

# 化学计量学方法在烟草分析化学中的应用研究

## 中文摘要

化学计量学 (Chemometrics) 是 20 世纪 70 年代以后发展起来的一种新兴化学分支学科, 它应用数学和统计学方法, 借助计算机技术, 设计和选择最优的测量程序和实验方法, 并且通过解释化学数据而获得最大限度的化学信息。本论文将化学计量学方法引入烟草化学研究领域, 利用偏最小二乘法 (partial least squares, PLS) 技术对烟草样品的常规成分与近红外光谱之间的定量模型进行了研究, 提出了基于相关系数、小波变换以及无信息变量消除 (uninformative variable elimination, UVE) 的三种数据压缩和变量筛选方法, 提高了 PLS 的运算速度, 改善了 PLS 的定量预测精度。

偏最小二乘法 (PLS) 是近年发展起来的多元分析方法。本章采用 PLS 建立了近红外光谱与卷烟样品常规成份之间的关系模型, 该模型可用于卷烟样品常规成份的定量预测。为了进一步对光谱的波长进行了选择, 以消除与建模无关的波长, 提出了基于相关系数的波长选择方法, 结果表明该方法即可提高 PLS 的运算速度, 又可得到较高的预测精度。

小波变换在分析化学领域已得到广泛应用, 本章将小波变换 (WT) 用于烟草样品近红外光谱 (NIR) 的数据压缩, 并结合偏最小二乘法 (PLS) 对压缩后的数据进行建模, 建立了 WT-PLS 方法。与直接采用 PLS 相比, WT-PLS 可有效地压缩原始谱图的数据, 去除谱图中噪声和背景的干扰, 降低所建模型的随机性, 从而大大提高了运算速度, 并获得了较高的预测精度。

近红外光谱与多变量校正技术相结合是复杂植物样品定量分析的良好手段。本章提出了结合小波变换数据压缩和无信息变量消除 (UVE) 变量筛选的一种新型算法。首先将近红外光谱经过小波变换进行数据压缩, 然后采用 UVE 方法进一步去除与建模不相关的小波系数, 并最终利用压缩和筛选后的小波系数建立校

正模型。结果表明所建立的方法可大量压缩和去除与建模不相关的光谱数据，与 PLS 和 WT-PLS 相比，可进一步提高运算速度，并进一步改善定量预测的精度。

采用微波辐射对烟草(烟叶或烟丝)样品进行处理，并采用固相微萃取(SPME)-GC-MS、热解-GC-MS 以及电子显微镜技术研究了微波辐射对烟叶和烟丝样品的物理作用和化学作用。结果表明：微波辐射对烟气的化学成分有显著影响，使许多化学成分，特别是低沸点烃类化合物的含量明显降低；微波辐射对烟叶表面的物理结构也有较大影响，微波辐射处理后烟叶的表面组织变得疏松，可在一定程度上改善烟草制品的内在质量和吸食效果。

# Development and Applications of Chemometrical Methods in Tobacco Analytical Chemistry

## Abstract

Chemometrics is new branch discipline of chemistry since 1970s, which has been widely used in chemical studies. In this thesis, chemometrical methods related to building the quantitative model of near-infrared spectroscopy (NIR) and the routine components of tobacco samples using the partial least square (PLS) technique were developed, including the data compression and variable selection methods based on relationship coefficients, wavelet transform (WT) and uninformative variable elimination (UVE). By using these methods, both the computation speed and the prediction precision of the PLS model can be improved.

PLS was employed for building the model of near-infrared spectroscopy (NIR) and the routine components of tobacco samples, which can be used to predict the quantitative information of these routine components. A method of wavelength selection was also developed for near-infrared spectroscopy to minimize the interference of the irrelevant information for multivariate calibration. This procedure can improve both the calibration speed and the accuracy of the prediction.

Wavelet transform (WT) was applied to compress the near infrared spectra of tobacco samples, PLS was used to build the regression model on the compressed wavelet coefficients. Compared with the PLS algorithm, the WT-PLS can compress the original data into much less data points, and get rid of the interrupt of noise and background. Therefore, both the quality of the PLS model and the speed of the computation can be improved. As a result, higher predicting precision can be achieved.

An algorithm is proposed to extract the relevant information from near-infrared (NIR) spectra for multivariate calibration of routine components in tobacco samples. The algorithm is a combination of wavelet transform (WT) data compression and a procedure of uninformative variable elimination (UVE). After compression of the NIR spectra by WT, the UVE approach is used to eliminate the irrelevant wavelet coefficients. By the remaining wavelet coefficients, a calibration model is built finally for prediction. It is proven that the model based on the extracted relevant features is better than those obtained with full-spectrum data, because the irrelevant information can be removed from spectra for multivariate calibration. Furthermore, both the prediction precision and calculation speed can be improved.

A method for irradiation of tobacco samples with microwave was proposed. By using the SPME (solid-phase microextraction)-GC-MS, pyrolysis GC-MS and SEM (scanning electron microscope) techniques, the physical and chemical effects of the microwave irradiation on the tobacco samples were studied. It was shown that the chemical composition of the cigarette smoke was significantly changed after the microwave irradiation, especially those hydrocarbons with low boiling point. New chemical composition was also detected. On the other hand, effects of the microwave irradiation on the physical structure of the tobacco leaf sample were observed by SEM technique. From the pictures of the tobacco leaf surface, it is clear that the structure of the tissue was loosened than that of the comparison sample. Furthermore, it was also indicated by comparison smoking that the smoking quality can be slightly improved by the microwave irradiation.

## 致 谢

本论文是在导师邵学广教授的精心指导下完成的。邵老师在科研知识和技术上的指导令我收益非浅。从论文的选题，实验的设计，程序的编制到文章的撰写，都得到邵老师的精心指导帮助。邵老师在科研工作中所具有的敏锐洞察力、敏捷的思路，使我认识到一个优秀的科研工作者应具备的基本素质。邵老师对待科研工作认真负责，孜孜以求的态度给我留下了深刻的印象。这种勤勉治学的态度及脚踏实地的敬业精神永远值得我学习。

苏庆德教授、何友昭教授、吴守国老师、宛寿康老师、张良老师等在学习期间也给予我很大的帮助，使我收益良多，他们的学识和工作态度让我钦佩，在此表示感谢。

感谢本实验室陈达、王国庆、马超雄等同学的合作和帮助。

感谢中国科学技术大学的多年培养以及郑州烟草研究院的大力支持。

# 第一章 烟草分析化学与化学计量学方法的应用

## 1.1 烟草分析化学

烟草是一种叶用经济作物,其主要用途是将其叶片加工成各种各样的烟制品供人们吸食。同人类所有的吸食、饮用品一样,烟草的品质受到人们的广泛关注。烟草及其制品的品质主要由其内在化学成分组成及含量所决定。烟草化学成分的变化极为复杂,烟草类型不同,化学成分存在差异。烟草种植的生态环境、栽培措施、烟草成熟后的采收调制、发酵加工及陈化过程等都对作为卷烟原料的烟叶的化学成分产生影响。在烟制品的加工制造过程中,加工工艺、加香加料等各项技术加之烟叶原料的品质,决定了烟制品的质量。所有上述因素,所形成的化学成分及其含量的综合,通过人们的吸食过程——烟制品的燃烧过程,最终决定了烟气的化学构成,从而形成了人们对烟草的吸食需求和质量评价,也决定了烟草对人体健康的影响。

### 1.1.1 烟叶的主要化学成分

烟叶中的主要化学成分和其它高等植物一样,基本元素也是由碳、氢、氧、氮四种元素组成。这些元素在烟草中以不同比例、不同方式相互结合形成多种不同性质的化合物。烟叶的化学成分与烟草类型、栽培技术、调制方法和加工工艺有密切关系。因此,对烟叶化学成分的研究,在烟草工农业生产上都有重要的意义。例如,碳水化合物和含氮化合物在烟叶燃烧时,它们的热解产物的反应完全相反,糖呈酸性,含氮化合物呈碱性。因此,判断烟叶吃味的优劣,一般需要将总糖、总氮和烟碱的含量结合起来进行分析,它们之间要有适当的比例,也就是适度的酸碱平衡关系<sup>[1]</sup>。

#### 1. 碳水化合物

碳水化合物是由碳、氢、氧三种元素组成的多羟基醛或多羟基酮。它是烟叶中主要的一类化学成分。碳水化合物占调制后烟叶干物重量的 25%~50%, 含量

的多少随烟草的类型而异。烟叶中的化合物按性质、结构可分为单糖、低聚糖和高聚糖三类。总糖(又称水溶性糖)是影响吃味的主要化学成分之一,其含量的高低将对烟质产生不同的影响。

## 2. 含氮化合物

烟草中含氮化合物种类大体可分为蛋白质、游离氨基酸、生物碱、硝酸盐以及烟草中特有的亚硝酸胺等。烟草中含氮化合物对烟草的吃味和吸烟者的健康都有重要影响。

烟草植株从土壤中吸收各种无机含氮化合物(主要是铵盐和硝酸盐),在烟草根系或绿叶中,这些物质被还原成氨基-NH<sub>2</sub>,再经同化作用而形成各种氨基酸和蛋白质等有机含氮化合物。

## 3. 烟草生物碱

生物碱是一类存在于植物体内碱性的含氮有机化合物。这类化合物大部分含有氮杂环,对人体具有一定的生理作用,并具有弱碱性。烟草生物碱属于吡啶衍生物类。已鉴定的有 20 多种,从分子结构来看,主要有两类。一类是吡啶环与氢化吡咯环相结合,如烟碱、去甲基烟碱、去氢去甲基烟碱、二烯烟碱、去甲基二烯烟碱等;另一类是吡啶环与吡啶或氢化吡咯环相结合,如异烟碱、新烟碱、甲基新烟碱、去氢新烟碱、N-甲基去氢新烟碱、2,3-二吡啶等。

烟草中主要生物碱有烟碱、降烟碱、新烟碱和异烟碱,其中烟碱俗称尼古丁(Nicotine)在烟叶中占总生物碱含量的 95%以上。

## 4. 无机成分

烟叶中的无机成分主要包括水分和矿质元素两大类。由于水分在整个烟叶的生长发育、调制、流通、加工及消费全过程中都是一个重要因素,因此烟草中水分含量比较大。

烟草在生长过程中,不断地从周围环境中吸收二氧化碳和水分,同时不断地从土壤中吸收矿物质,以维持正常的生长和生理活动。这些被吸收的无机物就成为烟叶细胞的组成部分。在烟叶充分燃烧后,剩余的不能挥发的残烬物就是灰分。

矿质元素主要以金属氧化物、氯化物、难溶性碳酸盐、硫酸盐等形式存在于灰分之中。烟叶中主要有 K、Ca、Mg、Cl，其次为 P、S、Al、Fe、Mn、Si 等。

### 5. 其他组分

烟草中的化学成分，除了上述几类化合物之外，尚有多种其他组分，例如烃、醇、醛、酮、醚、酚、色素、萜类化合物，这些化合物对于烟叶的质量同样具有重要作用。

酚类是一类羟基直接与芳香环相连接的化合物，包括酚及其衍生物，它是由烟草体内的三大物质(糖类、脂类、蛋白质)之间的代谢形成的一类重要物质。烟叶中的酚类大多数与其它物质结合复杂分子的化合物，如色素、树脂等，也有以单独形式存在的各种酚或多酚类物质。酚类物质对烟叶香气、吃味、颜色等均有重要影响。

烟草色素属于天然色素中的植物色素，它是指在新鲜的烟叶和花中能用肉眼看到的有色物质或本来无色经化学变化而显色的物质。

### 1.1.2 烟草分析化学

烟草化学是最近几十年逐步发展形成的一门基础学科。20 世纪 50 年代以前，有关烟草化学与分析方面的问题主要通过农业化学、植物生理学和一般的化学理论与方法来研究解决，还未形成一门独立的基础学科。直到 20 世纪 60 年代初期，由于在世界范围内掀起“吸烟与健康”问题的论战，在时起时伏的禁烟浪潮冲击下，人们对烟草本质的揭示和烟气化学成分的研究，才达到一个空前的广度和深度。在整个烟叶生产的各个环节，都要通过化学分析作为理论基础和技术依据。烟草科学家和生物、医学等许多关心人类健康的科学家们，运用当代最新检测技术，分析检测出烟叶和烟气中的化学成分，形成了一门独立的基础学科。

据 1982 年 M. F. Dube 等人报道，烟叶中已鉴定的化学成分有 2549 种，烟气中有 3875 种，其中 1135 种为烟叶和烟气所共有，单独存在于烟叶中的为 1414 种，烟叶和烟气中已鉴定的化学组分总共有 5289 种。因此，通过研究烟草分析化学，可以对烟叶的质量进行检验，保证生产优良的产品；同时随着人们对生产过程的深入探讨，可借助于分析测定的数据，为烟制品生产部门开发、试制新的



烟制品、改进烟制品的工艺、改革生产过程的控制和管理以及提高卷烟安全性提供重要依据。

全面地揭示烟草和烟气的化学成分是一项浩繁的工作,需要借助现代化的检测手段,特别是色谱、光谱、质谱等现代分析技术。色谱(包括其联用技术)分析与光谱(特别是近红外光谱)分析是烟草化学中已经和正在获得成功应用的现代分析技术。借助于这些现代化的分析技术,烟草化学家获得了越来越多的各种条件下烟草及其制品中化学成分的组成与含量的信息。

### 1. 气相色谱在烟草化学研究中的应用

借助不同色谱技术的高效分离能力及与各种联用的检测技术,有可能对烟草这样的复杂样品进行分离分析,确定特定状况下烟草中的化学成分和含量。色谱法中的高效液相色谱(High performance liquid chromatography, HPLC)、气相色谱(Gas chromatography, GC)、毛细管电色谱(Capillary electrophoresis, CE)和裂解气相色谱(Pyrolysis gas chromatography, Py GC)等在烟草化学中均有成功的应用。

GC 在烟草化学中的应用已经有 40 多年的历史<sup>[2]</sup>,并且至今仍是活跃的研究领域。在样品处理技术、样品导入系统、二维色谱技术、检测器以及数据处理方法等方面均已开展了大量工作。

由于烟草样品组成的复杂性,在对特定的组分进行色谱分析之前,常需要进行不同的样品预处理。传统的水蒸汽蒸馏、溶剂萃取、索氏提取等方法,在提取某类目标化合物时也会有其它类型的化合物组被提取,对目标化合物分析产生干扰。

固相萃取(Solid phase extraction, SPE)是一种经济、省时、洁净,又具有较好重复性的样品处理技术,可以用于 GC 和 HPLC 的样品预处理,可用于许多不同类型样品的分析,如烟草中的糖类、植物碱的同时分析<sup>[3]</sup>。在溶剂萃取之后,通过洁净的硅酸镁载体(florisil)<sup>[4]</sup>或者与顶空进样结合的固相微萃取(Solid phase micro extraction, SPME)装置<sup>[5]</sup>,可以分析烟草中的脂肪酸和脂肪酯。采用 SPE-GC/MS 可以测定卷烟中的烟碱和其它微量的生物碱<sup>[6]</sup>。

样品须导入到色谱柱系统才可以完成其分离、检测及定性与定量处理。现代 GC 中的进样技术主要有顶空进样技术和液体样品进样技术。

顶空进样技术是一种特别适合于挥发性样品的进样技术,而烟草及其制品中含有大量的挥发性成分。已有文献综述了早期的顶空技术及其应用<sup>[7]</sup>,报道了利用半自动、全自动的动态顶空进样技术(吹扫捕集系统)用于水中挥发性有机物的测定<sup>[8]</sup>。并通过测试顶空气相成分评价不同类型的烟草质量。

对于液体样品,如烟草的溶剂提取物或其它液体样品,主要有以下几类进样方法:连续分流进样、不分流进样<sup>[9]</sup>、柱上进样(on-column)<sup>[10]</sup>和程序升温气化<sup>[11]</sup>。

虽然毛细管柱气相色谱具有很高的分离效率,但烟草化学家仍然面临着某些组分采用单一色谱柱不能有效分离的困难,例如在一个 GC/MS 峰中包含一个或多个的较小含量的次要成分,但这些被包裹着的微量成分则可能对烟草样品的芳香性或品味具有较为显著的影响。二维色谱可以缓解需要对单一柱分离时混合峰分离的压力,有文献比较了 3 种不同的毛细管色谱柱采用柱切换技术用于烟草香精油二维色谱的分析结果<sup>[12]</sup>。

热导池检测器(Thermo conductivity detector, TCD)和氢火焰离子化检测器(Flame ionization detector, FID)是 2 种较为常用的检测器。采用检测器的类型,取决于具体的应用、试验的设计、样品的复杂程度和使用的场合。GC 检测器可分为 3 种类型:离子化型、传导性和光谱性。

氮磷检测器(Nitrogen and phosphorous detector, NPD)是对含氮和磷的有机化合物具有很高响应的选择性检测器,可以成功地用于烟草生物碱的检测<sup>[13]</sup>。载气或燃烧气中正常的杂质,如果未经气体洗涤器去除,将会降低检测器的灵敏度和选择性。即使在理想状态下, NPD 的检测结果也可能不如 FID 检测的稳定性。

火焰光度检测器(Flame photometric detector, FPD)是对含磷、含硫的化合物有高选择性和高灵敏度的一种色谱检测器。NPD 可在香料分析中发挥重要作用<sup>[14]</sup>。文献<sup>[15]</sup>综述了光致电离检测器(Photoionization detector, PID)的原理及应用。电子捕获检测器(Electron capture detector, ECD)只对具有电负性的物质(如含有卤素、硫、磷、氮、氧的物质)有响应,可用于烟草农药残留的检测<sup>[16]</sup>。质谱检测器(Mass spectrometry detector, MSD): MSD 是目前重要的 GC 检测器,主要有离子阱、四极杆质谱,可以工作于总离子流或选择离子流模式。红外检测器(FT-IR detector): 作为检测器,它可以给出从色谱柱中流出物的 FT-IR 光谱信息,并且根据标准的 FT-IR 谱库进行定性,也可以与同时得到的 MS 谱图信息作相关性分

析。

数据处理系统除了能够给出传统的色谱数据积分处理报告外,结合数据库技术(如 MS 谱库和 FT-IR 谱库)的工作站软件还可以实现联机数据库检索,借助强大的 MS 或 FT-IR 的定性能力,可对未知色谱峰进行定性,使不借助标准对照品的色谱定性成为可能。根据美国分析仪器协会(Analytical Instrument Association, AIA)所制定的分析数据交换协议,已经有多个分析仪器制造商所采用的色谱和质谱分析数据交换协议(Analytical Data Interchange Protocols, Andi Protocols),采用 netCDF 格式进行数据存储(已接纳为 ASTM 标准),较好地实现了跨仪器平台的 GC/MS 数据共享或交换<sup>[17]</sup>。

## 2. HPLC 用于烟草和烟气的成分分析

HPLC 在烟草化学中的应用可以追溯到 20 世纪 70 年代。由于分离、定性、定量的自动化技术的提高,使得 HPLC 较之经典的柱色谱、纸色谱、薄层色谱等具有更高的分离效率和效果。

文献[18]综述了烟气中的多环芳烃的 HPLC 分离方法,文献[19]报道了利用 HPLC 和 GPC 进行烟气冷凝物成分分析。在特定烟草组分分析中,HPLC 发挥了极其重要的作用。

烟草中含有多种生物碱,称为烟草生物碱(Tobacco alkaloids),已经得到鉴定的就有 20 余种,包括吡啶与氢化吡啶环相结合的化合物(如烟碱、去甲基烟碱等)和吡啶环与吡啶或氢化吡啶环相结合的化合物(如新烟碱、甲基新烟碱等)。采用 HPLC 技术,可以测定烟草<sup>[20-22]</sup>、烟气<sup>[23]</sup>、烟叶中致敏物质提取物的分析<sup>[24]</sup>、尼古丁的对映体<sup>[25]</sup>。有大量的文献报道了不同样品类型的烟碱及其衍生物的分析。烟草中的生物碱是潜在的亚硝胺先导物<sup>[26]</sup>,烟草制品中亚硝胺的含量通常在 0.1~100 ppm,因此需要对样品进行富集处理。

文献[27]报道了 HPLC 用于分析烟叶加工过程中多胺成分的变化。离子色谱可以用于来自烟草中 DNA 水解产物的 DNA 碱基<sup>[28]</sup>分析。采用荧光检测器,反相液相色谱可以检测烟草中的氨基酸<sup>[32]</sup>。碳水化合物是烟草中大量存在的一类化合物,是构成植物的骨架和网络的结构物质。HPLC 在烟草中糖类物质的分析中也得到了广泛应用<sup>[30]</sup>。烟气中酚类和甲酚类化合物的直接 HPLC 测定<sup>[31]</sup>和对苯

二酚与苯磷二酚经显色衍生化反应后的检测<sup>[32]</sup>也有文献报道。

此外,利用 HPLC 在不同的色谱条件下还可以对色素<sup>[33]</sup>, 羰基化合物<sup>[34]</sup>, 环己烷提取物<sup>[35]</sup>, 离子型(无机离子和有机离子)化合物<sup>[36]</sup>, 添加剂(如甘草酸<sup>[37]</sup>、薄荷醇<sup>[38]</sup>、山梨酸酯<sup>[39]</sup>、保鲜剂<sup>[40]</sup>)等进行分离和检测。

### 3. 毛细管电泳色谱(CE)在烟草化学中的应用

CE 是以高压电场为驱动力,以毛细管为分离通道,依据样品中各组分间淌度和分配行为上的差异而实现分离、分析的一类技术,CE 与 HPLC 互为补充和完善。CE 的应用范围广,除分离生物大分子(肽、蛋白、DNA、糖等)外,还可用于小分子(氨基酸、药物等)及离子(无机、有机),甚至各种颗粒(如硅胶颗粒等)。近年来 CE 在烟草化学中的应用也日渐增多,Deng 等<sup>[41]</sup>报道了 CE 法分离检测烟草种子中醇溶性蛋白质,文献[42, 43]报道了 CE 用于烟草生物碱的检测方法。

### 4. 热解气相色谱(Py-GC)法在烟草化学中的应用

Py-GC 是在热裂解和气相色谱两种技术基础上发展起来的一种分析技术。借助 Py-GC 法可以考查一些难挥发物质(聚合物或分子)在一定温度下的热解行为,或考查不同组分间(裂解产物或预混合体系)的反应情况。

杨叶昆等<sup>[44]</sup>对烟草中 8 种主要非挥发性有机酸在不同温度下的热裂解行为进行了研究,讨论了它们对烟气品质的影响及 Py-GC 在烟气热解研究中的应用。

朱大恒等<sup>[45]</sup>综述了国内外有关烟气成分与烟叶主要化学成分关系的研究成果,结果表明:烟叶在燃吸过程中,烟气有害成分与烟叶化学成分之间存在密切关系,大部分烟气的有害成分是由烟叶中各种化学成分经高温裂解反应而形成的,某些烟气的有害成分是由烟叶中相应的有害成分直接进入烟气而形成。由于 Py-GC 的特点,它必将在烟气成分与烟叶或烟气化学成分之间关系的研究中发挥更大的作用。

Py-GC 可用来研究雪茄烟不同裂解阶段产物、烟碱在热解期间和燃烧的卷烟中的最终产物。有文献报道用同位素标记法辅助 Py-GC 分析,通过热解烟草添加剂,预测它们在卷烟过程中的行为<sup>[46]</sup>。用同位素标记法结合 Py-GC 可帮助阐明碳水化合物的热解机制<sup>[47]</sup>,在同位素气体存在下通过研究烟草燃烧期间 CO 的

形成解释其反应顺序<sup>[48]</sup>。

Py-GC 可用来评价烤烟总粒相物和烟草的凝聚能<sup>[49]</sup>，建立阴燃卷烟热解过程的数学模型<sup>[50]</sup>、燃烧和裂解中微粒形成的模型<sup>[96]</sup>等。

国内外对烟支或烟草中化学成分裂解产物的研究涉及到大量的物质，如多酚类化合物、胶质、纤维素、木质素、儿茶酚、固醇、糖酯和蔗糖、烟碱和生物碱、氨基酸和聚氨基酸、某些聚合物、绿原酸和芦丁、亚硝酸胺、潜香化合物和香味添加剂、烟草增香剂、农药残留等，王素方等对此进行了综述<sup>[51]</sup>。

### 5. 近红外光谱(NIR)技术在烟草化学中的应用

烟草化学分析在烟草科研中起着相当重要的作用，但常规的化学分析方法费时、费力、操作烦琐复杂。NIR 是一种无损分析技术，采用透射或漫反射方式可以直接对样品分析，样品处理简单(无需处理或仅需研磨成一定大小的颗粒以保证样品的均匀性)，已广泛用于食品、石油、化工、医药等行业，因而 NIR 也正在成为烟草化学成分分析的理想技术。

NIR 的波长范围是 780-2500 nm，通常又将这个范围划分为近红外短波区(780-1100 nm)和近红外长波区(1100-2500 nm)。近红外光谱源于化合物中含氢基团，如 C—H, O—H, N—H, S—H 等振动光谱的倍频及合频吸收，烟草中的总植物碱、总挥发碱、总糖、还原糖和总氮等都包含了这些基团。NIR 应用烟草工业的第一篇文献报道于 1968 年，对烟叶的透光率进行无损检测<sup>[52]</sup>。

NIR 分析不是一个直接测定方法，而是将未知样品测得的光谱通过训练集样品得到的校正模型来预测其组成或性质，因此训练集样品的组成或性质的适用范围、基础数据的准确性及选择化学计量学方法的合理性，都将直接影响最终的分析结果。NIR 是由无典型特征的、高度重叠的、宽的低吸收带组成，直到多变量校正技术发展以前，近红外光谱本身固有的这些缺点妨碍了该技术的广泛应用<sup>[53,54]</sup>。早期应用逐步多元线性回归进行化学成分<sup>[55]</sup>的预测，而现在 PLS 和主成分回归(PCR)是 NIR 技术中常用的两种化学计量学方法<sup>[56,57]</sup>，可以基于特征变量或对全部光谱区域进行计算而不采取预先的特征选择，但近来人们发现，有效的特征选择，可以极大地改善模型的预测能力和减少模型的复杂性<sup>[58,59,60]</sup>。

对 NIR 数据处理通常有基于波长、傅里叶变换(FT)系数和小波变换(WT)三

种形式。在烟草化学中, NIR 技术早期多限于单一类型物质的预测。

基于波长的 NIR 首次应用于烟草成分的测定是总还原糖的测定<sup>[61]</sup>, 建立了一个由 9 个参数组成的校正模型, 模型基于二次导数光谱, 所选择的波长点是 1458, 1476, 1516, 1790, 1914, 2028, 和 2178 nm。尼古丁一般占烟草中总植物碱的 98%以上, 早期研究表明<sup>[62]</sup>, 对于二次导数光谱, 预测尼古丁含量的较为可靠稳定的波长点是 2250, 2130, 2454, 2575 和 2144 nm。对于总氮的预测, NIR 原始光谱和二次导数光谱均可得到满意的结果<sup>[65]</sup>, 由 6~8 个参数组成的线性校正模型具有大致相同的预测精度。多酚含量的预测也可由 NIR 的二次导数光谱数据得到<sup>[63]</sup>, 其中绿原酸是主要的多酚物质, 占总多酚的 75%~95%。文献[64]报道了基于二次导数光谱的 7 参数模型用于烟支中焦油含量的测定。烟草样品中的钾等无机离子本身没有近红外吸收, 但其与一些有机基团结合组成的有机盐或其它类型的化合物的含量与近红外吸收之间具有一定的相关性, 因而有可能通过近红外光谱数据的分析, 测定或预测其含量<sup>[65]</sup>。

借助快速傅里叶变换(FFT)可以将 NIR 光谱转换为傅里叶变换(FT)系数也可以基于 FT 系数对样品数据进行处理, 并且可以减少多重共线性、校正颗粒大小不规则带来的影响。可以由 FT 系数预测特定组分的含量、计算 NIR 的二次导数光谱、缩短建模时间(相当于基于波长计算时间的 2%)<sup>[65]</sup>。

已有将小波变换(WT)应用 NIR 数据处理的文献报道, 借助 WT 可同时完成 NIR 数据的压缩、平滑、滤噪、求导<sup>[66-69]</sup>。

目前借助 PLS、PCR 等多元校正技术, NIR 技术可以同时测定烟草中的多种组分的含量, 应用于烟草化学成分、焦油和烟气的单一或多组分的同时测定<sup>[70-72]</sup>。

## 1.2 化学计量学及其在烟草化学研究中的应用

化学计量学(chemometrics)是 70 年代以后发展起来的一种新兴化学分支学科, 它应用数学和统计学方法, 借助计算机技术, 设计和选择最优的测量程序和实验方法, 并且通过解释化学数据而获得最大限度的化学信息。80 年代以来, 化学计量学已在许多领域中得到了广泛应用, 例如化学化工过程的过程分析与过程控制、考古科学中的出土文物鉴别、农林牧业产品的质量分析与质量控制等等。

近几年, 化学计量学用于中草药的成分分析与质量控制、农业植物样品的成

分分析等研究工作也相继出现。这些研究突破了传统成分分析的概念与方法，基于混合体系的化学测量信号与待测成分或质量参数的建模进行成分分析或质量控制，解决了传统方法无法解决的疑难问题。但在烟草行业中，化学计量学的研究工作刚刚起步，无论在国内还是在海外均没有系统的研究工作。

化学计量学方法在烟草化学中的研究开始于 20 世纪 80 年代，我国的烟草行业在 20 世纪 90 年代前后才开始有所报道。在烟草化学研究中常用的化学计量学方法可以总结如下。

### 1.2.1 聚类分析

聚类分析的基本原理是多维空间中同类对象之间的“距离”小而不同类的对象之间的“距离”大。尽管与许多其它化学计量学方法相比聚类分析的理论尚还不尽完善，但在许多实际问题中已发挥了重要作用。在聚类分析中，“距离”的概念有不同的定义，如明氏(Minkowski)距离、曼哈顿(Manhattan)距离、马氏(Mahalanobis)距离等。将对象按照一定的级别进行分级聚类是聚类分析的一种常用方法，分析对象根据它们之间的距离或相似性聚集到一起。最常用的聚集过程是将每个对象逐步聚集到一起，开始时每个对象作为一类，随着聚集的进展逐步形成更大的对象群体。

目前聚类分析是烟草行业中应用较多的一类化学计量学方法，研究工作主要集中在区域划分<sup>[73-78]</sup>、病虫害分析<sup>[79-86]</sup>和品种资源分析<sup>[87-89]</sup>等等。这些研究为种植区划、品种布局和引种提供了参考依据，为选种、育种提供了理论依据，对系统认识烤烟的品种资源具有重要意义，有利于在病虫害防治中采取针对性措施。

NIR 中定性的含义与 IR 进行结构鉴定不同，主要是判断所分析的样品是否适合所建立的校正模型，或确定样品的类别，严格地讲是一种归类判别或分类，目前常用的方法是 PCA 结合马氏距离的方法。农副产品及食品的分析是 NIR 最大的应用领域，除了可以对烟草中的总植物碱、总氮、总糖、总还原糖、无机离子等进行建模预测外，NIR 还可用于烟草的分类<sup>[90]</sup>。此外，蔡昆争等<sup>[91]</sup>采用平均聚类法、模糊聚类法、主成分分析法对广东不同地区土壤养分含量进行聚类分析，此种作法对烟草种植栽培具有一定的借鉴意义。王鑫等<sup>[92]</sup>对龙岩烟区植烟土壤的养分进行了聚类分析。Liu G 等<sup>[93]</sup>用聚类分析找出烟草酸性成份与烟叶颜色

的关系。Peral Fernandez J.L.<sup>[94]</sup>用聚类分析法讨论、比较了烟草样品无机化方法。

### 1.2.2 回归分析

回归分析是使用最广泛的数理统计方法之一，通常可分为一元回归、多元回归以及基于因子分析的主成分回归等。

一元回归用于建立一个自变量( $x$ )和一个因变量( $y$ )之间的数学模型，一般采用最小二乘法得到一元回归方程 ( $y = ax + b$ )。一元回归在烟草研究的应用较为广泛，例如用来考察烟叶中烟碱与烟气中烟碱、总氮与蛋白质的关系。多元回归是一元回归的推广，用于建立多个自变量和多个因变量之间的关系。

王允白<sup>[95]</sup>分析了 60 种烤烟样品中的 23 种化学指标，利用烟叶内在化学成分与 TPM (烟气总粒相物)的对应关系建立多元回归方程来预测卷烟的 TPM 和烟碱含量。Tso T.C.<sup>[96]</sup>运用多元回归方法建立了烟叶特征与化学成份之间的关系，结果表明烟草中很多成份与外形、品质不成线性关系，在使用多元回归时要慎重选择多元回归模型<sup>[97]</sup>。

如果自变量之间存在相关关系，在回归分析中这些因子不仅增加计算的工作量，而且会增加回归方程的不稳定性。此时，就需要在回归分析之前进行变量的筛选。逐步回归是较早用于变量筛选的方法之一，它在计算的每一步中都对回归方程中所含的因子逐个进行检验，最终使回归方程包含一切对因变量作用显著的因子，排除对自变量作用不显著的因子。逐步回归有较强的实用性，对结果解释比较直观，在农业上用于分析气象状况、施肥情况、化学物质等因素对病虫害的影响<sup>[98-101]</sup>以及施肥与烟叶生长之间的关系<sup>[102]</sup>。在工业上用于分析化学成份对总粒相物的关系<sup>[103,104]</sup>以及评估化学成份对卷烟品质的影响<sup>[105,106]</sup>。

### 1.2.3 偏最小二乘法

偏最小二乘法(PLS)是建立在主成分分析和主成分回归基础上发展起来的一种化学计量学方法。随着近红外光谱法的发展，偏最小二乘法在分析化学，特别是在工业分析化学中的应用越来越广泛。目前在烟草行业的应用还不多，但发展很快。

PLS 可用于建立近红外光谱与烟草中化学成分间的关系模型，运用建好的数



学模型根据近红外光谱可快速测定烟草中的化学成份。各个仪器生产厂家的近红外软件包中都含有偏最小二乘法。盛良全等<sup>[107]</sup>将 PLS 应用于直接测定烟叶样品中微量的  $\text{Cu}^{2+}$ 、 $\text{Pb}^{2+}$ 、 $\text{Cd}^{2+}$ 、 $\text{Co}^{2+}$ ，解决了四种离子的络合物光谱重叠难以检测的问题。Lebeau B.<sup>[108]</sup> 根据烟草中易挥发成份，用偏最小二乘法计算烟草配方，得到了较好的效果。Li H.<sup>[109]</sup>用偏最小二乘法处理间隔扫描吸收数据来计算烟草中烟碱和苯的浓度。

#### 1.2.4 主成分和因子分析

因子分析(FA)是多元统计分析技术，其核心为将原变量表达为新变量的线性组合，可以压缩数据、提高数据质量，建立物理模型、预测未知数据点，快速解决各种类型的复杂问题。主成分分析(PCA)是因子分析的基础，其中心目的是将数据降维以排除化学信息中相互重叠的信息。

Karen Rossiter<sup>[110]</sup> 用 PCA 分析了酯类的水果香气与其结构参数之间的关系。Hayato Hasebe<sup>[111]</sup>提出了利用烟草的顶空气相成份的主成分分析可将烟草不同类型区分开，并可将同类烟叶不同质量的烟叶分开。

#### 1.2.5 人工神经网络

人工神经网络(artificial neural network, ANN)，有时简称为神经网络，是在现代生物学研究人脑组织所取得成果的基础上提出的，它利用大量简单的处理单元广泛连接组成的复杂网络，来模拟人类大脑的神经网络结构和行为。它的研究成果显示了人工神经网络具有人脑功能的基本特征：学习、记忆、概括、归纳和抽取等，从而解决了人工智能研究中的某些局限性。它不同于以前人工智能领域中普遍采用的基于逻辑和符号处理的理论和方法，开辟了崭新的途径。

人工神经网络在烟草中的应用涉及到与烟草质量相关的各个方面，将其与人工嗅觉系统相结合<sup>[112-114]</sup>用于评价卷烟的内在品质，根据已知模型识别各种卷烟。克服了人受环境、状态等因素的影响，能够更客观的反应卷烟的内在品质。人工神经网络与视觉系统结合使用可用于烟叶分级和质量分选<sup>[115,116]</sup>。人工神经网络在烟草中另一个重要应用是质量分析<sup>[117-119]</sup>，根据烟草中的化学成分数据，运用人工神经网络建立模型预测出卷烟内在质量指标。在烟叶发酵<sup>[120]</sup>和烘烤<sup>[121]</sup>中运

用神经网络与模糊控制相结合，解决工艺过程中的难点和实现过程的自动控制。此外，人工神经网络在烟草设备的故障诊断上也显示了一定的优越性<sup>[122,123]</sup>。虽然很多工作在数据采集和实用性上还有待进一步完善，但为神经网络在烟草中的应用提供了很多有益的思路。

### 1.2.6 化学计量学方法用于烟草的质量评价

为了判断烟叶质量，需对烟叶进行烟气感官评吸，但感官评吸带有主观性，且很难定量，Hasebe 等<sup>[124]</sup>通过烟叶顶空分析，获得评价烟叶质量的分值，采用多元回归分析建立起定量评价烟叶质量的方法。陈建忠<sup>[125]</sup>应用数理统计方法探讨三峡库区烟叶化学特性在全国烤烟中的地位。王允白等<sup>[126]</sup>通过对 218 份晒红烟烟叶样品化学成分与评吸鉴定数据的简单相关分析、化学成分与评吸结果各项的逐步回归分析和化学成分、化学成分比值及其派生值与评吸结果的典型相关分析研究了晒红烟化学成分与评吸结果间关系研究。White 等<sup>[127]</sup>研究了裂解温度对烟气冷凝物诱变性的影响。

## 参考文献

- [1] 金闻博, 戴亚, 横田拓等. 烟草化学, 2000, 53-59.
- [2] Shaw WG, Stephens RL, Weybrew J. *Tob Sci*, 1960, 4: 179-181.
- [3] Švob Troje Z, Fröbe Z, Perovic D. *J Chromatogr A*, 1997, 775(1): 101-107.
- [4] Wang MX, Moldoveanu SC, Lauterbach JH, Weeks WW, *Proceedings of the 48th Tobacco Chemists Research Conference*, Greensboro, NC, September 1994.
- [5] Clark TJ, Bunch J, *J Chromatogr Sci*, 1997, 35: 209.
- [6] Wu WJ, Ashley DL, Watson CH. *Anal Chem*, 2002, 74: 4878-4884.
- [7] McNally ME, Grob RL. *Amer Lab*, 1985, 17(2): 106-114.
- [8] Sellar TA, Lichtenberg JJ. The volatile organic compounds at ug/L level in water by gas chromatography. National Environmental Center, Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH 45268 (EPA-670/4-74-009).
- [9] Arrendale RF, Severson RF, Chortyc OT. *Beitr Tabakforsch*, 1984, 12: 186-197.
- [10] Freeman RR, Augenblick KB, Philip RJ. On-column injection with open tubular columns. Technical Paper No 88 (1980). Hewlett-Packard Company, Avondale, PA 19311.
- [11] Poy F, Visani S, Terrosi F. *J Chromatogr*, 1981, 217: 81-90.
- [12] Gordon BM, Rix CE, Borgerding MF. *J Chromatogr Sci*, 1985, 23: 1-10.
- [13] Severson RF, McDuffie KL, Arrednale RF, et al. *J Chromatogr*, 1981, 211: 111-121.
- [14] Parliament TH, Spencer MD. *J Chromatogr Sci*, 1981, 19: 435-438.
- [15] Davenport JN, Adlard ER. *J Chromatogr*, 1984, 290: 13-32.
- [16] Cai JB, Liu, BZ, Zhu, XL, Su, QD. *J Chromatogr*, 2002, 964(1-2): 205-211.
- [17] Nelson DC, Dauciunas V, Lysakowski R, Duff M. *American Laboratory*, 1992, 24 (14):52.
- [18] Haebeler AF, Chortyk DT. High pressure liquid chromatography of polynuclear aromatic hydrocarbon constituents. In *Recent Advances in Tobacco Science*, 1:

- 72-96, 29th Tob Chem Res Conf, Beltsville – College Park Md (1975).
- [19] Bell JH, Liquid and gel permeation chromatography. In Recent Advances in Tobacco Science, 1: 51-71, 29th Tob Chem Res Conf, Beltsville – College Park Md (1975).
- [20] Makarinec MP, Manning DL, Oldham P. *J Liq Chrom*, 1981, 4: 31-39.
- [21] Meyer R, Wegner KG. *Anal Biochem*, 1985, 148: 269-276.
- [22] Watson ID, *J Chromatogr*, 1977, 143: 203-206.
- [23] Green CR, Conrad FW, Bridle KA, Borgerding MF. *Beitr Tabakforsch*, 1985, 13: 11-16.
- [24] Sudan BJ, Brouillard LC, Strehler C, et al. *J Chromatogr*, 1984, 288: 415-422.
- [25] Cundy KC, Crooks PA. *J Chromatogr*, 1983, 281: 17-33.
- [26] Hoffman D, Lavoie EJ, Hechl SS. *Cancer Letters*, 1985, 26: 67-75.
- [27] Madsen JP, Bush LP, Gay SL. *J Agric Food Chem*, 1985, 33: 1182-1185.
- [28] Citti L, Gervasi PG, Turchi G, et al. *J Chromatogr*, 1983, 261: 315-319.
- [29] Crouse WE, Bell JH. The determination of free amino acids on tobacco. 39th Tobacco Chemists' Research Conference, Montreal, Quebec. 1985.
- [30] Blaschek W. *J Chromatogr*, 1983, 256: 157-163.
- [31] Jeanty G, Massé J, Bersot P, et al. *Beitr Tabakforsch*, 1984, 79: 1026-1032.
- [32] Morée-Tesla P. *Ann de Tabac SEITA*, 1981, 18: 67-74.
- [33] Maynor HW, Lauterbach JW. Automated ion chromatography method for the determination of nitrate other organic acid anions in cigarette paper. 38th Tobacco Chemists' Research Conference, Atlanta, Georgia, 1984.
- [34] Perini FR, Bell JH. Determination of 5-hydroxymethylfurfura in tobacco as a measure of browning and quality. 37th Tobacco Chemists' Research Conference, Washington DC, USA, 1983.
- [35] Wahlberg I, Aiodt R, Wallin I, et al. *Acta Chem Scand B*, 1984, 38: 21-30.
- [36] Harvey WR, Baker PG. The analysis of tobacco leaf, filter and cigarette paper by ion chromatography. 38th Tobacco Chemists' Research Conference, Atlanta, Georgia, 1984.
- [37] Bell JH. *Tob Sci*, 1980, 24: 126-129.

- [38] Core MT, Logue M, Haut SA. Separation of menthols and menthones by normal phase high performance liquid chromatography, 34th Tobacco Chemists' Research Conference, Richard, Va, 1980.
- [39] Brunnemann KD, Posset L. *J Chromatogr*, 1980, 202: 494-499.
- [40] Battard JC. *Ann de Tabac SEITA*, 1981, 18: 75-82.
- [41] Deng B, Lu JP, Yan JH.. *J Chromatogr A*, 1998, 203: 321-327.
- [42] Suresh R. *J Chromatogr B*, 1997, 695: 117-129.
- [42] Frank-Micheal M. *J Chromatogr A*, 1999, 853: 27-34.
- [43] Lochmann H, Bazzanella A, Kropsch S, Bächmann K. *J Chromatogr A*, 2001, 917:311-317.
- [44] 杨叶昆; 陈章玉; 罗丽莉. 烟草科学研究, 2000, (2): 48-50.
- [45] 朱大恒, 李彩霞, 张爱忠, 宋润斌, 余健. 烟草科技, 1999, (4): 25-27.
- [46] Stotesbury SJ. *Bei Zur Tab Int*, 2000, 19(2): 55-64.
- [47] Paine III JB, Naworal J, Thomas CE. Isotopic labeling to elucidate pyrolysis mechanisms. *CORESTA*, 2001.
- [48] Bakker RR. Formation of carbon oxides during tobacco combustion. Pyrolysis studies in the presence of isotopic gases to elucidate reaction sequence. *CORESTA*, 1983, (4): 139-145.
- [49] Albo JP. Evaluation of tobacco condensate potential through pyrolysis. *TA*, 1979, (4): 243-247.
- [50] Chen P. A mathematical model of cigarette smoldering process. *CORESTA*, 2001.
- [51] 王素芳, 苏庆德. 中国烟草学会 2002 年学术年会论文集, 广州. 2002.12: 784-787.
- [52] McClure WF. *Tobacco Science*, 1968, 12: 232-235.
- [53] Thomas EV, Haaland DM. *Anal Chem.* , 1990, 62: 1091-1099
- [54] Geladi P, Kowalski BR. *Anal. Chim. Acta* , 1986, 185: 1-17
- [55] Ben-Gera, Norris KH. *J Food Sci*, 1968, 33: 64-67.
- [56] Thomas EV, Haaland DM. *Anal Chem.* , 1990, 62: 1091-1099
- [57] Geladi P, Kowalski BR. *Anal. Chim. Acta*, 1986, 185: 1-17

- [58] Leardi R, Lupiáñez González A. *Chemometrics Intell. Lab. Syst. [J]*, 1998, 41: 195-207
- [59] Leardi R. *J Chemometrics*, 2000, 14: 643-655
- [60] 蔡文生, 邵学广, 赵贵文, 张懋森. *分析化学*, 1997, 25(2): 231-237
- [61] McClure WF, Norria KH, Weeks WW. *Beitr Tabakforsch*, 1977, 9: 13-17.
- [62] Hamid A, McClure WF, Weeks WW. *Beitr Tabakforsch*, 1978, 9: 267-274.
- [63] Willimson RE, McClure WF. *Beitr Tabakforsch*, 1982, 11(6): 219-227.
- [64] Willimson RE, Chaplin JF, McClure WF. Near-infrared spectrophotometry of tobacco leaf for tar yield of smoke. 40th tobacco chemists' research conference (1986, Knoxville, Tennessee).
- [65] McClure WF. Status of near-infrared technology in the tobacco industry. *Advances in the analytical methodology of leaf and smoke, 40th tobacco chemists' research conference* (1986, Knoxville, Tennessee): 3-53
- [66] SHAO Xue-Guang, LEUNG A K-M, CHAU, FT. *Acc. Chem. Res.* 2003, 36(4): 276-283.
- [67] Trygg J, Wold S. *Chemometrics and Intelligent Laboratory Systems*, 1998, 42: 209-220.
- [68] 邵学广, 庞春艳, 孙莉. *化学进展*, 2000, 12(3): 233-244.
- [69] 邵学广, 庞春艳. *计算机与应用化学*, 2000, 17(3): 247-250.
- [70] 王东丹, 李天飞, 吴玉萍等. *云南大学学报(自然科学版)*, 2001, 23(2): 135-137.
- [71] 王芳, 陈达, 邵学广. *烟草科技*, 2002, (5): 23-26
- [72] Tomas Bebese, Carlos Gastellu. *Tobacco Science*, 1983, 27: 92-94
- [73] 王立德, 李扬立. *安徽烟草科技*, 1993, 2: 10
- [74] 陶方玲. *中国烟草昆虫研究-理论与实践(一)*: 5
- [75] 钱时祥. *安徽农业大学学报*, 1994, 1: 21
- [76] 陈建忠. *四川农业大学学报*, 1999, 1: 50
- [77] JOHNSON P.H.S. The use of cluster analysis in flue-cured tobacco farm management in zimbabwe. *CORESTA*, 1984, 1: 121
- [78] JOHNSON P.H.S. Cluster analysis of tobacco variables: Its application to the

- tobacco soils of the TSUNGWESI Intensive Conservation Area.  
TA,1983,7-8:425
- [79] 陈家骅. 河南农业大学学报,1990,4:428
- [80] 高正良. 植物保护,1998,5:18
- [81] 张应球. 湖南烟草,1996,2:32
- [82] 陈伟群; 张天宇. 中国烟草学报,1997,3:30
- [83] 巫厚长; 孟庆雷. 安徽农业大学学报,1999,1:68
- [84] 喻盛甫等. 中国农业科学,1999,5:55
- [85] 王志德等. 中国烟草科学,2000,1:33
- [86] 张修国等. 菌物系统,2000,1:39
- [87] 牛佩兰; 王元英. 中国烟草科学,1991,2:22
- [88] 何川生等. 中国烟草学报,2000,2:21
- [89] ANN D.J.; KIM Y.D. Varietal classification on the basis of cluster analysis in  
local tobacco.CORESTA,1982,3-4:74
- [90] 汪尔康. 21 世纪的分析化学. 北京: 科学出版社, 1999. 75-91.
- [91] 蔡昆争, 段舜山, 陈荣均. 烟草科技, 1999, (4): 90-93.
- [92] 王鑫, 上官克攀, 邱标仁, 奚振邦, 徐四新. 烟草科技, 2001, (10): 43-45
- [93] Liu G; Shi H. The relationship of acidic components to leaf colours of oriental  
tobacco by fuzzy cluster analysis.54th tobacco science research conference,  
1996
- [94] Peral Fernandez J.L.. Comparison of mineralization methods for tobacco samples  
by use of cluster analysis. Anal Abs,1988,50(5):5G4
- [95] 王允白. 中国烟草学报,1998,2(4):1
- [96] Tso T.C. TA,1978,2:110
- [97] Van Der Merwe A.J. TA, 1984,5-6:184
- [98] 王鑫; 沈焕梅.中国烟草科学, 1999, 2:27
- [99] 杨录明等. 烟草科学研究, 2001, 1:47
- [100] 邓建华等. 西南农业大学学报, 2000, 5:448
- [101] 孟国玲等. 湖北农学院学报, 2000, 1:16

- [102] 刘好宝. 中国烟草学报, 1998, 1:60
- [103] 郭家明. 合肥工业大学学报, 1997, 5:121
- [104] 王允白. 烟草研究与管理, 1997, 4:19
- [105] 闫克玉等. 烟草科技, 2001, 10:5
- [106] 王允白;王宝华. 中国烟草科学, 1997, 1:17
- [107] 盛良全. 烟草科技, 1996:5:19
- [108] Lebeau B.; Hammers W.E. J Chromatoger, 1991, 596(2):285
- [109] Li H.; Sun X.Q. Anal Abs, 1993, 55(9):9E148
- [110] Karen Rossiter. Perfumer & Flavorist, 1996, 21:33
- [111] Hayato Hasebe.T. Bei Zur Tabak Int, 1999, 5(18):213
- [112] 殷勇等. 农业机械学报, 1999, 6:88
- [113] 邹小波, 方如明, 蔡健荣. 江苏理工大学学报, 2000, 3:1
- [114] 高大启, 吴守一. 农业机械学报, 1998, 4:167
- [115] 张建平. 农业工程学报, 1997, 4:179
- [116] 蔡健荣, 方如明, 张建平, 张世庆. 农业工程学报, 2000, 3:118
- [117] 高大启, 吴守一. 农业机械学报, 1999, 1:58
- [118] 殷勇, 吴守一. 农业机械学报, 1999, 3:71
- [119] 高大启, 吴守一. 分析化学, 1998, 10:1174
- [120] 刘文定, 谢克明. 电力学报, 1998, 5:25
- [121] 刘兆魁, 吴建胜. 农业机械学报, 1998, 2:151
- [122] 吴和平. 机械, 1996, 3:19
- [123] 骆德汉, 郎文辉. 烟草科技, 1999, 5:25
- [124] 洗可法(译). 测试顶空气相成分评价不同类型烟草质量. 烟草科技. 2001, (3): 27-33.
- [125] 陈建忠. 四川农业学报, 1999, (1): 50-54.
- [126] 王允白, 王宝华. 中国烟草学报, 1997, (1): 17-20.
- [127] White J.L., Conner B.T., Perfetti T.A., et al. *Food and Chemical Toxicology*, 2001, 39: 499-505.



## 第二章 相关系数用于近红外光谱数据的变量筛选

### 2.1 引言

利用近红外光谱对烟草样品进行定量分析是目前国际烟草界的一个热门课题<sup>[2-4]</sup>。近红外光谱区的信息量远远大于紫外光谱区，但近红外光谱严重重叠，难以精确判定近红外谱图的归属。同时，分析仪器只能提供矢量型数据，采用常规的数据处理方法难以确定谱图与物质浓度之间的关系。因此，我们引入化学计量学方法用于近红外光谱的数据处理。由于化学计量学应用数学和统计学方法并借助计算机技术，可用于矢量型或矩阵型化学数据处理，并得到常规方法无法获得的结果，在化学各领域中已发挥了重要作用<sup>[5]</sup>。

在化学计量学中，常用的建模方法有偏最小二乘法(PLS)、逐步回归法(SRA)、主成分回归法(PCR)等<sup>[5]</sup>，其中 PLS 的计算方法简单且预测精度较高。我们采用 PLS 对近红外光谱与烟草样品的常规成分进行了相关建模研究，同时提出了一种波长选择方法，该方法可删除与建模无关的波长，既提高了运算速度，又提高了模型的预测精度。

### 2.2 原理与方法

#### 2.2.1 偏最小二乘法

PLS 是主成分回归(PCR)的发展。在 PCR 中，只对光谱矩阵  $X$  进行主成分分析，以抽象因子进行回归，但对浓度矩阵  $Y$  未加任何处理。事实上， $Y$  矩阵中也包含了非有用的信息，也应做同样的处理。PLS 就是基于这种思想，同时对  $X$  和  $Y$  矩阵进行主成分分解，并以主因子进行回归，进一步提高了方法的可靠性。

先对  $X$ 、 $Y$  矩阵进行分解，得：

$$X = TP + E \quad (2-1)$$

$$Y = UQ + F \quad (2-2)$$

式中， $T$  和  $U$  分别为  $X$  矩阵和  $Y$  矩阵的得分矩阵， $P$  和  $Q$  分别为  $X$  矩阵和  $Y$  矩阵的载荷(即主成分)矩阵， $E$  和  $F$  分别为用 PLS 模型拟合  $X$  和  $Y$  时所引进的误

差。

然后，将  $T$  和  $U$  作线性回归，计算关联系数矩阵  $B$ ，得：

$$U = TB \quad (2-3)$$

$$B = T'U(T'T)^{-1} \quad (2-4)$$

公式(2-4)中  $T'$ 表示矩阵  $T$  的转置。

最后，根据公式(2-1)得到的  $P_{校正}$ 和未知样品的光谱矩阵  $X_{未知}$ 求出未知样品  $X$ 矩阵的  $T_{未知}$ ，从而计算出未知样品的浓度  $Y_{未知}$ 。即：

$$Y_{未知} = T_{未知} BQ \quad (2-5)$$

有关 PLS 的详细算法步骤可参考有关文献<sup>[5]</sup>。计算程序采用 Matlab 语言编写，于 PentiumIV 计算机上运行。

## 2.2.2 波长选择

采用全谱计算时，不仅计算量大，而且在某些光谱区域样品的光谱信息很弱，或与样品的组成和性质间缺乏相关关系，会造成模型的精度降低甚至错误。因此，在光谱分析中，波长点的选择非常重要<sup>[6,7]</sup>。

为了找出对建模最有效的光谱区域，本文将化学指标的浓度值与样品的光谱数据进行关联，求出浓度与波长的相关系数  $R$ 。取相关系数较大的波长点用来建立模型。

对于向量  $x$  和  $y$ ，其相关系数  $R$  计算如下：

$$R = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-6)$$

其中， $\bar{x}$ 、 $\bar{y}$ 分别为  $x$ 、 $y$  的均值。

本文采用预置的相关系数阈值，自动选择那些相关系数大于阈值的波长点建立校正模型。阈值根据所建模型在新的预测集上的均方根误差(root of mean squared error of prediction,  $RMSEP$ )值来确定。取不同的阈值对波长进行选择并建模，可得到相应的  $RMSEP$  值，对应于最小  $RMSEP$  值的阈值即为所求。 $RMSEP$  计算如下：

$$RMSEP = \left( \sum_{i=1}^n (y_i - \hat{y}_i)^2 / n \right)^{1/2} \quad (2-7)$$

其中,  $y_i$  为第  $i$  个样本的实测值,  $\hat{y}_i$  为第  $i$  个样本的估计值,  $n$  为预测集的样本数。

## 2.3 实验

### 2.3.1 样品来源

共收集国内烤烟型卷烟产品和各种不同等级的烤烟烟叶约 500 个样品, 以尽可能大的覆盖各种烟草样品的常规化学成分(总糖、总植物碱、总氮、氯和蛋白质)的不同含量。

### 2.3.2 仪器与试剂

PERKIN ELMER Spectrum Identichack FT-NIR System (波数:  $4000\text{cm}^{-1}$  ~  $9000\text{cm}^{-1}$ ) 和各分析方法中所列仪器。

### 2.3.3 分析方法

#### (一) 试样的制备和水分测定

##### 1 原理

将烟草或烟草制品经过低温烘干, 研磨成一定粒度的烟末, 在一定条件下烘干试料, 由烘干前后的重量差求出试样的水分含量。

##### 2 仪器设备

烘箱(LC-223, 上海爱斯佩克);

粉碎机;

筛网孔径 0.45mm;

##### 3 结果的计算

试样的水分重量百分含量, 由下式得出:

$$\text{水分 \%} = \frac{m_1 - m_2}{m_1 - m_0} \times 100 \quad (2-8)$$

式中:  $m_0$  — 称量皿重量, g;

$m_1$  — 烘干前称量皿重量, g;

$m_2$  — 烘干后称量皿重量, g。

## (二) 水溶性糖含量的测定

### 1 原理

用 5%醋酸水溶液萃取烟草样品, 萃取液中的糖(水溶性总糖测定时应水解)与对羟基苯甲酸酰肼反应, 在 85℃的碱性介质中产生一黄色的偶氮化合物, 其最大吸收波长为 410nm, 用比色计测定。

### 2 试剂

Brij35 溶液(聚乙氧月桂醚);

0.5mol·L<sup>-1</sup> 氢氧化钠溶液;

0.008mol·L<sup>-1</sup> 氯化钙溶液; 5%醋酸溶液;

0.5mol·L<sup>-1</sup> HCl 溶液; 1.0mol·L<sup>-1</sup> HCl 溶液;

1.0mol·L<sup>-1</sup> NaOH 溶液;

5%对羟基苯甲酸酰肼溶液(HOC<sub>6</sub>H<sub>4</sub>CONHNH<sub>2</sub>);

D-葡萄糖。

### 3 仪器设备

连续流动分析仪(AA3, 德国 BRAN+LUBBE);

天平, 感量 0.0001g(AE163, 瑞士 Mettler);

振荡器。

### 4 水溶性糖的计算

以干基计的水溶性糖的含量, 以葡萄糖计, 由下式得出:

$$\text{总(还原)糖 \%} = \frac{c \times v}{m \times (1 - w)} \times 100 \quad (2-9)$$

式中： $c$  — 样品液总(还原)糖的仪器观测值，mg/mL；  
 $v$  — 萃取液的体积，mL；  
 $m$  — 试料的重量，mg；  
 $w$  — 试样的水分含量。

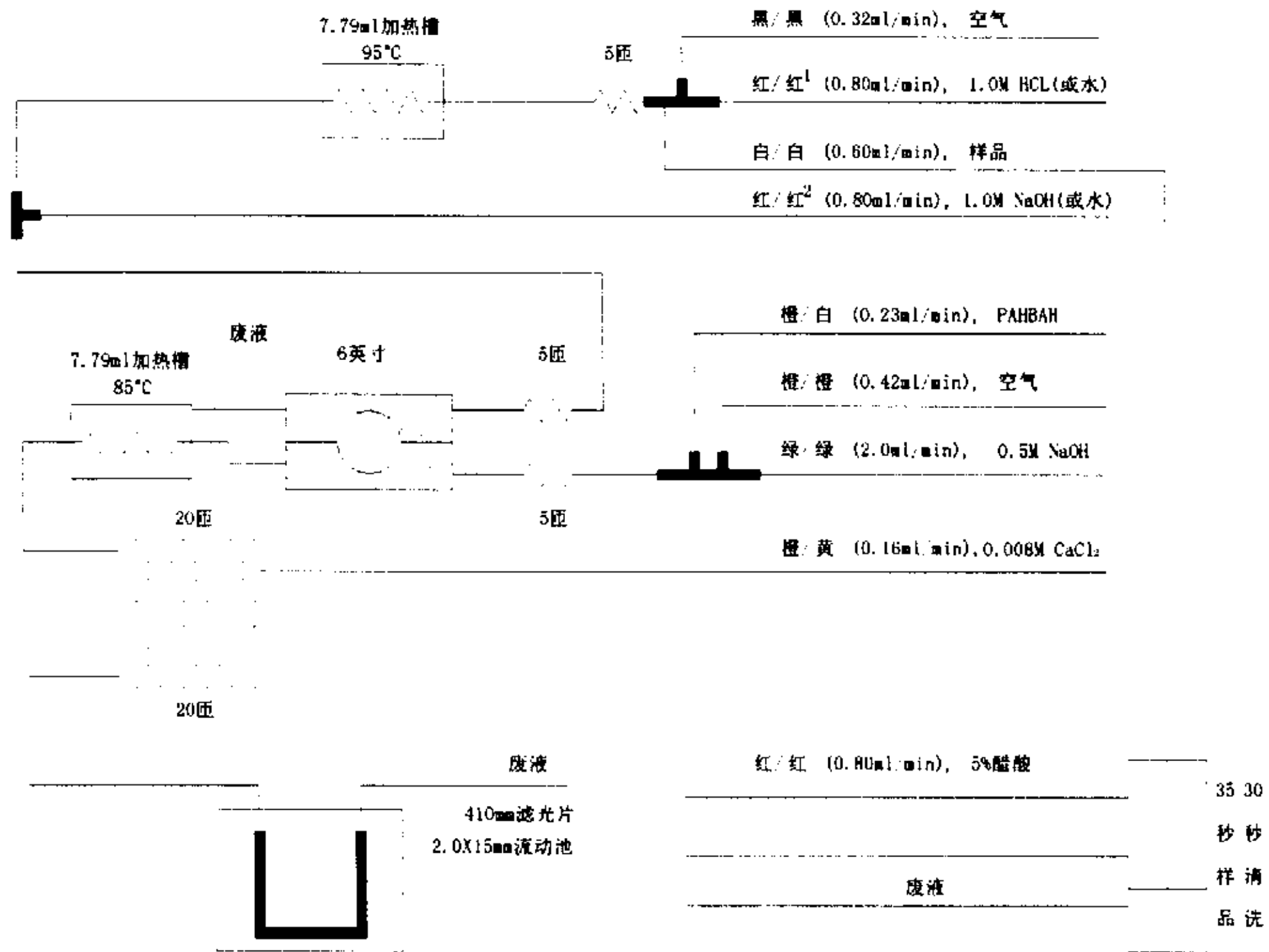


图 2-1 水溶性糖测定管路图

### (三) 总植物碱含量的测定

#### 1 原理

用水萃取烟草样品，萃取液中的总植物碱(以烟碱计)与对氨基苯磺酸和氯化氰反应，氯化氰由氰化钾和氯胺 T 在线反应产生。反应产物用比色计在 460nm 测定。

#### 2 试剂

Brij35 溶液(聚乙氧月桂醚)；

缓冲溶液 A: 2.35g 氯化钠(NaCl)和 7.60g 硼酸钠( $\text{Na}_2\text{B}_4\text{O}_3 \cdot 10\text{H}_2\text{O}$ )溶于 1L 水中；

缓冲溶液 B：26g 磷酸氢二钠 ( $\text{Na}_2\text{HPO}_4$ )，10.4g 柠檬酸 [ $\text{COH}(\text{COOH})(\text{CH}_2\text{COOH})_2 \cdot \text{H}_2\text{O}$ ]和 7g 对氨基苯磺酸( $\text{NH}_2\text{C}_6\text{H}_4\text{SO}_3\text{H}$ ) 溶于 1L 水中；

2% 氯胺 T 溶液；

0.4% 氰化钾溶液；

纯烟碱。

### 3 仪器设备

连续流动分析仪(AA3, 德国 BRAN+LUBBE)；

天平，感量 0.0001g(AE163, 瑞士 Mettler)；

振荡器。

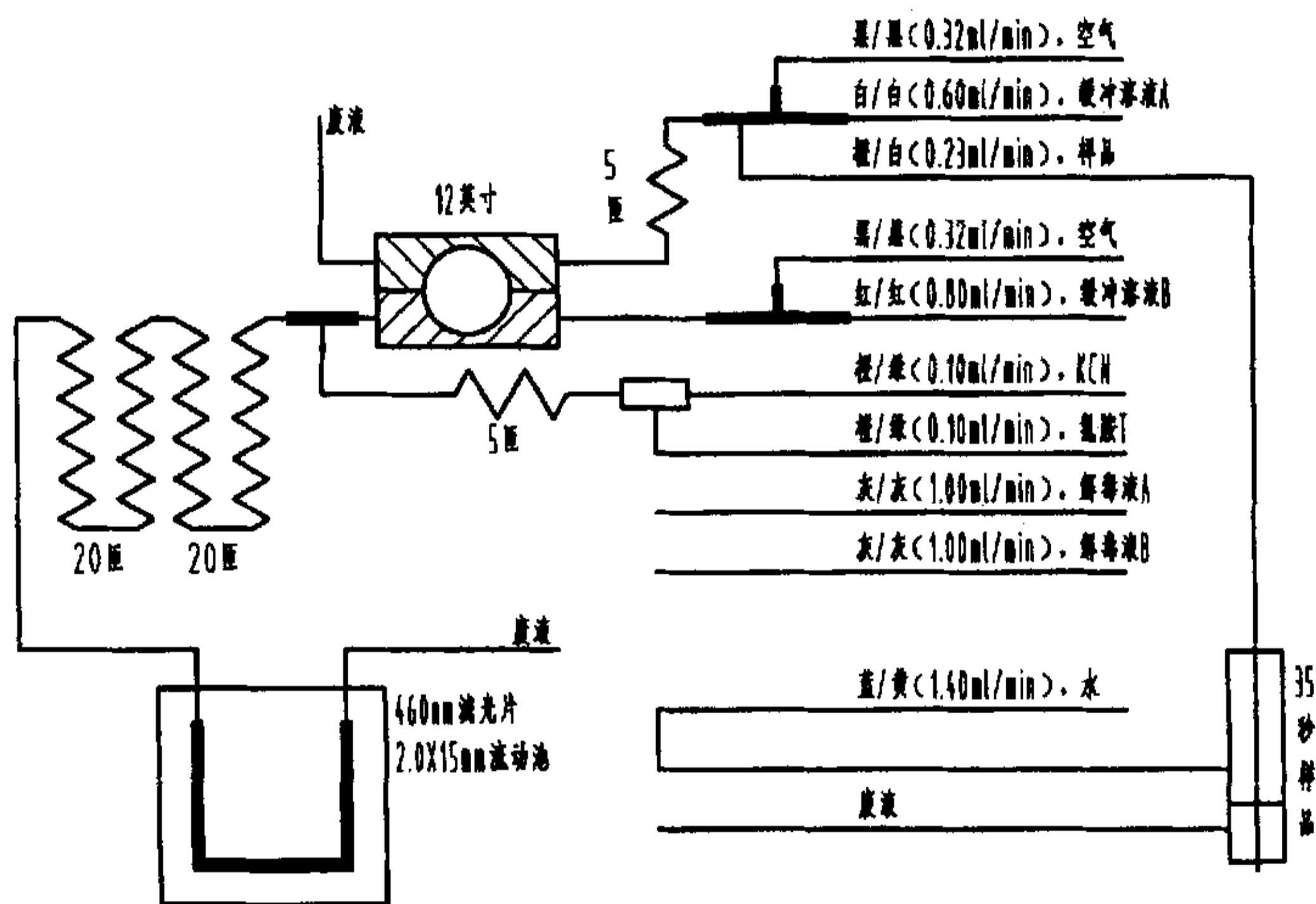


图 2-2 总植物碱测定管路图

### 4 总植物碱的计算

以干基计的总植物碱的含量，由下式得出：

$$\text{总植物碱 \%} = \frac{c \times v}{m \times (1 - w)} \times 100 \quad (2-10)$$

式中:  $c$  — 样品液总植物碱的仪器观测值, mg/mL;

$v$  — 萃取液的体积, mL;

$m$  — 试料的重量, mg;

$w$  — 试样的水分含量。

#### (四) 总氮含量的测定

##### 1 原理

有机含氮物质在浓硫酸及催化剂的作用下, 经过强热消化分解, 其中的氮被转化为氨。在碱性条件下, 氨被次氯酸钠氧化为氯化胺, 进而与水杨酸钠反应产生一靛蓝染料, 在 660nm 比色测定。

##### 2 试剂

Brij35 溶液(聚乙氧月桂醚);

15%次氯酸钠溶液;

氯化钠/硫酸溶液: 10.0g 氯化钠和 7.5mL 浓硫酸溶于 1L 水中;

水杨酸钠/亚硝基铁氰化钠溶液: 75.0g 水杨酸钠( $\text{Na}_2\text{C}_7\text{H}_5\text{O}_3$ )和 0.15g 亚硝基铁氰化钠 ( $\text{Na}_2\text{Fe}(\text{CN})_5\text{NO} \cdot 2\text{H}_2\text{O}$ )溶于 500mL 水中;

缓冲溶液: 25.0g 酒石酸钾钠( $\text{NaKC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$ ), 17.9g 磷酸氢二钠( $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ )和 27.0g 氢氧化钠( $\text{NaOH}$ ), 溶于 500mL 水中;

红色氧化汞( $\text{HgO}$ );

固体硫酸钾;

浓硫酸。

##### 3 仪器设备

连续流动分析仪(AA3, 德国 BRAN+LUBBE);

天平, 感量 0.0001g(AE163, 瑞士 Mettler);

##### 4 总氮含量的计算

以干基计的总植物碱的含量, 由下式得出:

$$\text{总氮}\% = \frac{c}{m \times (1-w)} \times 100 \quad (2-11)$$

式中:  $c$  — 样品液总氮的仪器观测值, mg;

$m$  — 试料的重量, mg;

$w$  — 试样的水分含量。

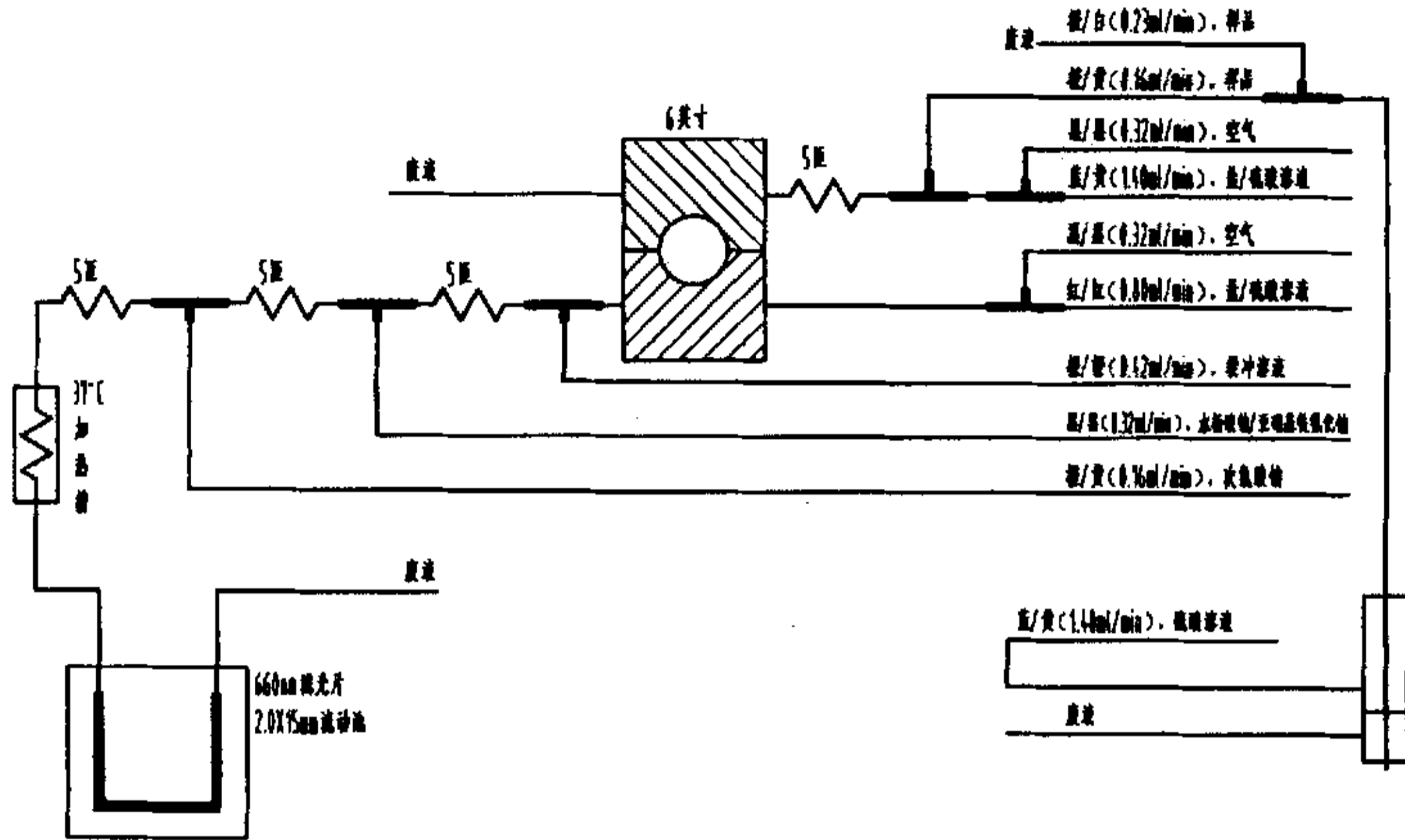


图 2-3 总氮测定管路图

## (五) 氯含量的测定

### 1 原理

用水萃取烟草样品中的氯，氯与硫氰酸汞反应，释放出硫氰酸根进而与三价铁反应形成络合物，反应产物在 480nm 比色测定。

### 2 试剂

Brij35 溶液(聚乙氧月桂醚):

a. 硫氰酸汞储备液: 2.1g 硫氰酸汞 $[\text{Hg}(\text{SCN})_2]$ 溶于 500mL 甲醇中;

b. 硝酸铁储备液: 101.0g 硝酸铁 $[\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}]$ 和 15.8mL 硝酸 $(\text{HNO}_3)$ 溶于 500mL 水中;



显色剂:  $a$  和  $b$  各 300mL 用水定容至 1000mL;

基准氯化钠。

### 3 仪器设备

连续流动分析仪(AA3, 德国 BRAN+LUBBE);

天平, 感量 0.0001g(AE163, 瑞士 Mettler);

振荡器。

### 4 氯含量的计算

以干基计的氯含量, 由下式得出:

$$\text{氯}\% = \frac{c}{m \times (1-w)} \times 100 \quad (2-12)$$

式中:  $c$  — 样品液氯的仪器观测值, mg;

$m$  — 试料的重量, mg;

$w$  — 试样的水分含量。

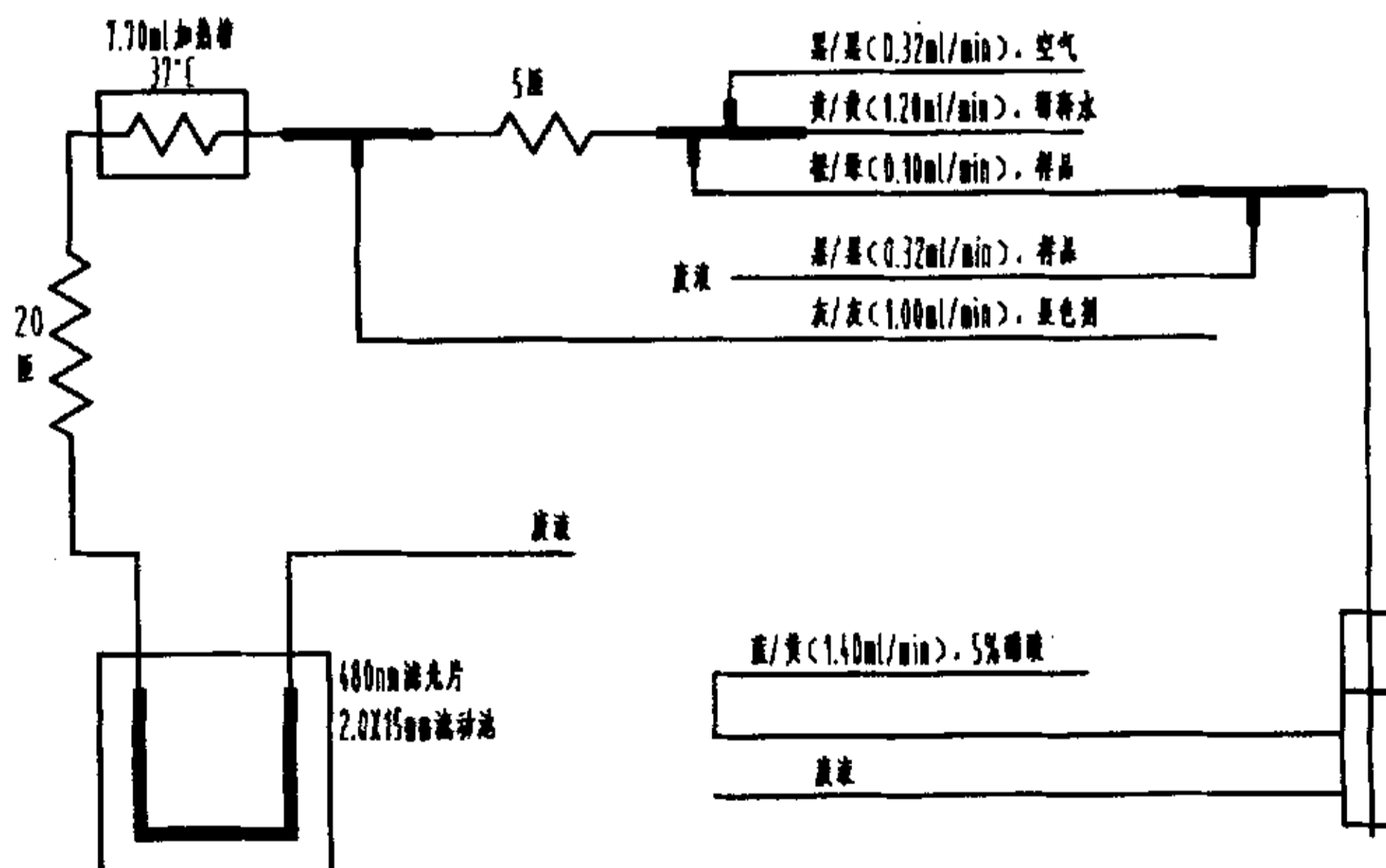


图 2-4 氯测定管路图

## (六) 总蛋白质含量的测定

### 1 原理

用醋酸沉淀烟草样品中的蛋白质氮，然后在浓硫酸及催化剂作用下转化为硫酸铵。强碱蒸馏释放出氨，用标准酸液接收，反滴定。

### 2 试剂

锌粒；

浓硫酸  $d^{20} = 1.84\text{g/ml}$ ；

红色氧化汞；硫酸钾；

氢氧化钠溶液-硫代硫酸钠溶液：500g 氢氧化钠和 40g 硫代硫酸钠 ( $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ )溶于 1L 水中；

0.5%醋酸；

0.1 mol·L<sup>-1</sup> 硫酸铵溶液；

硫酸标准滴定溶液， $C(0.5\text{H}_2\text{SO}_4) = 0.1 \text{ mol}\cdot\text{L}^{-1}$ ；

氢氧化钠标准滴定溶液， $C(\text{NaOH}) = 0.1 \text{ mol}\cdot\text{L}^{-1}$ ；

甲基红指示剂，0.1%。

### 3 仪器设备

天平，感量 0.0001g(AE163，瑞士 Mettler)；

抽滤装置，直径 120mm 布氏漏斗，真空抽滤装置；

蒸馏装置。

### 4 总蛋白质含量的计算

以干基计的蛋白质含量，由下式得出：

$$P(\%) = 6.25 \times \frac{[(V_1 - V_0) \times C_1 - V_2 \times C_2] \times 14}{m \times (1 - W) \times 1000} \times 100 \quad (2-13)$$

式中：6.25 — 蛋白质中氮的转换系数；

$V_0$  — 空白滴定消耗硫酸标准溶液的体积(mL)；

$V_1$  — 硫酸标准溶液的体积(mL)；

$V_2$  — 滴定耗用的氢氧化钠标准溶液的体积(mL)；

- $C_1$  — 硫酸标准溶液浓度,  $\text{mol}\cdot\text{L}^{-1}$ ;  
 $C_2$  — 氢氧化钠标准溶液浓度,  $\text{mol}\cdot\text{L}^{-1}$ ;  
 $m$  — 烟草样品质量(g);  
 $W$  — 烟草样品的水分百分含量(%)。

### (七) 近红外(NIR)光谱的测定

相同的烟末样品一式两份,一份用于上述常规成分的测定,另一份用于近红外测定。

用 FT-NIR 仪测定烟样的 NIR,称取烟末样品 15g,用压锤将烟末表面压制平实,在温度  $20^{\circ}\text{C}\sim 21^{\circ}\text{C}$  测定 NIR,测试范围为  $4000\text{cm}^{-1}\sim 9000\text{cm}^{-1}$ ,每隔一个波数取点,共 5001 个数据点。每隔 4 个样品重新校正背景。

#### 2.3.4 数据处理

首先将光谱测试数据转换为 ASCII 文件,并构成数据矩阵  $X$ ,将常规成分的测量结果构成矩阵  $Y$ ,然后计算采用自编的程序完成。

在建立模型的过程中,首先要排除奇异样本,因为奇异样本对模型的预测精度影响很大,甚至可导致模型崩溃。本文使用交叉验证的方法排除奇异样本。其次,在选择校正集和预测集时,为了保证预测集的物质浓度在校正集的物质浓度范围内,先将样本中最高和最低浓度所对应的样本取出,然后随机地从余下的样本中选取 11 个样本为预测集,剩余样本和前 2 个样本合起来为校正集。

## 2.4 变量筛选对预测结果的影响

### 2.4.1 全谱计算结果

采用 PLS 对原始谱图与烟草样品的常规成份进行相关建模,得到的结果如表 2-1 所示。

从表 2-1 可见所建模型的预测值与实际测量值之间的差异。其中,总糖、总氮和蛋白质的平均相对误差分别为 2.94%、3.35%和 3.52%,其相对均方根误差(RMSEP 与预测样本平均值之比)分别为 3.58%、4.03%和 4.26%,它们的预测结

果与实验结果基本吻合，说明这些化学成分在近红外区域有对应的吸收，从而

表 2-1 使用全谱的校正结果

项目	RMSEP	平均相对误差( $E\%$ ) <sup>1)</sup>	相关系数 $R$
总糖	0.795	2.94	0.954
总植物碱	0.0762	5.06	0.889
总氮	0.0718	3.35	0.978
氯	0.0426	5.56	0.967
蛋白质	0.448	3.52	0.960

注：1)平均相对误差： $E(\%) = \frac{1}{n} \times \left( \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\%$ ，式中， $n$ 为样本个数， $y_i$ 为第 $i$ 个样本的实际测量值， $\hat{y}_i$ 为第 $i$ 个样本的预测值。

可得到较好的预测结果。总植物碱和氯的平均相对误差分别为5.06%和5.65%，其相对均方根误差分别为5.73%和6.26%，它们的预测误差较大，需要进一步提高模型预测精度。从总的结果而言，各个常规成分的相关系数均大于0.85，说明所建的模型具有良好的线性关系。

#### 2.4.2 波长选择计算结果

从使用全谱校正的结果来看，尽管可以得到较好的预测结果，但某些成分，如总植物碱、氯等预测误差仍然较大，因此仍有必要对谱图进行波长选择，消除与建模无关的波长，以提高模型的预测能力。

对波长与化学成分的浓度进行关联，获得波长与化学成分的相关系数  $R$ 。然后采用相关系数阈值预置，取那些相关系数大于阈值的波长点，删除对建模贡献较小的波长，从而得到优化的模型。也就是说，对于特定的化学成份，相关系数越小的波长，说明在此波长下该化学成份的吸收越弱甚至没有，引进这样的波长会降低模型的预测精度，故在建模前应删除这些无用的波长，以达到优化模型的目的。以总糖为例，取阈值为0.08，删除那些相关系数小于0.08的波长点并建立模型。表2-2为本文所采用的阈值、删除的波长数目以及优化后的模型的预测结果。表中所采用的阈值较小，这是因为烟草中含有种类繁多的有机物质，在建模时容易互相干扰，导致了各种常规成份与测量值的相关性较差。

从表2-1与表2-2的比较可见，采用波长选择后的校正模型预测精度明显高于

未优化的数据，有效地减小了模型的 $RMSEP$ 值，降低了平均相对误差，提高了线性相关系数。这表明波长选择是一种有效的优化模型方法。表2-3列出了部分实测值与2种不同方法所得到的预测值比较。

表2-2 波长选择计算结果

常规成份	阈值	删除的波长数目	$RMSEP$	平均相对误差 ( $E\%$ )	相关系数 $R$
总糖	0.08	1587	0.721	2.72	0.964
总植物碱	0.06	717	0.0728	4.73	0.899
总氮	0.04	953	0.0697	3.20	0.981
氯	0.01	174	0.0421	5.34	0.970
蛋白质	0.19	986	0.435	3.34	0.963

表 2-3 实测值与两种不同方法的预测值比较

总糖 (%)			总植物碱 (%)			总氮 (%)			氯 (%)			蛋白质 (%)		
实测值	优化前	优化后	实测值	优化前	优化后	实测值	优化前	优化后	实测值	优化前	优化后	实测值	优化前	优化后
18.46	17.95	18.16	1.39	1.43	1.45	1.81	1.83	1.81	0.39	0.37	0.37	10.32	10.14	9.96
20.84	19.19	19.26	1.05	1.13	1.12	1.92	1.92	1.89	0.70	0.73	0.74	10.51	10.89	10.77
19.48	20.31	19.84	1.10	1.16	1.17	1.75	1.83	1.80	0.56	0.52	0.53	9.68	10.25	10.05
24.58	25.72	25.53	1.18	1.28	1.28	1.91	1.84	1.87	0.76	0.82	0.82	10.23	10.56	10.59
17.83	18.98	19.36	1.18	1.14	1.14	1.86	1.84	1.85	0.95	0.93	0.89	10.37	10.15	10.19
19.74	19.86	19.91	1.58	1.53	1.52	2.06	1.95	1.93	0.45	0.44	0.48	11.22	11.00	10.94
19.58	19.64	19.86	1.53	1.58	1.55	1.91	1.87	1.80	0.89	0.84	0.85	10.30	10.70	10.54
27.14	26.26	26.54	1.28	1.23	1.24	2.98	3.02	3.04	0.88	0.93	0.97	16.99	16.82	16.90
22.96	22.50	22.83	1.22	1.27	1.24	1.72	1.65	1.67	0.81	0.73	0.76	9.41	8.89	9.12
24.54	24.83	24.88	1.32	1.23	1.19	1.74	1.63	1.69	0.67	0.70	0.67	10.88	10.18	10.12
26.31	27.38	26.83	1.32	1.31	1.30	1.70	1.66	1.64	0.57	0.55	0.56	10.63	10.71	10.72
23.16	22.63	23.49	1.44	1.39	1.37	1.50	1.50	1.47	0.71	0.66	0.68	9.38	9.41	9.34
20.08	19.93	20.47	1.36	1.33	1.34	1.56	1.57	1.59	1.03	1.02	1.07	9.75	9.83	9.80
20.29	21.56	21.46	1.50	1.35	1.38	1.50	1.52	1.51	0.54	0.48	0.53	10.88	9.96	9.99
22.60	22.00	22.06	1.32	1.29	1.32	1.65	1.54	1.54	0.72	0.71	0.75	9.69	9.92	9.83
26.46	25.35	26.38	1.62	1.52	1.52	1.64	1.50	1.52	0.67	0.69	0.68	10.25	9.63	9.66
22.24	22.11	21.81	1.43	1.57	1.55	1.66	1.70	1.67	0.58	0.61	0.62	10.38	10.40	10.20
22.41	22.06	22.59	1.40	1.39	1.38	1.54	1.68	1.65	0.58	0.62	0.62	9.63	10.57	10.65
23.59	23.98	23.66	1.00	1.10	1.08	1.68	1.63	1.65	0.61	0.68	0.68	10.50	10.17	10.24
20.97	22.00	21.63	1.36	1.47	1.42	1.50	1.44	1.48	0.53	0.55	0.55	9.38	9.08	8.95

在使用关系模型预测化学成分含量时，影响预测精度的原因有多方面。一方

面, 在建立模型的过程中, 应该要涵盖要分析样品的含量范围。在所测的浓度或性质范围内, 样品的个数应该均匀分布。这样, 不论是高浓度还是低浓度范围的样品才能得到同样的精度。如果模型中没有包含与待测样本类似的样本, 则该样本的预测精度有可能较低。另一方面, 分析样品的基体必须相同(如pH值或含水率等), 否则背景干扰将会非常严重, 导致模型适用性变差或根本不能用。此外, 如果训练集样本的变化范围越大(包括浓度变化及形态变化), 建立模型的校正模型的适用范围也就越广, 但分析精度可能变差。如果模型的适用范围较小, 则分析结果的精度相对较高, 但适用面会变窄。在控制分析实验中, 有时必须建立两种模型(宽范围和窄范围)。

## 2.5 结论

采用偏最小二乘法建立了近红外光谱与卷烟常规化学成分的相关模型。并在此基础上, 提出了采用相关系数为判据对波长进行优选的方法, 删除对建模无用的干扰波长, 从而得到优化的模型, 有效地提高了模型的预测能力, 并获得了较好的预测精度。

## 参考文献

- [1] 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
- [2] Craftsbrandner S. J., Collins M., Sutton T. G., *et al.*, Effect of leaf maleic hydrazide concentration on yield and dry-matter partitioning in burley tobacco (Nicotiana-glauca-L)[J], Field Crops Research, 1994, 37 (2): 121-128.
- [3] Diffie John T. Tobacco analysis by NIR spectroscopy[J], Pract. Spectrosc., 1992, (13): 433-473.
- [4] 张建平, 谢雯燕, 束茹欣, 等。烟草化学成分的近红外快速分析研究[J], 烟草科技, 1999, (3): 37-38.
- [5] 邵学广, 蔡文生 化学信息学[M], 北京: 科学出版社, 2001.
- [6] Bakken G.A., Houghton T.P., Kalivas J.H. Cyclic subspace regression with analysis of wavelength-selection criteria[J], Chemom. Intell. Lab. Syst, 1999, (45): 225-239.
- [7] Smith B. M., Gemperline P. J., Wavelength selection and optimization of pattern-recognition methods using the genetic algorithm[J], Anal. Chim. Acta, 2000, (423): 167-177.

### 第三章 WT-PLS 用于烟草样品常规成分的定量预测研究

#### 3.1 引言

在植物样品(如中药样品、烟草样品等)的研究中,常规化学成份的分析对于了解样品的性质具有非常重要的意义。常规化学成份一般采用化学分析法,分析过程比较复杂,费时、费力、成本高,且重现性差。近红外光谱(NIR)是近年来发展起来的一种快速、简便、非破坏性的分析方法<sup>[1]</sup>。用近红外光谱对植物样品化学成份进行定量预测,是目前分析化学的一个重要课题<sup>[2-3]</sup>。然而,对于含有大量化学物质的植物样品,其近红外光谱严重重叠,且吸收较弱,易受噪音和背景的干扰。在实际应用中,常采用化学计量学方法对近红外光谱进行处理。为保证获得最大的信息量,在采样时常采用密集取点,造成建模时波长点数远远多于样本个数,从而易使模型具有随机性<sup>[4]</sup>。为克服近红外光谱的以上缺点,需要寻找一种既能滤噪又能压缩数据的有效方法。

研究表明,小波变换是一种高效的数据压缩技术<sup>[5-6]</sup>。本文利用小波变换,对植物样品的近红外谱图进行压缩,并将压缩后的数据采用PLS建模,得到了满意的结果。

#### 3.2 算法原理

小波定义为满足一定条件的函数通过平移和伸缩产生的一个函数族,即:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in R, a \neq 0 \quad (3-1)$$

其中,  $a$  用于控制伸缩(dilation),称为尺度参数(scale parameter);  $b$  用于控制位置(position),称为平移参数(translation parameter);  $\psi(t)$  叫作小波基或小波母函数。

定义

$$a = a_0^m \quad (a_0 > 1, \quad m \in Z) \quad (3-2)$$



$$b = nb_0 a_0^m \quad (b_0 \in R, \quad n \in Z) \quad (3-3)$$

得到

$$\psi_{m,n}(t) = a_0^{-m/2} \psi(a_0^{-m} t - nb_0) \quad (3-4)$$

一般取  $a_0 = 2$ ,  $b_0 = 1$ , 此时称为二进小波。

小波变换的定义为:

$$Wf(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) dt \quad (3-5)$$

由卷积的定义

$$f * g(x) = (f(u) * g(u))(x) = \int_{-\infty}^{+\infty} f(u) g(x-u) du = \langle f(u), g(x-u) \rangle \quad (3-6)$$

上式可写成:

$$W_a f(b) = \langle f(t), \psi_a(t-b) \rangle \equiv f * \tilde{\psi}_a(b) \quad (3-7)$$

其中,  $\tilde{\psi}_a(x) = \psi_a(-x)$ 。

相应的离散小波变换定义为:

$$Wf(m,n) = \langle f(t), \psi_{m,n}(t) \rangle = a_0^{-m/2} \int_{-\infty}^{+\infty} f(t) \psi(a_0^{-m} t - nb_0) dt \quad (3-8)$$

因此, 小波变换的实质可理解为函数  $f(t)$  在小波的空间投影或理解为函数  $f(t)$  经滤以器  $\tilde{\psi}_a(x)$ , 滤波后的结果。

某信号小波分解以后将从原来的空间投影到小波空间, 由于小波变换的特点, 在小波空间的系数将有一部分特别小, 对信号的表示没有意义。如果我们将小的系数去除, 在重构的信号中将不会丢失有意义的信息。因此, 小波变换可用于数据的压缩。

从小波变换的原理可知, 小波变换是一种线性变换。对原始数据进行小波压缩后, 可采用小波系数代替原始数据, 利用多元校正方法建立模型。取较大的小波系数组成新的数据矩阵代替原始数据矩阵, 可以大大降低数据量, 既能有效地去除噪音, 又能提高多元校正的速度。同时, 采用较少的变量建模, 有利于减少模型的随机性并提高预测精度。

本文中, 对烟草样品的近红外光谱采用 WT-PLS 算法对其常规成份(总糖、总植物碱、总氮、氯和蛋白质)进行定量预测。设近红外光谱矩阵为  $X$ , 常规成

份的浓度矩阵为  $Y$ ，WT-PLS 算法的主要步骤如下：

(1) 对光谱矩阵  $X$  进行小波压缩，按绝对值的大小对每个样品的近红外光谱取相同数量的小波系数，得到小波系数数据矩阵  $X_w$ ；对浓度矩阵  $Y$  不加任何处理。

(2) 采用 PLS 对  $X_w$ 、 $Y$  矩阵进行分解，即：

$$X_w = TP + E \quad (3-9)$$

$$Y = UQ + F \quad (3-10)$$

其中， $T$  和  $U$  分别为  $X_w$  矩阵和  $Y$  矩阵的得分矩阵， $P$  和  $Q$  分别为  $X_w$  矩阵和  $Y$  矩阵的载荷(即主成分)矩阵， $E$  和  $F$  分别为用 PLS 模型拟合  $X_w$  和  $Y$  时所引进的误差。然后，PLS 将  $T$  和  $U$  作线性回归，计算关联系数矩阵  $B$ ，即：

$$U = TB \quad (3-11)$$

$$B = T'U(T'T)^{-1} \quad (3-12)$$

(3) 对于未知样本的光谱数据采用与步骤(1)相同的方法进行小波压缩，得到未知样本的小波系数数据矩阵  $X_{w\text{未知}}$ 。根据  $X_{w\text{未知}}$  和公式(3-9)得到的  $P_{\text{校正}}$  求出  $T_{\text{未知}}$ ，从而计算出未知样品的浓度  $Y_{\text{未知}}$ ，即：

$$Y_{\text{未知}} = T_{\text{未知}} BQ \quad (3-13)$$

### 3.3 实验部分

#### 3.3.1 仪器

PERKIN ELMER Spectrum Identichack FT-NIR System;

AutoAnalyzer III(德国 BRAN+LUBBE);

分析天平(瑞士 Mettler);

振荡器;

真空抽滤装置;

蒸馏装置。

#### 3.3.2 试剂

(1) 总糖的测定:

Brij35 溶液(聚乙氧月桂醚);

0.5 mol·L<sup>-1</sup> 氢氧化钠溶液;  
0.008 mol·L<sup>-1</sup> 氯化钙溶液;  
5%醋酸溶液;  
0.5 mol·L<sup>-1</sup> HCl 溶液;  
1.0 mol·L<sup>-1</sup> HCl 溶液;  
1.0 mol·L<sup>-1</sup> NaOH 溶液;  
5%对羟基苯甲酸酰肼溶液(HOC<sub>6</sub>H<sub>4</sub>CONHNH<sub>2</sub>);  
D-葡萄糖。

(2) 总植物碱的测定:

Brij35 溶液(聚乙氧月桂醚);

缓冲溶液 A: 2.35g 氯化钠(NaCl)和 7.60g 硼酸钠(Na<sub>2</sub>B<sub>4</sub>O<sub>3</sub> · 10H<sub>2</sub>O)溶于 1L 水中);

缓冲溶液 B: 26g 磷酸氢二钠(Na<sub>2</sub>HPO<sub>4</sub>), 10.4g 柠檬酸 [COH(COOH)(CH<sub>2</sub>COOH)<sub>2</sub> · H<sub>2</sub>O] 和 7g 对氨基苯磺酸(NH<sub>2</sub>C<sub>6</sub>H<sub>4</sub>SO<sub>3</sub>H) 溶于 1L 水中;

2%氯胺 T 溶液;

0.4%氰化钾溶液;

纯烟碱。

(3) 总氮的测定:

Brij35 溶液(聚乙氧月桂醚);

次氯酸钠溶液;

氯化钠/硫酸溶液: 10.0g 氯化钠和 7.5mL 浓硫酸溶于 1L 水中;

水杨酸钠/亚硝基铁氰化钠溶液: 75.0g 水杨酸钠(Na<sub>2</sub>C<sub>7</sub>H<sub>5</sub>O<sub>3</sub>)和 0.15g 亚硝基铁氰化钠 (Na<sub>2</sub>Fe(CN)<sub>5</sub>NO · 2H<sub>2</sub>O)溶于 500mL 水中;

缓冲溶液: 25.0g 酒石酸钾钠(NaKC<sub>4</sub>H<sub>4</sub>O<sub>6</sub> · 4H<sub>2</sub>O, 17.9g 磷酸氢二钠 (Na<sub>2</sub>HPO<sub>4</sub> · 12H<sub>2</sub>O)和 27.0g 氢氧化钠(NaOH), 溶于 500mL 水中;

红色氧化汞(HgO);

固体硫酸钾；

浓硫酸。

#### (4) 氯的测定：

Brij35 溶液(聚乙氧月桂醚)；

a. 硫氰酸汞储备液：2.1g 硫氰酸汞 $[\text{Hg}(\text{SCN})_2]$ 溶于 500mL 甲醇中；

b. 硝酸铁储备液：101.0g 硝酸铁 $[\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}]$ 和 15.8mL 硝酸 $(\text{HNO}_3)$ 溶于 500mL 水中；

显色剂：a 和 b 各 300mL 用水定容至 1000mL；

基准氯化钠。

#### (5) 蛋白质的测定：

锌粒；

浓硫酸  $d^{20} = 1.84\text{g/ml}$ ；

红色氧化汞；

硫酸钾；

氢氧化钠溶液-

硫代硫酸钠溶液：500g 氢氧化钠和 40g 硫代硫酸钠 $(\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O})$ 溶于 1L 水中；

0.5%醋酸；

0.1mol·L<sup>-1</sup> 硫酸铵溶液；

硫酸标准滴定溶液， $C(0.5\text{H}_2\text{SO}_4) = 0.1 \text{ mol} \cdot \text{L}^{-1}$ ；

氢氧化钠标准滴定溶液， $C(\text{NaOH}) = 0.1 \text{ mol} \cdot \text{L}^{-1}$ ；

甲基红指示剂，0.1%。

### 3.3.3 样品

共测试了 72 个牌号的烤烟型卷烟产品。按照 YC/T31-1996 制备成粉末样品。分别进行近红外光谱测试和常规化学成分(总糖、植物碱、总氮、氯和蛋白质)的

测定。近红外测试范围从  $4000\text{cm}^{-1}\sim 9000\text{cm}^{-1}$ ，每隔一个波数取点，共取 5001 个数据点。将全部样本分为校正集和预测集。为了保证预测集的物质浓度在校正集的物质浓度范围内，先将样本中最高和最低浓度所对应的样本取出，然后随机地从余下的样本中选择 20 个样本作为预测集，剩余样本和前两个样本合起来为校正集。

### 3.3.4 分析方法

常规化学成份的测定方法均采用行业标准或 ISO 标准，具体方法如下：

YC/T 159 烟草及烟草制品 水溶性糖的测定 连续流动法

YC/T 160 烟草及烟草制品 总植物碱的测定 连续流动法

YC/T 161 烟草及烟草制品 总氮的测定 连续流动法

YC/T 162 烟草及烟草制品 氯的测定 连续流动法

YC/T 164 烟草及烟草制品 蛋白质的测定

用 FT-NIR 仪测定烟草样品的 NIR，称取烟末样品 15g，用压锤将烟末表面压制平实，在温度  $20^{\circ}\text{C}\sim 21^{\circ}\text{C}$  测定 NIR，测试范围为  $4000\text{cm}^{-1}\sim 9000\text{cm}^{-1}$ ，每隔一个波数取点，共 5001 个数据点。每隔 4 个样品重新校正背景。

## 3.4 小波变换对近红外光谱的数据压缩

### 3.4.1 最佳小波基、分解尺度及压缩倍数的确定

我们分别对常用的正交小波基 Daubechies(db)，Symmlet(sym) 及 Coiflet(coif) 进行了比较，采用 PLS 方法建模，以验证集的预测平方和 (PRESS) 为评价标准，取对应于最小 PRESS 值的小波基。PRESS 值计算如下：

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3-14)$$

式(3-14)中， $n$  为样本个数， $y_i$  为样本实测值， $\hat{y}_i$  为样本预测值。

先固定压缩倍数为 40 倍，通过采用不同小波基时验证集预测结果的比较可得：对于总糖、总植物碱、总氮、氯和蛋白质，尽管所得到的最佳小波基不太一

致,但是差别较小,因此,我们取  $\text{coif3}(N=18)$ ,  $N$  表示小波滤波器的长度)为最佳小波基。

通过不同分解尺度时预测结果的比较可得:总糖、总植物碱、总氮、总氯、蛋白质的最佳分解尺度分别为 6、5、4、3 和 5。各个组分近红外吸收强度差别较大,可能是导致分解尺度不一致的原因。含量越高和近红外吸收越强的组分,对应的近红外吸收值较大,小波变换后的细节对其影响较小,其分解尺度较大;含量较低而近红外吸收又较弱的组分,对应的近红外吸收值较小,小波变换后的细节对其影响较大,它们的相当一部分信息包含在细节中,为了保证它们的信息最大限度地保留,故分解尺度较小。

为了考察压缩倍数对预测结果的影响,表 3-1 列出了采用上述最佳小波基和最佳分解尺度,不同压缩倍数时验证集的  $PRESS$  值。从表 3-1 可见,总糖、总植物碱、总氮、氯、蛋白质的最佳压缩倍数分别为 50、45、15、10、55。不同组分之间的压缩倍数差别较大,但与分解尺度之间有一定关系。含量高和近红外吸收强的组分分解尺度高,其压缩倍数也高;而含量低和近红外吸收较弱的组分分解尺度低,它们的压缩倍数也相应较低。

表 3-1 不同压缩倍数时验证集的  $PRESS$  值

Table 3-1 The  $PRESS$  of the prediction results at different compression ratio

压缩倍数 Compression times	总糖 Total sugar	总植物碱 Alkaloid	总氮 Total nitrogen	氯 Total chloride	蛋白质 Protein
5	4.48	0.1083	0.466	0.017	2.72
10	4.24	0.1082	0.167	0.016	2.71
15	3.60	0.1082	0.047	0.019	2.68
20	5.09	0.1079	0.056	0.049	2.84
25	3.61	0.1078	0.097	0.031	2.90
30	8.90	0.1069	0.124	0.021	2.91
35	12.9	0.1001	0.115	0.119	3.01
40	17.8	0.1053	0.119	0.135	3.17
45	2.16	0.0795	0.112	0.139	1.95
50	2.15	0.0858	0.123	0.148	2.03
55	9.93	0.1500	0.124	0.160	1.81
60	9.48	0.1635	0.115	0.168	2.13
65	8.13	0.1594	0.120	0.169	2.31

### 3.4.2 WT-PLS 与 PLS 结果的比较

表 3-2 为采用 PLS 对全谱进行建模和采用 WT-PLS 进行建模时预测集的预测结果比较。从表中的结果比较可见, 采用 WT-PLS 所建的模型明显优于采用 PLS 所建的模型。其中, 总糖、总植物碱、氯和蛋白质的平均相对误差分别下降了 0.96%、0.53%、0.15%、0.58%, 总氮的平均相对误差略有提高(-0.04%), 但幅度很小; 同时, 各组分的线性相关系数也有较大的提高。以上结果说明通过小波压缩可以消除噪音与背景干扰, 从而提高了模型的预测精度和线性相关性; 与采用全谱相比, 通过压缩自变量, 不同组分对应的小波系数也各不相同, 增加了各个组分之间的区别, 可有效地减少模型的随机性。另一方面, 通过小波压缩, 可有效地减少建模数据量, 大大地加快建模速度。一旦采用 WT-PLS 算法建立模型后, 预测时的模型计算时间在数秒之内, 可以满足快速分析的要求。同时, 模型的预测精度在误差允许的范围内, 其线性关系良好, 说明 WT-PLS 有望成为植物样品快速定量分析的一种有力工具。

表 3-2 PLS 与 WT-PLS 的预测结果比较

Table 3-2 The comparison of the prediction results by PLS and WT-PLS

常规成份 Routine components	数据点 Data points	平均相对误差( $E\%$ ) <sup>*</sup> Average relative error	相关系数 $R$ Correlation coefficient
<b>PLS</b>			
总糖 Total sugar	5001	2.79	0.935
总植物碱 Alkaloid	5001	4.91	0.926
总氮 Total nitrogen	5001	3.35	0.978
氯 Total chloride	5001	5.70	0.970
蛋白质 Protein	5001	3.52	0.960
<b>WT-PLS</b>			
总糖 Total sugar	76	1.83	0.975
植物碱 Alkaloid	111	4.38	0.950
总氮 Total nitrogen	333	3.39	0.979
总氯 Total chloride	500	5.55	0.971
蛋白质 Protein	90	2.94	0.976

\*: 平均相对误差:  $E(\%) = \frac{1}{n} \times \left( \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\%$ , 式中,  $n$  为样本个数,  $y_i$  为

第  $i$  个样本的实际测量值,  $\hat{y}_i$  为第  $i$  个样本的预测值。

### 3.5 结 论

采用小波变换用于植物样品近红外光谱数据的压缩, 结合 PLS 方法进行建模, 不仅有效去除了建模中噪音和背景的干扰, 同时减少了变量个数和模型的随机性, 提高了模型的运算速度, 改善了模型的预测精度和线性相关性, 为植物样品常规化学成份的快速分析提供了一种有力手段



## 参考文献

- [1] Lu Wanzhen(陆婉珍), Yuan Hongfu(袁洪福), Xu Guangtong(徐广通), Qiang Dongmei(强冬梅). *Analytical Technology of Near Infrared Spectroscopy* (现代近红外光谱分析技术), Beijing(北京): Chinese Petrochemical Industry Press(中国石化出版社), 2000: 1~13
- [2] Diffie, John T. *Pract. Spectrosc.*, 1992, (13): 433~473
- [3] Kittas C, Baille A, Giaglaras P. *J. Agr. Eng. Res.*, 1999, 73( 4):341~351
- [4] Riccardo Leardi, Amparo Lupiáñez González. *Chemometr. Intell. Lab. Syst.*, 1998 (41):195~207
- [5] Shao Xueguang(邵学广), Gu Hua(顾华), Cai Wensheng(蔡文生), Pan Zhongxiao(潘忠孝). *Spectroscopy and Spectral Analysis* (光谱学与光谱分析), 1999, 19(2): 158~160
- [6] Shao X G, Cai W S. *Rev. Anal. Chem.*, 1998, (4): 235~285

## 第四章 WT-UVE-PLS 用于烟草样品常规成分的定量预测研究

### 4.1 引言

植物样品(如烟草, 茶, 中药)的化学成份分析具有相当重要的意义。通常要对这些产品进行研究和利用就必须对它们的化学成分有充分的了解。传统的湿法(溶液法)化学分析方法不仅费用高, 而且费时费力, 分析结果的重现性也存在一定问题。

实践表明, 近红外光谱提供了一种植物样品的廉价且快速的检测方法<sup>[1-3]</sup>。利用近红外光谱和多变量校正技术对复杂植物样品进行定量分析是一种良好的选择。然而, 在植物样品的近红外光谱分析中仍然存在谱峰宽、吸收弱、归属不明确且严重重叠等问题<sup>[4]</sup>。另外, 光谱数据矩阵常常很大, 为尽可能得到更多的光谱信息, 近红外光谱往往包括几千个波长点的数据, 而样品数也往往很大, 有时为数百个, 测量数据中存在大量的冗余信息<sup>[5]</sup>。因此, 如何压缩这些光谱数据, 并从压缩后的数据中得到更多有用的信息具有非常重要的意义。

小波变换是分析数据压缩的一个有力工具, 它将原始数据变换到小波域得到小波系数。那些幅值非常小的小波系数往往只表示了原始信号中的无用信息, 如噪声, 删除这些系数对原始信号的有用信息将没有显著影响。而保留下来的小波系数可以充分表示原始信息, 可用于进一步的分析<sup>[6-7]</sup>。

关于 PLS 分析中的变量筛选方法, 目前已有许多报道<sup>[8-15]</sup>。其中, D. Jouan-Rimbaud 等<sup>[8]</sup>提出了一种非常有意义的方法, 称为 RCE-PLS (relevant component extraction for PLS), 它可以从相关信息中分离出不相关的信息和噪音。RCE-PLS 是在 UVE-PLS (uninformative variables elimination for PLS)方法<sup>[10]</sup>的基础上发展起来的一种方法, 将原始数据矩阵压缩为小波系数再用 RCE 作更进一步的筛选。然而这种方法应用于可能包含有几千个变量的大型数据集时, 就存在一个功效的问题, 因为它采用了留一法交叉验证(leave-one-out cross-validation)用于去除不相关的变量, 而留一法交叉验证需要反复的迭代运算。

本章的目的是提出一种有效的方法用于大量植物样品的近红外数据集的校正。与 RCE-PLS 算法不同的是,在本方法中我们采用了 UVE 算法来替代 RCE 对整个小波系数集进行进一步的压缩,这种方法称为 WT-UVE-PLS。由于它的运算速度比较快,因此可适用于数据量庞大的光谱数据分析。在此项工作中,我们用 WT-UVE-PLS 方法处理了 70 个样品具有 5001 数据点的近红外光谱。数据的处理的结果证明了这种方法能够提取庞大近红外数据集的有用信息,同时预测速度和精度均得到了改善。

## 4.2 WT-UVE-PLS 的原理与算法

为了对近红外光谱进行数据量压缩和筛选,本文采取了数据压缩和变量筛选两种手段。首先作为一种很有效的压缩工具<sup>[6-7]</sup>,小波变换用于对大型 NIR 数据矩阵进行压缩。在数据压缩过程中,小波系数中不包含有用信息的部分将被剔除。但是即使经过这一步骤后,所保留下来的的小波系数中还可能含有与多元校正不相关的信息。因此还必须采取更进一步的方法来去除那些信息量小或对校正模型无贡献或贡献较小的系数。为了实现这个目的,我们采取了 UVE 算法,并使用“稳定性”(stability)<sup>[10]</sup>作为参照标准来从以上所得到的的小波系数中进一步提取有用信息。我们将这个先用小波变换预压缩再结合 UVE 算法提取有用信息的方法称为 WT-UVE。

WT-UVE 具体算法过程如下:

1. 采用小波变换对光谱数据集进行压缩:作为一种有效的数据处理工具,小波变换已广泛的应用于数据压缩<sup>[16,17]</sup>。在这些报道的工作中,每条谱线选用一个小波基进行变换,并将变换后所得到的的小波系数构建为数据矩阵。每一列中变异较大的将被保留。在本工作中,由于还需要进一步的变量筛选,所以在第一步中我们采用了较简单的小波算法对数据矩阵进行初步压缩。在该算法中,校正集的光谱的平均光谱首先被压缩来获得最大小波系数的索引,然后对每一条光谱进行变换,并将变换后的小波系数和上一步所得到的索引值保留下来。在整个运算过程中,小波滤波器、分解尺度和压缩比需要进行优化来获得最佳的压缩效果。
2. 确定 PLS 模型使用的因子数:因子数( $f$ )是 PLS 模型的一个关键的参数,它影响到模型预测的准确性和运算速度,尤其是对于需要用留一法交叉验证的

UVE 方法。尽管已报道了多种确定数据矩阵因子数的方法，在本文中我们采用了留半交叉验证法(leave-half cross-validation)。在该方法计算过程中，将校正集的数据分为两部分，最优的因子数通过以其中一部分数据当训练集，另一部分当预测集来获得。

3. 筛选小波系数中不含有用信息的部分：在该步骤中采用了 UVE 方法，首先在小波系数矩阵  $W(n, p)$  中添加了一个随机变量矩阵  $N(n, p)$  得到矩阵  $WN(n, 2p)$ ，其中  $n$  和  $p$  分别为样品数和小波系数的数据点数。然后，我们通过使用矩阵  $WN(n, 2p)$  和因子数  $f$  来对下列模型

$$y = WN \cdot \beta + E \quad (4-1)$$

进行留半交叉验证，其中模型中的  $y$  是浓度向量， $\beta$  是回归系数向量，它可以通过以下公式计算：

$$\beta = \sum_{i=1}^f b_i h_i q_i \quad (4-2)$$

其中

$$\begin{cases} h_i = w_i & \text{for } i = 1 \\ h_i = \prod (I - w_{i-1} p_{i-1}) & \text{for } i > 1 \end{cases} \quad (4-3)$$

在公式(4-2)和(4-3)中  $p$ 、 $q$ 、 $b$  和  $w$  分别为在 PLS 模型的光谱因子、浓度因子、内部回归系数和权重向量。

回归系数矩阵  $B = [\beta_1, \dots, \beta_n]$  可以通过交叉验证获得，其中前面一半  $(1, \dots, p)$  对应于  $p$  个小波系数，剩下的一半  $(p+1, \dots, 2p)$  对应于随机变量。因为  $\beta_i$  中的每一个系数代表相应变量对模型的贡献。所以，每一个变量的可靠性可以通过公式(4-4)所得的稳定性来进行定量计算：

$$s_i = \text{mean}(\beta_i) / \text{std}(\beta_i) \quad i = 1, \dots, 2p \quad (4-4)$$

其中  $\text{mean}(\beta_i)$  和  $\text{std}(\beta_i)$  分别为交叉验证中第  $i$  变量的回归系数的平均值和标准偏差。很明显，可以认为那些稳定性比随机变量还低的变量对模型没有贡献，可以被剔除。在实际运算中，我们采用了以下方法得到一个阈值：

$$\text{cutoff} = k \times \max(\text{abs}(s_{\text{noise}})) \quad (4-5)$$

其中  $k$  介于 0.7~0.9 之间。

4. 用已经过筛选的小波系数建立 PLS 模型。

5. 通过步骤(1)-(3)使用同样的参数获得预测集的小波系数并进行筛选, 然后通过步骤(4)所建立的模型进行预测。

### 4.3 实验部分

本实验采用的近红外光谱仪为 PERKIN ELMER Identichack FT-NIR。共测量了 70 个烟草样品, 采用  $1\text{cm}^{-1}$  的间隔记录了  $4000\text{-}9000\text{cm}^{-1}$  光谱数据。烟草样品中的常规组分(TS, AS 和 TN)采用连续流动分析仪 AA3 按照标准方法进行测定。其中 TS 表示总糖含量, 主要由葡萄糖、果糖和少量的蔗糖组成。AS 表示植物碱, 主要成分是尼古丁, 也包括少量的其它植物碱类。TN 表示总含氮化合物, 主要组成是蛋白质和少量的氨基酸和有机胺。将 70 个样品的数据随机地分为两部分, 其中 50 个样品光谱数据用于校正集, 另外 20 个样品的光谱数据用作预测集。

### 4.4 结果与讨论

#### 4.4.1 小波变换与数据压缩

小波变换用于分析化学信号的数据压缩已有许多报道<sup>[6,7,16,17]</sup>。在本章的方法中, 小波滤波器、分解尺度和压缩比等三个参数需进行优化。我们采用留半交叉验证法所得到的 *RMSEP*(预测值均方误差)作为优化的标准。*RMSEP* 可通过以下公式计算:

$$RMSEP = \left( \sum_{k=1}^3 \sum_{i=1}^{50} (\hat{y}_{k,i} - y_{k,i})^2 / n \right)^{1/2} \quad (4-6)$$

其中  $y_{k,i}$  和  $\hat{y}_{k,i}$  分别为  $i$  个样品第  $k$  组分的测量浓度和预测浓度。样品 1-25 的  $\hat{y}_{k,i}$  通过以样品 26-50 作为校正集获得。样品 26-50 的  $\hat{y}_{k,i}$  则以样品 1-25 作为校正集获得。

首先, 计算使用不同小波滤波器(Daubechies, Symmlet, Coiflet)和分解尺度(1-10)以不同的组合方式将光谱数据集进行处理得到 *RMSEP*, 结果表明所得到的 *RMSEP* 没有明显差别, 均在  $4.00 \pm 0.38$  范围内, 可以看出使用不同参数所得的结果没有明显区别。因此, 在本方法中我们采用了最简单的“dbl”或 Haar 小波

滤波器，并将分解尺度设定为 5。

小波压缩算法中的另一个关键参数是压缩比，对于数据压缩本身的要求来说通常总是期望有较高的压缩比。但压缩程度过高往往会导致部分有用信息的丢失。此外在本章的工作中，由于还使用 UVE 对小波系数中的无用信息进行筛选，因此保留更多有用信息比纯粹的追求高的压缩比更显得重要。图 4-1 所示为校正数据集的平均光谱和以 140、150 和 160 个小波系数重构所得到的光谱。很明显，使用 160 个系数进行重构所得的谱图几乎没有信息丢失。而使用 140 和 150 个系数进行重构无法得到完整的谱图。

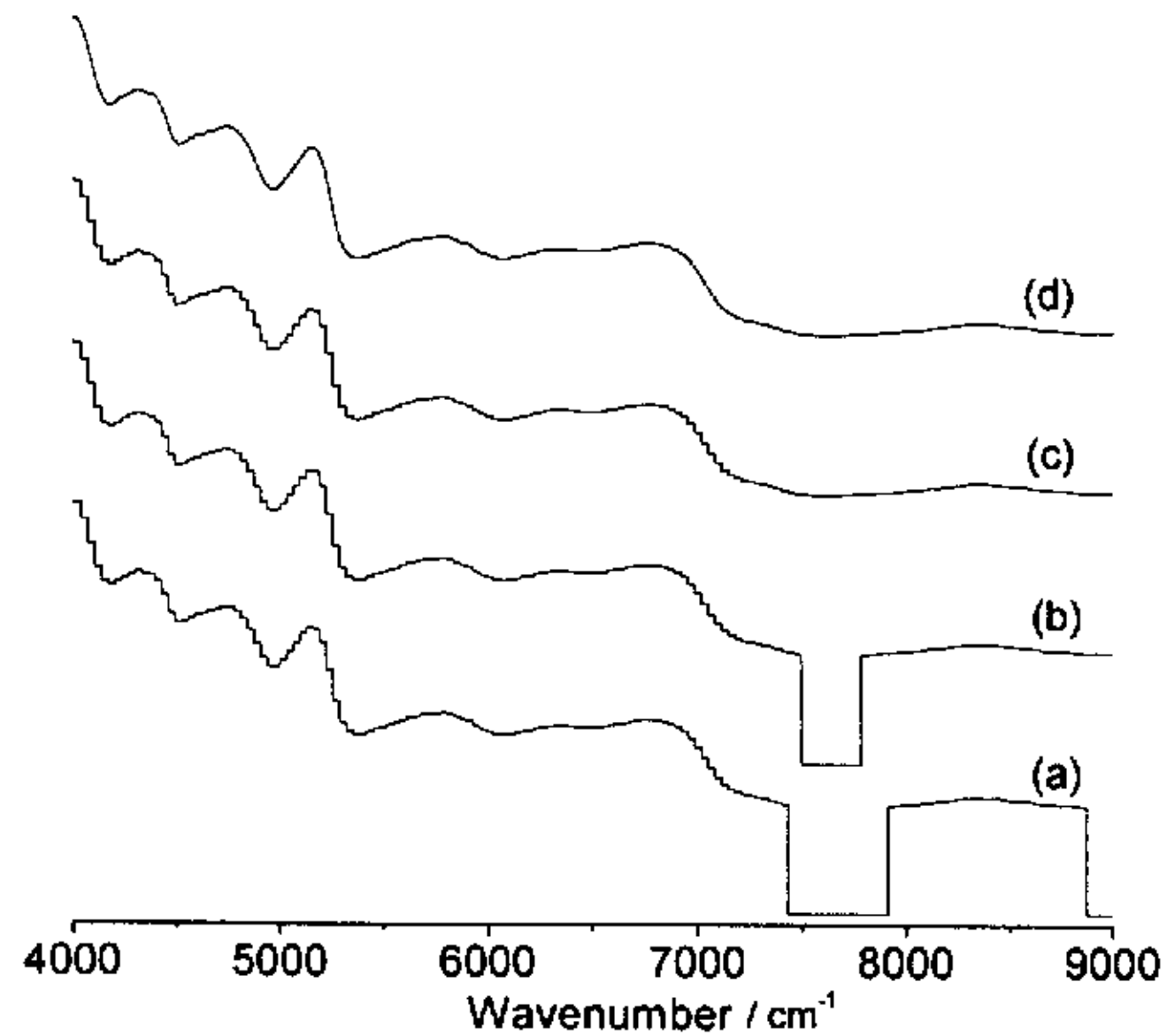


图 4-1 通过 140 (a), 150 (b) 和 160 (c) 个小波系数重构的光谱以及与原始平均光谱(d)的比较

#### 4.4.2 因子数的确定

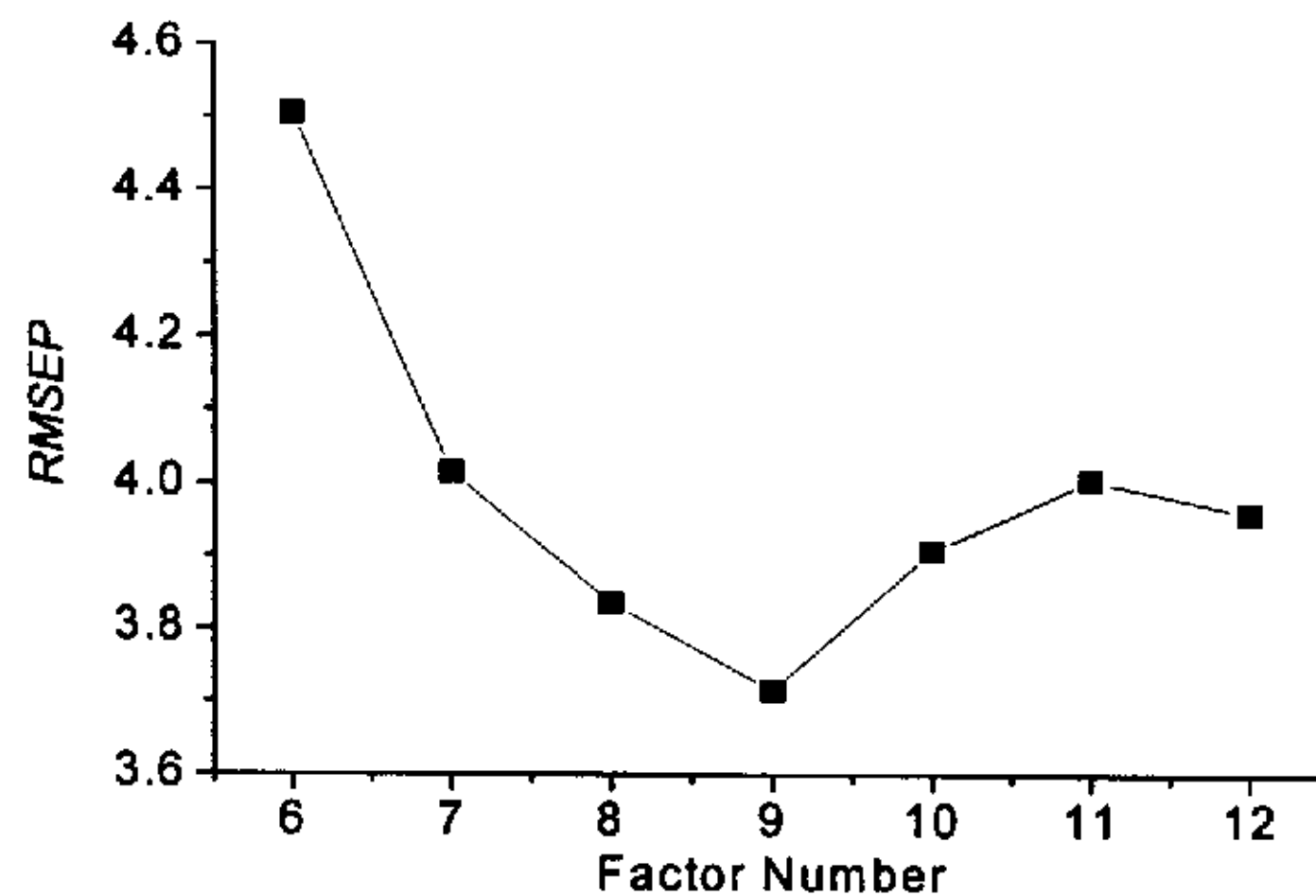


图 4-2  $RMSEP$  与因子数的关系

校正模型中所用的因子数同时影响到预测的正确性和计算速度。为了得到 PLS 模型的合适因子数，我们同样采用了留半交叉验证法。使用不同因子数所得到的 *RMSEP* 结果与因子数的关系如图 4-2 所示。很明显，因子数为 9 时可以得到最优的结果。

#### 4.4.3 UVE 变量筛选

通过使用 UVE 方法，我们分别计算了小波系数矩阵和随机数矩阵的稳定性。作为一个示例，图 4-3 显示的是预测组分 TS 的稳定性，AS 和 TN 组分的稳定性图形与之相似。图中虚线为  $k = 0.7$  所得的上下边界。所有稳定性在边界以内的系数，可以认为它们对校正模型无贡献，从而可以被去除。经过 UVE 方法筛选后，三种组分 TS，AS 和 TN 保留的系数分别为 97、63 和 78。

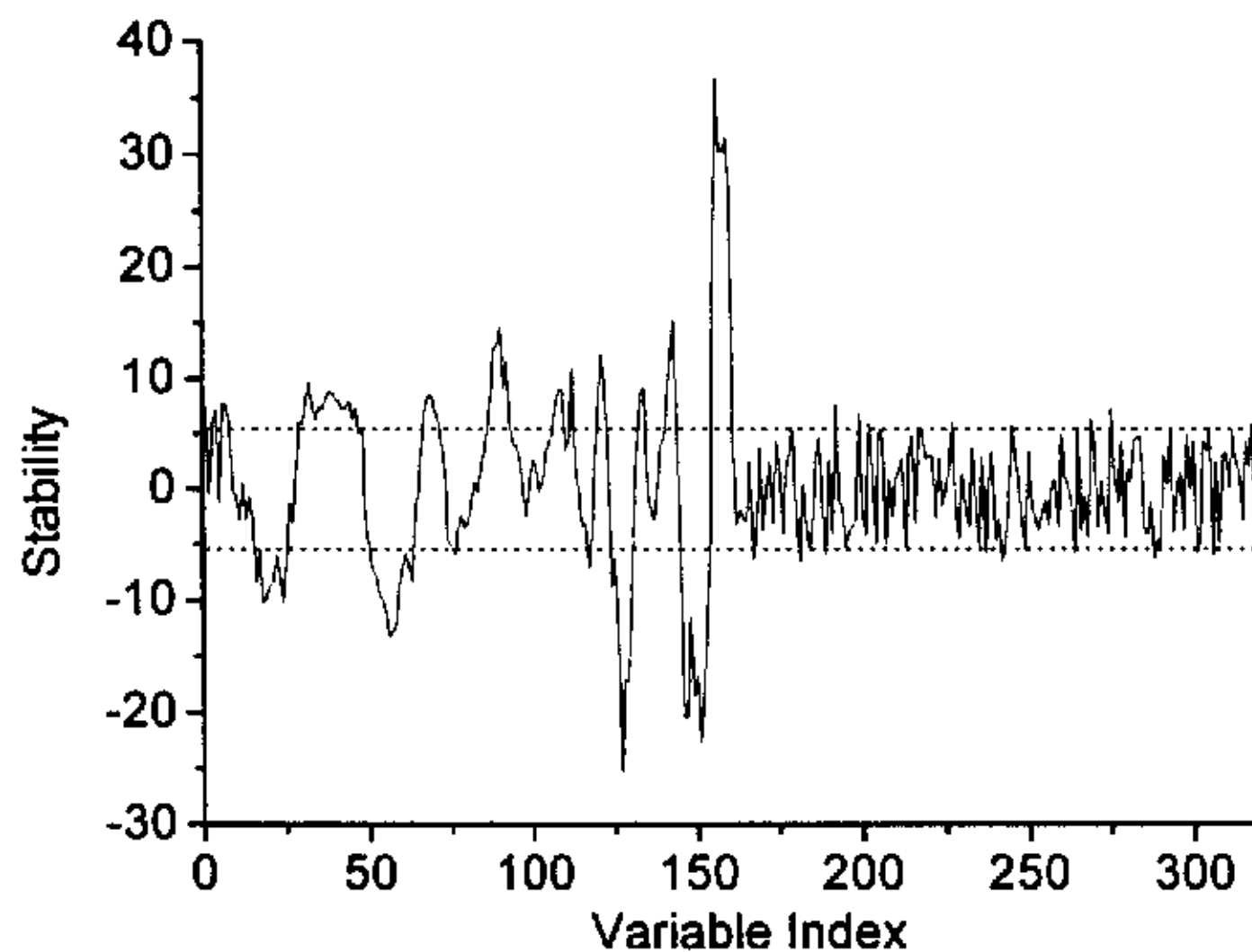


图 4-3 小波系数矩阵和随机数矩阵的稳定性  
(虚线为  $k = 0.7$  所得的上下边界)

为了检验变量筛选的效果，我们将保留下来的小波系数进行重构。图 4-4 (a)、(b)和(c)分别显示了对应于 TS、AS 和 TN 三种预测组分重构谱图。由于小波压缩中使用了 Haar 小波，重构所得图形有些像柱状图，其中柱形区域表示校正模型选用了这些波长，而平整区域的波长则表示它们没有被选用。

对这些图形进一步分析，可以得到比较有趣的结果。首先从三个图都可以看出，原始光谱中的平坦区域经过 UVE 方法后几乎全部被消除。这主要是由于光谱中这些平坦区域不包含有用的信息。另外，通过三个图形之间的比较，对于不

同组分的分析所保留的区域明显不同。

尽管很难将光谱区域的归属与特定的化学结构一一对应起来,但是我们发现在用于分析 TS 的重构光谱中包含着可归属为 C-H 和 C-C 吸收的区域。同样地, AS 和 TN 的光谱中包含 N-H 的吸收区域。另一方面,从这些结果我们可以看出,变量筛选过程对于近红外光谱的多元校正具有相当重要的意义。而且对于不同的组分分析建立不同的模型可以得到更好的结果。

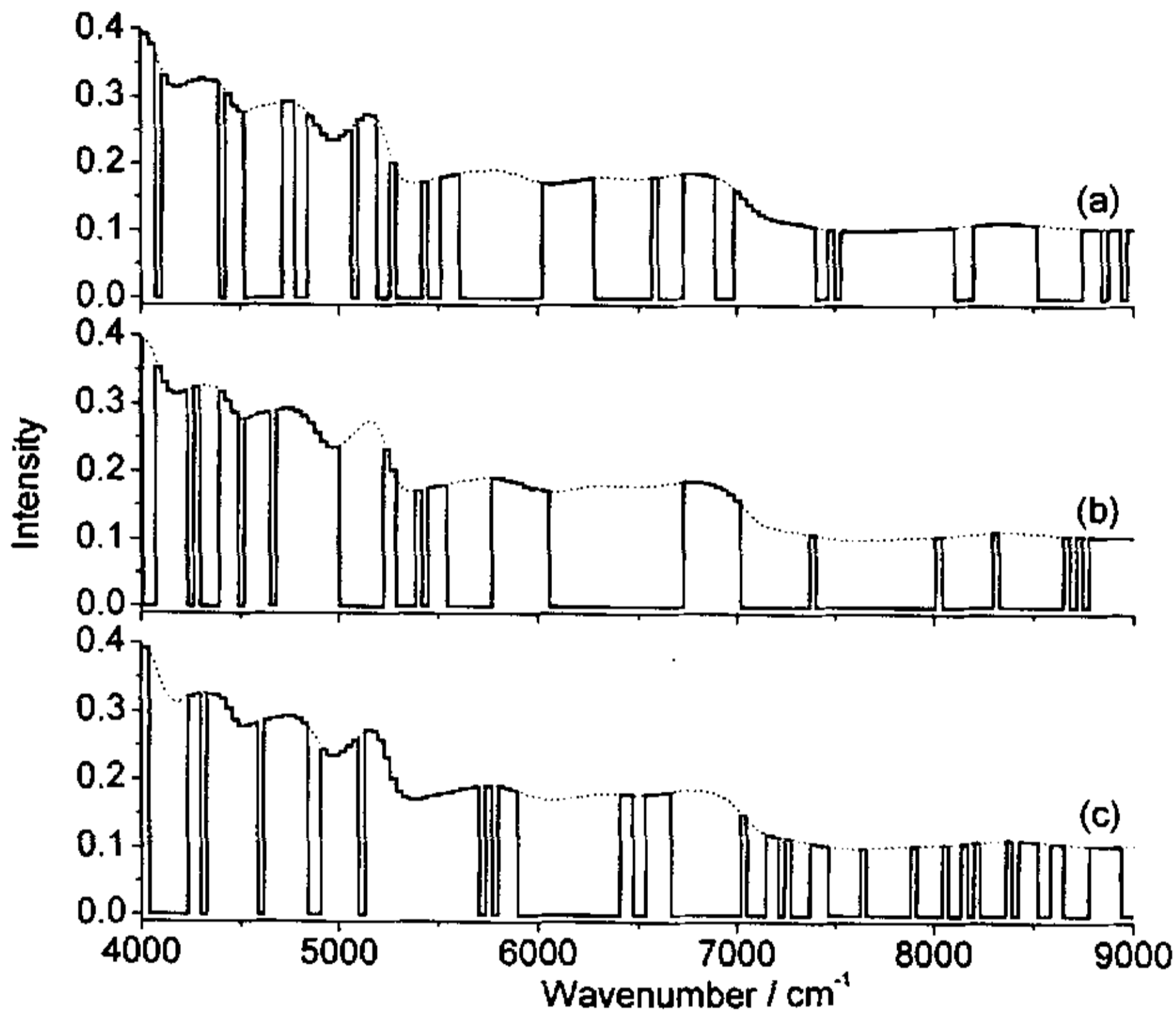


图 4-4 变量筛选后对应于 TS(a)、AS(b)和 TN(c)三种预测组分重构谱图

#### 4.5 预测结果

图 4-5 显示了三种预测成分的测量浓度和 WT-UVE-PLS 模型的预测浓度之间的关系曲线,图中(a)、(b)和(c)分别对应于 TS、AS 和 TN,  $a$ 、 $b$  和  $r$  分别为由最小二乘回归所得到的截距、斜率和相关系数,20 个预测样品的  $RMSEP$  由方程(4-6)计算得到。以星号标注的两个点分别对应于最高和最低的回收率。通过这些曲线图和数据可以明确地看出,本方法可以比较准确地用于烟草样品中 TS、AS 和 TN 三种组分的含量预测。



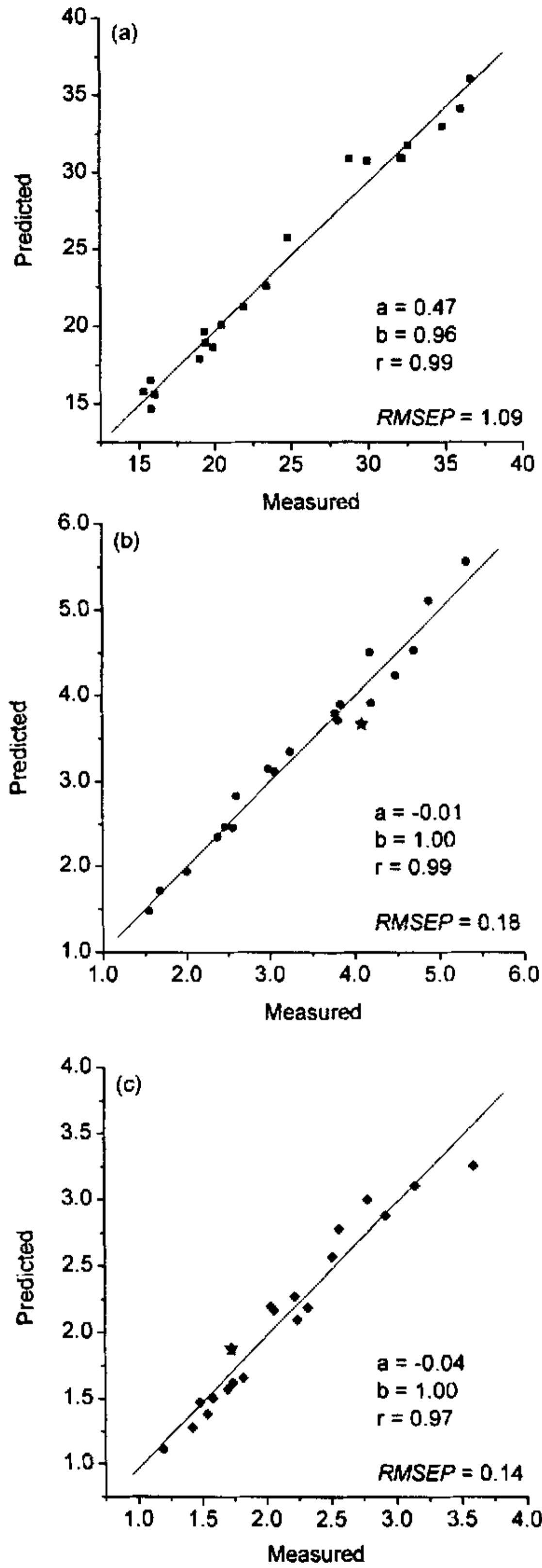


图 4-5 TS (a), AS (b)和 TN (c)的实验值与 WT-UVE-PLS 方法预测值之间的关系  
 (星号标注的两个点分别对应于最高 108.8%和最低的回收率 90.0%)

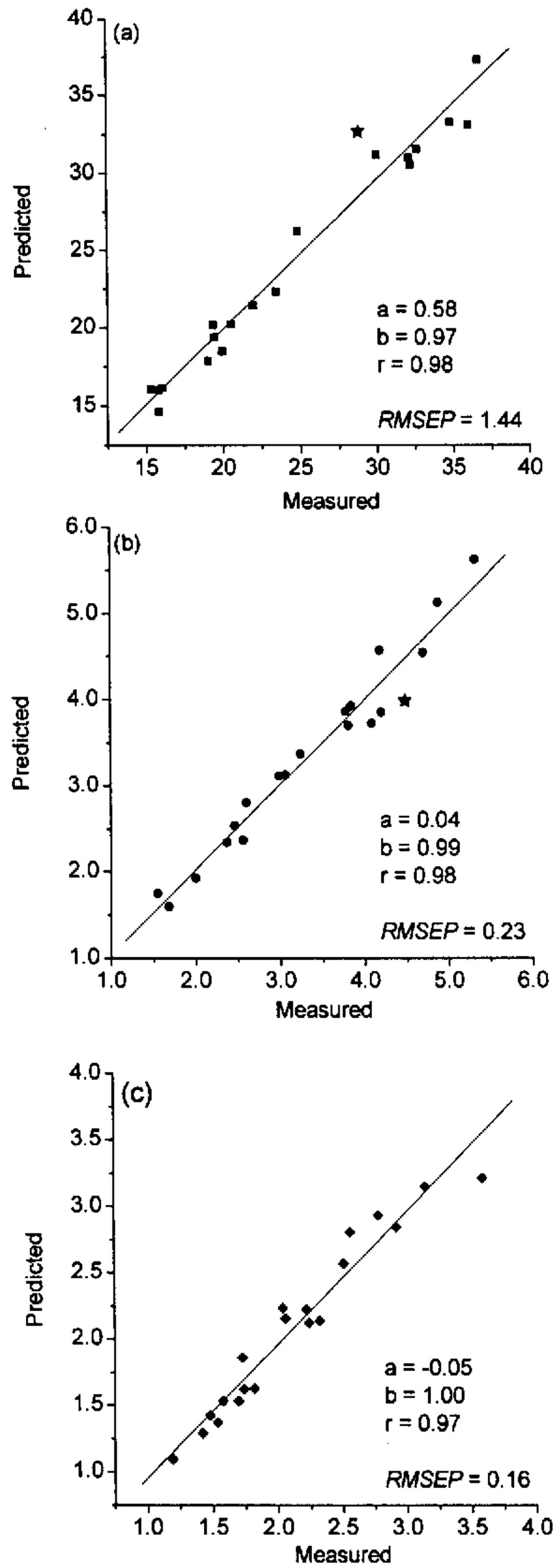


图 4-6 TS (a), AS (b)和 TN (c)的实验值与 PLS 方法预测值之间的关系  
(星号标注的两个点分别对应于最高 113.3%和最低的回收率 88.9%)

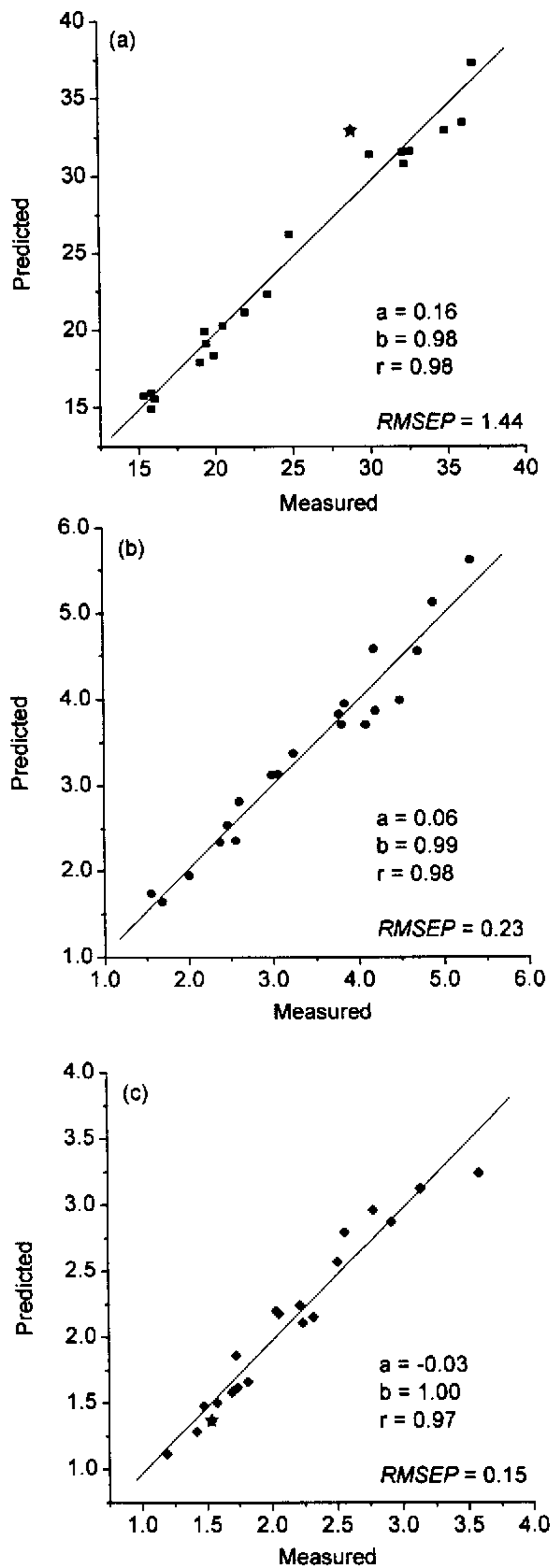


图 4-7 TS (a), AS (b)和 TN (c)的实验值与 WT-PLS 方法预测值之间的关系  
 (星号标注的两个点分别对应于最高 114.1%和最低的回收率 88.9%)

为了对本方法的预测结果进行检验，我们还列出了以 PLS 算法和 WT-PLS 算法所得的结果，分别如图 4-6 和图 4-7。可以看出，引入小波压缩方法仅可以提高运算速度，对预测结果改善并没有帮助，而使用 WT-UV-PLS 方法处理数据不仅可以获得很高的压缩比，同时还大大改善了预测结果。

## 4.6 结论

在本文中，我们提出了一种基于 WT 和 UVE 的数据压缩和变量筛选方法用于处理近烟草样品的红外光谱数据。经过该算法的压缩和筛选，原始的具有几千个数据点的谱图可以缩减到少于 100 个点。由于 UVE 算法可以将校正模型中不相关的变量剔除，因此预测结果比使用全谱或直接使用小波系数进行校正更精确。此外，对于不同的预测组分，使用 UVE 方法进行变量筛选时可以得到不同的选择结果，即所得到的光谱区域不同。因此，对于每个组分使用不同的校正模型可以得到更好的预测结果。本方法不仅为近红外光谱的数据处理提供了一种有效手段，同时还可以应用于其它复杂植物样品的分析。

## 参考文献

- [1] M. Ratcliffe, J. F. Panozzo, *J. Inst. Brew.*, 105 (1999) 85.
- [2] A. M. Bruno-soares, I. Murray, R.M. Paterson, J.M.F. Abreu, *Animal Feed Sci. Tech.*, 75 (1998) 15.
- [3] T. Borjesson, B. Stenberg, B. Linden, A. Jonsson, *Plant and Soil*, 214 (1999) 75.
- [4] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Pages, *Chemom. Intell. Lab. Syst.*, 50 (2000) 75.
- [5] A. Hoskuldsson, *Chemom. Intell. Lab. Syst.*, 55 (2001) 23.
- [6] J. Trygg, S. Wold, *Chemom. Intell. Lab. Syst.*, 42 (1998) 209.
- [7] X.G. Shao, W.S. Cai, *Rev. Anal. Chem.*, 17 (1998) 235.
- [8] D. Jouan-Rimbaud, B. Walczak, R. J. Poppi, O. E. deNoord, D. L. Massart, *Anal. Chem.*, 69 (1997) 4317.
- [9] B.K. Alsberg, A.M. Woodward, M.K. Winson, J.J. Rowland, D.B. Kell, *Anal. Chim. Acta*, 368 (1998) 29.
- [10] V. Centner, D.L. Massart, O.E. deNoord, S. deJong, B.M. Vandeginste, C. Sterna, *Anal. Chem.*, 68 (1996) 3851.
- [11] L. Pasti, D. Jouan-Rimbaud, D.L. Massart, O.E. deNoord, *Anal. Chim. Acta*, 364 (1998) 253.
- [12] R.K.H. Galvao, M.F. Pimentel, M.C.U. Araujo, T. Yoneyama, V. Visani, *Anal. Chim. Acta*, 443 (2001) 107.
- [13] T. Eklov, P. Martensson, I. Lundstrom, *Anal. Chim. Acta*, 381 (1999) 221;
- [14] M.C.U. Araujo, T.C.B. Saldanha, R.K.H Galvao, T. Yoneyama, H.C. Chame, V. Visani, *Chemom. Intell. Lab. Syst.*, 57 (2001) 65.
- [15] F. Lindgren, P. Geladi, A. Berglund, M. Sjostrom, S. Wold, *J. Chemom.*, 9 (1995) 331.

- [16] Walczak B., Massart D. L., *Chemom. Intell. Lab. Syst.*, 36 (1997) 81;
- [17] Leung A.K.M., Chau F.T., Gao J.B., Shih T.M., *Chemom. Intell. Lab. Syst.*, 43 (1998) 69.

## 第五章 微波辐射对烟草样品的物理与化学作用研究

### 5.1 微波辐射与物质的相互作用

微波是频率大约在 300MHz-300GHz, 即波长在 100cm 至 1mm 范围内的电磁波。它位于电磁波谱的红外辐射(光波)和无线电波之间。

在许多科学领域, 微波得到了成功地应用。在化学领域中, 微波在无机合成、有机合成、试样处理、环境化学等方面的研究中获得了成功地应用<sup>[1]</sup>。实践证明微波技术可以极大地提高化学反应速度。关于微波的作用机理, 可总结为两个主要方面, 即: ①微波的热效应。由于物质分子的偶极振动同微波振动具有相似的频率, 在快速振动的微波磁场中, 物质分子吸收微波的能量而产生热能。一般来说, 物质的介电常数越大, 对微波的耦合作用越强, 极性分子同微波有较强的耦合作用, 非极性分子同微波产生弱耦合作用或不产生耦合作用。微波加热特点在于加热速度快和“内加热”(由内而外加热)。②微波具有“特殊效应”, 这种效应可降低化学反应的活化能, 从而对化学反应具有催化作用<sup>[2]</sup>。

在烟草研究中, 微波技术的应用研究也日益展开。首先, 微波作为一种能源可用于烟丝的快速干燥, 不仅可以快速加热, 而且还可以克服常规加热干燥的某些缺陷<sup>[3]</sup>。由于微波辐射的特点, 当采用一定频率的微波对烟丝进行辐照后, 烟气中的总粒相物(T.P.M)、焦油量以及总糖的含量均有明显下降, 但香气和吸味没有明显变化<sup>[4]</sup>。在烟叶调制期间进行微波辐照对烟叶中的致癌物质——亚硝胺类化合物的含量也有显著影响<sup>[5]</sup>。另外, 微波可对烟草中的氨基酸和糖类物质的影响也有报导, 一方面微波的照射可促进氨基酸的降解, 另一方面氨基酸和糖类发生的 Maillard 反应<sup>[6]</sup>。

本文提出了一种采用微波辐射技术对烟叶或烟丝样品进行处理的方法, 通过烟气粒相物、气相物的成分分析和烟叶表面的组织结构分析, 探讨了微波辐射对烟叶和烟丝样品的物理作用和化学作用, 并对评吸结果进行了比较。

## 5.2 实验部分

### 5.2.1 材料与仪器

1. 烟叶和烟丝：红旗渠配方中的贵州 C<sub>3</sub>L，红旗渠配方烘丝处理后的叶丝。
2. 试剂：二氯甲烷(A.R.)。
3. 仪器设备：微波炉(2450MHz, 900W)；吸烟机(JJ-3A 型, 郑州烟草研究院)；旋转蒸发仪(RE-52A, 上海亚荣生化仪器厂)；气质联用仪(Shimadzu GCMS-QP5000, Agilent GCMS-5973N)；电子显微镜(Hitachi, X-650)；热解进样器(SGE)；固相微萃取进样器(SUPELCO, SPME fiber assembly 5-7312, 5-7300)。

### 5.2.2 实验方法

1. 烟叶或烟丝样品的微波辐射处理：将烟叶或烟丝样品放置于微波炉的托盘上，时间分别设置为 1s, 10s, 20s 进行辐照处理。

2. 烟气的收集与分析：首先将微波处理前后的烟丝样品制成卷烟样品，然后经称重、吸阻检测以及外观评价等严格挑选品质相似的卷烟样品 20 支，于吸烟机上采用剑桥滤片过滤对气相物进行收集。存放 2 小时后采用 SPME-GC-MS 进行成分分析，固相微萃取时间为 30min。GC-MS 分析中，色谱柱为 DB-WAX 毛细管柱(30m, i.d. 0.25mm)；载气为氦气；进样口温度为 230°C；离子源温度为 250°C；柱温升温程序为 50°C(2min) - 4°C/min - 200°C(10min) - 5°C/min-220°C。

3. 烟气粒相物的收集与分析：首先采用上述方法进行样品的制备与选择，并于收集剑桥滤片上的总粒相物。采用 20ml 二氯甲烷分别对每一滤片进行萃取，萃取时间为 40min。萃取液混合后，利用旋转蒸发仪浓缩至 2~3ml。最后采用 GC-MS 进行成分分析，GC-MS 的测试条件同上。

4. 烟丝热解产物的分析：首先将微波处理前后的烟丝样品分别制备为粉末，然后利用热解进样器对粉末样品的热解产物进行 GC-MS 分析。在 GC-MS 分析中，色谱柱为 HP35 毛细管柱；载气为氦气；分流比为 1:1；热解温度为 600°C；



柱温升温程序为 35°C(3min)-3°C/min-80°C(1min)-12°C/min-200°C(8min)。

5. 烟叶的形貌分析：根据外观评价选取同一烟叶叶片上外观平整的部位并分为两部分，一部分进行微波处理，另一部分作为对照。取两个样品的相似部位约 10~20mm<sup>2</sup> 水平置于样品架上，首先在样品的整个尺寸范围内进行观测，比较表面结构的差异，然后选取代表性部位，在相同条件下分别拍摄两个样品表面的组织结构照片。实验中，加速电压为 1KV~40KV，探针电流为 10<sup>-6</sup>-10<sup>-12</sup>A。

### 5.3. 微波辐射对烟气成分的影响

#### 5.3.1 微波对烟气粒相物组成成分的影响

为了考察微波辐射对烟气组成成分的影响，首先采用 GC-MS 手段对粒相物的组成成分进行了分析。在对照样品和微波辐照后样品的 GC-MS 色谱图中，保留时间小于 36min 的区间共有 8 种主要成分，表 5-1 为结构解析的结果。从谱图的比较可以看出，8 个色谱峰的相对峰高或峰面积无明显变化。这说明微波辐照对这 8 种化合物在粒相物中的含量并无明显影响。

表 5-1 烟气粒相物中沸点较低的主要成分

编号	化合物名称	编号	化合物名称
1	乙酸	5	苯酚
2	1,2-丙二醇	6	3-苯基-2-丙烯醛
3	3-(1-甲基-2-吡咯烷基)-吡啶 (尼古丁)	7	3,4-二氢-2H-吡咯-5-醇
4	1-十八碳炔	8	1,2,3-丙三醇三乙酸酯

图 5-1 为微波处理前后烟气粒相物 GC-MS 色谱图在保留时间区间 36~50min 的“指纹图”比较。可以看出，两张谱图中色谱峰数目较多且存在重叠现象。尽管通过色谱条件的优化等技术手段可以对谱图进行改善，但很难对每一种组成成分进行解析。因此，我们只对此两张谱图进行了指纹比较。可以发现，微波处理后，色谱峰的数目有所减少，即微波处理使粒相物的组成成分有所简化，但变化不大。

### 5.3.2 微波对烟气气相物组成成分的影响

气相物是烟气的主要组成部分(约占总烟气量的 92%),也是影响卷烟产品评吸指标的主要因素。因此,微波对烟气气相物的影响是本文的主要研究内容之一。为了对烟气气相成分的全面考察,我们采用了两种不同极性的固相微萃取进样器,即:极性较强的 5-7312 型和极性相对较弱的 5-7300 型,对烟气气相成分中的极性较强和极性较弱的化合物分别进行了 GC-MS 分析。

在强极性化合物的色谱图中,共检测到 40 余个色谱峰,去除强度较低的色谱峰后,得到了 26 种化合物。表 5-2 为部分谱峰的解析结果,其中最后一列的“含量变化”是根据对照样品和微波处理(10s)样品的谱图比较得到的相对变化。从表中的数据可以看出,微波处理后有 12 种化合物的含量几乎不发生变化,6 种化合物几乎完全消失,5 种化合物含量减少,1 种化合物含量增加,新增加了 2 种化合物。

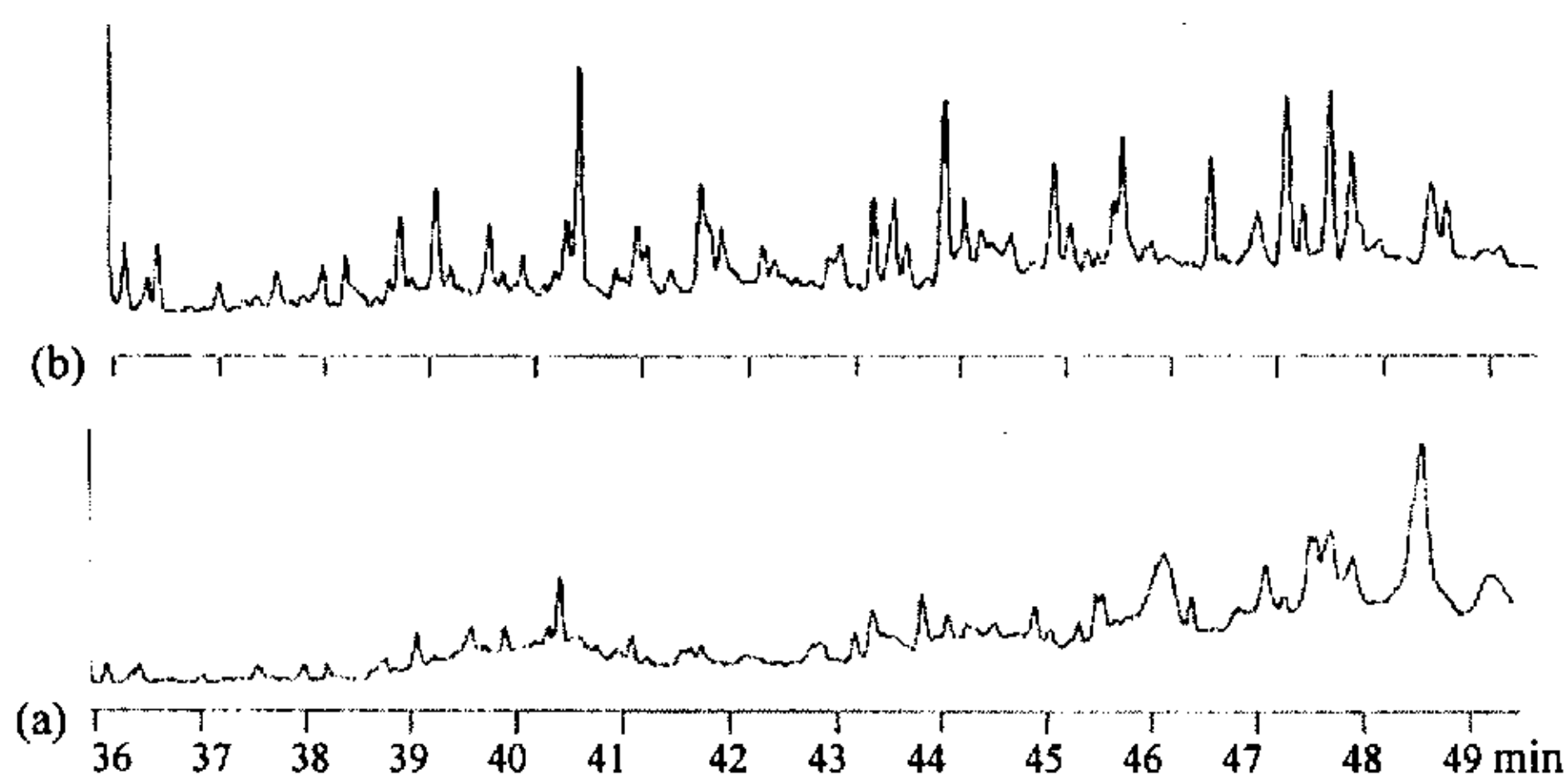


图 5-1 微波处理前后烟气粒相物 GC-MS 色谱图的“指纹图”比较

(a)对照样; (b)微波处理(10s)

图 5-2 为弱极性化合物 GC-MS 分析的色谱图,其中谱图(a)为对照样品的色谱图,谱图(b)和(c)分别为微波处理 1s 和 10s 后样品的色谱图,(b)对应的处理时间较短,(c)对应的处理时间较长。表 5-3 总结了三张谱图中部分谱峰的结构解析结果。从表中的数据可以看出,微波处理时间较短时,有 6 种化合物的含量不发生变化,20 种化合物几乎完全消失,新增加了 2 种化合物。但随着微波处

理时间的增加,只有3种化合物的含量不受影响,23种化合物几乎完全消失,新增加的2种化合物中也几乎完全消失。

为了进一步验证上述实验结果,我们采用热解进样的GC-MS对热解产物进行了分析。表5-4列出了微波处理前后有明显变化的某些芳香烃类化合物及其衍生物。从表中数据可以看出,大多数芳香烃类的化合物均有不同程度的减少或完全消失。

表 5-2 微波处理对烟气气相物中强极性组分的影响

编号	化合物名称	含量变化 <sup>(a)</sup>	编号	化合物名称	含量变化
1	乙酸	=	14	2-甲基-1-戊醇	-
2	草酸	-	15	1,6-庚二烯-4-醇	=
3	2-甲基-2-丁烯酸	=	16	4-环戊烯-1,3-二酮	D
4	4-戊烯酸	D	17	3-甲基丁酸	=
5	丁酸乙烯酯	=	18	3-甲基-2,5-咪喃二酮	=
6	1,1-乙二醇二乙酸酯	D	19	2-羟基吡啶	=
7	烯丙基(2-甲基-烯丙基)醚	=	20	3,4-二氢-6-甲基-2H-吡喃-2-酮	-
8	烯丙基乙烯基醚	-	21	2-甲基四氢咪喃	=
9	4-戊烯醛	=	22	2-甲基-环戊酮	=
10	2-己烯醛	=	23	3,5-双(1,1-二甲基乙基)-3,5-环己二烯-1,2-二酮	N
11	3-甲基-4-甲烯基-2-己酮	D	24	苯并噻唑	+
12	3-己烯-2,5-二酮	-	25	4-吗啉乙胺	D
13	1,2-乙二醇	N	26	4-吗啉乙醇	D

<sup>(a)</sup>注释:“=”表示微波处理后该化合物的含量没有明显变化,“-”表示微波处理后该化合物的含量降低,“D”表示微波处理后该化合物消失,“+”表示微波处理后该化合物的含量增加,“N”表示微波处理后新增加的化合物。

根据以上结果不难看出,微波辐射对烟气气相物的组成有较大影响,使一部分组分的含量减少甚至完全消失,同时也产生了少量的新组分。结合表5-2、表5-3和表5-4中化合物的类型和含量变化情况,我们可以总结出以下结果:

(1) 微波辐照后含量减少或消失的化合物有两种情况,即:

- (1.1) 沸点较低的化合物;
- (1.2) 具有一定极性的醇、酮、醛、醚以及取代杂环化合物;
- (2) 结构稳定且极性较小的化合物一般不发生变化;
- (3) 微波辐照后新产生的化合物一般具有结构稳定的结构且分子量较大、沸点较高。

因此,我们可以认为微波对烟叶或烟丝样品产生了两种作用。一种作用为加热作用,使烟叶中低沸点的化学成分挥发而含量减少或消失;另一种作用则是所谓的“特殊效应”,使烟叶中的某些化学成分之间发生了某些化学反应。

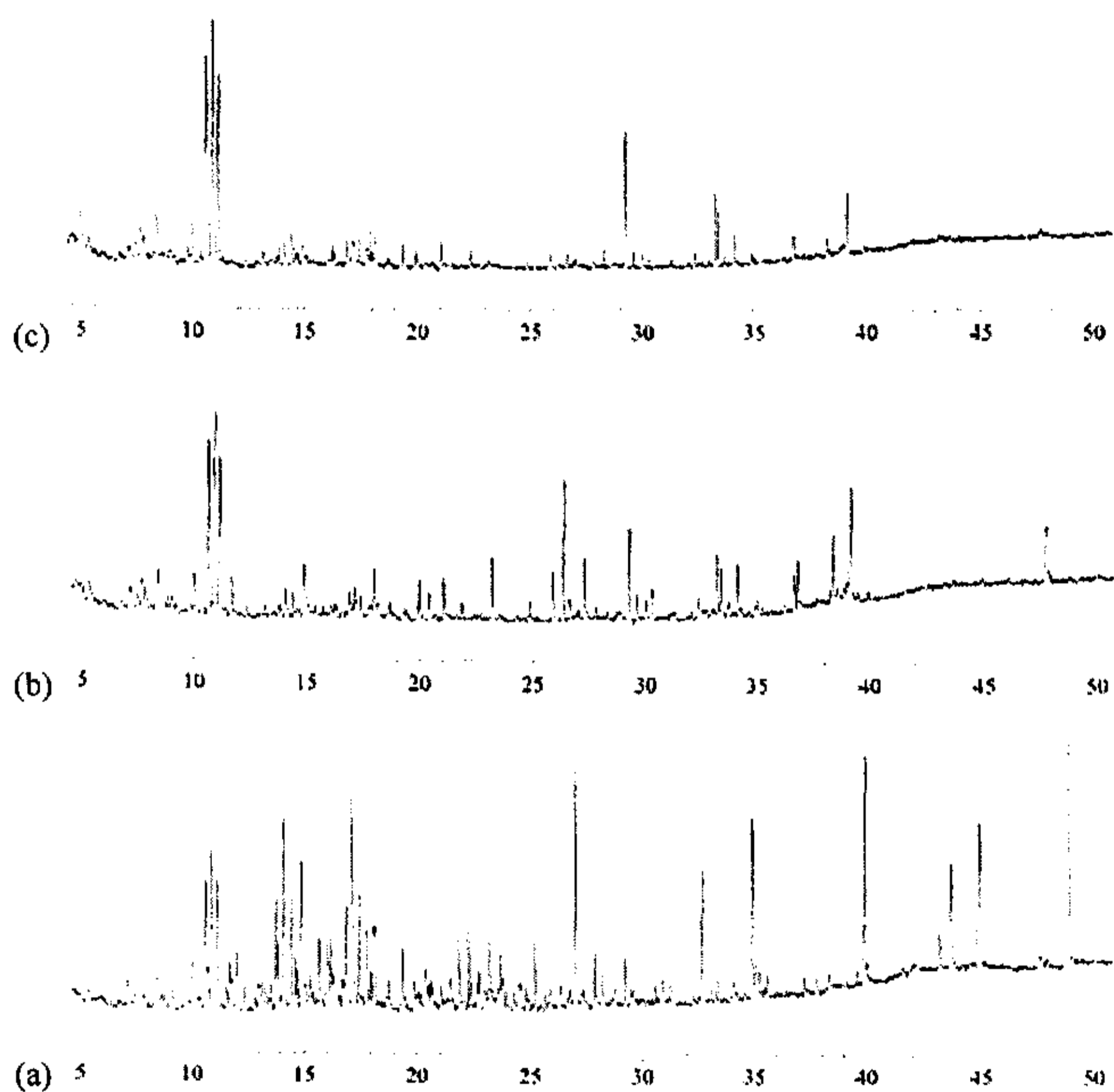


图 5-2 微波处理前后弱极性烟气气相物 GC-MS 色谱图的比较

(a)对照样; (b)微波处理(1s); (c)微波处理(10s)

### 5.3.3 微波对烟叶表面组织结构的影响

由于微波的加热效应,微波辐照对物质的物理结构也会产生影响<sup>[7]</sup>。为了探

讨论微波辐照对烟叶表面组织结构的影响,我们采用电子显微镜技术对微波辐照前后烟叶表面的组织结构进行了分析和比较。图 5-3 为微波辐照前后烟叶表面的组织结构照片比较。可以看出,微波辐照前,烟叶表面的组织结构致密且较平滑,微波辐照后,烟叶表面的组织结构变得疏松且褶皱增加。此结果表明微波的辐照使烟叶的组织结构发生了变化。造成此结果的原因可能是由于微波的快速“内加热效应”,烟叶内部的挥发物向外运动导致了物理结构的变化。因此,微波辐照可使烟叶的结构更加蓬松,增加了烟叶的填充性能和燃烧性能。

表 5-3 微波处理对烟气气相物中弱极性组分的影响

编号	化合物名称	含量变化 <sup>(a)</sup>	编号	化合物名称	含量变化
1	十二烷	DD	15	2,4-二甲基-1-(1-甲基乙基)苯	DD
2	十三烷	DD	16	1,2,4,5-四甲基苯	DD
3	十四烷	DD	17	1,2,3,4-四甲基苯	DD
4	2,3,4-三甲基己烷	=D	18	萘	=D
5	2,6-二甲基辛烷	DD	19	芴	ND
6	2,6-二甲基壬烷	DD	20	菲	ND
7	2,8-二甲基十一烷	DD	21	1-庚烯-3-炔	DD
8	2,6,11-三甲基十二烷	DD	22	2-己基-1-辛醇	DD
9	2-甲基十七烷	DD	23	6-十一胺	DD
10	7-甲基-1-十一烯	==	24	1-氯十八烷	DD
11	2,4,6,8-四甲基-1-十一烯	==	25	2-庚烯酸乙基酯	DD
12	1,4-二乙基苯	DD	26	4-戊酮	DD
13	1-乙基-2,3-二甲基苯	DD	27	1-甲基-2-亚硝基苯	DD
14	4-乙基-1,2-二甲基苯	=D	28	苯并噻唑	==

(a) 注释:“=”表示微波处理后该化合物的含量没有明显变化,“D”表示微波处理后该化合物消失,“N”表示微波处理后新增加的化合物。第一列表示微波处理时间短(1s),第二列表示微波处理时间长(10s)。“=D”表示短时间处理没有明显变化,但处理时间长时消失,“ND”表示处理时间短时新产生的化合物在长时间处理后又消失。

### 5.3.4 卷烟样品的评吸结果

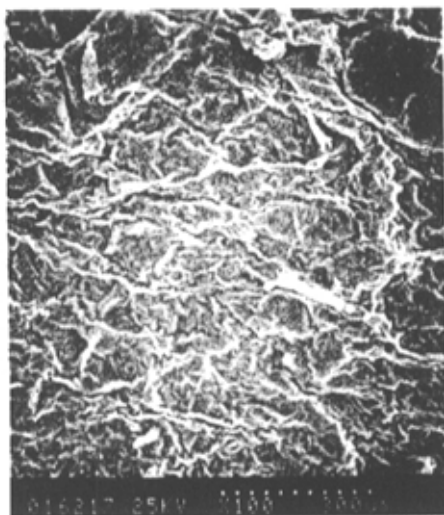
为了比较微波辐照前后样品内在质量的变化,我们将微波处理前后的烟丝卷

制成卷烟样品。经过挑选后进行了评吸,评吸结果表明:香气质有所改善,杂气减少,刺激性减弱,余味变的更加舒适。此结果与文献<sup>[4]</sup>的结果有所不同。

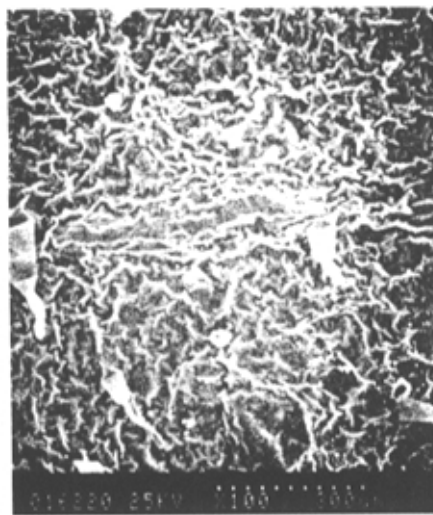
表 5-4 热解-GC-MS 对芳香烃及其衍生物的分析结果

编号	化合物名称	含量变化 <sup>(a)</sup>	编号	化合物名称	含量变化
1	苯	-	13	3-甲基苯酚	D
2	甲苯	-	14	联苯	-
3	乙苯	-	15	(4-甲基苯基)苯	D
4	(1-甲基乙基)-苯	D	16	2,3-苯并呋喃	-
5	1,4-二甲苯	-	17	1,2-苯并吡咯	D
6	1,2-二甲苯	-	18	萘	-
7	1-甲基-4-乙烯基苯	-	19	2-甲基萘	-
8	丙烯基苯	D	20	2,3-二甲基萘	-
9	烯丙基苯	D	21	1,6-二甲基萘	-
10	丙炔基苯	D	22	2-乙烯基萘	-
11	苯酚	D	23	亚联苯	D
12	2-甲基苯酚	D	24	蒽	-

<sup>(a)</sup>注释:“-”表示微波处理后该化合物的含量降低,“D”表示微波处理后该化合物消失。



(a) 对照样品的表面结构



(b) 微波处理后烟叶样品的表面结构

图 5-3 微波处理对烟叶表面组织结构的影响

## 5.4 结论

本文探讨了微波辐射对烟叶和烟丝样品的物理与化学作用,结果表明:经过微波的辐照处理,烟气的化学成分有明显变化;微波辐照对烟叶表面的组织结构有较大影响;微波辐照对卷烟的内在质量有所改善。通过本文的实验结果,我们可以得到如下结论:

(1) 微波对烟叶或烟丝具有快速加热作用,使低沸点的化学成分挥发而使烟气中这些成分的含量减少甚至消失。

(2) 由于微波的加热是“内加热”,内部成分向外部的运动对烟叶的组织结构有较明显的作用,使烟叶的结构蓬松,对增加烟叶的填充性和燃烧性具有一定意义。

(3) 微波对结构稳定且极性较小的化学成分影响较小。

(4) 微波的辐照可能导致了烟叶内部的某些化学反应,使烟气中产生了一些新的化合物。

(5) 微波辐照对卷烟的内在质量有所改善。

总之,本文的研究表明了微波辐射对烟叶或烟丝具有较明显的物理和化学作用,但由于本文只做了初步的研究,加之烟草体系的复杂性,此课题仍有许多问题需要进一步的工作,例如微波对烟气中化学成分含量变化的定量研究,变化量与微波辐照量的关系,化学成分含量变化的机理以及微波对烟叶内部组织结构的影响等等。

## 参考文献

- [1] 金钦汉, 戴树珊, 黄卡玛, 微波化学, 2001, 科学出版社(北京)
- [2] 樊兴君, 尤进茂, 谭干祖等, 化学进展, 1998, 3(3):285~295
- [3] 邱纪青, 傅淑英, 郑新章等, 烟草科技, 1999, (6):3~5
- [4] 张明高, 杨经圆, 冉瑞红等, 烟草科技, 1999, (1):6~8
- [5] J.R. Williams, Tobacco products having reduced nitrosamine content, United States Patent, No.6135121, Oct. 24, 2000
- [6] W. Coleman, T. Perfetti, Beitrage zur Tabakforschung International, 1997, 17(3):75~95
- [7] 熊健, 叶君, 梁文芷等, 华南理工大学学报(自然科学版), 2000, 28(3):84~89



附录：攻读硕士学位期间完成论文情况

- [1] 王芳, 陈达, 邵学广  
近红外光谱与卷烟样品常规成份的关系模型研究  
草科技, 2002, (5):23-26
- [2] 王芳, 陈达, 邵学广  
小波变换—PLS 用于植物样品成份的定量预测研究  
烟草科技, 2003, (*In Press*)
- [3] 王国庆, 王芳, 陈达, 苏庆德, 邵学广  
近红外光谱技术用于复杂植物样品中无机离子测定的新方法  
光谱学与光谱分析, (*Accepted*)
- [4] 陈达, 王芳, 邵学广, 苏庆德  
近红外光谱与烟草样品总糖含量的非线性模型研究  
光谱学与光谱分析, (*Accepted*)
- [5] 庄亚东, 张映, 王芳, 邵学广  
卷烟中多酚类物质的分析  
烟草科技, (*Accepted*)
- [6] Da Chen, Fang Wang, Xueguang Shao and Qingde Su  
Elimination of interference information by a new hybrid algorithm for  
quantitative calibration of near infrared spectra  
Analyst, 2003, 128(9):1200-1203
- [7] Xueguang Shao, Fang Wang, Da Chen, Qingde Su  
A Method for near-infrared spectral calibration of complex plant samples with  
wavelet transform and uninformative variables elimination  
Analytical and Bioanalytical Chemistry, (~~Revised~~)  
*Accepted*