10	4.01
25	3.44	2.50	55	3.45	3.77
26	2.98	1.76	56	2.46	6.75
27	3.75	3.51	57	2.96	5.26
28	5.02	4.64	58	2.47	4.09
29	5.14	5.50	59	5.08	2.09
30	2.32	2.71	60	3.78	3.09

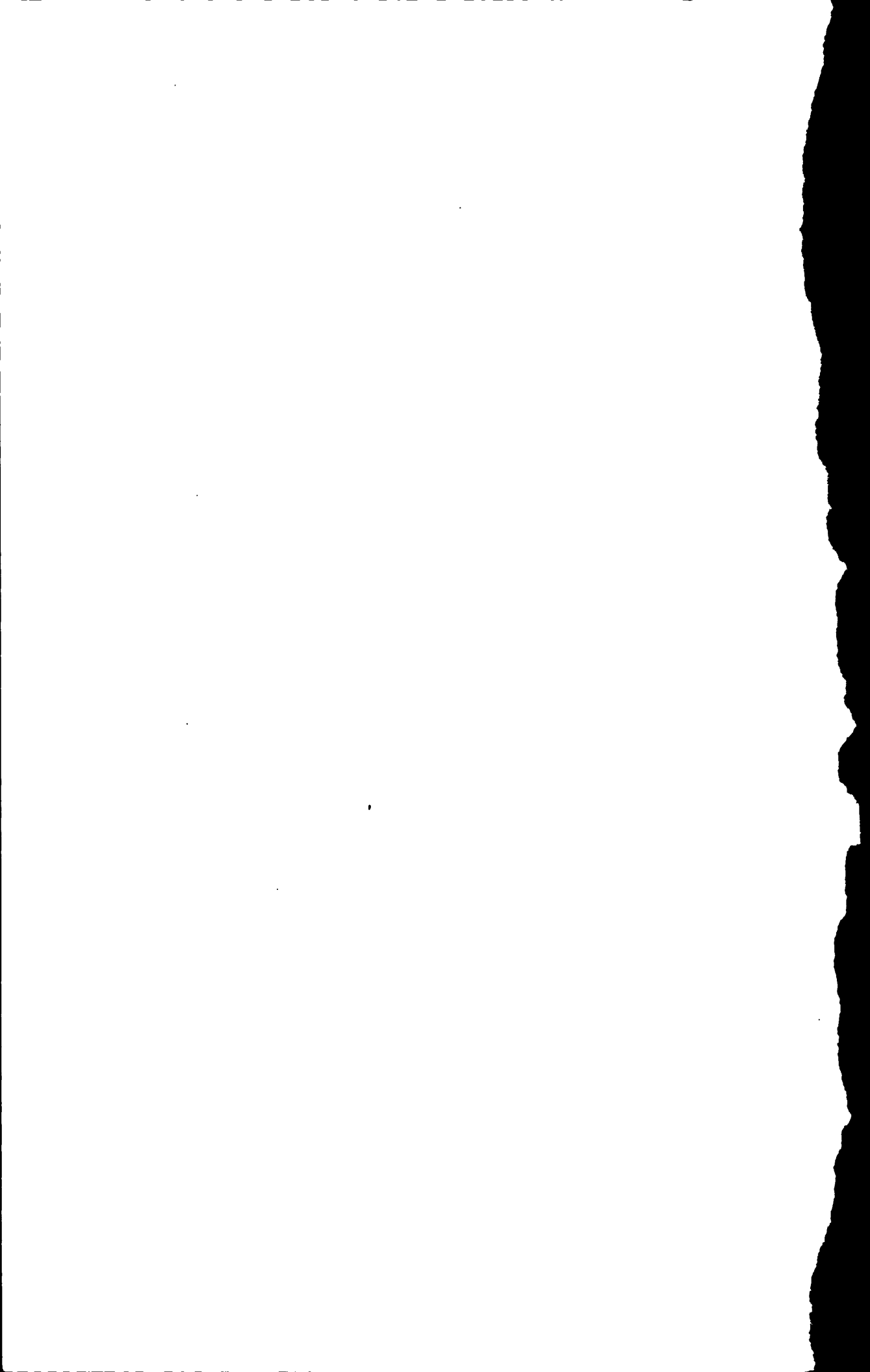
注：样品化学值单位：g/100g

附录 B 50 个奶粉样品的化学检测

奶粉标号	名称	脂肪	蛋白
1	伊利幼儿配方奶粉 1-3 岁幼儿适用	18.99	16.21
2	伊利较大婴儿配方奶粉 6-12 个月婴幼儿适用	22.82	16.37
4	伊利全脂奶粉	28.23	21.89
5	三鹿助长奶粉	18.09	14.32
6	伊利全脂甜奶粉	25.55	17.13
12	雅士利幼儿配方奶粉 1-3 周岁幼儿适用	21.44	17.96
13	正元钙多维学生奶粉 (摇篮公司)	22.51	15.08
14	光明全脂甜奶粉	15.08	18.25
15	光明全脂甜奶粉	18.62	16.11
16	三鹿慧幼 1 婴儿配方奶粉 0-12 个月适用	15.38	10.72
17	雀巢全脂甜奶粉	20.06	17.60
18	伊利全脂奶粉	24.03	19.09
19	雀巢全脂甜奶粉	34.44	17.04
21	三鹿中老年奶粉	6.17	18.55
23	雀巢高钙甜奶粉	21.12	15.23
24	雀巢学生甜奶粉	16.32	12.60
25	飞鹤婴儿钙+锌奶粉 0-6 个月婴儿适用	19.55	12.88
26	雀巢高钙甜奶粉	26.97	19.00
27	三鹿慧幼 3 幼儿配方奶粉 1-3 周岁适用	13.61	13.46
28	三鹿全脂加糖奶粉	29.52	15.55
29	伊利中老年多维高钙奶粉	14.73	16.51
30	伊利全脂甜奶粉	26.84	18.75
31	伊利全脂奶粉	26.27	23.50
32	雅士利优怡女士奶粉	9.49	20.98
34	飞鹤高钙低糖奶粉	8.79	14.88
36	摇篮全脂补钙速溶鲜奶粉	17.53	14.21
37	完达山全脂奶粉	18.03	19.54
38	完达山高乳钙中老年奶粉	14.38	19.18
39	完达山全脂加糖奶粉	18.80	7.87
41	伊利全脂加糖奶粉	23.95	14.57
42	雀巢能多健婴儿配方奶粉	19.04	10.41
43	雀巢能多健较大婴儿及幼儿配方奶粉	17.15	16.77
51	森永加糖全脂奶粉	19.91	19.31
54	光明全脂奶粉	27.60	20.41
56	光明全脂甜奶粉	26.04	18.48
57	多美滋/多乐加婴儿配方奶粉	35.05	9.85
59	伊利全脂奶粉	25.45	22.04
61	光明中老年奶粉益生元	23.32	16.48
62	贝因美/美丽健 (浙江) 中老年配方奶粉	13.55	16.12
66	三鹿准妈妈奶粉	11.12	20.96

68	森永学生奶粉	12.01	21.52
69	雀巢力多精奶粉	17.73	13.24
70	南山倍慧 AA 金维多婴幼儿配方奶粉	22.23	17.33
71	二鹿较大婴儿及幼儿配方奶粉	23.67	17.41
72	贝因美出生婴儿奶粉 (0~6 个月)	27.28	14.45
73	美赞臣安儿宝幼儿奶粉(1~3 岁幼儿)	12.81	16.51
74	蒙牛婴儿配方奶粉 (0~6 个月)	18.85	17.24
75	雅士利婴儿配方 II 奶粉 (0~6 个月)	33.74	12.32
77	摇篮能聪婴儿配方奶粉 (0~1 周岁)	17.90	10.80
80	二鹿因宝配方奶粉 (0~6 个月)	23.56	11.42

注：样品化学值单位：g/100g



应用近红外光谱分析技术在线检测牛奶成分的研究

作者：邹贤勇
学位授予单位：江苏大学

本文链接：http://d.g.wanfangdata.com.cn/Thesis_Y1994560.aspx