



GRADUATION DESIGN (THESIS)

专业班级: 统计学 1302 班

2017 年 6 月

当归指纹图谱数据分析与建模

摘要

当归指的是为伞状科植物当归的干燥根，是传统中医常见药材，在现代医学中有着公认的疗效。由于当归的内部化学成分较为复杂且在生长周期中所受到诸多环境因素和人为因素的影响，所以对当归的采摘时间以及采摘的量效考核很难控制。本文对采摘自甘肃岷县的 27 个不同生长期当归进行代谢组学的系统研究。

本文采用核磁共振（HNMR）技术以及气质联用（GCMS）技术两种代谢图谱分析方法相互耦合分析的方法，并结合 OPLS 偏最小二乘法和统计检验对不同采收期的当归次生代谢产物的动态积累情况进行分析。结果显示出：Z 蒿本内酯、阿魏酸、欧当归内酯 A 等 5 种是三年生长周期内变化较为活跃的药效成分，还有当归多糖、15 烷醇、丁烯基苯酞等非药效成份变化也很明显，这些可作为评价药材品质和控制采收时期标志化学成分；此外还结合图谱描述了他们在生长周期的变化情况，并结合以上化合物含量变化时的所处时期以及相应变化量，发掘次生代谢物对当归生长作用；在实验过程中还发现，在当归传统采收时节内并非药效成分的含量达到最大，而是 Z 蒿本内酯等药效成分与丁烯基苯酞等非药效成分之间的比例达到一个合理程度。这为当归药材的资源的合理利用、采收期确定提供理论依据。

本文在第五章中对比了核磁共振和气相色谱两种代谢组学的实验分析方法。先从统计检验和多元统计的实验结果出发，探索实验的设计自身的缺陷。如阿魏酸沸点高，无法在 GCMS 实验中的挥发油中存在，而 15 烷醇不溶于 DMSO 溶剂无法在核磁共振中检测。随后有对两种实验分析方法进行灵敏性测试，探索他们对当归中的化合物变化是否敏感。最后依据 GCMS 数据库的优势，在 HNMR 谱中成功定位 Z 蒿本内酯以及 E 蒿本内酯的化学位移，实现了代谢图谱之间的对比耦联分析。

关键词：当归 NMR GCMS 指纹图谱 OPLS 代谢组学 图谱耦联分析
统计检验

Analysis and model of fingerprint data of Angelica

ABSTRACT

Angelica refers to the dry roots of *Angelica sinensis*, an herb of traditional Chinese medicine, and has a commonly curative effect in modern medicine. Because the internal chemistry components of *Angelica sinensis* are very complex and are affected by many environmental factors and human factors during the growth cycle, it is difficult to control the plucking time and the dose effect of angelica's picking. In this paper, a systematic study on metabonomics of *Angelica sinensis* at 27 samples from Minxian, Gansu Province which is in the different growth stages was conducted.

In this paper, nuclear magnetic resonance (HNMR) technique and gas chromatography-mass spectrometry (GCMS) technology combined with the orthogonal partial least squares (OPLS) method and statistical test has analysed the dynamic accumulation of angelica metabolites in different harvest season. The results show that: Z-ligustilide, ferulic acid, and three other kinds are active pharmaceutical ingredients in three years of growth period, and the changes of polysaccharide, 15 alkanol, n-butylidenephthalide and other nonpharmacodynamic components are also obvious, which can be the symbolic chemical compositions used to evaluate the quality of angelica and control of the harvesting time period; In addition, according to the period when those changes happen and the corresponding amount of those changes, this paper has explored effects of metabolites on *Angelica sinensis* in growth cycle; In the experimental process, We has found that the medicinal ingredients in angelica sinensis don't reach the maximum content in the traditional harvest season, but the ratio between the Z-ligustilide and other drug ingredients with n-butylidenephthalide reached a reasonable level. That provides a theoretical basis for determining angelica harvest and the rational use of the angelica resource.

In the fifth chapter, we compare the difference of the HNMR method and GCMS method. Starting from the experimental results of statistical tests and multivariate statistics, the paper tries to find out the defects in the designs of two experiments. For example, ferulic acid has a high boiling point and can not be found in the volatile oil in the GCMS experiment, while 15 alkyl alcohol that insolubles in DMSO solvent can not be detected in nuclear magnetic resonance. Then we tested the sensitivity of the two experimental methods to find out whether they were sensitive to changes about the compounds in angelica. Finally, according to the advantages of GCMS database, the chemical shifts of the Z-ligustilide and E-ligustilide were successfully located in the HNMR spectra. So a comparative coupling analysis between metabolic spectra was realized.

Key words: *Angelica sinensis* NMR GCMS Fingerprint Metabolomics

Spectral coupled analysis statistical test

目录

第一章 绪论.....	1
1 选题背景及意义	1
2 文献综述	2
2.1 国外相关研究综述.....	2
2.2 国内相关研究综述.....	2
2.3 综述总结.....	2
3 当归药材的植物学概况	3
3.1 药材性状.....	3
3.2 样本产地分布.....	4
4 研究方法.....	4
5 主要创新点.....	4
第二章 不同生长期当归的代谢组学实验研究.....	5
1 代谢组学指纹图谱技术概述.....	5
1.1 核磁共振指纹图谱.....	5
1.2 GCMS 气相联用指纹图谱.....	6
2 当归样品的采集.....	6
3 不同生长期当归指纹图谱的测定.....	6
3.1 ^1H NMR 指纹图谱.....	7
3.1.1 实验结果及分析.....	7
3.1.2 当归样品 ^1H NMR 指纹图谱的归属.....	8
3.1.3 当归 NMR 图谱数据处理过程.....	9
3.2 GCMS 气相色谱联用指纹图谱.....	10
3.2.1 GCMS 气质联用实验结果.....	10
3.2.2 GCMS 代谢图谱的代谢组学处理与化合物鉴别.....	10
3.2.3 当归 GCMS 指纹图谱数据处理过程.....	13
第三章 基于方差分析模型的指纹图谱变量筛选.....	13
1 HNMR 核磁共振谱图方差分析模型	14
1.1 正态性检验—Epps-Pulley 检验.....	14
1.1.1 方法及原理.....	14
1.1.2 验证结果与分析.....	14
1.2 方差齐性检验—修正的 Bartlett 检验.....	15
1.2.1 方法及原理.....	15
1.2.2 验证结果与分析.....	16
1.3 方差分析统计模型分析结果.....	16
1.3.1 方法及原理.....	16
1.3.2 验证结果与分析.....	17
2 GCMS 气质联用图谱多总体方差分析模型	18
2.1 正态性检验—Epps-Pulley 检验结果与分析.....	18
2.2 方差齐性检验—修正的 Bartlett 检验结果与分析.....	19

2.3	方差分析统计模型分析结果与分析.....	19
第四章	基于 ^1H NMR 指纹图谱的当归 OPLS 判别分析.....	21
1	正交偏最小二乘回归原理与方法.....	21
1.1	基本原理.....	21
1.2	偏最小二乘的理论.....	22
1.3	正交偏最小二乘计算方法.....	22
1.4	正交成分个数的确定.....	23
2	核磁共振指纹图谱不同生长期当归 OPLS 分析结果.....	23
2.1	数据压缩结果.....	23
2.2	到模型的距离.....	25
2.3	主成分得分图.....	25
2.4	主成分荷载值.....	26
2.5	核磁共振化学位移的显著性检验.....	27
2.5.1	两个正态总体方差比的 F 检验.....	27
2.5.2	$\sigma_1 = \sigma_2 = \sigma$ 但未知时的两样本 t 检验.....	29
2.6	基于 HNMR 图谱数据 OPLS 模型预测与回判.....	32
3	气质联用总离子流图图谱不同生长期当归 OPLS 分析结果.....	33
3.1	代谢物的 OPLS-DA 分析.....	33
3.2	挥发性成分的动态变化.....	35
3.3	重要化合物依据年份的显著性 T 检验.....	36
3.3.1	两个正态总体方差比的 F 检验.....	36
3.3.2	$\sigma_1 = \sigma_2 = \sigma$ 但未知时的两样本 t 检验.....	37
4	结论.....	38
第五章	GC-MS 与 HNMR 两种代谢组学方法差异比较.....	39
1	代谢图谱物质定性差异.....	39
2	代谢指纹图谱对的敏感度分析敏感度.....	40
3	保留时间和核磁共振特征峰之间的关联分析.....	41
3.1	灰色关联分析法的建模过程.....	41
3.2	灰色关联分析结果.....	42
第六章	结果讨论.....	44
1	结论.....	44
2	体会与展望.....	45
致谢	46
参考文献	54

第一章 绪论

1 选题背景及意义

中药作为我国中医医学理论体系中实现治疗的药效基础，中药材的真伪，质量的优劣，临床用药的安全性，以及阐明其中药材药效基础及其作用机理，进行中药的质量控制和评价对于我国国粹之一的中医药走向世界，实现国际化有着重大意义。然而中药材的药用成分大多数为次生代谢产物，化学成分繁复，生物活性各异，疗效以及理化性质也迥然不同。而且植物药材的化学成分对于气候、年限、理化环境、采收时节或炮制方法等因素相当敏感。传统中药的安全性、有效性方面的现代科学研究方法论证的欠缺，也使使得中药的国际化受到一定的阻碍。

本论文所采用的当归药材是甘肃省的道地药材，占有全国当归药材的近 90%。而甘肃岷县是当归药材的主产区。随着经济的发展，人们对保健、疾病预防等越发重视，因此传统的当归药材在保健和医疗上的应用也就更加广泛。大量的医学保健用途的需求也使得对于当归药材深入研究尤为重要。就目前来看，对当归药材的研究很多，但大多都集中在对其单一化学成分的挖掘和研究上，对其整个生长周期内次生代谢物的积累变化情况的研究情况较少。由于当归药材的质量优对于生长环境，加工手法等诸多方面的因素较为敏感。所以对当归资源在生长周期上的整体利用将大大减少对道地药材资源的浪费，提高当归的使用效率。

而将 NMR 核磁共振技术和 GCMS 气质联用技术的代谢组学实验方法应用在传统的当归药材上，会在微观分子层面上对当归在生长周期内的化合物积累情况给予更明确的显示。并且结合采用了多元统计分析方法，对产自甘肃岷县不同生长期的当归样品代谢图谱分别进行较为系统的挖掘分析。还以 NMR 指纹图谱与 GCMS 指纹图谱为基础，对当归药材中的代谢成分随年份的积累变化情况进行跟踪分析，对当归药材的合理采摘和优种优育提供理论参考；最后还将 GCMS 指纹图谱和 ^1H NMR 指纹图谱两种技术对比分析，科学比对不同图谱分析技术的优劣，且两种指纹图谱作为相互补充，为当归药材的生长周期内的质量控制提供参考数据，对提高道地药材的利用率，减少中药资源的浪费具有重要的研究意义。

2 文献综述

2.1 国外相关研究综述

Fiehn 研究小组^[38]先是利用质谱色谱 GCMS 技术,对 4 中不同基因型笋瓜的韧皮部的 433 种次生代谢产物进行了代谢组学分析。然后利用化学计量学手段对植物的表现型进行归类,发现了亚油酸、葡萄糖、柠檬酸和果糖等四种代谢产物对笋瓜的年份分类有着重大贡献。Ward^[36]利用核磁共振技术,对产自不同生长环境下的拟南芥挥发油的指纹谱图进行了分析,利用 PCA 分析法对所有的拟南芥样品进行关于生长环境的分类,并发现对于环境有较大敏感性的代谢产物为氨基酸、有机酸和糖等。Tarachiwin 等^[5]基于 ^1H NMR 指纹谱图和多元统计分析方法,实现了对西瓜的质量评价体系和预言能力;Kang J 等^[6]为区别韩国药材和中国药材的差异,利用核磁共振技术对中药材和韩药材的内部化学结构进行详细阐述;Zulak 等^[42]利用精准定量的 ^1H NMR 代谢组学手段,分析罂粟的生长周期中初生代谢产物和次生代谢产物的相互演化关系;

2.2 国内相关研究综述

指纹图谱法在中医药领域拥有着更为广泛的应用。王玉兰等^[22]基于 ^1H NMR 指纹谱图并结合多元统计分析方法揭示不同生长周期的迷迭香挥发油的次生代谢物的含量变化;韩明霞等^[50]对不同产地的葛根进行了 NIR 红外指纹图谱分析,对不同纬度,经度和海拔的葛根制定出了一个完整的葛根产地质量评价体系。刘琳娜^[10]等对岷县当归药材样本的挥发油进行了质谱色谱联用分析,鉴定出了挥发油 47 个离子峰,并鉴定出了其中的 40 种有效化合物,其中含量较高的有 Z 藁本内酯 (54.3%)、2-乙基-1-己醇(7.0 %)、十三烷(4.8%)等 7 个主要成分,共约占总量的 95.5%。袁久荣等^[11]采用气质联用技术分析岷归的干燥根挥发油,共分离出 91 种组分,在 NIST11 数据库中鉴定了 59 个成分,占挥发油总组分的 94%以上,且含量较高的成分与刘琳娜的研究结果高度一致。袁久荣等^[11]又以毛细管 GCMS 技术测定当归饮片提取物的组成,GCMS 结果显示当归饮片能够检测到 88 个离子流量,鉴定出其中 42 种成分,占当归饮片的 51.34%。潘新坡等^[12]采用近红外 (NIR) 指纹图谱技术,利用联合区间间隔偏最小二乘法从近红外指纹图谱中筛选出子空间,对比了代谢产物在不同的当归部位中的异同、并对当归饮片所选用的当归部位进行分类与预测。

2.3 综述总结

当前,从国内外文献综述中发现,对于植物代谢组学研究主要集中在以下方面^[15-17]:

- (1) 对选定的植物代谢组学的研究。对所选定植物的某个部位,利用多元统计方法,对其代谢成分、含量进行分析,探索不同部位的差异所带来的变化规律^[18-22];
- (2) 对某些植物代谢组学研究。这一种研究是通过采集生长在不同环境下的药材,探究外界生长环境的差异对胺类、生物碱、萜类、醌类、黄酮类、强心苷和皂苷等次生代谢化合物的作用 and 影响^[23, 24];
- (3) 利用代谢组学图谱鉴定植物的提取物中相应的代谢物质,并研究代谢物质在植物体中含量,分布以及运输方向。并根据这些化合物的代谢方向,来判断初生代谢物和次生代谢物对植物的生长周期都有哪些影响。

关于代谢指纹图谱在中药方面的研究,大多集中于药效成分的形成和自身的生长情况,以及生化、生理代谢活动。这些研究大都是中药材的主要活性成分,对中药材利用和疾病预防治疗有着很好的启发作用。然而较少的文献探讨次生代谢产物在药用植物体内的合成积累时间以及机制,且对于不同的化学分析手段获得的当归代谢图谱之间的相互对比和相互印证方面的文献并未涉及。整篇文章将对此些问题进行深入地探讨,以求对于控制植物药材的采集时期、药用控制起到启发性的作用。

3 当归药材的植物学概况

3.1 药材性状

当归药材为伞形科植物当归的干燥根,药品略成圆柱状,上端根茎部称“归头”,主根称“归身”,下部有支根 3~5 条或更多,称为“归尾”,全身称“全归”,长 15~25 cm。表面黄棕色至棕褐色,具纵皱纹及横长皮孔样突起。质柔韧,断面黄白色或淡黄棕色,皮部厚,形成层环黄棕色,有裂隙及多数棕色点状分泌腔,木部色较淡,根茎处有髓或呈空腔。根头(归头)具环纹,上端圆钝,直径 1.5~4 cm,有紫色或黄绿色的茎或叶鞘的残基;主根(归身)表面凹凸不平;支根(归尾)上粗下细,有少数须根痕,直径 0.3~1 cm。有浓郁的香气,柴性大、干枯无油或断面呈绿褐色者不可供药用,味甘、辛、微苦^[1,25,26]。



图 1-1 第一年当归



图 1-2 第二年当归



图 1-3 第三年当归



图 1-4 当归药材

3.2 样本产地分布

甘肃岷县栽培当归历史悠久，甘肃岷县又有“中国的当归之乡”的美誉，是著名的“岷归”基地，由于“岷归”主要药效成分含量比国内其他产地和日本以及韩国的当归要高得多，因此岷县当归不仅畅销全国，而且在国际市场上享有很高声誉^[2]。本次试验研究的当归样本也就采样于岷县的 27 个当归样本。

4 研究方法

- (1) 本论文分别采用 GCMS 气质联用和 NMR 核磁共振技术对采自甘肃岷县不同生长期的当归药材的提取物进行测定分析，通过观测其相对含量随生长时期的积累变化，探讨当归药材各年份的代谢成分含量积累情况，并通过每一个代谢产物在当归的个生长时期的变化，探索代谢产物对当归生长的作用。为当归药材的年限利用开发提供理论参考。

- (2) 建立当归 NMR 与 GCMS 指纹图谱, 尝试利用方差分析模型来筛选在当归生长时期内差异较大的特征峰和化学保留时间, 据此来判别在三年当归的生长时期差异较大的物质。再根据正交偏最小二乘判别分析 OPLS-DA 等多元统计方法, 并结合相对含量法对谱图数据全面系统的进行挖掘分析和方法对比。探寻当归在生长周期中对年份变化较为敏感的代谢成分, 以及利用 t 检验检测当归在自然生长过程当中不同年份之间的代谢成分的差异情况。并尝试利用当归指纹图谱的数据所构建的 OPLS 偏最小二乘模型进行当归的年份分类与预测。
- (3) 对比气质联用色谱和核磁共振图谱两种计量化学分析手段。从两种化学计量学方法的图谱数据上分析各化合物在两种化学分析手段上是否有显示, 敏感程度。并通过灰色关联分析来相互定性化合物成分。再通过两者的数理统计结果来鉴别两者之间的统计实验判断结果是否有差异, 通过对比探索是否两种试验方法分析出的对年份分类影响重要的化合物都相同。

5 主要创新点

- (1) 采用了代谢组学分析中的多元分辨方法^[51]。该方法可以将复杂的重叠色谱峰和核磁共振谱峰分辨成多个单一的色谱与光谱图进行识别, 从而实现对于两种图谱的重叠峰的定性定量分析。
- (2) 利用了改进的 PLS-DA 模式识别法, 即为有监督的正交偏最小二乘模型多元统计方法。对当归代谢组学数据进行深入挖掘实现了样品类别的识别和划分, 对比了几种多元统计方法的优劣并通过建立的多参数模型对未知的样本进行预测判别分析。
- (3) 利用多个图谱处理软件如 Chemosolv, 化学计量学软件 SIMCA-p 统计分析软件如 Matlab, R 软件进行分析, 数据和分析结果的可视化相对完善。
- (4) 采用多种植物代谢组学实验数据进行数理统计检验与多元统计分析方法。借用检验与统计分析结果, 相互对比印证当归生长周期内的次生代谢成分即其相对含量的变化。根据当归代谢数据开创性地并将两种代谢组学实验方法的优劣性通过进行比照。
- (5) 对于当归在两种化学分析手段上产生的 GCMS 和 HNMR 指纹代谢图谱进行相互对比和验证, 来更加立体直观的展现不同图谱对当归药材内部物质所展现的差异, 并相互补充说明, 以弥补单一化学分析手段的不足。

第二章 不同生长期当归的代谢组学实验研究

1 代谢组学指纹图谱技术概述

1.1 核磁共振指纹图谱

核磁共振(NMR) 是一种基于有自旋性质的原子核在其核外磁场作用下, 吸收射频辐射后产生能级跃迁的分析技术。NMR 于 20 世纪 70 年代最先应用到生物医学的研究当中。利用分辨率较高的 NMR 技术对植物器官内的许多微量代谢成分进行检测, 能获得相应生物体代谢产物的大量信息, 研究这些代谢成分的 NMR 图谱, 全面阐释这些信息反映的生物学意义, 可以诠释生物体代谢的一般途径。对于新药的研发、临床应用、植物药材的代谢组学等方向有这重大意义。NMR 技术对样品损伤性小, 不破坏内部结构性质, 且需要的样品量很少、样品的前处理简单、提供信息量较大等优点, 很适合研究植物药材的复杂成分^[27]。

1.2 GCMS 气相联用指纹图谱

相比 NMR 来讲 GCMS 气相色谱联用技术具有高分辨、快速扫描能力。且关于 GCMS 气相色谱质谱联用具有大量标准代谢物数据库, 而且在测定相对分子质量及推算分子式时, 质谱是无可取代的。然而会对试验样品存在损伤性, 无法进行多次的测量与对比。但往往使用一种分析技术往往很难获得植物体中所有的化合物结构, 组成和含量等方面的综合信息。因此多种现代分析技术的联用, 相互印证和确认是代谢组学分析的发展方向^[13]。

2 当归样品的采集

本研究对采自甘肃岷县不同生长期的 27 个当归样品进行检测分析, 建立当归 NMR 和 GCMS 指纹图谱, 为当归药材的质量控制和代谢组学研究提供基础。以下表 2-1 为当归采集的样品信息。

表 2-1 不同生长期当归样品信息

实验编号	生长年限	采样时间	实验编号	生长年限	采样时间
A1	1 年	08. 30	B11	2 年	09. 25
A2	1 年	09. 13	B12	2 年	10. 05
A3	1 年	10. 01	B13	2 年	10. 14
A4	1 年	10. 08	B14	2 年	10. 24
B1	2 年	03. 29	B15	2 年	11. 07
B2	2 年	04. 04	C1	3 年	04. 30
B3	2 年	05. 29	C2	3 年	05. 29
B4	2 年	06. 15	C3	3 年	06. 28
B5	2 年	06. 28	C4	3 年	07. 26
B6	2 年	07. 12	C5	3 年	08. 30
B7	2 年	07. 26	C6	3 年	09. 13
B8	2 年	08. 15	C7	3 年	09. 30
B9	2 年	08. 27	C8	3 年	10. 29
B10	2 年	09. 14			

3 不同生长期当归指纹图谱的测定

3.1 ^1H NMR 指纹图谱

3.1.1 实验结果及分析

利用 ^1H NMR 指纹图谱样品制备技术,对采自甘肃岷县的不同生长时期的 27 个当归样品进行测定,将同生长时期的当归谱图利用 Matlab 2013a 软件叠加到一起进行宏观上的观测比较,结果见图 2-1—2-3。

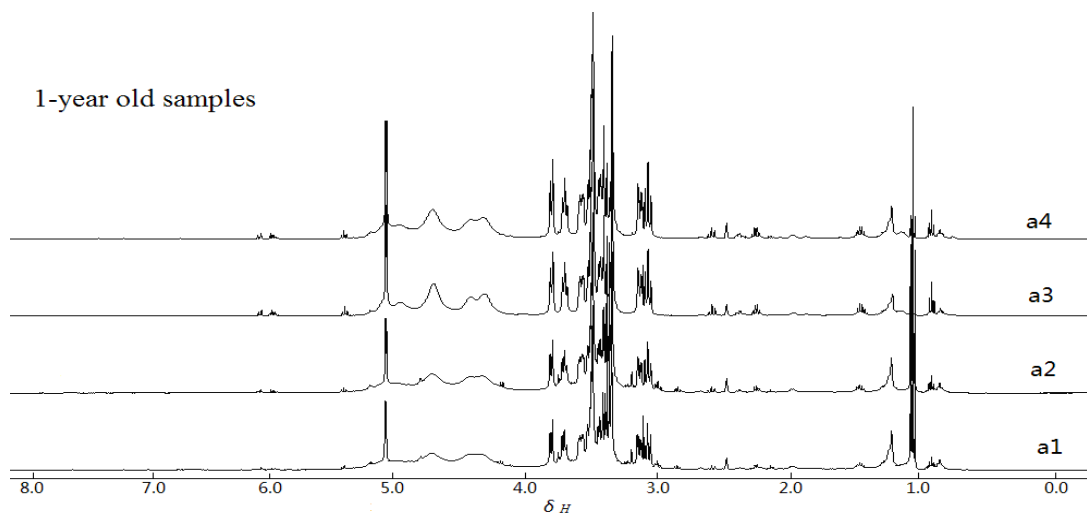


图 2-1 一年生甘肃岷县当归 ^1H NMR 指纹图谱

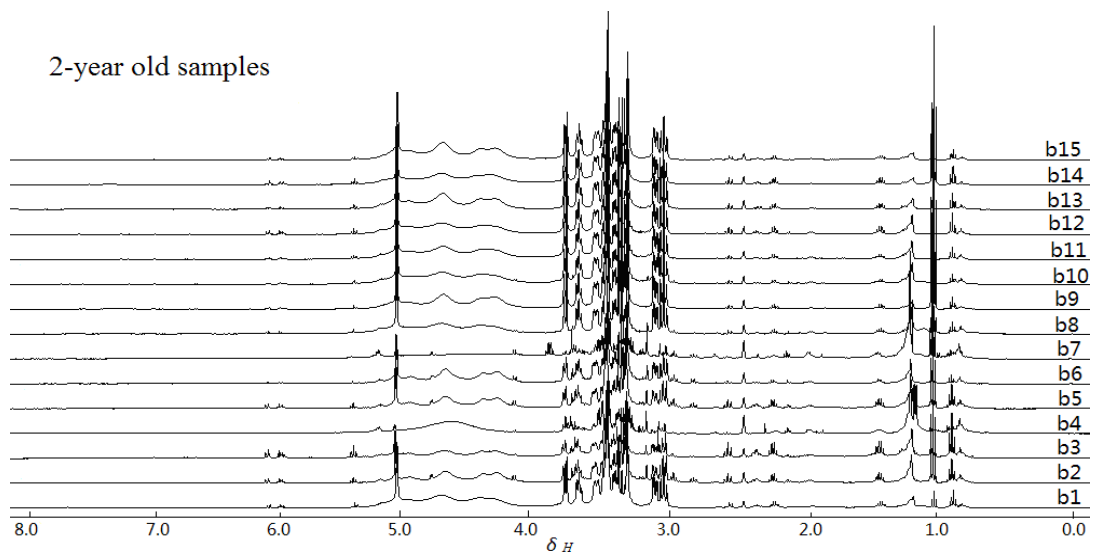


图 2-2 二年生甘肃岷县当归 ^1H NMR 指纹图谱

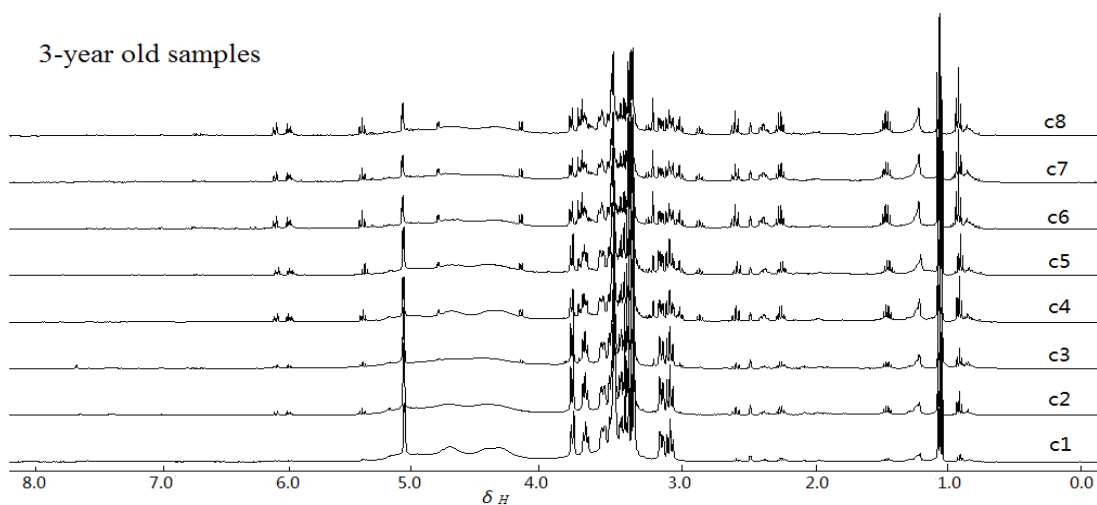
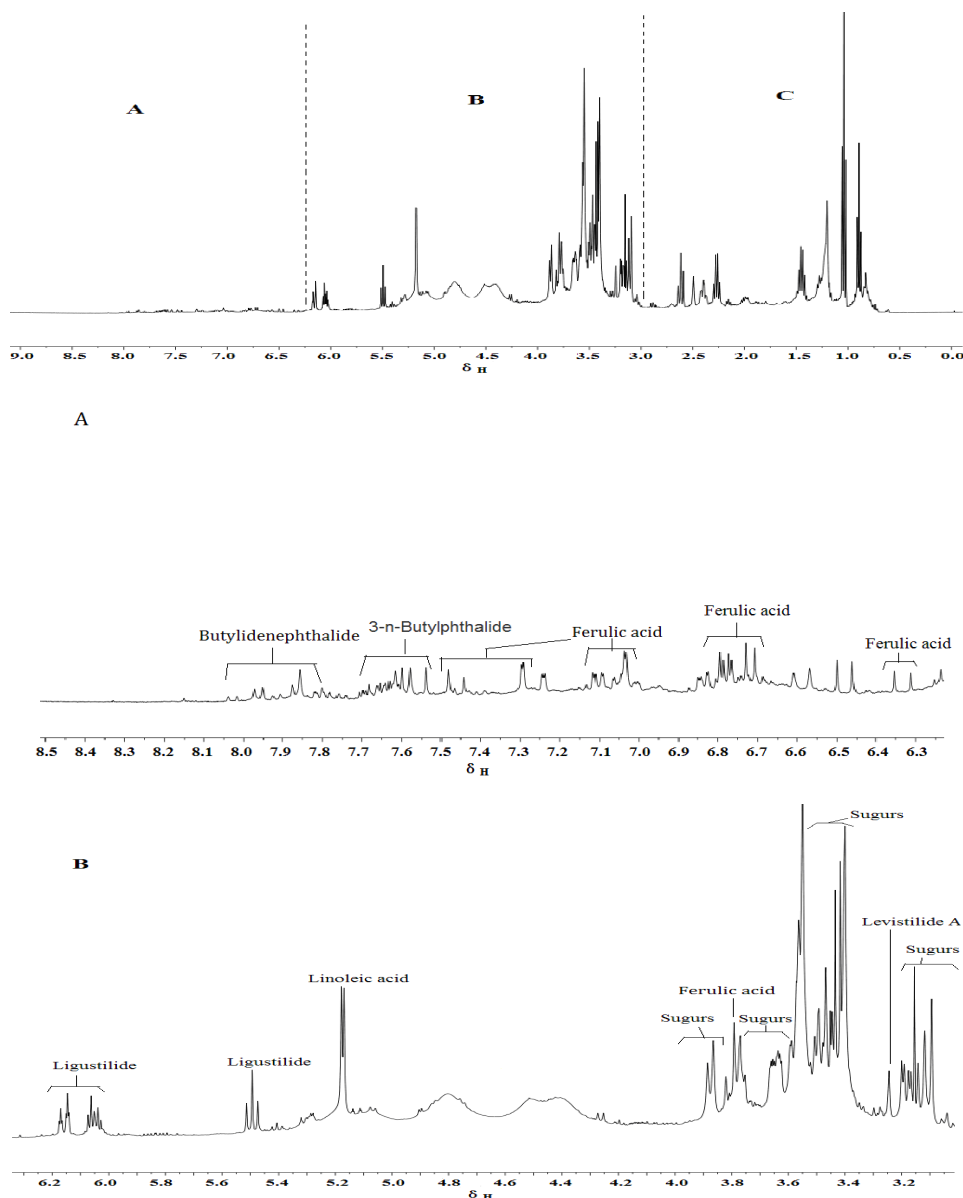


图 2-3 三年生甘肃岷县当归 ^1H NMR 指纹图谱

由图 2-1—2-3 可以看到, 27 个样品的 ^1H NMR 指纹图谱形状具有较好的相似性, 在每一个年份上分别在化学位移 δ 1.0、 δ 2.0~ δ 3.0、 δ 3.0~ δ 4.0、 δ 4.0~ δ 5.5、 δ 6.0 左右检测出明显的特征峰。对于一年生当归样品谱图, 他们的内部差异也可以说成是季节差异主要体现在 δ 4.0~ δ 5.5 和 δ 6.0 位移的特征信号峰上; 两年生当归样品图谱差异主要集中在 δ 3.0~ δ 4.0、 δ 4.0~ δ 5.5 和 δ 6.0 位移区间; 三年生当归样品谱图差异主要体现在 δ 4.0~ δ 5.5 和 δ 6.0 的特征峰上。而不同年份之间的当归样品的图谱之间的差异主要直观地表现为 δ 3.0~ δ 4.0 和 δ 4.0~ δ 5.5 的特征峰高度上。

3.1.2 当归样品 ^1H NMR 指纹图谱的归属

采用对照品实验的方法来对比当归样品的谱峰差距, 并查阅参考相关文献^[28-31], 对当归药材中的药效成分在图谱上进行初步的特征峰归属, 结果见图



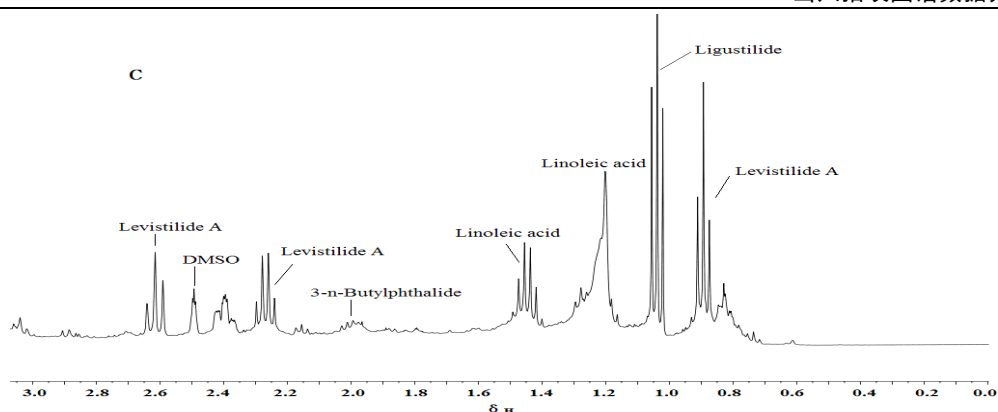


图 2-4 当归样品 HNMR 指纹图谱的谱峰归属图

由图 2-2 的核磁共振信号峰的定性来看, δ 1.22 和 δ 1.46 为亚油酸的特征信号峰当归多糖特征信号峰主要在 δ 3.25、 δ 3.38、 δ 3.36、 δ 3.57、 δ 3.64、 δ 5.20; 藁本内酯的特征信号峰在 δ 1.03、 δ 5.47、 δ 6.03、 δ 6.19; δ 3.78、 δ 6.35、 δ 6.78、 δ 7.07 和 δ 7.27 为阿魏酸的特征信号峰位移区间; δ 0.83、 δ 2.29、 δ 3.26 为欧当归内酯 A 的特征信号峰; δ 7.86 为丁烯基苯酞的位移区间; δ 2.1 和 δ 7.61 为 3-正丁基苯酞的特征信号峰特征信号峰。

3.1.3 当归 NMR 图谱数据处理过程

本研究采用 Chemosolv 软件将 TMS 峰定位为 δ 0.00, 并调整基线, 然后手动调相位后。对所有的 ^1H NMR 谱图进行自动积分, 积分区间为 δ 0.6~8.5, 积分间隔设为 δ 0.01。将积分区间 δ 1.2~1.5 和 δ 2.46~2.55 的积分值设为零, 避免剩余 DMSO 的溶剂峰对分析带来的影响。生成一个 27×750 的矩阵, 矩阵的行代表不同的样品, 列代表不同的化学位移段积分量。

3.2 GCMS 气相色谱联用指纹图谱

3.2.1 GCMS 气质联用实验结果

采用 GC-MS 气质联用测定 27 批不同生长期的当归药材, 利用 Chemosolv 对谱峰进行去噪、基线校准等预处理^[32]。并采用多元曲线分辨法 (MCR) 中的基于差异分析的迭代目标转换因子分析法 (D-ITTEFA) 解析严重的重叠质谱峰^[22, 23], 获取纯质谱图。

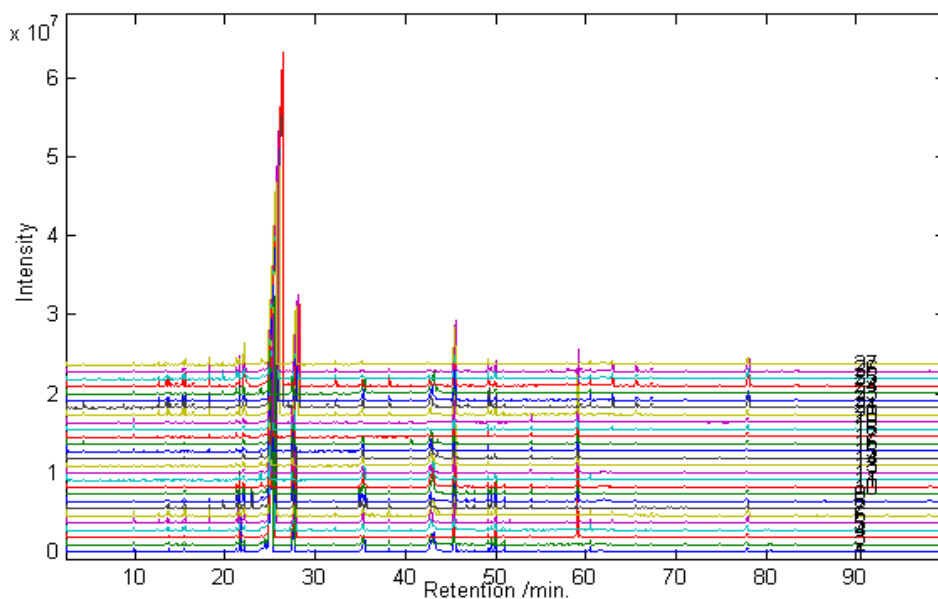


图 2-5 不同生长期当归 GC-MS 总离子流图叠加

3.2.2 GCMS 代谢图谱的代谢组学处理与化合物鉴别

应用化学计量学的分析方法处理 GC-MS 分析出不同生长期当归的挥发油获得的大量数据，并根据对照品的 GC-MS 指纹图谱并结合 NIST2011 气相色谱数据库对当归的挥发油中的次生代谢物进行鉴定。从当归挥发油中定性出挥发性和差异较大 44 个化合物，见表 2-2^[13]。

表 2-2 当归挥发性成分鉴定表

编号	Rt / (min)	化合物名称	分子量	分子式	保留指数	匹配度
1	3.153	Bicyclo[3.1.1]hept-2-ene, 2,6,6-trimethyl-	136	C ₁₀ H ₁₆	940	96
2	4.436	1,3,6-Octatriene, 3,7-dimethyl-, (E)-beta.-cis-Ocimene	136	C ₁₀ H ₁₆	1030	93
3	5.433	Undecane	156	C ₁₁ H ₂₄	1100	93
4	6.595	n-pentylbenzene	148	C ₁₁ H ₁₆	1158	99
5	6.626	6-butyl-1,4-cycloheptadiene	150	C ₁₁ H ₁₈	1691	99
6	7.304	Dodecane	170	C ₁₂ H ₂₆	1200	92
7	7.560	1-methyl-4-[1-methylethenyl]-2-Cyclohexen-ol	152	C ₁₀ H ₁₆ O	<1295	83
8	8.825	6-Undecanone	170	C ₁₁ H ₂₂ O	1274	93
9	9.408	Tridecane	184	C ₁₃ H ₂₈	1300	97
10	10.031	2-Methoxy-4-vinylphenol	150	C ₉ H ₁₀ O ₂	1315	87

11	10.958	1,4-Cyclohexadiene-1,2-dicarboxylic anhydride	150	C ₈ H ₆ O ₃	1329	97
12	12.007	Tetradecane	198	C ₁₄ H ₃₀	1400	97
13	12.351	Vanillin	152	C ₈ H ₈ O ₃	1405	95
14	12.730	1H-3a,7-Methanoazulene, octahydro-3,8,8-trimethyl-6-methylene-, [3R-(3.alpha.,3a.beta.,7.beta.,8a.alpha.)]-	204	C ₁₅ H ₂₄	1423	93
15	13.225	Thujopsenedeca	204	C ₁₅ H ₂₄	1429	99
16	13.476	1H-Cycloprop[e]azulene, hydro-1,1,7-trimethyl-4-methylene-	204	C ₁₅ H ₂₄	1437	99
17	13.849	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-	204	C ₁₅ H ₂₄	1459	94
18	14.001	1H-Cycloprop[e]azulene, decahydro-1,1,7-trimethyl-4-methylene-, [1aR-(1a.alpha.,4a.beta.,7.alpha.,7a.beta.,7b.alpha.)]	204	C ₁₅ H ₂₄	1454	83
19	14.228	(+)-Epi-bicyclosesquiphellandrene	204	C ₁₅ H ₂₄	1471	94
20	14.735	Spiro[5.5]undec-2-ene, (-)-3,7,7-trimethyl-11-methylene-	204	C ₁₅ H ₂₄	1472	98
21	14.823	1,6-Cyclodecadiene, [s-(E,E)]-1-methyl-5-methylene-8-(1-methylethyl)	204	C ₁₅ H ₂₄	1487	98
22	15.359	gamma.-Elemene	204	C ₁₅ H ₂₄	1465	80
23	15.685	l-b-Bisabolene	204	C ₁₅ H ₂₄	1500	96
24	16.245	Cyclohexene, [S-(R*,S*)]-3-(1,5-dimethyl-4-hexenyl)-6-methylene-,	204	C ₁₅ H ₂₄	1524	95
25	16.624	1H-Benzocycloheptene, 2,4a,5,6,7,8-hexahydro-3,5,5,9-tetramethyl-, (R)-	204	C ₁₅ H ₂₄	1527	94
26	18.378	(+)-spathulenol	220	C ₁₅ H ₂₄ O	1568	99
27	22.260	1(3H)-Isobenzofuranone, 3-propylidene-	174	C ₁₁ H ₁₀ O ₂	1575	85
28	24.352	n-butylidenephthalide	188	C ₁₂ H ₁₂ O ₂	1675	88

29	26.398	Z-ligustilide	190	C ₁₂ H ₁₄ O ₂	1741	94
30	27.208	n-Pentadecanol	228	C ₁₅ H ₃₂ O	1775	97
31	27.517	E-ligustilide	190	C ₁₂ H ₁₄ O ₂	1797	93
32	30.764	Pentadecanoic acid	242	C ₁₅ H ₃₀ O ₂	1878	99
33	31.562	1-Hexadecene	224	C ₁₆ H ₃₂	1883	96
34	33.498	Hexadecanoic acid, methyl ester	270	C ₁₇ H ₃₄ O ₂	1927	91
35	35.560	n-Hexadecanoic acid	256	C ₁₆ H ₃₂ O ₂	1975	90
36	38.388	Falcarinol	244	C ₁₇ H ₂₄ O	2028	91
37	39.443	9-Octadecen-1-ol, (Z)-	268	C ₁₈ H ₃₆ O	2068	93
38	40.888	Z,E-2,13-Octadecadien-1-ol	266	C ₁₈ H ₃₄ O	2076	95
39	42.911	9,12-Octadecadienoic acid (Z,Z)-, methyl ester	294	C ₁₉ H ₃₄ O ₂	2093	97
40	43.126	9,12-Octadecadienoic acid, methyl ester	294	C ₁₉ H ₃₄ O ₂	2101	94
41	49.214	Linoleic acid	280	C ₁₈ H ₃₂ O ₂	2136	91
42	51.746	(Z)-9,17-Octadecadienal	264	C ₁₈ H ₃₂ O	2297	96
43	54.148	Phenol, 2,2'-methylenebis	340	C ₂₃ H ₃₂ O ₂	2389	97
44	93.771	Vitamin E	430	C ₂₉ H ₅₀ O ₂	3130	93

3.2.3 当归 GCMS 指纹图谱数据处理过程

一共有27个取自不同采收时期的当归样品，每一个样品在0-100min的保留时间内一共有17195个扫描点，即每个样本在每个保留时间的扫描点都进行了质量范围为500aum的扫描471次的扫描。将同一保留时间扫描点的471次扫描所得的离子数相加，即形成该样本在相应保留时间的总离子数。因此在GCMS指纹图谱中提取的数据构成了一个27×17195的矩阵，矩阵的行代表不同的样品，列代表不同的保留时间下的总离子密度数。

第三章 基于方差分析模型的指纹图谱变量筛选

获得核磁共振 HNMR 和 GC-MS 气相色谱实验数据后,通过数据清洗和整理后,我们分别得到一个 27×750 阶的核磁共振谱数据矩阵和一个 27×17195 阶的气相质谱数据矩阵。对于核磁共振图谱来说, $\delta 0.6-8.5$ 的化学位移上,存在着众多特征谱峰。而对于气质联用谱来讲, $0 \sim 100 \text{min}$ 的保留时间里也有着众多的离子峰。想要判断各个年份的当归总体在某一气相色谱的保留时间的离子密度或核磁共振的化学位移上的强度是不是存在差异,并且是不是达到统计学中的显著水平,进而从众多变量中挑选出差异较大的化学位移或保留时间做进一步分析。在此章节我们应用数理统计中的显著性检验的 ANOVA 模型来验证三个不同年份的总体在 HNMR 核磁共振谱和 GCMS 气相色谱图中在哪些化学保留时间或化学位移在统计学中达到了具有显著差异。并筛选出这些变量。单因子方差模型要满足三个假设:独立性,正态性,方差齐性。因此首先验证变量是否满足这些条件。

1 HNMR 核磁共振谱图方差分析模型

1.1 正态性检验—Epps-Pulley 检验

1.1.1 方法及原理

Epps-Pulley 检验对多种备择假设有较高的效率,其出发点是利用样本的特征函数与正态分布的特征函数的差的模的平方产生一个加权积分得到的。我们在这里简单地介绍该检验方法。

设 x_1, \dots, x_m 是来自 $N(\mu_1, \sigma_1^2)$ 的样本, EP 检验统计量的定义为:

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n} \sum_{i=2}^n \sum_{j=1}^{i-2} \exp\left\{\frac{-(x_i - x_j)^2}{2s_*^2}\right\} - \sqrt{2} \sum_{i=1}^n \exp\left\{\frac{-(x_i - \bar{x})^2}{4s_*^2}\right\} \quad (3-1)$$

其中 \bar{x} , s_*^2 就是前述的样本均值和 (除以 n 的) 样本方差。其拒绝域为 $\{T_{EP} \geq T_{1-\alpha, EP}(n)\}$, $T_{1-\alpha, EP}(n)$ 是样本容量为 n 时 EP 检验统计量 (在原假设下的分布) 的 $1-\alpha$ 分位数^[45]。

1.1.2 验证结果与分析

因为当归代谢组学数据是在标准互不影响的独立环境下采取适当的单位得到的测量值，因此数据组之间的独立性得以保障。为检验总体分布是否正态，采用 EP 检验。以第二年份的当归总体为例做 EP 正态性检验。检验第二年份的当归总体在每一化学位移上是否都符合正态分布。经过 Matlab2013a 编程计算^[46]，在显著水平 $\alpha=0.01$ ，拒绝域为 $\{T_{EP} \geq 0.560\}$ 得到下表（显示部分化学位移的检验结果）：

表 3-1 二年生当归总体正态性 EP 检验

化学位移波段/ δ	所属化合物	单侧临界值	T_{EP}	是否显著
0.890	欧当归内酯 A	0.560	0.7821	是
1.020	蒿本内酯	0.560	0.8352	是
1.200	亚油酸	0.560	0.2188	否
1.450	亚油酸	0.560	0.0344	否
2.000	3-正丁基苯酐	0.560	0.5245	否
3.250	欧当归内酯 A	0.560	0.1425	否
3.650	当归多糖	0.560	0.1876	否
5.190	当归多糖	0.560	0.0485	否
6.180	蒿本内酯	0.560	0.2650	否
6.340	阿魏酸	0.560	0.5006	否
7.280	阿魏酸	0.560	2.0486	否
7.600	3-正丁基苯酐	0.560	1.1231	是
7.850	丁烯基苯酐	0.560	0.4819	否

经过 EP 检验证明在三个年份当归总体中都具有正态经的化学位移变量共有 493 个然后将这 493 个具有正态性的化学位移进行下一步的方差齐性检验。

1.2 方差齐性检验—修正的 Bartlett 检验

1.2.1 方法及原理

Bartlett 证明了：大样本场合， $\ln(MS_e/GMS_e)$ 的某个函数近似服从自由度为 $r-1$ 的 χ^2 分布^[45]。根据 Bartlett 的结论，可取：

$$B = \frac{1}{c} [f_e \ln MS_e - \sum_{i=1}^r f_i \ln s_i^2] \quad (3-2)$$

作为统计量，对于给定的显著水平 α ，检验的拒绝域为：

$$W = \{B \geq \chi_{1-\alpha}^2(r-1)\} \quad (3-3)$$

在诸样本量 m_i 均不小于 5 时使用上述检验。针对样本量低于 5 时不能使用 Bartlett 检验的缺点，博克斯提出修正的 Bartlett 检验统计量：

$$B' = \frac{f_2 BC}{f_1(A-BC)} \quad (3-4)$$

其中 B 与 C 如式与式所示，且

$$f_1 = r-1, \quad f_2 = \frac{r+1}{(c-1)^2}, \quad A = \frac{f_2}{2-c+2/f_2} \quad (3-5)$$

在原假设 $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2$ 成立下，Box 还证明了统计量 B' 的近似分布是 F 分布 $F(f_1, f_2)$ ，对于给定的显著水平 α ，该检验的拒绝域为：

$$W = \{B' \geq F_{1-\alpha}(f_1, f_2)\} \quad (3-6)$$

1.2.2 验证结果与分析

我们用修正的 Bartlett 对三年份当归关于 493 个具有正态性的化学位移做方差齐性检验。经过 Matlab2013a 编程计算。在显著水平 $\alpha=0.05$ 下，在 F 分部分位表中查得拒绝域为：

$$W = \{B' \geq F_{0.95}(2, 562)\}$$

$$W = \{B' \geq 3.00\}$$

得到以下检验表格（只显示部分重要化合物位移检验结果）：

表 3-2 三年份当归总体的方差齐性检验

化学位移波段/ δ	所属化合物	$F_{0.95}(2,562)$	B'	是否显著
2.280	欧当归内酯 A	3.00	0.7240	否
5.480	蒿本内酯	3.00	0.6552	否
1.200	亚油酸	3.00	0.3176	否
1.450	亚油酸	3.00	2.2559	否
2.000	3-正丁基苯酐	3.00	2.1632	否
3.250	欧当归内酯 A	3.00	1.7672	否
3.650	当归多糖	3.00	4.2144	是
5.190	当归多糖	3.00	5.2527	是
6.180	蒿本内酯	3.00	1.5630	否
6.340	阿魏酸	3.00	4.2161	是
7.280	阿魏酸	3.00	2.1705	否
7.850	丁烯基苯酐	3.00	1.4970	否

上表 3-2 中由于除 $\delta 3.650$, $\delta 5.190$, $\delta 6.340$ 波段的所有其他的化学位移变量在个总体中的 $B' < 3.00$, 故在上表中其余的 7 个波段内接受原假设 H_0 , 即认为除过 $\delta 3.650$, $\delta 5.190$, $\delta 6.340$ 波段的所有表 3-2 内的其他化学位移在三个年份水平下的方差无显著差异。纵观由 EP 检验筛选出的化学位移在三个总体下的方差齐性情况, 经过修正的 Bartlett 检验后, 满足正态且方差齐性的总共有 315 变量。

1.3 方差分析统计模型分析结果

1.3.1 方法及原理

本章我们的工作是比较各年份的当归, 在自己所属的年份下的各化学位移波段的均值是否相同, 即要对如下的一个假设进行检验:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_r \quad (3-7)$$

需要比较各组分的偏差平方和 Q ，来比较各组分的差异性，偏差平方和 Q 的大小又与数据个数（或自由度）有关，一般来说，数据越多，其偏差平方和越大。为了便于在诸偏差平方和间进行比较，统计上引入了均方的概念。定义为： $MS = \frac{Q}{f_Q}$ 其意为平均每个自由度上有多少平方和。

如今要对因子平方和 S_A 与误差平方和 S_e 之间进行比较，用其均方 $MS_A = \frac{S_A}{f_A}$ ， $MS_e = \frac{S_e}{f_e}$ 进行比较合理，因为均方排除了自由度不同所产生的干扰，故用 $F = \frac{MS_A}{MS_e} = \frac{S_A/f_A}{S_e/f_e}$ 作为检验 H_0 的检验统计量^[45]，拒绝域为：

$$W = \{F \geq F_{1-\alpha}(f_A, f_e)\} \quad (3-8)$$

1.3.2 验证结果与分析

为检验三年份的当归总体关于 315 个满足正态且方差齐性的化学位移变量均值是否在统计学中有明显的差异。采用方差分析检验。经过 Matlab2013a 编程计算，在显著水平 $\alpha=0.05$ ， F 分部分位表中查得拒绝域为：

$$W = \{F \geq F_{0.95}(2, 24) = 3.40\} \quad (3-9)$$

得到以下检验表格（显示部分标志性化合物）：

表 3-3 三年当归总体的方差分析结果

化学位移波段/ δ	所属化合物	$F_{0.95}(2, 24)$	F	是否显著
2.280	欧当归内酯 A	3.40	4.2902	是
5.480	蒿本内酯	3.40	6.1475	是
1.200	亚油酸	3.40	2.8398	否
1.450	亚油酸	3.40	5.8230	是
2.000	3-正丁基苯酐	3.40	3.7871	是
3.250	欧当归内酯 A	3.40	0.23196	否
6.180	蒿本内酯	3.40	5.2842	是
7.280	阿魏酸	3.40	3.6290	是
7.850	丁烯基苯酐	3.40	1.5429	否
3.390	当归多糖	3.40	1.2352	否
3.750	当归多糖	3.40	2.4953	否

经过方差分析的检验结果我们可知,在显著水平 $\alpha=0.05$ 的条件下,在上表中三个年份的当归样品的 F 值在 $\delta 2.280$, $\delta 5.480$, $\delta 1.450$, $\delta 2.000$, $\delta 6.180$, $\delta 7.280$ 波段是落在拒绝域上的,所对应的化合物分别为欧当归内酯,蒿本内酯,亚油酸,3-正丁基苯酞和阿魏酸。

结合整个 315 个化学位移波段来说,经过方差分析模型我们共鉴定出了在三个年份的当归中差异较大的 103 个化学位移波段,根据 $^1\text{H NMR}$ 核磁共振谱的可鉴定成分,解析了可能的几个显著差异化合物。如下表。也就是说,表中次生代谢产物在整个当归代谢过程中,这几种化合物在当归体内代谢旺盛,转化积累变化范围大。其中蒿本内酯和阿魏酸又是主要的药用成分,控制当归的采收时间,把握好这几种化合物的比例,有助于当归的道地性的最大表现。如下表 3-4 所示的差异化学位移波段一共检验出了 103 个,保留这 103 个变量,重新构成一个 27×103 阶矩阵进行下面分析。

表 3-4 差异化学位移所属化合物

化学位移波段/ δ	所属化合物	化学位移波段/ δ	所属化合物
0.720-0.860	欧当归内酯 A	4.370-4.880	未知化合物 a
1.430-1.750	亚油酸	5.140-5.200	当归多糖
2.070-2.960	3-正丁基苯酞	5.430-5.600	蒿本内酯
3.130-3.230	欧当归内酯 A	5.610-5.960	未知化合物 b
3.420-3.660	当归多糖	5.970-6.140	蒿本内酯
3.680-3.810	阿魏酸	6.300-6.750	阿魏酸

2 GCMS 气质联用图谱多总体方差分析模型

2.1 正态性检验—Epps-Pulley 检验结果与分析

因为当归代谢组学数据是在标准互不影响的独立环境下采取适当的单位得到的测量值,因此数据组之间的独立性得以保障。为检验总体分布是否正态,采用 EP 检验。经过 Matlab2013a,将合并矩阵拆开分为三个不同年份总体,编程计算,对 17195 个化学保留时间下的三个不同年份当归的总体,在显著水平 $\alpha=0.01$,拒绝域分别为一年生当归总体 $\{T_{EP} \geq 0.256\}$,二年生当归总体 $\{T_{EP} \geq 0.560\}$,三年生当归总体为 $\{T_{EP} \geq 0.526\}$ 。在同一个化学离子保留时间内只有在三个年份的总体中全部满足正态关系才能保留该化学离子保留时间,否则在不具备正态的保留时间内,总体的变化不够稳定,可能是受

月份，记录误差等其他因素影响，不能代表一个年份内当归在那个化学离子保留时间波峰 μ 的整体状况，因此需要删除那样的保留时间。进行正态性检验，得到下表（展示部分实验结果）：

表 3-5 不同年份当归总体正态性 EP 检验

化学离子保留时间/min	代表化合物	一年 T_{EP}	二年 T_{EP}	三年 T_{EP}	是否显著
10.043	2-甲氧基-4-乙烯基苯酚	0.1390	0.7576*	0.2140	是
24.352	丁烯基苯酐	0.0172	0.1571	0.0145	否
26.398	Z-藁本内酯	0.1221	0.4812	0.0483	否
28.397	E-藁本内酯	0.0300	0.2372	0.0136	否
33.498	棕榈酸	0.0166	0.1661	0.1362	否
43.214	亚油酸	0.2426	0.5366	0.3130	否
51.746	反油酸	0.1735	0.1416	0.2299	否
76.093	1,21-二十二碳二烯	0.0559	0.3291	0.0194	否
92.839	二十碳烷	0.5731*	0.2978	0.1730	是
93.177	维生素 E	0.4247*	0.2955	0.0958	是

经过 EP 检验证明在每一个化学离子保留时间里，三个年份的当归是否为全部具有正态性。不全部具有正态性的保留时间表明该化合物不仅只与年份有关，还受一些其他变量或误差的影响，因此，我们只保留在三个年份中都为正态的保留时间。符合正态性的保留时间可以进行下一步的方差齐性检验。

2.2 方差齐性检验—修正的 Bartlett 检验结果与分析

我们用修正的 Bartlett 对 GCMS 指纹图谱三年份当归关于具有正态性的化学保留时间做方差齐性检验。将 GCMS 代谢指纹图谱按照年份拆成三个总体，经过 Matlab2013a 编程计算。在显著水平 $\alpha=0.05$ 下，在 F 分部分位表中查得拒绝域为：

$$W = \{B' \geq F_{0.95}(2,562)\} \quad (3-10)$$

得到以下检验表格(只列举部分实验结果):

表 3-6 三年份当归总体的方差齐性检验

化学离子保留时间/min	所属化合物	$F_{0.95}(2,562)$	B'	是否显著
24.352	丁烯基苯酐	3.00	1.2043	否
26.398	Z-藁本内酯	3.00	1.0845	否
28.397	E-藁本内酯	3.00	0.7240	否
33.498	棕榈酸	3.00	0.0450	否
43.214	亚油酸	3.00	0.5659	否
51.746	反油酸	3.00	0.4154	否
54.148	苯氧基-2,2'-亚甲基	3.00	1.5067	否
76.093	1,21-二十二碳二烯	3.00	3.4135*	是
81.193	29 碳烷	3.00	1.5303	否
93.177	环辛烷	3.00	3.3562*	是

化学离子保留时间检验结果若是三个年份总体的 $B' < 3.00$ ，故接受原假设 H_0 ，即认为该气相色谱的保留时间所对应的化合物的波动变化，即方差在三个年份水平下的无显著差异。而对于拒绝了原假设的化学保留时间来说，所对应的化合物在三个年份的当归中波动幅度极不相同，并不稳定，不利于当做区分辨别三种不同年份的标准代谢化合物，例如环辛烷和 1,21-二十二碳二烯在方差齐性检验中在三个总体中方差并不相同，不稳定，不利于作为区分不同年份的标准化合物，因此将未通过方差齐性检验的保留时间去除，而留下在三个总体中方差相同的保留时间进行下面的方差分析。

2.3 方差分析统计模型分析结果与分析

为检验三年份的当归总体关于的均值是否在统计学中有明显的差异。采用方差分析检验。经过 Matlab2013a 编程计算，在显著水平 $\alpha=0.05$ ，F 分部分位表中查得拒绝域为：

$$W = \{F \geq F_{0.95}(2,24) = 3.40\} \quad (3-11)$$

得到以下检验表格（显示部分实验结果）：

表 3-7 三年份当归总体方差分析部分结果

化学离子保留时间/min	所属化合物	$F_{0.95}(2,24)$	F	是否显著
24.357	丁烯基苯酞	3.40	7.0248	是
26.398	Z-藁本内酯	3.40	4.4521	是
28.397	E-藁本内酯	3.40	5.8398	是
33.498	棕榈酸	3.40	3.4623	是
49.914	亚油酸	3.40	3.7871	是
51.746	反油酸	3.40	11.9488	是
54.148	苯氧基-2,2'-亚甲基	3.40	11.0408	是
64.132	未知化合物	3.40	2.1421	否
81.193	29 碳烷	3.40	3.1049	否
93.442	未知化合物	3.40	1.3124	否

经过方差分析的检验结果我们可知，在显著水平 $\alpha=0.05$ 的条件下，三个年份的当归样品的 F 值在保留时间为：24.357，26.398，28.397min 等保留时间是落在拒绝域上的，所对应的化合物分别为：丁烯基苯酞，Z-藁本内酯等主要药效成分，但苯氧基-2,2'-亚甲基等非主要药效成分也通过数理统计的检验，是在不同生长年份当归差异比较大的。在控制当归的采收时间，需要把握好这几种此生代谢化合物的比例，有助于当归的道地性的最大表现。经过方差分析验证后在三个年份的当归总体中存在差异的 4812 个化学离子保留时间，总体构成一个 27×4812 阶的实验矩阵。

第四章 基于 ^1H NMR 指纹图谱的当归 OPLS 判别分析

实验方法分析,只局限于其中一部分化合物的含量变化,无法利用代谢图谱的数据对不同年份样本的变化及其代谢物变化做一个全面地了解,并且数据处理过程复杂,操作困难,实验偏差过大。然而当化学分析结合统计学中的多元统计方法,可以对当归整体成分的变化方便、快速并全面地进行研究^[12-16]。在代谢组学研究中,由核磁共振谱产生的高维数据可以应用偏最小二乘方法进行分析,解决降维、分类可视化和特征筛选等问题。但由于代谢组学数据中的很多代谢物质(包括离子的碎片)都具有较高的相关性,且与分类信息无关,利用 PLS 法无法将分类的信息集中在前 2 个主成分中,结果使分类及可视化效果不好^[39]。

OPLS 正交偏最小二乘是一种新型的多元统计数据分析方法,它由 Johan trygg 等^[40]人于 2002 年首次提出。近十年来,OPLS 方法在理论上得到了迅猛的发展,并在计量化学中有了大量的应用 OPLS 是一种多因变量对多自变量的回归建模方法,其最大特点是去除自变量 X 中与分类变量 Y 正交无关的数据变异,使分类信息主要集中在所提取的主成分中,模型变得简单和易于解释,其判别效果及主成分得分图的可视化效果更加明显。本文在简单介绍 OPLS 的基本思想和原理的基础上,通过模拟实验探索 OPLS 方法的适用条件,并利用方差分析统计模型筛选后的 103 个化学位移和 4812 个保留时间变量分别进行 OPLS 正交偏最小二乘实证分析。

1 正交偏最小二乘回归原理与方法

1.1 基本原理

OPLS (Orthogonal PLS) 是从筛选出化学位移或保留时间的 X 矩阵中移除与系统响应变量正交的部分,从而 X 矩阵中的正交变量与非正交变量区别。OPLS 方法利用年份哑变量 Y 中的分类信息把 X 分成三部分:

$$X = T_p P_p^T + T_0 P_0^T + E \quad (4-1)$$

其中, $T_p P_p^T$ 表示预测组分, T_p 表示对 X 得分矩阵, P_p^T 表示对 X 的载荷矩阵, T_0 为 X 与 Y 的正交组分 (OPLS 成分) 的得分阵, $T_0 P_0^T$ 是与 Y 正交的部分, P_0^T 表示 Y 的载荷矩阵, E 为残差矩阵^[41]。

OPLS 方法的实现大体分为两步走:

第一步, 在 X 数据矩阵中将与 Y 正交的变量进行剔除, 即

$$X_p = X - T_0 P_0^T \quad (4-2)$$

其中 T_0 是正交成分的得分矩阵, P_0^T 表示对应的载荷矩阵。

第二步, 对 X_p 进行偏最小二乘分解。对于新的样本矩阵, 先进行正交成分的提取, 再进行预测。

1.2 偏最小二乘的理论

偏最小二乘完全根据原始数据, 通过迭代算法, 找到各个变量之间完全线性关系, 而且能估计出全部隐变量的值。它能够在小样本的情况下实行多变量对单变量的回归建模, 且偏最小二乘方法不用对数据作任何分布的假定。再有, PLS 由于收敛速度快, 因此适用于较大, 较复杂的结构方程模型。计算效率高。

设原始特征数据 X_0 为 $N \times m$ 维矩阵, Y_0 为 $N \times n$ 维矩阵, 即共有 N 个样本对, X_0 中样本特征为 m 维, Y_0 中样本特征为 n 维, 而 X 和 Y 是原始数据经过标准化之后生成的数据。设 X 和 Y 的第一个主成分轴向量分别为 w_1 ($m \times 1$ 维) 和 c_1 ($n \times 1$ 维)。则由 w_1 和 c_1 可以表示 X 和 Y 第一对主成分 t_1 和 u_1 , 其中,

$$t_1 = X \times w_1, \quad u_1 = Y \times c_1 \quad (4-3)$$

PLS 的求解思想, 即

$$Cov(t_1, u_1) = \sqrt{Var(t_1)Var(u_1)}Corr(t_1, u_1) \rightarrow \max$$

可求得 w_1 是对称矩阵 $X^T Y Y^T X$ 的最大特征值对应的特征向量, c_1 是 $Y^T X X^T Y$ 的最大特征值对应的特征向量。

根据主成分回归思想可以把 X 和 Y 分别对他们的主成分 t_1 和 u_1 进行回归建模如下:

$$X = t_1 p_1^T + E \quad (4-4)$$

$$Y = u_1 q_1^T + G \quad (4-5)$$

利用 t_1 和 u_1 之间的相关性这一情况把 Y 改为对 X 的主成分 t_1 进行回归建模：

$$Y = t_1 r_1^T + F \quad (4-6)$$

对上面三个回归方程，可以用最小二乘法计算出 p_1, q_1, r_1 如下：

$$p_1 = \frac{X^T t_1}{\|t_1\|^2}, \quad q_1 = \frac{Y^T u_1}{\|u_1\|^2}, \quad r_1 = \frac{Y^T t_1}{\|t_1\|^2} \quad (4-7)$$

其中 w_1 是 X 投影出 t_1 的方向向量，之后将 X 中主成分 t_1 不能解释的残差部分 F 作为新的 Y，按照前面的方法进行回归，循环往复^[42]。最终可将原始 X，Y 表示为：

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E \quad (4-8)$$

$$Y = u_1 q_1^T + u_2 q_2^T + \cdots + u_k q_k^T + G \quad (4-9)$$

1.3 正交偏最小二乘计算方法

将试验所得到的 X 数据矩阵进行标准化等于处理后，进行下列算法：

- (1) $w^T = Y^T X / (Y^T Y)$ ，估计 X 的权重向量；
- (2) $w = w / \|w\|$ ，将权重向量归一化；
- (3) $t = Xw / (w^T w)$ ，计算 X 数据得分向量 t；
- (4) $c^T = t^T Y / (t^T t)$ ，算出 Y 矩阵的载荷 c；
- (5) $u = Yc / (c^T c)$ ，生成 Y 矩阵得分向量。
- (6) $p^T = t^T X / (t^T t)$ ，产生 X 矩阵的载荷向量。
- (7) $w_{\perp} = p - [w^T p / (w^T w)]w$ ，估计正交的权重向量；
- (8) $w_{\perp} = w_{\perp} / \|w_{\perp}\|$ ，对 (7) 所得的向量进行归一化；
- (9) $t_{\perp} = Xw_{\perp} / (w_{\perp}^T w_{\perp})$ ，产生 X 正交矩阵得分向量；
- (10) $p_{\perp}^T = t_{\perp}^T X / (t_{\perp}^T t_{\perp})$ ，生成 X 矩阵的正交载荷；
- (11) $E_{opls} = X - t_{\perp} p_{\perp}^T$ ， E_{opls} 为剔除了 OPLS 正交成分（ t_{\perp} 为正交部分）；
- (12) 更多的正交成分重新到步骤 (3)，并让 $X = E_{opls}$ ，否则继续下一步 (13)；
- (13) 对于新的样本 X 通过校正模型的 W_{\perp} 和 P_{\perp} 来校正；
- (14) 对已剔除掉的与 Y 正交的成分后的 E_{opls} 进行偏最小二乘法提取所需成分

[41]

1.4 正交成分个数的确定

OPLS 使用特征值判别和交叉验证相互参考的方法得出正交成分和主成分个数。特

征值计算方法为：

$$\lambda = \|p - [w^T p / (w^T w) w]\| / \|p\| \quad (4-10)$$

当 λ 接近 0 时，在 X 中没有正交变量。于是 λ 值为正交成分的选取提供了重要的参考依据。而关于交互检验方法则和 PLS-DA 的做法一致。OPLS 数据处理中，联用特征值法和交互检验方法是确定正交成分个数的最好途径。

2 核磁共振指纹图谱不同生长期当归 OPLS 分析结果

2.1 数据压缩结果

利用化学计量学软件 SIMCA-p 化学计量学软件进行建模后，OPLS 模型自动拟合后得到了 2 个组分，并去掉了与 Y 年份哑变量正交的核磁共振化学位移的 6 个组分。

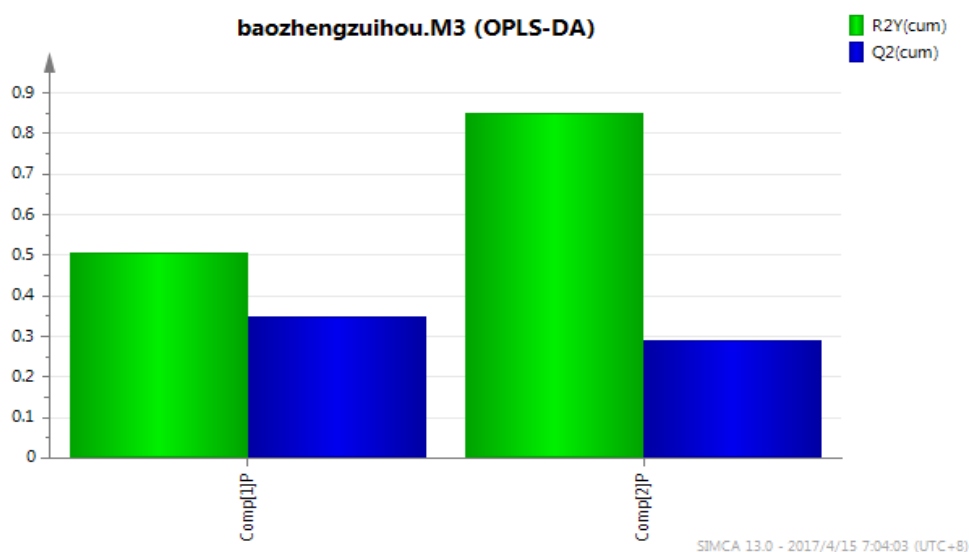


图 4-1 OPLS 正交主成分提取结果

OPLS-DA 模型均具有较高的 R2Y 和 R2X 值，即 X 与 Y 变量的信息值。并且 OPLS 模型体现出一个更高的 Q2(cum)值，达到 0.3 左右。在只有两个主成分情况下 OPLS-DA 的 R^2 就已经达到了 0.853，相较 PLS-DA 的主成分更少，因此 Q^2 的值比较大，也就是说，主成分个数减少，过拟合也得到了相对的改善。

现在的模型分析表展示出：表中数据作为与 Y 变量相关的两个信号组分，并且表中的数据均为 Y 变量的正交的六个组分。累积 $R^2 = 0.853$ ， $Q^2 = 0.30$ 。对 Y 变量当归年份的解释程度和预测能力都较高，并且过拟合现象有所改观。但模型的效果依旧达不到理想。过拟合现象依旧过于明显。

表 4-1 OPLS 主成分提取结果

Component	R2X	R2X(cum)	R2	R2(cum)	Q2	Q2(cum)
-----------	-----	----------	----	---------	----	---------

Model	0.974	0.853	0.30			
Predictive	0.276	0.853	0.30			
P1	0.243	0.243	0.508	0.508	0.348	0.348
P2	0.0321	0.276	0.344	0.853	-0.0579	0.29

表 4-2 正交组分提取表

Orthogonal in X	R2X	R2X(cum)	Eigenvalue
O1	0.311	0.311	8.4
O2	0.206	0.517	5.57
O3	0.0806	0.598	2.18
O4	0.045	0.643	1.21
O5	0.0291	0.672	0.787
O6	0.026	0.698	0.701

在上表中可以看到，则 X 中的正交成分 O6 的 $\lambda = 0.701$ ，接近 0，说明在 X 中已没有正交变量，所以停止提取正交成分。

2.2 到模型的距离

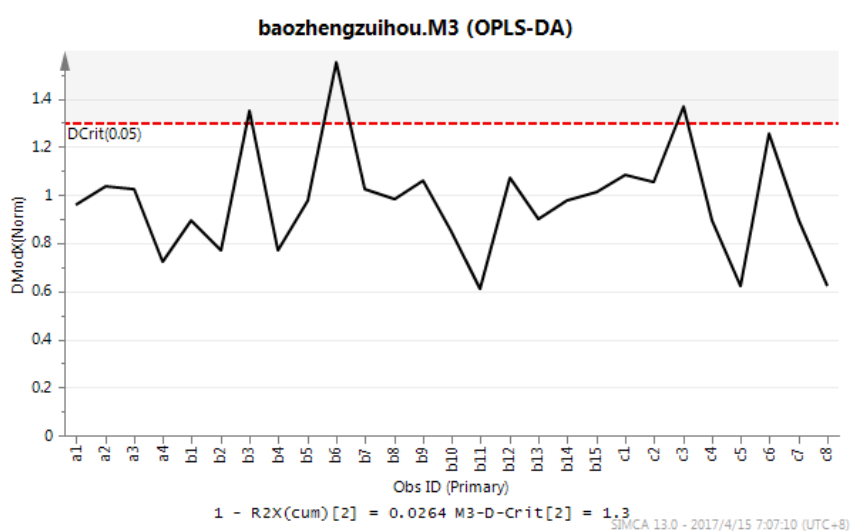


图 4-2 样品到 OPLS 模型距离

大多数样品的“distance to the model”小于临界距离，意味着数据的均一性得到了

保证。这是因为当归 HNMR 指纹图谱数据经过提出与年份哑变量正交提取后，是留下的数据集与 Y 变量的关系更加高，因此样本具模型的距离得到了缩短，数据集的均一性得到了保证。但想获得更为均一的数据可以对 NMR 图谱数据进行更加复杂的信号处理。

2.3 主成分得分图

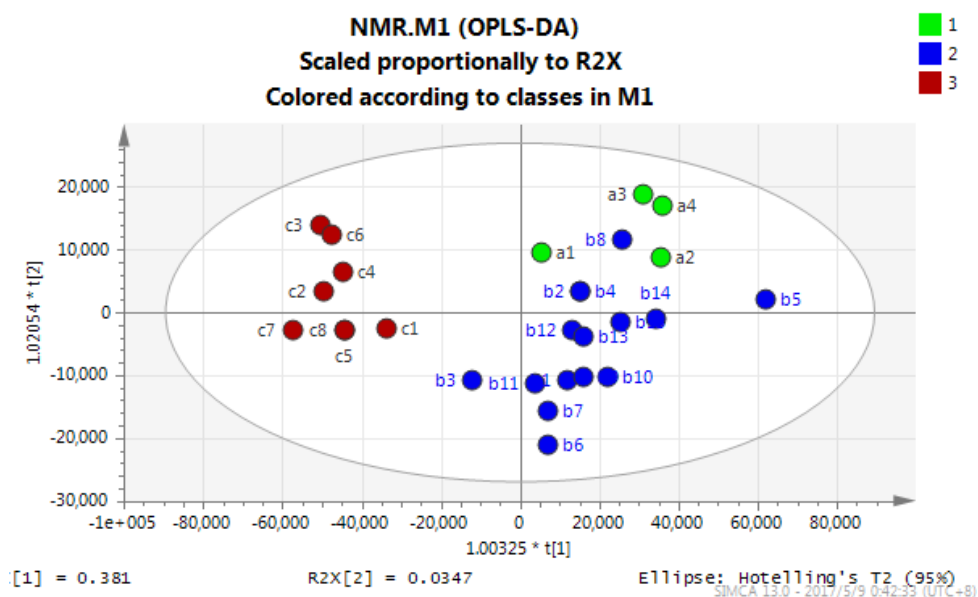


图 4-3 OPLS 主成分得分分类图

如图 4-3 显示，将当归 HNMR 指纹图谱中与所属年份正交的成分别除后，所达到的分类效果较好。不同生长时期的当归样品中，样品基本都能聚集在自己相应的区域内，都被很好地区分。第一年生当归所处部分为第一象限，第二年生当归大部分分布在第四象限，有少量的样品与一年生当归有重叠。第三年生当归分布在第二象限。可以通过所处象限得知，1，2 年通过第 2 主成分区分。1，3 年通过第一象限区分。2，3 年通过第 1、2 主成分分开。分类效果理想，且能通过具体主成分来区分，因此该模型是比较值得信赖的。

2.4 主成分荷载值

OPLS 分析不光是为了区分样品之间的年限分类，还有一个重要的目的，就是在大量化学位移中，找到引起当归差异的代谢物质。这些物质能为探究当归的内在差异、寻

求不同生长时期当归样品的生物标志物、研究药材的药理等提供了很重要的线索。

这些结果最终是以载荷图给出的，载荷图主要是反映了检测变量（如 $^1\text{H-NMR}$ 中的化学位移）的分布情况，载荷图中里面变量的分布与主成分得分图中的样品分布位置一一对应。因此，利用载荷图分析，可以较为清晰地检测到不同年份的样品中造成它们分类差异的化合物，然后再对这些差异物进行定向分析，减少了数据分析的工作量，提高了分析效率^[33]。

而且由于在主成分得分图上可以看到，有年份监督的正交偏最小二乘对样本的分类较为成功，这使载荷分析更为可信。这是 OPLS 的最大的优点之一，即它可以使荷载值具有解释能力。通过荷载值的解释我们就可以看到，区分不同年份当归的 $^1\text{H-NMR}$ 波段位于何处，进而通过对参照物的特征峰比较来发现决定不同生长年限的当归的重要次生代谢物。

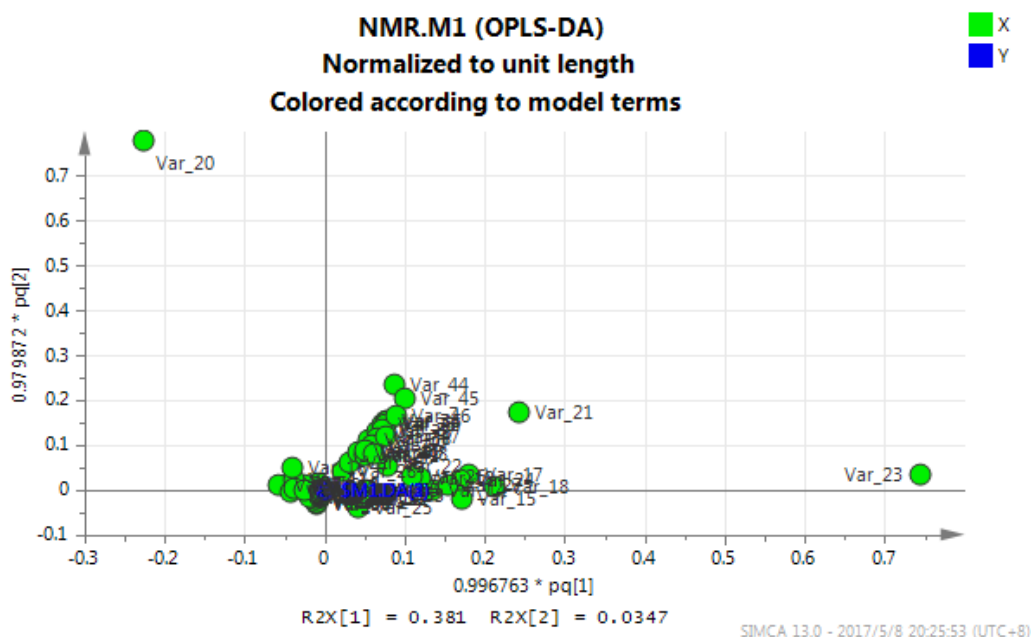


图 4-4 OPLS 模型载荷图

图 4-4 中展示了对区分不同生长年限当归样品有贡献的代谢成分的变化情况。OPLS-DA 的体现代谢成分差异较为明显。变量较为离散则为载荷较大的化学位移，可以看到引起当归样品分类的代谢物的变量主要为 Var_{14-18} , Var_{20-23} , Var_{31-38} , Var_{44-47} , Var_{96-103} 左右。三部分所对应 $^1\text{H-NMR}$ 谱的化学位移分别为 $\delta 3.130-\delta 3.230$, $\delta 3.460-\delta 3.580$, $\delta 4.40-4.46$, $\delta 4.82-4.87$, $\delta 6.34-6.75$ 段。在载荷图中看出对 PC1 贡献最大的代谢成分为化学位移为 $\delta 3.130-3.230$, $\delta 3.580$, $\delta 3.53$, $\delta 6.34-6.75$ 通过参考相关文献和与对照物核磁共振谱对比发现，为欧当归内

酯 A, 当归多糖和阿魏酸。对 PC2 贡献最大的代谢成分为化学位移为 δ 4.40–4.46, δ 3.46, δ 3.53 为未知化合物, 蒿本内酯和当归多糖。

2.5 核磁共振化学位移的显著性检验

通过 OPLS 的载荷图我们可以得出几个区别当归年份的具有累积差异波段, 为了确认波段的实际显著性, 且能够具体的比较不同年份当归的总体之间是否存在差异, 进而确定是不是达到统计学中的差异显著水平, 并以此结果来解析当归在生长周期中次生代谢物的转化、积累和分配情况。因此在此章节我们应用数理统计中的显著性检验来验证在由 OPLS 载荷图筛选出的有着对生长年份较为重要影响的 29 个化学位移上, 三个不同年份的当归总体之间是否在统计学标准中也达到了具有显著差异。

2.5.1 两个正态总体方差比的 F 检验

要比较两个不同年份组份的当归在某一化学位移上的波段是否有差异, 首先要确定两总体的方差是否在可接受的显著水平上相等。那么就要用到检验正态总体的方差比 F 检验。

①方法与原理

设 x_1, \dots, x_m 是来自 $N(\mu_1, \sigma_1^2)$ 的样本, y_1, \dots, y_n 是来自 $N(\mu_2, \sigma_2^2)$ 的样本。考虑如下三个假设检验问题。

$$\begin{aligned} \text{I} \quad & H_0: \sigma_1^2 \leq \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 > \sigma_2^2. \\ \text{II} \quad & H_0: \sigma_1^2 \geq \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 < \sigma_2^2. \\ \text{III} \quad & H_0: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 \neq \sigma_2^2. \end{aligned} \quad (4-11)$$

此处 μ_1, μ_2 均未知, 记 s_x^2, s_y^2 分别是由 x_1, \dots, x_m 算得的 σ_1^2 的无偏估计和由 y_1, \dots, y_n 算得的 σ_2^2 的无偏估计 (两个都是样本方差), 则可建立如下检验统计量:

$$F = \frac{s_x^2}{s_y^2} \quad (4-12)$$

当 $\sigma_1^2 = \sigma_2^2$ 时, 上式的 $F \sim F(m-1, n-1)$, 由此给出三个检验问题随影的拒绝域一次是

$$W_I = \{F \geq F_{1-\alpha}(m-1, n-1)\}$$

$$W_{II} = \{F \leq F_{\alpha}(m-1, n-1)\}$$

$$W_{III} = \{F \leq F_{\alpha/2}(m-1, n-1) \text{ 或 } F \geq F_{1-\alpha/2}(m-1, n-1)\} \quad (4-13)$$

记 $F_0 = \frac{s_x^2}{s_y^2}$ 是由样本计算得到的检验统计量的值^[45]。

② 验证结果与分析

在 matlab2013a 上进行上述算法，对三年份的当归之间的核磁共振波段方差差异进行判断。首先进行了一年生总体与二年生总体关于所选出 29 个可能有差异的核磁共振波段检测。经检验在显著水平 $\alpha=0.05$ 条件下，查表可知

$$F_{0.975}(3,14) = 4.24, \quad F_{0.025}(3,14) = \frac{1}{F_{0.975}(14,3)} = 0.07$$

则拒绝域为：

$$W = \{F \leq 0.07 \text{ 或 } F \geq 4.24\}$$

检验结果如下表：

表 4-3 一二年当归总体方差比的 F 检验

化学位移波段/ δ	F_0	是否显著	化学位移波段/ δ	F_0	是否显著
3.130	0.9510	否	4.460	0.7284	否
3.180	0.7664	否	4.470	0.4912	否
3.190	1.0453	否	4.820	3.3115	否
3.210	0.6134	否	4.850	1.7459	否
3.230	0.7102	否	4.860	1.1482	否
3.460	2.4526	否	4.870	0.7220	否
3.530	0.2328	否	6.330	0.2598	否
3.560	0.7672	否	6.340	0.7401	否
3.580	0.6618	否	6.370	0.2539	否
4.400	1.6404	否	6.380	0.7577	否
4.410	1.6974	否	6.470	0.5326	否
4.420	1.6052	否	6.710	0.5710	否
4.430	1.4481	否	6.720	0.3516	否

4.440	1.2437	否	6.750	0.3789	否
4.450	0.9858	否			

可以看到在第一年和第二年份的当归样品中在这 29 个化学位移上方差均没有差别。随后又进行了一年生与三年生之间和二年生与三年生之间的关于 14 个化学位移方差差异检验，其中一年生与三年生当归的判别域为

$$F_{0.975}(3,7) = 5.89, \quad F_{0.025}(3,7) = \frac{1}{F_{0.975}(7,3)} = 0.068$$

则拒绝域为：

$$W = \{F \leq 0.068 \text{ 或 } F \geq 5.89\}$$

其中二年生与三年生当归的判别域为：

$$F_{0.975}(14,7) = 4.60, \quad F_{0.025}(14,7) = \frac{1}{F_{0.975}(14,7)} = 0.30$$

则拒绝域为：

$$W = \{F \leq 0.30 \text{ 或 } F \geq 4.60\}$$

判别结果都不显著，即可以看做在所有的 29 个化学位移波段的方差差异均不显著，可以认为在显著水平 $\alpha=0.05$ 下认为三各年份的当归在这 29 个变量中是同方差的，这与方差分析的结果是相同的。因此可以继续向下进行 t 检验判断均值差异。

2.5.2 $\sigma_1 = \sigma_2 = \sigma$ 但未知时的两样本 t 检验

① 方法及原理

设 x_1, \dots, x_m 是来自 $N(\mu_1, \sigma_1^2)$ 的样本， y_1, \dots, y_n 是来自 $N(\mu_2, \sigma_2^2)$ 的样本。考虑如下三个假设检验问题^[45]。

$$\begin{aligned} \text{I} \quad & H_0: \mu_1 - \mu_2 \leq 0 \quad \text{vs} \quad H_1: \mu_1 - \mu_2 > 0. \\ \text{II} \quad & H_0: \mu_1 - \mu_2 \geq 0 \quad \text{vs} \quad H_1: \mu_1 - \mu_2 < 0. \\ \text{III} \quad & H_0: \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1: \mu_1 - \mu_2 \neq 0. \end{aligned} \quad (4-14)$$

在 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 但未知时，首先 $\bar{x} - \bar{y} \sim N(\mu_1 - \mu_2, (\frac{1}{m} + \frac{1}{n})\sigma^2)$ ，其次，由于：

$$\frac{1}{\sigma^2} \sum_{i=1}^m (x_i - \bar{x})^2 \sim \chi^2(m-1) \quad , \quad \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 \sim \chi^2(n-1) \quad (4-15)$$

记:

$$s_w^2 = \frac{1}{m+n-2} [\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2], \quad (4-16)$$

于是有

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s_w \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2). \quad (4-17)$$

这就给出了检验统计量为:

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s_w \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (4-18)$$

对于检验问题III, 检验的拒绝域为:

$$W_{III} = \{|t| \geq t_{1-\alpha/2}(m+n-2)\} \quad (4-19)$$

② 验证结果与分析

经过三个总体的两两对比方差, 我们得知可以在 $\alpha=0.1$ 的显著水平下, 不同年份的当归可以看做是同方差的。因此, 我们在 matlab2013a 上进行上述 t 检验算法, 对三年份的当归之间的核磁共振波段的均值是否差异进行判断。首先进行了一年生总体与二年生总体关于所选出 29 个可能有差异的核磁共振波段检测。经检验在显著水平 $\alpha=0.1$ 条件下, 查表可知:

$$t_{0.90}(17) = 1.3334$$

则拒绝域为:

$$W = \{|t| \geq 1.3334\}$$

表 4-4 1、2 年当归总体显著性 t 检验

化学位移波段/ δ	对应化合物	t_0	化学位移波段/ δ	所属化合物	t_0
3.130	欧当归内酯 A	0.7949	4.460	未知化合物	2.285*
3.180	欧当归内酯 A	0.5381	4.470	未知化合物	2.151*
3.190	欧当归内酯 A	0.8173	4.820	未知化合物	2.046*
3.210	欧当归内酯 A	0.7258	4.850	未知化合物	1.625*

3.230	欧当归内酯 A	0.5818	4.860	未知化合物	1.325*
3.460	蒿本内酯	2.407*	4.870	未知化合物	0.9389
3.530	当归多糖	2.125*	6.330	阿魏酸	0.1259
3.560	当归多糖	2.141*	6.340	阿魏酸	0.308
3.580	当归多糖	0.3177	6.370	阿魏酸	0.5332
4.400	未知化合物	2.462*	6.380	阿魏酸	0.2995
4.410	未知化合物	2.411*	6.470	阿魏酸	0.4982
4.420	未知化合物	2.391*	6.710	阿魏酸	0.5414
4.430	未知化合物	2.38*	6.720	阿魏酸	0.2039
4.440	未知化合物	2.365*	6.750	阿魏酸	0.2042
4.450	未知化合物	2.336*			

表中 t_0 带有‘*’的为显著变量

如上表所看来，经过总体均值差 t 检验后，在第一年份的当归和第二年份的当归中差异较大的成分主要有：蒿本内酯，未知化合物以及当归多糖等成分。

运用以上的检测方法，又进行了第一年与第三年当归总体的检测和第二年与第三年的检测分析查表可知则拒绝域分别为： $W_{13} = \{|t| \geq 1.372\}$ ， $W_{23} = \{|t| \geq 1.323\}$

表 4-5 1、3 年当归总体显著性 t 检验

化学位移波段/ δ	对应化合物	t_0	化学位移波段/ δ	所属化合物	t_0
3.130	欧当归内酯 A	4.038*	4.460	未知化合物	4.845*
3.180	欧当归内酯 A	4.474*	4.470	未知化合物	4.679*
3.190	欧当归内酯 A	1.525*	4.820	未知化合物	0.543
3.210	欧当归内酯 A	3.664*	4.850	未知化合物	0.322
3.230	欧当归内酯 A	3.907*	4.860	未知化合物	0.468
3.460	蒿本内酯	0.7201	4.870	未知化合物	0.286
3.530	当归多糖	4.75*	6.330	阿魏酸	2.153*
3.560	当归多糖	4.509*	6.340	阿魏酸	1.678*
3.580	当归多糖	3.834*	6.370	阿魏酸	1.401*
4.400	未知化合物	4.067*	6.380	阿魏酸	1.563*
4.410	未知化合物	4.318*	6.470	阿魏酸	1.997*
4.420	未知化合物	4.562*	6.710	阿魏酸	1.307

4.430	未知化合物	4.66*	6.720	阿魏酸	1.51*
4.440	未知化合物	4.822*	6.750	阿魏酸	2.006*
4.450	未知化合物	4.862*			

表中 t_0 带有‘*’的为显著变量

表 4-6 2、3 年当归总体显著性 t 检验

化学位移波段/ δ	对应化合物	t_0	化学位移波段/ δ	所属化合物	t_0
3.130	欧当归内酯 A	2.462*	4.460	未知化合物	3.386*
3.180	欧当归内酯 A	2.949*	4.470	未知化合物	3.26*
3.190	欧当归内酯 A	2.323*	4.820	未知化合物	2.713*
3.210	欧当归内酯 A	2.545*	4.850	未知化合物	1.163
3.230	欧当归内酯 A	2.613*	4.860	未知化合物	1.043
3.460	蒿本内酯	2.926*	4.870	未知化合物	1.768
3.530	当归多糖	1.606*	6.330	阿魏酸	1.259
3.560	当归多糖	1.724*	6.340	阿魏酸	1.514*
3.580	当归多糖	1.129	6.370	阿魏酸	1.552*
4.400	未知化合物	3.055*	6.380	阿魏酸	1.263
4.410	未知化合物	3.243*	6.470	阿魏酸	1.339*
4.420	未知化合物	3.384*	6.710	阿魏酸	2.709
4.430	未知化合物	3.421*	6.720	阿魏酸	2.612
4.440	未知化合物	3.474*	6.750	阿魏酸	3.184
4.450	未知化合物	3.466*			

表中 t_0 带有‘*’的为显著变量

通过对各年份当归样本关于 29 个高载荷差异变量的两两 t 检验发现,在显著性 $\alpha=0.1$ 的条件下,1 年生和 2 年生的当归样品主要的化合物差异在蒿本内酯,未知化合物以及当归多糖上。而 1 年生和 3 年生的当归样品主要差异是在欧当归内酯 A,未知化合物,当归多糖和阿魏酸上。2 年生和 3 年生的样品体现在蒿本内酯,欧当归内酯 A,未知化合物,当归多糖以及阿魏酸上。通过此 t 检验,查阅参考资料以及成分含量试验,我们大致可以推断当归三年内根部的代谢途径和物质积累方向。

综上,通过 HNMR 核磁共振谱分析出的,区分不同生长时期当归样品的主要代谢

成分阿魏酸、藁本内酯、欧当归内酯 A，当归多糖和一些未知化学成分。其中，1，2 年限当归的代谢化合物差异较大为藁本内酯、当归多糖以及一些未知化合物。区分 1，3 年限当归的代谢化合物为欧当归内酯 A，当归多糖和阿魏酸，区分 2，3 年限当归的代谢化合物为藁本内酯、当归多糖、欧当归内酯 A、阿魏酸以及一些未知化合物。

综合主成分得分图和载荷图分析结果我们可以初步的出结论，当归在第一年到第二年生长过程中，植物体中主要增长的代谢化合物为藁本内酯，第二年到第三年生长过程中主要消耗的是藁本内酯，导致主要药效成分——藁本内酯丧失，并打破合理的化合物比例，导致药用价值大大降低。而阿魏酸和当归多糖一直在生长过程中增长。根据传统采收期，当归中并不是所有的成分含量都达到最大，主要代谢成分的含量之间存在一定的比例关系。而鉴于阿魏酸和当归多糖一直在生长过程中增长，可以判断一定量的阿魏酸和多糖有助于当归中藁本内酯药效的发挥，达到一定的比例最好。

2.6 基于 HNMR 图谱数据 OPLS 模型预测与回判

利用 27 个当归样本所构建的 OPLS-DA 模型对 27 个样品进行回代判别，并对 6 个当归的测试样本进行判别预测。回代预测正确率为 96.29%，判别预测的正确率为 66.67%，由此可见，根据 OPLS-DA 模型和 NMR 指纹图谱数据对当归药材进行所属年份判断，其效果较为一般。各当归测试样本的判别和预测结果见表。

表 4-6 当归样品回判与预测结果

索引号	Y1Pred	Y2Pred	Y3Pred	原所属类	预测类
a1	0.751599	0.141245	0.107156	1	1
a2	0.595119	0.497195	-0.0923141	1	1
a3	0.942207	0.0138984	0.0438947	1	1
a4	0.929595	0.154914	-0.0845086	1	1
b1	-0.155062	1.11801	0.0370529	2	2
b2	-0.0221325	0.903614	0.118519	2	2
b3	0.303243	0.620913	0.0758445	2	2
b4	-0.0221325	0.903614	0.118519	2	2
b5	0.148148	1.00712	-0.155268	2	2
b6	-0.0702667	1.17751	-0.107242	2	2
b7	-0.0800162	0.993793	0.0862237	2	2
b8	0.491949	0.477928	0.0301229	2	1*
b9	0.260747	0.877356	-0.138103	2	2
b10	-0.012805	1.06096	-0.0481523	2	2

b11	-0.0662051	0.918829	0.147376	2	2
b12	0.0775492	0.869414	0.0530369	2	2
b13	0.00900248	0.853801	0.137197	2	2
b14	-0.0790777	1.11816	-0.0390842	2	2
b15	0.0243111	0.997225	-0.0215357	2	2
c1	-0.0474805	0.310273	0.737208	3	3
c2	-0.00854956	-0.0194956	1.02805	3	3
c3	0.177312	-0.208291	1.03098	3	3
c4	0.0207272	-0.0574042	1.03668	3	3
c5	-0.121725	0.199045	0.92268	3	3
c6	0.077213	-0.167905	1.09069	3	3
c7	-0.0015439	0.0392396	0.962304	3	3
c8	-0.121725	0.199045	0.92268	3	3
yc9	0.09573	0.822711	0.0815586	3	2*
yc10	-0.0341221	0.335905	0.698217	3	3
yc11	0.215299	-0.256408	1.04111	3	3
yb16	-0.0526058	0.615146	0.437459	2	2
yb17	0.579454	0.464005	-0.0434587	2	1*
ya5	0.545768	0.191679	0.262554	1	1

‘*’ 为误判样本，‘y’ 为预测样本

由上表可以看出，由 27 个当归样品所构建的 OPLS-DA 模型，对 6 个测试样本的分类还是比较准确，但对第二年和第三年的样本判断不是很清晰。查询 b17 样本年份为 1 月份样本，比较接近第一年。主要代谢成分的“消长”变化随季节变化也比较明显，所以当归样品可能在随不同年份区分的同时，还随不同季节有一定的差异，这就使得某一年份当归样品所在时段是季节成分占据主导而不是年限主导，容易被模型判断错误。

3 气质联用总离子流图图谱不同生长期当归 OPLS 分析结果

本节采用经过方差分析验证后在三个年份的当归总体中存在差异的 4812 个化学离子保留时间，利用正交偏最小二乘（OPLS）从整体分析不同生长年份的当归次生代谢物的变化，对变化较大的代谢物做定量分析，为将来对当归代谢途径研究提供研究基础。随后的特征变量整理和 OPLS-DA 分析及模型验证由 SIMCA-P 13.0 软件实现^[43,44]。

3.1 代谢物的 OPLS-DA 分析

在本论文中，我们采用 OPLS-DA 方法分析在整个成长周期中挥发性成分整体变化

趋势。图 14-5 显示的利用一二主成分得分进行分类的采集于相同年份样品点的聚集度高，说明植物的代谢变化与其生长期相关。

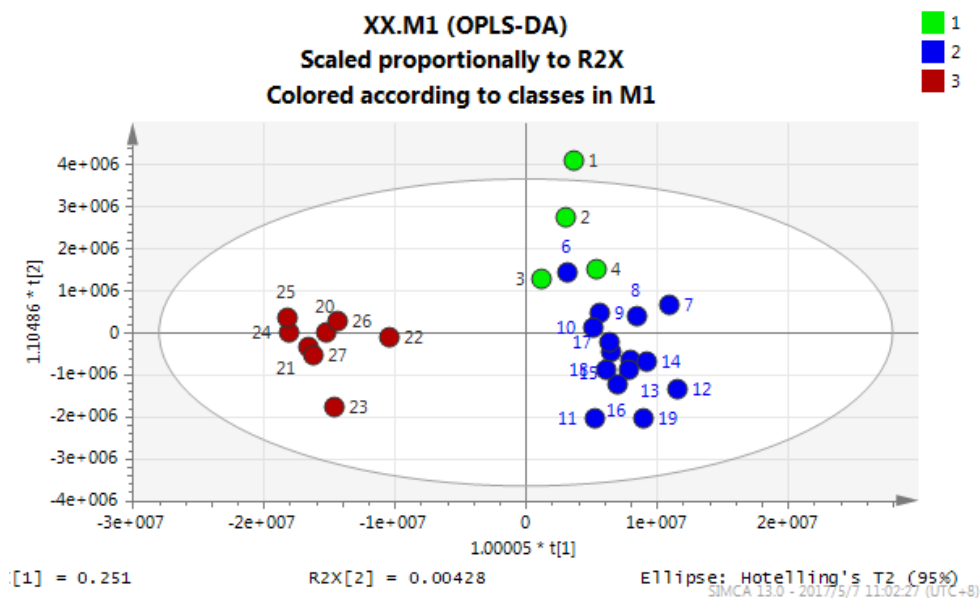


图 4-5 不同生长期当归 OPLS 分析的等分图

如图 4-5 显示，提出三个正交成分后的 OPLS 的主成分得分图按聚集情况可以分为三部分，第一部分包含第一年的 8 月 30 日至 10 月 8 日之间的样本（样本 1-4），第二部分包括采收于第二年全部的样本（样本 5-19），和第三年的 7 月 26 日和 10 月 29 日样本（样本 23 与 27）；第三部分包括采集与第三年全部的样本（样本 11 和 23），可以看到，除去第三年 23 号和 27 号样本外，利用 GCMS 质谱色谱图不同年份的当归被很好地聚集在一起，由此可以说明建立的正交偏最小二乘模型比较具有可靠性的，下表为 OPLS 的分析结果：

表 4-7 OPLS 主成分提取结果

Component	R2X	R2X(cum)	R2	R2(cum)	Q2	Q2(cum)
Model		0.99		0.775		0.606
Predictive		0.255		0.775		0.606
P1	0.251	0.251	0.531	0.531	0.431	0.431
P2	0.00428	0.255	0.244	0.775	0.175	0.606

Orthogonal in X(OPLS)		0.735		0
wO1	0.651	0.651	0	0
O2	0.0598	0.711	0	0
O3	0.0155	0.726	0	0
O4	0.00858	0.735	0	0

在表 4-7 中可以看出，在抽取了 4 个正交成分后，有提取了两个主成分，模型拟合效果较好，可以看到对 4812 个预测变量的累计解释能力已经达到了 0.775，解释说明能力较强。其中在预测变量矩阵中，提取了与 Y 正交成分达到 0.735。而模型的交叉检验结果为 0.606，交叉检验效果较好，模型比较值得信赖。因此我们有理由相信 loading 载荷图中所显示的高载荷变量是对三个年份当归有较强的影响和识别能力的变量即化学保留时间，其所对应的化合物则是在不同年份当归变化较为显著的部分，且与年份有着密切关联的。

图 4-7 显示。在 OPLS-DA 载荷图中，我们规定在第一主成分上载荷值的绝对值达到 0.1 以上或在第二主成分上载荷值的绝对值达到 0.1 以上的为高载荷值保留时间。得到的分别为在 1.198~1.7098min, 2.2855min, 24.3559~24.8735min, 26.1748~26.7562min, 27.5196~27.5662min, 49.9855~50.9741min 上的 39 个保留时间。在当归的不同生长期內载荷值较大的离子保留时间在 NIST2011 数据库中对应的这些化合物包括：

未知化合物 a (大约在 1.198~1.7098min)，未知化合物 b(2.2855 min)，丁烯基苯酐(24.3559~24.8735min)，Z-蒿本内酯(26.1748~26.7562min)，十五烷醇(27.5196~27.5662min)，亚油酸(49.9855~50.9741min)，对于未知化合物由于 NIST11 数据库中并没有显示该保留时间和质荷比的化合物为哪种成分，所以我们暂且以英文字母进行标识。从 loading 图中得出的这些对当归代谢轨迹具有重要影响的化合物也可充当评价不同时期当归的质量的指标性成分。

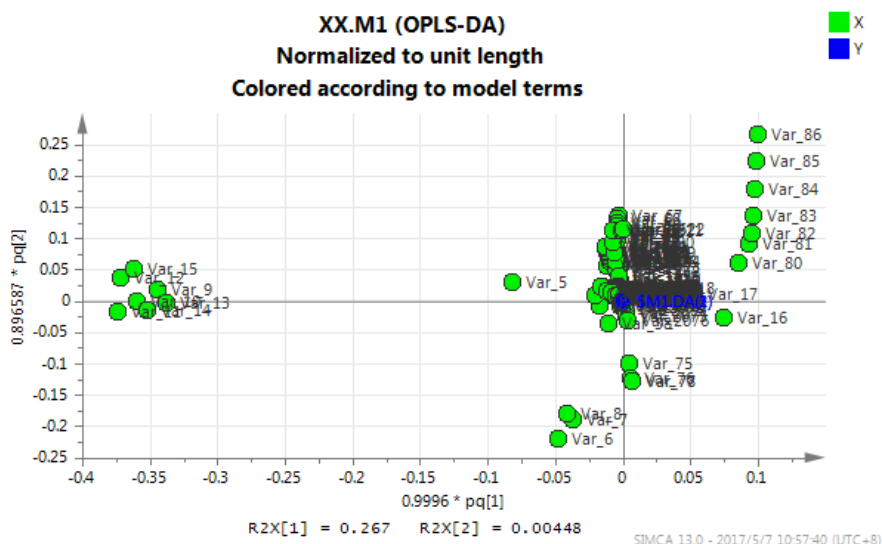


图 4-6 不同生长期当归 OPLS 分析的 loading 图

3.2 挥发性成分的动态变化

不同生长期当归之中有差异挥发性成分的正交偏最小二乘分析结果表明，在当归的生长期中有 6 个化合物的含量有较大的变化，且每个成分的相对含量变化情况在图 4-7 中显示。Z-藁本内酯的成分变化趋势显示，在第一年的样本中含量普遍较高；在第二年 7 月，8 月和第三年的 7 月的含量最低（样本 11，12，13 和 23），这可能与季节的代谢情况有关。同时，可以发现未知化合物 b 的变化趋势和 Z-藁本内酯的变化情况相似，在第二年和第三年的 7 月 8 月含量减少，推测该成分也有可能是当归的药效成分。亚油酸在第二年和第三年的 7 月 26 日含量最高，作为能源物质，此化合物存在于采收期间（样本 17-19），但储存量明显减少。十五烷醇的变化情况比较明显，可以看到在植物体内的相对含量中前一二年中保持了相对较高水平，而到了第二年的 10 月份（样本 19）后含量急剧减少，这恰好与当归根的生物量减少有关，可以推断十五烷醇与根的膨胀有关；丁烯基苯酐的含量在第二年的 5 月到第二年 9 月的样本中相对含量逐渐减小，随后随着时间相对含量开始逐渐增加，由于此时的当归正处于叶盛期，可以推断丁烯基苯酐可能与地上部分的增长有关。总而言之，在当归的整个生长期中，其根中的很多成分显示出较大的变化。

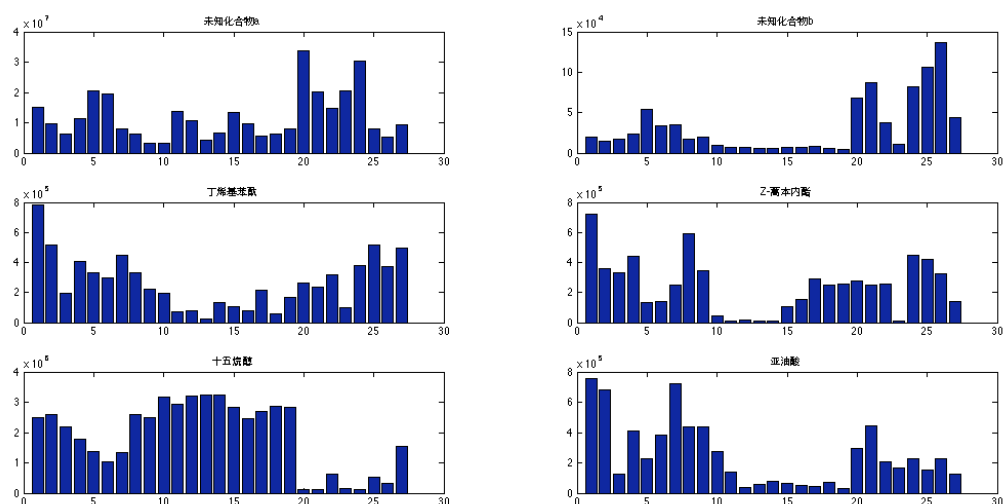


图 4-7 不同生长期当归标志代谢物动态含量变化

3.3 重要化合物依据年份的显著性 T 检验

通过前两章的方差分析和正交偏最小二乘法的结果可以看到追踪到了六种在年份上含量有明显差距，且对当归的生长时期有极大影响的次生代谢化合物。但方差分析只能考察一个因子在各个水平上的均值是否有差距，而不能具体判断两两年份之间的差距。为了更好地比较各个年份之间关于这六种化合物之间有无差距，采用数理统计学上的 t 检验，更为精确的比较化合物年份的变化。

3.3.1 两个正态总体方差比的 F 检验

在 matlab2013a 上进行上述算法，对三年份的当归之间的化学离子保留时间方差差异进行判断。首先进行了一年生总体与二年生总体关于所选出 6 个可能有差异的化学保留时间检测。经检验在显著水平 $\alpha=0.05$ 条件下，查表可知

$$F_{0.975}(3,14) = 4.24, \quad F_{0.025}(3,14) = \frac{1}{F_{0.975}(14,3)} = 0.07$$

则拒绝域为：

$$W = \{F \leq 0.07 \text{ 或 } F \geq 4.24\}$$

表 4-8 一二年当归总体方差比的 F 检验

化学保留时间/min	所属化合物	左侧临界	右侧临界	F_0	是否显著
1. 6284	未知化合物 a	0. 07	4. 24	0. 4646	否
2. 2855	未知化合物 b	0. 07	4. 24	0. 3779	否
24. 5362	丁烯基苯酐	0. 07	4. 24	4. 0146	否
26. 3982	Z-蒿本内酯	0. 07	4. 24	1. 2366	否
27. 5487	十五烷醇	0. 07	4. 24	0. 2575	否
49. 9855	亚油酸	0. 07	4. 24	1. 9257	否

可以看到在第一年和第二年份的当归样品中在这 6 个化学保留时间上方差均没有差别。随后又进行了一年生与三年生之间和二年生与三年生之间的关于 6 个化学保留时间方差差异检验，其中一年生与三年生当归的判别域为

$$F_{0.975}(3,7) = 5.89 \quad , \quad F_{0.025}(3,7) = \frac{1}{F_{0.975}(7,3)} = 0.068$$

则拒绝域为：

$$W = \{F \leq 0.068 \quad \text{或} \quad F \geq 5.89\}$$

其中二年生与三年生当归的判别域为：

$$F_{0.975}(14,7) = 4.60 \quad , \quad F_{0.025}(14,7) = \frac{1}{F_{0.975}(14,7)} = 0.30$$

则拒绝域为：

$$W = \{F \leq 0.30 \quad \text{或} \quad F \geq 4.60\}$$

判别结果都不显著，即可以看做在所有的 6 个化学保留时间波段的方差差异均不显著，这与方差齐性的检验结果相似，可以认为在显著水平 $\alpha=0.05$ 下认为三各年份的当归在这 6 个变量中是同方差的。因此可以继续向下进行 t 检验判断均值差异。

3.3.2 $\sigma_1 = \sigma_2 = \sigma$ 但未知时的两样本 t 检验

经过三个总体的两两对比方差，我们得知可以在 $\alpha=0.1$ 的显著水平下，不同年份的当归可以看做是同方差的。因此，我们在 matlab2013a 上进行上述 t 检验算法，对三年份的当归之间的化学保留时间的均值是否具有差异进行判断。首先进行了一年生总体与二年生总体关于所选出 6 个可能有差异的化学位移保留时间检测。经检验在显著水平 $\alpha=0.1$ 条件下，查表可知：

$$t_{0.90}(17) = 1.3334$$

则拒绝域为:

$$W = \{|t| \geq 1.3334\}$$

t 检验判别结果如下表:

表 4-9 1、2 年当归总体显著性 t 检验

化学保留时间/min	所属化合物	单侧临界值	t_0	是否显著
1.6284	未知化合物 a	1.3334	0.4836	否
2.2855	未知化合物 b	1.3334	0.5136	否
24.5362	丁烯基苯酐	1.3334	3.4085	是
26.3982	Z-蒿本内酯	1.3334	3.1273	是
27.5487	十五烷醇	1.3334	0.7848	否
49.9855	亚油酸	1.3334	2.3131	是

如上表所看来, 经过总体均值差 t 检验后, 在第一年份的当归和第二年份的当归中差异较大的成分主要有: 丁烯基苯酐, Z-蒿本内酯, 亚油酸等成分。

运用以上的检测方法, 又进行了第一年与第三年当归总体的检测和第二年与第三年的检测分析, 检验分析结果如下:

表 4-10 1、3 年当归总体显著性 t 检验

化学保留时间/min	所属化合物	单侧临界值	t_0	是否显著
1.6284	未知化合物 a	1.372	1.3102	否
2.2855	未知化合物 b	1.372	1.2448	否
24.5362	丁烯基苯酐	1.372	1.2802	否
26.3982	Z-蒿本内酯	1.372	2.0990	是
27.5487	十五烷醇	1.372	6.4726	是
49.9855	亚油酸	1.372	2.4168	是

表 4-11 2、3 年当归总体显著性 t 检验

化学保留时间/min	所属化合物	单侧临界值	t_0	是否显著
1.6284	未知化合物 a	1.323	1.0262	否
2.2855	未知化合物 b	1.323	2.5047	是

24.5362	丁烯基苯酐	1.323	2.7273	是
26.3982	Z-蒿本内酯	1.323	1.3454	是
27.5487	十五烷醇	1.323	7.3698	是
49.9855	亚油酸	1.323	0.3273	否

通过对各年份当归样本关于 6 个高载荷变量的两两 t 检验发现, 在显著性 $\alpha=0.1$ 的条件下, 1 年生和 2 年生的当归样品主要的化合物差异在丁烯基苯酐, Z-蒿本内酯, 亚油酸。而 1 年生和 3 年生的当归样品主要差异是在 Z-蒿本内酯, 十五烷醇和亚油酸上。2 年生和 3 年生的样品体现在未知成分化合物 b, 丁烯基苯酐, 亚油酸上。通过此 t 检验, 查阅参考资料以及成分含量试验, 我们大致可以推断当归三年内的代谢途径。

当归在生长期第一年第二年的生长过程中主要积累的化合物在 Z-蒿本内酯, 丁烯基苯酐, 亚油酸上且相对含量一直缓步在增长, 且随着当归的植物体生物量在增加, 使上述化合物的总含量增加。而第二年的生长过程中, 当归的 Z-蒿本内酯和亚油酸有相对含量有下降的趋势。而到了当归的第二年到第三年的交界生长时期, 未知化合物 b 的含量增加巨大, 根据当归生长时期来判断, 应该是与当归花和果实的生长有关。且主要药效成分—Z-蒿本内酯丧失, 并打破合理的化合物比例, 导致药用价值大大降低, 因此传统的当归采集时间第 2 年的 10 月左右是药用价值最大的时期。可以看出我国劳动人民在长期的经验积累过程中积攒起来的经验和智慧是值得现代人借鉴的。

4 结论

提取正交成分的 PLS 模型是更加具有现实的解释意义, 而且仅两个主成分和一个解释性的荷载值图。通过消除与 Y 年份哑变量正交的指纹图谱信息, 留下的数据集与 Y 变量的关系更加高, 因此样本距模型的距离得到了缩短, 数据集的均一性得到了保证。因此利用正交偏最小二乘分析法 (OPLS) 分析了当归在生长的各个时期中, 成分变化较大且与生长年份有关的次生代谢物, 所达到的分类效果较好。

对于不同生长时期的当归样品中, 在主成分载荷图中各个年份的样本样品基本都能聚集在自己的区域内, 都被很好地区分。在载荷图中除了发现了几个主要的药效成分的代谢变化较大。根据不同的指纹代谢图谱分析, 还将丁烯基苯酐和亚油酸等多个物质作为影响当归生长时期的指标性成分。因此我们可以在研究不同时期的当归, 可以选择不同的对照物质, 不仅仅局限于蒿本内酯和阿魏酸;

其次, 利用数理统计中的总体均值的 T 检验, 根据两不同的代谢图谱相对比结合,

发现当归在第一年到第二年生长过程中, 植物体中主要增长的代谢化合物为 Z-蒿本内酯, 第二年到第三年生长过程中主要减少的是 Z-蒿本内酯, 而阿魏酸和当归多糖一直在生长过程中增长。根据传统采收期, 可以初步判断蒿本内酯是当归的主要药效成分。而鉴于阿魏酸和当归多糖一直在生长过程中增长, 可以判断一定量的阿魏酸和多糖有助于当归中蒿本内酯药效的发挥。OPLS-DA 模型的回判效果和预测效果不是很理想, 可能由于每一个年度的生长过程中都要经过春夏秋冬四个季节, 其主要代谢成分的“消长”变化随季节变化也比较明显^[49], 所以当归样品可能在随不同年份区分的同时, 会随不同季节有一定的差异, 而错判的样品接近年份之间的交替月份, 所以季节因素可能占据主导, 年份原因不明显。随后, 结合了当归内各化合物的含量变化与变化所出的时期来大胆推测次生代谢物可能具有的功效。

最后, 在对不同生长年限当归重要代谢物质含量进行分析时, 可以发现在传统采收期时, 并非其药用成分相对含量是最高的时期, 再结合生物量在传统采摘时期达到最大。因此可以猜测, 可能是某几种成分之间达到一种比例使得当归具有最大药效;

第五章 GC-MS 与 HNMR 两种代谢组学方法差异比较

对于复杂混合物的剖析, 气相色谱方法和核磁共振法都是重要方法, 他们也在一般有机物的结构分析中发挥着越来越大的作用。这两种代谢组学分析方法各有优缺点。例如核磁共振技术, 它的样品不需要繁琐处理, NMR 的优点在于能够无创性、无偏向地检测样品, 对所有化合物的灵敏度相同, 且可提供化合物的结构信息, 具有良好的客观性和重现性, 它可对样品进行无需标准品的原位定量分析, 有效减少预处理的损失和溶剂的干扰^[7,8]。相比 NMR 来讲 GCMS 气相色谱联用技术具有高分辨、快速扫描能力。且关于 GCMS 气相色谱质谱联用具有大量标准代谢物数据库, 对于测定相对分子质量及推算分子式时, 气质谱是无可取代的。往往使用一种分析技术往往难以获得全部未知物的化合物结构, 组成和含量等各方面综合信息。所以多种现代分析技术的, 联用和相互印证确认, 使得能获得更为准确的混合物定性定量的剖析结果。在上述两章中我们对利用气相色谱法获得的 GC-MS 当归指纹图谱和利用核磁共振法获得的 HNMR 指纹图谱分别进行了关于当归挥发油的成分及相对含量的测定, 且对得到的指纹图谱进行关于不同生长时期有重要影响的化合物数据挖掘, 得到结果不论是实验结果还是数据挖掘

结果都有一定的差异，因此在本章我们重点讨论挖掘 GC-MS 代谢图谱与 NMR 代谢图谱的异同，以期弥补单一分析技术的不足，发挥多种优势，探索多种分析技术的耦合的代谢组学研究，用于提高代谢产物的检测灵敏度和通量。

1 代谢图谱物质定性差异

相比存在 NIST11 数据库的 GC-MS 气相色谱质谱指纹图谱，HNMR 核磁共振图谱单单仅凭借对照品实验的指纹图谱来确定各化合物的标志特征峰。所能确定的标志峰较少，且操作复杂。因此所能确定的化合物也比较少，只有标志性的几个化合物。例如：E-蒿本内酯，棕榈酸，反油酸等。但通过对照品实验 HNMR 核磁共振谱确定了几种当归的主要化合物相应的化学位移是 GC-MS 指纹图谱无法确定的，例如阿魏酸，3-正丁基苯酐、欧当归内酯等化合物是 GC-MS 气相色谱质谱图上无法确定的。下表显示当归中主要成分在两种图谱中的存在情况。

表 5-1 当归分析的代谢图谱差异化合物

HNMR 核磁共振图谱	共同化合物	GC-MS 气相色谱质谱
3-正丁基苯酐	Z-蒿本内酯	E-蒿本内酯
欧当归内酯 A	亚油酸	棕榈酸
阿魏酸	当归多糖	反油酸
	丁烯基苯酐	Z-芥酸酰胺
		二十九烷
		2-甲氧基-4-乙烯基苯酚

由于实验条件不同，核磁共振技术在当归中提取的混合物的方法是利用 DMSO 二甲基亚砜溶剂提取的，所以像上表 5-1 中 GCMS 气相色谱技术中可以检测到的 Z-芥酸酰胺，棕榈酸以及二十九烷等不溶于 DMSO 溶剂的将无法用实验手段检验出来^[30]。而反观 GCMS 气相色谱技术，在毛细色谱柱内采用的是在当归的干燥根中通过升温到 250℃ 所得到的挥发油。但很不幸的事，两种当归内有药用价值的重要化合物欧当归内酯 A 以及阿魏酸的物理性质稳定，沸点分别为 298℃ 以及 352℃，实验设计本身的缺陷，使得两种重要化合物无法被气化，气相色谱技术将无法检验到这两类物质的存在^[54]。而相对 HNMR 核磁共振谱图来说由于 GCMS 具有 NIST 2011 数据库检索，可以更多的确定在 HNMR 谱图上无法被确定的化合物。所以两者谱图的相互对照，可以相互耦联地说明在当归中年生长年份有着敏感变化的重要代谢成分。

2 代谢指纹图谱对的敏感度分析敏感度

不管是对 GCMS 气相色谱质谱图还是 HNMR 核磁共振图谱，都对其构建了方差分析统计模型，从两个指纹图谱中分别甄别出在各年份当归中有差异的成分。对于两个图谱都可以鉴别出来的代谢化合物，在每个图中的产生依据年份变化的相对变化量敏感度又不尽相同，能否很好的反应重要代谢化合物的年份变化是一种剖析技术的关键。下表反映了两图谱对共有化合物生长时期的变化反映的灵敏度：

表 5-2 ANOVA 年份变化灵敏度化合物

HNMR 灵敏化合物	共同灵敏化合物	GCMS 灵敏化合物
欧当归内酯 A	蒿本内酯	丁烯基苯酐
3-正丁基苯酐	亚油酸	棕榈酸
当归多糖		反油酸
未知化合物 a		苯氧基-2,2'-亚甲基
阿魏酸		

HNMR 与 GCMS 指纹图谱对反应化合物依年份变化均敏感的有蒿本内酯，亚油酸等重要化合物。对丁烯基苯酐来说，GCMS 气相色谱质谱图对其年份变化的反应能力要强于 HNMR 核磁共振谱。而对当归多糖的反应来说，由于多糖物质大多是挥发性较强的，所以在气相分析中，在挥发油中的含量几近饱和⁰，所以在 GCMS 中不论年份怎样变化，多糖类物质的相对含量变化并不明显，而 HNMR 核磁共振谱选择溶剂为 DMSO 二甲基亚砷溶剂，当归多糖对此溶剂溶解度适中，对多糖相对含量的年份反应很明显。而阿魏酸和欧当归内酯 A 是 GCMS 气相色谱质谱所不具有的，故理解两图谱的差异对当归挥发油的各技术的耦联分析至关重要。

在对 loading 载荷图分析上显示了对不同生长时期当归有极大影响的次生代谢产物，但两张代谢图谱在片对偏最小二成分析上的结果也有不同，具体差异如下表：

表 5-3 loading 载荷图分析差异化合物

HNMR 重要化合物	共同灵敏化合物	GCMS 重要化合物
欧当归内酯 A	蒿本内酯	丁烯基苯酐
当归多糖	亚油酸	十五烷醇
未知化合物 a		未知化合物 b
阿魏酸		未知化合物 c

在对共同敏感的化合物的三年中的代谢过程分析中,不论是偏最小二乘的载荷图中还是数理统计检验模型中代谢变化都没有差距。但对于一些代谢化合物而言两者的敏感度就有些差距。根据正交偏最小二乘法的 loading 图中,两种代谢图谱对蒿本内酯和亚油酸的反应很灵敏,通过把握两种物质在图谱上的相对特征峰面积即可推断当归的相应年份。但对各年份当归有重大影响的物质也有不尽相同,HNMR 核磁共振谱图中显示,通过欧当归内酯 A,当归多糖以及阿魏酸的相对物质含量即可有效的把握当归的生长年份。而在 GCMS 指纹图谱中只要测定丁烯基苯酞和十五烷醇的相对含量就可把握当归的生长时期。因此研究探索各个分析手段的指纹图谱对物质变化的灵敏度,是对分析化学的一个重大推动,也是对耦联各个剖析技术对未知物进行分析的重要的一步。

3 保留时间和核磁共振特征峰之间的关联分析

Z-蒿本内酯是当归中重要活性成分^[2],也是公认的用于鉴定当归和评定其质量的特征指标之一,然而 HNMR 核磁共振图谱中只按照对照品的实验谱峰分析出在 δ 1.02、 δ 5.48、 δ 6.02、 δ 6.18 为蒿本内酯的特征信号峰,并未像 GCMS 气质联用图谱有标准数据库可以鉴定区分出 Z-蒿本内酯与 E-蒿本内酯,若能够利用 GCMS 气质联用图谱中关于蒿本内酯的信息来定位 HNMR 核磁共振图谱上的每个特征峰的所属为 Z-蒿本内酯还是 E-蒿本内酯,便可利用 NMR 代谢图谱的无创性的特点来科学地利用当归中的重要药效成分了。本节利用灰色关联分析来比较判断两种类型的蒿本内酯与四个化学位移波段的关联性,进而确定这四个波段的所代表的蒿本内酯分别是哪一种类型的蒿本内酯。

3.1 灰色关联分析法的建模过程

灰色关联分析考虑到相关系数方法的种种弊病和不足,采用关联分析的方法来作系统分析。作为一个发展变化的系统,关联度分析事实上是动态过程发展态势的量化分析。即发展态势的量化比较分析。灰色关联分析法的基本思想是根据各比较数列集构成的曲线族与参考数列构成的曲线之间的几何相似程度来确定比较数列集与参考数列之间的关联度,比较数列构成的曲线与参考数列构成的曲线的几何形状越相似,关联度越大^[51]。

参考数据列应该是反映系统行为特征的数据序列,在两代谢图谱的蒿本内酯的灰色关联分析中是指 GCMS 气质联用图谱中,在 Z-蒿本内酯与 E-蒿本内酯的保留时间下的离子密度强度,利用灰色关联分析进行综合评价的步骤如下:

参考数据列常记为 x_0 , 记第 1 个时刻的值为 $x_0(1)$, 第 2 个时刻的值为 $x_0(2)$, 第 k 个时刻的值为 $x_0(k)$ 。因此,参考序列可表示为 $x_0 = (x_0(1), x_0(2), \dots, x_0(n))$ 关联分析中

被比较数列记为 x_1, x_2, \dots, x_n 。

对于一个参考数据列 x_0 ，有几个比较数列 x_1, x_2, \dots, x_n 的情况。可以用下述关系表示各比较曲线与参考曲线在各点（时刻）的差。

$$\xi_i(k) = \frac{\min_i(\Delta_i(\min)) + 0.5 \max_i(\Delta_i(\max))}{|x_0(k) - x_i(k)| + 0.5 \max_i(\Delta_i(\max))} \quad (5-1)$$

式中， $\xi_i(k)$ 是第 k 个时期的样本比较曲线 x_i 与参考曲线 x_0 的相对差值，它称为 x_i 对 x_0 在 k 时刻的关联系数。其中，0.5是分辨系数，记为一般在0与1之间选取；

$$\min_i(\Delta_i(\min)) = \min_i(\min_k |x_0(k) - x_i(k)|) \quad (5-2)$$

$$\max_i(\Delta_i(\max)) = \max_i(\max_k |x_0(k) - x_i(k)|) \quad (5-3)$$

这是由于作关联度计算的数列的量纲最好是相同的，当量纲不同时要化为无量纲。此外还要求所有数列有公共交点。为了解决这两个问题，计算关联系数之前，先将数列作初值化处理，即用每一个数列的第一个数 $x_i(1)$ 除其它数 $x_i(k)$ ，这样既可使数列无量纲又可得到公共交点 $x_i(1)$ 即第1点。

关联系数的数很多，信息过于分散，不便于比较，为此有必要将各个时刻关联系数集中为一个值，求平均值便是做这种信息处理集中处理的一种方法。关联度的一般表达式为：

$$r_i = \frac{1}{N} \sum_{k=1}^N \xi_i(k) \quad (5-4)$$

3.2 灰色关联分析结果

根据以上算法，以 GCMS 气质联用图谱中在 Z-蒿本内酯与 E-蒿本内酯的保留时间下的离子密度强度为参考数据列，以核磁共振谱 δ 1.02、 δ 5.48、 δ 6.02、 δ 6.18 化学位移的特征峰强度作为比较数列，通过 matlab2013a 进行编程计算。得到如下结果：

表 5-4 保留时间与化学位移之间关联度

	δ 1.02	δ 5.48	δ 6.02	δ 6.18
Z-蒿本内酯	0.1121	0.4706	0.5598	0.2318
E-蒿本内酯	0.5164	0.4604	0.2825	0.2053
Z 蒿+E 蒿	0.3421	0.6201	0.4013	0.3967

表 5-5 Pearson 相关系数表

	δ 1.02	δ 5.48	δ 6.02	δ 6.18
Z-蒿本内酯	0.5054	0.6506	0.7598	0.3318
E-蒿本内酯	0.7203	0.5264	0.5234	0.3552
Z 蒿+E 蒿	0.6403	0.7028	0.6604	0.5385

从表 5-4 与 5-5 中我们不难发现每一个的核磁共振图谱的化学位移波段与 Z-蒿本内酯或 E 蒿本内酯的离子强度密度的灰色关联度都有所不同。且经过 Pearson 相关分析后，可以看到，Pearson 相关系数与灰色关联度的结果的相似度大致相同。在两种关联度表格中表，都可以发现在核磁共振谱中化学位移为 δ 1.02 的特征峰与 E 蒿本内酯的离子流量峰的相关系数和关联度最高，而 δ 6.02 的特征峰和 Z 蒿本内酯的相关系数和关联度最高，而 δ 5.48 位移处的特征峰则与 Z、E 蒿本内酯的混合离子密度峰的相关性最强。因此可以得出如下结论， δ 1.02 的化学位移为 E 蒿本内酯的特征峰，而 δ 6.02 化学位移的特征峰为 Z 蒿本内酯的特征峰。 δ 5.48 处的特征峰为两化合物均有的特征峰。也可以确定在核磁共振对比试验中，所采用的蒿本内酯的对照品为两种蒿本内酯的混合物，但若想更准确地发掘 HNMR 核磁共振谱的 Z-蒿本内酯和 E-蒿本内酯特征峰，还需另找对照品做对照试验实验。

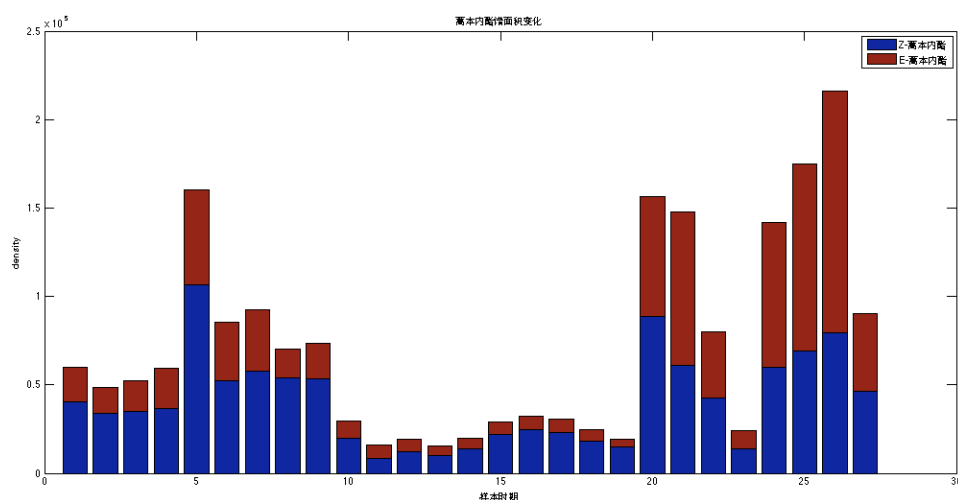


图 5-1 蒿本内酯综合含量变化

上图 5-1 是经由 GCMS 气质联用图谱所测定的 Z 蒿本内酯和 E 蒿本内酯的离子流量总和，可以看到蒿本内酯在整个三年的周期中的变化趋势。可以发现，在当归的生长时期中，前一两年当归中挥发油中是 Z-蒿本内酯占据主导，大致的相对含量 Z-蒿本内酯：E 蒿本内酯=3：1，而到第三年中转变为 E-蒿本内酯占据主导，但大致比例为 1：1。在当归药用采摘时期时可以发现是恰恰是 Z-蒿本内酯相对含量在蒿本内酯中占据最大的时期。

第六章 结果讨论

1 结论

本论文主要利用代谢组学的两种实验方法——核磁共振技术（HNMR）和气质联用技术（GCMS），并结合了统计学的专业知识——统计差异检验和 OPLS 多元统计分析，按照点-面-点的方式来探究当归生长过程中生长期对当归药用部分中主要药效成分所造成差异，以及这些药效成分的代谢变化和运输方向，用来控制当归的采收时节，科学地把控当归的代谢物质含量和鉴定当归的生长年份。经过在两张指纹代谢图谱中的藕联分析的结果表明，在不同年的相同生长期内，好些化合物具有相同的变化趋势，而且部分化合物彼此之间存在消长关系（如 Z 蒿本内酯与 E 蒿本内酯之间就有较强的消长现象）。在当归生长的三年中，蒿本内酯，欧当归内酯，阿魏酸（HNMR 检测）丁烯基苯酐，当归多糖以及十五烷醇（GCMS 检测）等变化较为明显。并研究了其中的重要代谢成分例如 Z 蒿本内酯，亚油酸，丁烯基苯酐等药效化合物的相对含量变化。在第二部分中，通过数理统计中的显著性 t 检验分析出各个生长年份的当归之间差异性较大的代谢物质，一次说明在生长过程中随各个生长时期的发育需要，以及引发代谢物积累、分解、转移等，表征出代谢变化；并且为当归人工养殖以及开发利用提供依据。在第三部分，对比两种代谢组学分析方法的差异，分析两种代谢指纹图谱在分析当归生长时期未知变化构

成差异的机制,探索出了两种化学分析方法对当归中某些物质的灵敏性以及多元统计分析方法得适用性。最后还利用 GCMS 图谱拥有 NIST11 数据库的优势在核磁共振图谱上成功定位了对照实验无法准确把握的 Z 蒿本内酯与 E 蒿本内酯的化学位移。代谢组学 GC-MS 和 HNMR 分析是可以搞效监控生长于不同环境和时期的代谢物的影响方法,在以分子组成为基础中草药质量控制,优培优育方面有着广泛的应用前景。同时,基于代谢组学方法还可以预测植物提取物的活性甚至毒性,在药物研究领域是一种很有潜力的技术分析工具。

2 体会与展望

本文基于 HNMR 和 GC-MS 的代谢组学方法对当归研究中,由于各方面的限制的影响因素,对于当归的年份差异的分析仍然存在以下值得进一步挖掘和探讨的问题:

- (1) 代谢组学研究是对实验设计要求较为苛刻的,在不同生长期当归研究的实验中;虽然尽可能的要求对生长环境的统一,但难免存在人为因素的影响,造成一定的误差;但此次实验未达到 4-6 次的样品重复,所以未能消除环境、实验操作、人为因素所带来的误差^[35]。
- (2) 由于技术和实验条件的欠缺,不管对当归 HNMR 或是 GCMS 图谱来说都只是通过特征信号峰的积分面积来求出该特征峰的所属化合物的相对含量,并没有对其中的重要化合物进行一测多评的定量分析^[53],因此忽略了植物体在生长周期中的自重变化。导致对当归代谢物的年份变化的解释缺乏直观性和有效性。
- (3) 在对当归 GC-MS 分析时,缺少了荷质比的研究,导致部分的生物信息缺失,使得对当归成分的研究不够完整。所以当归被衍生化的物质有待进一步研究。
- (4) 由于代谢物的繁杂、庞大, NIST 数据库的容量以及分析技术局限,例如对物理性质比较稳定或不溶于某些溶剂或过饱和等性质,都会使含量分析上存在误差。对代谢途径的研究仍是个瓶颈问题。为了能全面的开发利用,这部分内容需进行进一步的研究。
- (5) 在偏最小二乘分析中利用其衍生算法正交偏最小二乘,预测效果和分类效果并未达到十分理想,对于想获得更为均一的数据的话,可以对 GCMS, NMR 图谱数据进行更加复杂的信号处理分析,例如正交信号矫正以及微波压缩等^[48]。
- (6) 在方差分析模型筛选变量时,可以对方差齐性的限制缩小,使模型可以筛选出均值有差异但变化幅度不一致的化学位移和保留时间。

结束语(或致谢)

在本论文完成之际，我谨向所有曾给予我关心和帮助的人致以最诚挚的谢意！由衷感谢导师许青松教授！在我完成本科生论文期间，给予的深切关怀、耐心指导、严格要求。本论文从选题、设计到完成得到了导师的悉心指导，倾注了许老师的大量心血。许导师严谨的治学态度、渊博的学识、开拓敏锐的思维给我极大的启迪。特别是导师对事业执着的追求所体现出来的学者风范，严谨求实的科研态度，一直激励着我，并将终生受益。

感谢数学与统计学院的各位恩师在我本科生四年的学习及论文完成过程中给予的教诲和帮助，他们几年来的教育是值得我一生珍惜的宝贵财富。

感谢我的舍友和同学等在课题完成过程中给予的热心帮助和支持。

衷心的感谢我的父母和其他亲朋好友对我的关心、支持和理解，他们对我的关心、鼓励和支持，是我努力学习的动力源泉。

最后，感谢曾经教育和帮助过我的所有老师。衷心地感谢为评阅本论文而付出宝贵时间和辛勤劳动的专家和教授们！

参考文献

- [1] 中华人民共和国药典(一部) [M].北京:中国医药科技出版社, 2010. 124-125.
- [2] 宋秋月,付迎波,刘 江,等.当归的化学成分研究[J]. 中草药, 2011, 42(10): 1 900-1 904.
- [3] 陈波,张巍,康海宁,等. 茶叶的 HNMR 指纹图谱研究[J]. 波谱学杂志, 2006, 23(2): 170-180.
- [4] 孙庆雷,林云良,王晓,等.黄芩的 HNMR 指纹图谱研究[J]. 波谱学杂志, 2007, 24 (2) : 163-168.
- [5] Tarachiwin L, Masako O and Fukusak E, Quality Evaluation and Prediction of Citrullus lanatus by ^1H NMR-Based Metabolomics and Multivariate Analysis. J. Agric. Food Chem.
2008, 56, 5827–5835.
- [6] Kang J, Choi M.Y, Kang S. M,etc, Application of a ^1H Nuclear Magnetic Resonance (NMR)

Metabolomics Approach Combined with Orthogonal Projections to Latent

Structure-Discriminant Analysis as an Efficient Tool for Discriminating between Korean and

Chinese Herbal Medicines, *J. Agric. Food Chem.* 2008, 56, 11589–11595.

- [7] Zulak K. G, Weljie A. M, Vogel H. J and Facchini P. J, Quantitative HNMR metabolomics reveals extensive metabolic reprogramming of primary and secondary metabolism in elicitor-treated opium poppy cell cultures, *BMC Plant Biology*, 2008, 8:5.

- [8] Xiao C. N, Dai H, Liu H, Wang Y. LAND Tang H R, Revealing the Metabonomic Variation of

Rosemary Extracts Using ¹HNMR Spectroscopy and Multivariate Data Analysis, *J. Agric.*

Food Chem. 2008, 56, 10142–10153.

- [9] Zhang L. X, Zhang W, etc. Resolving co-eluting chromatographic patterns by means of dissimilarity analysis in iterative target transformation factor analysis[J]. *Chromatogr.*

A

1218(2011) :7219-7225.

- [10] 刘琳娜,梅其炳,程建峰.当归挥发油的化学成分分析[J]. *Chinese Traditional Patent Medicine*(中

成药), 2005, 27(2): 204-206.

- [11] 袁久荣,容蓉,杨东.当归饮片精油成分的研究[J].*中国中药杂志*,1998, 23(10) : 601-605.

- [12] 潘新坡等. NIR 近红外光谱用于中药材鉴别的研究[J].*药物分析杂志*, 2012,15(4):15-20.

- [13] 萨日娜.基于 GC-MS 的当归生长过程代谢组学研究[D].甘肃中医药大学,2012.

- [14] 刘晖, 倪京满.当归地上与地下部分挥发油的比较研究[J].*西北药学杂志*.

- [15] Wei S.Y, Xu C.J, etc. Analytical comparison of different parts of Radix Angelicae Sinensis by gas chromatography coupled with mass spectrometry[J]. Chromatogr. A. 1187(2008): 232–238.
- [16] Liang X.M, Jin Y, etc. Qualitative and quantitative analysis in quality control of traditional Chinese medicines[J]. Chromatogr. A. 1216(2009): 2033–2044.
- [17] Na Jom K, Frank T, etc. A metabolite profiling approach to follow the sprouting process of mung beans (*Vigna radiata*)[J], Metabolomics. 7(2011):102–117.
- [18] Wishart D. S, Metabolomics: applications to food science and nutrition research, Trends[J]. Food. Sci. Tech.19(2008): 482–493.
- [19] Ge Y.L, Qian D.W, etc. Studies on the dynamic accumulations of Radix Angelica sinensis volatile constituents in different regions and harvest times and the appropriate harvest time, Chin[J]. Pham. Anal. 29(2009): 517-523.
- [20] Piao X.L, Park J. H, etc. Development of gas chromatographic/mass spectrometry-pattern recognition method for the quality control of Korean Angelica[J]. Pharm. Biomed. Anal.

(2007): 1163–1167.

- [21] Choi Y.H, Kim H.K, Hazekamp A, Erkelens C, Lefeber A.W.M, Verpoorte R. Metabolomic Differentiation of Cannabis sativa Cultivars Using HNMR Spectroscopy and Principal Component Analysis. J. Nat. Prod, 2004, 67(6): p953-957.
- [22] Wang Y.L, Tang H.R, Nicholson J K, Hylands P.J, Sampson J, Whitcombe I, Stewart C.G, Caiger S, Oru I, Holmes E. Metabolomic strategy for the classification and quality control of phytomedicine: A case study of chamomile flower (*Matricaria recutita* L.). Planta Med, 2004, 70 (3): 250-255.
- [23] 陈耀祖,张惠迪,蔡文涛等.当归化学成分分析研究-甘肃岷县当归中微量元素的分析[J].兰州大学化学学报 (自然科学版) ,1983,19:192-193.
- [24] 张敏,胡坪,罗国安等.当归水溶性成分 HPLC 指纹图谱研究[J].中成药,2007,29(5):628-630.
- [25] 阮健,王凤山,刘珂等.当归挥发油的 GC 指纹图谱研究[J].中草药,2006, 37(9):1338-1341.
- [26] 王亮.适宜当归栽培化学除草剂研究[D]. 延边:延边大学农学院, 2007.
- [27] 赵剑宇,颜贤忠.基于核磁共振的代谢组学研究进展 [J]. 国外医学药学分册,2004,31(5): 308-312.
- [28] 孙元琳,陕方,崔武卫等.当归多糖 ASP3 及其水解产物的 NMR 光谱分析[J].高等学校化学学报 2009,30(9): 1739-1743.
- [29] 孔丽娟,梁侨丽,吴启南等.黑三棱的化学成分研究[J].中草药, 2011, 42(3): 440-442.
- [30] 肖永庆,李丽,游小琳等.川芎化学成分的研究[J].中国中药杂志, 2002, 27(7):519-521.
- [31] 易进海,刘云华,陈燕,等.核磁共振波谱法测定藁本内酯对照品的含量[J]. 药物分析杂志, 2010, 30(4): 680-682.

- [32] 夏玮,刘书琦,张文清等.桑叶多糖 MP-3b 的结构性质研究[J]. 中成药, 2009, 31(3): 427-430.
- [33] 阿基业.代谢组学数据处理方法—主成分分析[J]. 中国临床药理学与治疗学, 2010,15(5):49-52
- [34] Zhang L. X, Dai H etc. Detection and use of selective ions for providing initial estimates of the concentration vector in iterative target transformation factor analysis[J]. Chromatogr. A. 1217(2010) :3128-3135.
- [35] Holmes, E.; Tang, H. R.; Wang, Y. L.; Seger, C. The assessment of plant metabolite profiles by NMR-based methodologies. Plant Med.2006, 72, 771–785.
- [36] William T, Graham A. M, Peter C, Richard G, An environmental HNMR metabolomic study of the exposure of the marine mussel *Mytilus edulis* to atrazine, lindane, hypoxia and starvation. Metabolomics, 2009, 5:33–43.
- [37] Sauter H, Lauer M, Fitsch H. Metabolic profiling of plants: a new diagnostic technique in: Synthesis and chemistry of agrochemicals11.American Chemical Society Symposium Series 443. American Chemical Society, Washington DC, p. 288-299.
- [38] Fiehn O. Metabolic networks of *Cucurbita maxima* phloem. Phytochemistry, 2003, 62 (6): 875-886.

- [39] Taylor J, King R D, Altmann T, Fiehn O. Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics*, 2002, 18(2): 241-248.
- [40] Trygg J, Wold S. Orthogonal projections to latent structure (O-PLS). *Journal of Chemometrics*, 2002, 16: 372-374.
- [41] 李俊南, 侯艳, 孙凤宇, 李康. OPLS 方法的原理及其在代谢组学数据判别分析中的应用. *中国卫生统计*, 2014(5): 765-769.
- [42] 许国旺, 路鑫, 杨胜利等. 代谢组学研究进展[J]. *中国医学科学院学报*. 11(6): 701-711.
- [43] 毛煜, 袁伯俊. 代谢组学的研究现状与展望[J]. *中国新药杂志*. 2007, 16(13): 1105-1110.
- [44] 宋秋月, 付迎波, 刘江等. 当归的化学成分研究[J]. *中草药*, 2011, 42(10): 1900-1904.
- [45] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 北京: 高等教育出版社: 2004.
- [46] 陈杰. MATLAB 宝典. 北京: 电子工业出版社: 2013.
- [47] 严辉, 段金廛, 钱大玮, 等. 我国不同产地当归药材质量的分析与评价[J]. *中草药*, 2009, 40(12): 1988-1992.
- [48] 温锦波. 基于 NMR 的代谢组学的数据预处理方法及其在糖尿病研究中的应用[D]. 福建: 厦门大学, 2007. 11-16.
- [49] 马华侨, 邸多隆, 邵士俊等. 当归 HPLC 指纹图谱研究[J]. *中草药*, 2004, 35(8): 931-932.
- [50] 吴燕燕, 尚明英, 蔡少青等. 当归的化学成分指纹图谱[J]. *药学学报*, 2008, 43(7): 728-732.
- [51] 段金廛, 严辉, 宿树兰等. 药材适宜采收期综合评价模式的建立与实践[J]. *中草药*, 2010, 41(11): 1755-1760.
- [52] 钟郁青, 徐礼. 银杏叶中双黄酮成分的提取分离及其 HPLC 法测[J]. *药学学报*, 1995, 30: 694-697.
- [53] 朱晶晶, 王智民, 匡艳辉, 等. 一测多评法同步测定人参和三七药材中多种人参皂苷的含量[J]. *药*

学学报, 43(12):1211-1216

[54] 蔺海明,刘学周,刘效瑞等. 栽培方式对当归干物质积累和生长动态影响的研究[J].中草药, 2007, 38(2): 257-261.

夏建飞,梁琼麟,罗国安等.代谢组学研究策略与方法的新进展[J].分析化学.2009,37(1):136~143.