

三维荧光结合化学计量学用于白术的产地溯源

陈伟¹, 吴海龙^{1*}, 王童^{1*}, 常月月¹, 陈瑶², 杨健³,
付海燕⁴, 杨小龙⁴, 李旭富⁵, 俞汝勤¹

1. 湖南大学化学化工学院, 化学生物传感与计量学国家重点实验室, 湖南长沙 410082
2. 湖南工业大学生命科学与化学学院, 湖南省生物医用纳米材料与器件重点实验室, 湖南株洲 412008
3. 中国中医科学院中药资源中心, 道地药材国家重点实验室培育基地, 北京 100700
4. 中南民族大学药学院, 湖北省民族药物现代化工程技术研究中心, 湖北武汉 430074
5. 北京同仁堂平江白术有限公司, 湖南平江 414500

摘要 白术是一种著名中药材, 其品质与其种植产地密切相关, 迫切需要寻找一种能快速对白术进行品质鉴定和产地溯源的方法。白术含有多种内源荧光物质, 可为三维荧光技术鉴定白术的产地提供依据。首先使用交替三线性分解(ATLD)对白术三维荧光光谱进行表征, 再将三维荧光光谱与偏最小二乘判别分析(PLS-DA)和 k 最近邻法(k NN)两种模式识别方法相结合对白术样品进行产地溯源。结果表明, PLS-DA和 k NN各自建立的分类模型能有效对三大白术产区(安徽、湖南和浙江)的样本进行区分, 对预测集中白术道地药材浙江白术的分类准确率(CCR)分别高达80%和90%。所建立方法为白术的产地溯源提供一种新颖、快速的解决方案, 有利于维护白术市场的秩序, 利于其产业发展。

关键词 白术; 化学计量学; 三维荧光; 交替三线性分解; 产地溯源

中图分类号: R282.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)09-2875-09

引言

近些年来, 中医药凭借其毒副作用小, 对某些疾病有着极佳的疗效以及环境友好等特点, 受到了越来越多国家的关注。目前, 已有多个国家创办了中医药研究机构和中医诊所。世界卫生组织在《迎接21世纪的挑战》报告中指出: “21世纪的医学不应该继续以疾病为主要研究领域, 而应当以人类的健康作为医学的主要研究方向^[1]”。更让中医药变成了热门话题, 也为中医药的发展提供了千载难逢的机遇。然而目前中医药还存在许多问题。道地药材作为中药材品质的标杆, 凭借其优良的品质和显著的疗效被广泛应用。《神农本草经》记载“土地所出, 真伪新陈, 并各有法”, 阐述了道地药材品质与地理条件的差异^[2]。由于目前中药材产地的多元化, 不同产地有着不同的栽培条件和不同的加工工艺, 直接导致了中药材产品中有效成分含量的差异, 导致市场上中药材产品质量参差不齐。更为棘手的是, 不同产地的药材在加工成成品后, 其外观较为相似, 很难通过外观分辨其品质的

优劣, 更无法准确分辨其产地。为了更好地推动中医药的发展, 对道地药材进行地理溯源非常重要。

白术(学名: *Atractylodes macrocephala* Koidz. 缩写AM), 著名中药材, 素有“北参南术”的美誉。本品来源为菊科苍术属植物白术的干燥根茎。白术的化学成分主要为挥发油、苷、糖、黄酮、聚乙炔及甾醇等物质。目前白术已被证明具有多种药理活性, 其中包括: 改善胃肠道功能^[3]、抗肿瘤^[4]、性腺激素调节^[5]、免疫调节^[6]等。白术产地主要分布在中国的浙江、湖南、安徽及河北等地, 自古以浙江为道地产区, 目前以安徽、浙江和湖南等产地为主^[7]。白术质量有明显的地域性差异, 生长环境, 土质以及气候对其品质、性状和药效影响非常大, 张龙开等在对白术的研究中已证实不同地区白术的化学组分及含量有着非常显著的差异^[8]。

目前, 已报道的白术产地溯源方法有Hu等利用稳定同位素比值法和多元元素分析法, 通过测定分析不同地区白术所含元素含量的差异进行地理溯源^[9]。上述方法在样本预处理时需要消耗大量时间进行消解, 且仪器操作流程繁琐。因此, 亟需发展一种快速精准的白术产地溯源方法。与上述方

收稿日期: 2021-05-25, 修订日期: 2021-08-24

基金项目: 国家重点研发计划项目(2020YFC1712700), 国家自然科学基金项目(21775039, 21521063)资助

作者简介: 陈伟, 2000年生, 湖南大学化学化工学院本科生 e-mail: wchen@hnu.edu.cn

* 通讯作者 e-mail: hlwu@hnu.edu.cn; wangtong@hnu.edu.cn

法相比, 荧光法具有检测速度快、灵敏度高、仪器设备易操作、能用于在线监测、样本预处理简单和成本较低等优点。特别是三维荧光(EEM)技术能够快速提供复杂体系详细的光谱信息, 大大提高了荧光技术的分析潜力。EEM与化学计量学方法相结合, 作为一种强有力的分析策略, 已成功运用于食用油^[10]和茶叶^[11]等食品的品质分类和真伪鉴别。本工作采用 EEM 结合偏最小二乘判别分析(PLS-DA)和 k 最邻近法(k NN)两种模式识别方法对白术样品进行产地溯源。首先使用交替三线分解算法(ATLD)对白术的三维荧光图谱进行解析, 对不同产地白术的内源荧光成分进行初步分析; 之后使用 PLS-DA 和 k NN 对白术样本数据进行分类, 通过交叉验证优化分类模型的复杂度参数; 最后使用该模型对测试集和预测集样本进行预测, 以此验证该方法的准确性。

1 实验部分

1.1 白术样品采集

使用的白术样品均为 2020 年 11 月实地采集, 前后时间不超过 5 d, 以严格控制采摘时间的一致性, 消除白术样品因采摘季节年份不同而引起的差异。实地采集样品共 125 份,

其中产自安徽省 45 份、湖南省 45 份, 浙江省 35 份, 采样地点和采样条件见图 1 和表 1。使用随机采样算法, 将不同产地白术样本分成训练集(总样本的四分之三, 95 个)和测试集(剩余的四分之一样本, 30 个)。另采集 10 个浙江白术样本作为预测集。

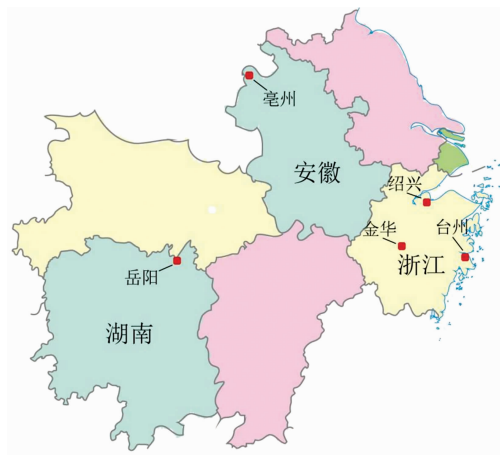


图 1 本工作中的白术采样点

Fig. 1 Sampling points of AM in this work

表 1 白术采样时的地理和天气条件以及样本数(不包括预测集样本)

Table 1 Geographical and weather conditions and the number of samples when picking AM
(The prediction set samples are not included)

	样本数	东经	北纬	海拔/m	温度/℃	湿度/%	天气
安徽省	亳州 1	115°37'39"—115°37'41"	33°44'7"—33°44'10"	15.59~26.61	11.5~16.4	37~43	阴
	亳州 2	115°40'54"	33°45'41"—33°44'43"	16.09~24.29	13.1~16.3	30~47	阴
	亳州 3	115°41'0"—115°41'3"	33°45'34"—33°44'35"	7.84~36.02	13.1~16.5	38~46	阴
湖南省	平江 1	113°35'38"—113°48'4"	28°39'13"—28°59'28"	75.5~463.3	10.2~11.2	68~76	阴
	平江 2	114°2'17"—114°2'21"	28°47'26"—28°47'32"	149.32~158.59	13.2~15.3	56~59	阴
	平江 3	114°2'16"—114°2'18"	28°47'25"—28°47'28"	150.00~168.32	11.6~12.7	65~72	阴
浙江省	磐安	120°23'26"—120°23'28"	28°56'46"—28°56'47"	60.08~411.71	8.3~10.2	70~91	阴
	天台	120°49'2"	29°12'5"	512.7	9.8~13.5	56~69	阴
	新昌	120°49'11"—120°49'13"	29°15'8"—29°15'26"	433.3~446.92	12.5~19	49~78	阴

注: 1, 2, 3 代表同一地区的不同采样点

Note: 1, 2 and 3 represent different sampling points in the same area

鲜白术的处理步骤: 将白术从土壤挖出后, 除去地上茎叶, 保留地下根茎并去除表面的泥土, 放入烘箱内, 每次放同一产地的白术。第一次烘焙时, 温度为 60~70 ℃, 烘焙 2~3 h 后上下翻动一遍, 使须根干透脱落。将除去须根的白术取出放到晾晒架上, 使其内部水分外溢, 外皮软化。24 h 后, 进行复烘, 在 60 ℃ 下烘焙 24 h, 然后再放到晾晒架上晾 1 d, 使其内部水分充分向外渗出, 再将返润的白术烘烤 24 h。重复烘烤晾晒, 直到白术表面不再有水分渗出, 通常一般需要反复烘焙三次。将烘干后的白术样品放入中药超细研磨机研成粉末, 再用 80 目的分样筛进行筛分, 将筛分后的粉末和残渣分别用塑封袋装好并标记相应名称备用。

1.2 萃取条件的优化和萃取步骤

选用有机试剂作为萃取剂萃取白术所含的有效成分。以《中国药典》中白术样品提取液的制备方法为基础, 对萃取溶

剂、溶剂的配比和超声时间三个萃取条件进行了优化。

优化步骤: 首先, 选用甲醇和乙醇两种试剂作为萃取剂, 并用超纯水稀释, 分别配制体积分数为 100%, 80%, 60%, 40%, 20% 和 0% 的甲醇/乙醇水溶液进行萃取, 离心、过滤后取上清液进行荧光扫描得到 EEM 谱图。为了直观地比较萃取条件的优劣, 从白术的荧光图谱中选取了两个荧光信号峰(激发/发射波长为 280/355 nm 和激发/发射波长为 355/430 nm)进行跟踪分析。结果如图 2 所示, 由图 2(a) 和 (b) 可以看出, 甲醇和乙醇萃取白术样品的荧光强度非常接近, 无明显差异。由于甲醇对人体的神经系统和血液系统有一定的毒害作用且甲醇价格比乙醇贵。综合考虑, 选择乙醇(绿色溶剂)作为萃取剂。然后对不同体积比的萃取剂进行分析, 其中 280/355 nm 荧光峰在 20% 的乙醇水溶液中荧光最强, 355/430 nm 荧光峰在 40% 乙醇水溶液中荧光最强, 考

考虑到不同产地白术的 280/355 nm 荧光峰信号值可能存在的差异,若信号值高于 10000 仪器将无法检测,选择体积分数为 40% 的乙醇水溶液作为萃取剂。最后,对超声萃取时间进行优化。从图 2(c)可以看出,两个荧光峰的强度均不随萃取时间的延长而发生明显变化。在不影响萃取结果的情况下,为了缩短实验时间,最终选择萃取时间为 5 min。

白术最终的萃取步骤如下:准确称取 400 mg 的白术粉末样品于 10 mL 离心管中,加入 8 mL 体积分数为 40% 的乙醇水溶液,放入超声机中超声 5 min 后,在 $4\ 000\ \text{r} \cdot \text{min}^{-1}$ 、

$4\ ^\circ\text{C}$ 的条件下离心 10 min,将上层清液用 $0.45\ \mu\text{m}$ 微孔滤膜过滤待用。在 EEM 测量前,将已过滤好的上清液用 40% 的乙醇水溶液稀释 10 倍再进行荧光扫描。

1.3 药品与试剂

本工作所使用的药品与试剂如下。药品列表及详细信息见表 2。

试剂:甲醇和乙醇(均为色谱纯)购买自 Sigma-Aldrich(圣路易斯,美国);超纯水($18.20\ \text{M}\Omega \cdot \text{cm}$)由 Milli-Q 超纯水系统(Millipore,美国)生产。

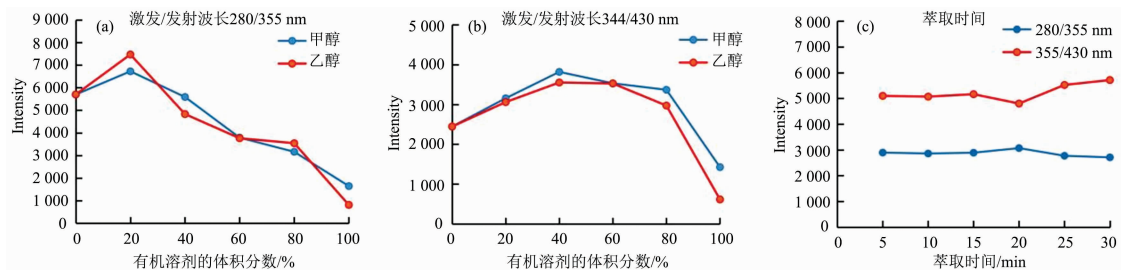
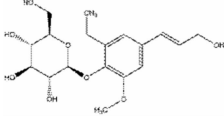
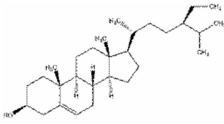
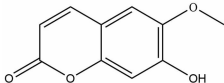
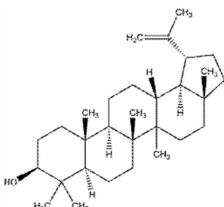


图 2 (a)激发/发射波长为 280/355 nm 的荧光强度与萃取剂种类和萃取剂比例的关系; (b)激发/发射波长为 355/430 nm 的荧光强度与萃取剂种类和萃取剂比例的关系; (c) 荧光强度随超声时间的变化

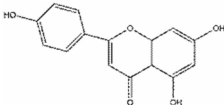
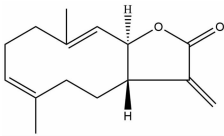
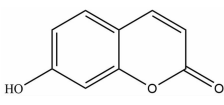
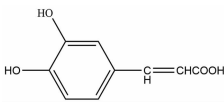
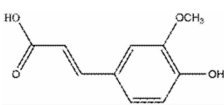
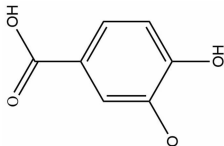
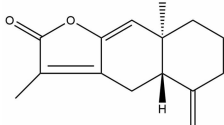
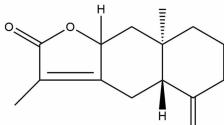
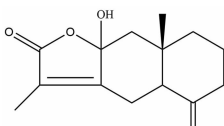
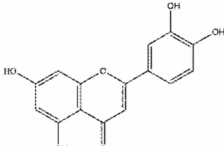
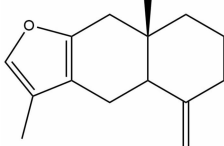
Fig. 2 (a) The relationship between fluorescence intensity of excitation/emission wavelength of 280/355 nm and extractant type and extractant ratio; (b) The relationship between fluorescence intensity of excitation/emission wavelength of 355/430 nm and extractant type and extractant ratio; (c) Variation of fluorescence intensity with ultrasonic time

表 2 白术中可能存在的 17 种标准物质的中英文名称、分子量、浓度、结构式、纯度和生产公司


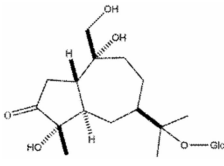
Table 2 Chinese and English names, molecular weights, concentrations, structural formulas, purity and production companies of 17 standard substances that may exist in AM

序号	名称	分子量	浓度/ ($\mu\text{g} \cdot \text{mL}^{-1}$)	结构式	纯度	公司
1	紫丁香苷 (Syringin)	372.4	0.17		$\geq 98\%$	博美生物科技有限公司
2	β -谷甾醇 (β -sitosterol)	414.7	9.7		$\geq 98\%$	博美生物科技有限公司
3	东莨菪内酯 (Scopoletin)	192.2	9.68		$\geq 98\%$	博美生物科技有限公司
4	羽扇豆醇 (Lupeol)	426.7	8.7		$\geq 98\%$	博美生物科技有限公司

续表 2

5	芹菜素 (Apigenin)	270.2	12.52		≥98%	博美生物科技有限公司
6	木香炔内酯 (Costunlide)	232.3	12.34		98%	博美生物科技有限公司
7	7-羟基香豆素 (7-Hydroxycoumarin)	162.1	0.176		≥98%	博美生物科技有限公司
8	咖啡酸 (Caffeic acid)	180.2	8.68		≥98%	博美生物科技公司
9	阿魏酸 (Ferulic acid)	194.2	9.82		≥98%	博美生物科技公司
10	原儿茶酸 (Protocatechuic acid)	154.1	9.04		≥98%	博美生物科技公司
11	白木内酯 I (Atractylenolide I)	230.3	9.5		≥98%	上海源叶生物科技有限公司
12	白木内酯 II (Atractylenolide II)	232.3	10.08		≥98%	上海源叶生物科技有限公司
13	白木内酯 III (Atractylenolide III)	248.3	8.46		≥98%	上海源叶生物科技有限公司
14	木犀草素 (Luteolin)	286.2	8.4		≥98%	上海源叶生物科技有限公司
15	苍术酮 (Atractylon)	216.3	8.42		≥98%	上海源叶生物科技有限公司

续表 2

16	苍术素 (Atractylodin)	182.2	7.6		99.5%	成都普思生物科技股份 有限公司
17	苍术苷 A (Atractylaside A)	448.5	4.68		≥98%	甄准生物科技公 司

1.4 偏最小二乘判别分析(PLS-DA)

偏最小二乘判别分析(PLS-DA)是一种线性分类方法,将偏最小二乘回归算法用于分类判别,详见文献[12]。

1.5 k 最邻近法(k NN)

k 最邻近法(k NN)是一种经典的有监督非线性分类算法,根据样本周围最近 k 个样本的最大类别概率来判断样本的归属类型,具体实现步骤如下:(1)计算测试样本数据与各训练样本数据之间的欧氏距离;(2)将获得的距离值从大到小进行排序;(3)选择距离值最接近的 k 个训练样本并记录其分类,通过留一交叉验证法选择最优的 k 值;(4)统计前 k 个点所在类别出现的频率;(5)返回前 k 个点出现频率最高的类别作为当前点的预测分类。

1.6 交替三线性分解算法(ATLD)

交替三线性算法(ATLD)于 1996 年由吴海龙等提出,是一种高效的“数学分离”算法,可对三线性数阵进行解析,并提供有明确物理化学意义的解析结果,基于所获得的各组分轮廓矩阵可进行更深入的定性和定量分析^[13]。ATLD 与传统平行因子分析(PARAFAC)相比,所需内存大大减少,运算的效率极大提高,是现存迭代三维校正算法中收敛速度最快的一种算法。详情见文献[14]。

1.7 仪器条件和软件

实验所有白术样品的 EEM 测量均在一台配备了 150 W 氙灯的 F-7000 荧光光谱仪(日立,日本)上进行,并连接一台电脑用于读取数据,测量时配合使用 10 mm 的石英比色皿。仪器相关参数:激发波长范围为 200~600 nm(步长 5 nm);发射波长范围为 200~650 nm(步长 5 nm);扫描速度为 30 000 nm·min⁻¹;检测电压为 650 V;狭缝宽度为 5.0 nm。数据均保存为 .txt 格式。本工作所有的数据处理均在 MATLAB 环境下进行,使用 classification-toolbox 5.1 中的 PLS-DA 和 k NN 对样本进行数据分析,可从网址 <http://www.michem.unimib.it/download/matlab-toolboxes> 免费获取^[12]。PLS-DA 方法的 LVs(最佳潜变量数)和 k NN 方法的 k (邻近样本数)均通过留一交叉验证所获得的最佳分类准确率(CCR)来确定。在 CCR 相同的情况下,优先选择最小的 LVs 和 k 值。CCR 可被视为总体精度,通过式(1)计算

$$CCR = \frac{\sum_{g=1}^G n_{gg}}{N} \quad (1)$$

式(1)中, n_{gg} 为原本属于第 g 类的样本被正确分到第 g 类的样本数量; N 为总样本数; G 为总类别数。

2 结果与讨论

2.1 白术样本的荧光光谱

各样本原始的 EEM 数据是在激发波长范围为 200~600 nm,发射波长范围为 200~650 nm 下采集的。为了简化运算并消除无用信息,选择一个包含白术样本主要荧光特征的较小荧光区域用于后续分析。该区域的激发波长范围为 200~560 nm,发射波长范围为 200~620 nm,每一个矩阵数据的大小为 73×85(激发波长数×发射波长数)。由于 EEM 光谱中存在严重影响数据三线性结构的瑞利散射和拉曼散射,因此必须对其进行预处理。每个样本矩阵都减去三个空白样本(体积分数为 40%的乙醇水溶液)矩阵数据的平均值,以此来消除轻微的拉曼散射,瑞利散射通过 Bro 等提出的插值法^[15]来消除。该算法首先直接把散射区域扣除,然后再通过插值法拟合缺失数据,如图 3(a,b,c)所示。

图 4(a,b,c)分别为预处理后不同地区白术的典型 EEM 谱图,从图 4 中可以看出不同地区的白术 EEM 谱图存在一定的相似性和差异性。例如在激发/发射波长 355/450 nm 处都存在一个强荧光峰。安徽白术样本在激发/发射波长 275/370 nm 附近有一个信号比较强的荧光峰,而浙江白术样本该处信号较弱,湖南白术样本在该处几乎没产生荧光信号,实验结果反映出不同地区白术的内源荧光成分和微环境有所差异。不同产地白术 EEM 谱图的强度、形状和位置的差异性和相似性为随后采用化学计量学方法分类提供一定的可能和挑战。

2.2 基于 ATLD 算法对白术荧光数据进行表征

首先,将 135 个白术样本荧光数阵沿着样本维方向进行堆叠,得到一个大小为 73×85×135(激发波长数×发射波长数×样本数)的三维数阵,再采用 ATLD 算法对三维数阵进行分解。在 ATLD 分解之前,通过核一致性诊断算法(CORCONDIA)^[16]对其组分数(N)进行估计,当 CORCONDIA 结果大于 60%时,可认为模型符合三线性。最终确定组分数 $N=3$,此时 ATLD 解析三维数阵得到的结果如图 5 所示。图 5(a,b,c)分别为归一化激发光谱图,归一化发射光谱图和相

对浓度图。从图 5(a,b)分别可以看出:组分 1(蓝色)的激发/发射波长为 365/440 nm;组分 2(绿色)的激发/发射波长为 405/470 nm;组分 3(红色)的激发/发射波长为 280/360 nm,其在激发波长 225 nm 和发射波长 450 nm 处各存在一个明显的肩峰。值得注意的是,这里所指的“组分”可能并不代表具体某一种物质,而通常是某一类荧光信号接近的物质(共因子)。为了更加准确直观地分析三组分各代表的物质,对白术中可能含有的 17 种物质进行了三维荧光扫描,随后对有明显荧光信号的七种内源荧光物质进行分析,得到的归一化激发光谱图 6(a)和归一化发射光谱图 6(b),如图 6 所示。通过比较分析图 5 和图 6 各分组的激发/发射波长,考虑白术

复杂的内源环境,做出以下推测:组分 1 所代表的物质可能为 7-羟基香豆素和东莨菪内酯;组分 3 所代表的物质可能为紫丁香苷、原儿茶酸和苍术酮;组分 2 无对应物质,暂定为未知物质。由图 5(c)可看出不同产地白术所含三组分浓度的差异,其中组分 1 在安徽、湖南和浙江白术中含量无明显差异,组分 2 含量在安徽、湖南和浙江白术中呈上升趋势,组分 3 在安徽白术中含量最高,在湖南和浙江白术中含量较低。通过分析各组分所可能代表的物质以及各组分在不同产地白术的含量差异,可以看出不同产地白术的荧光数据存在一定的差异,为其产地溯源提供了依据。

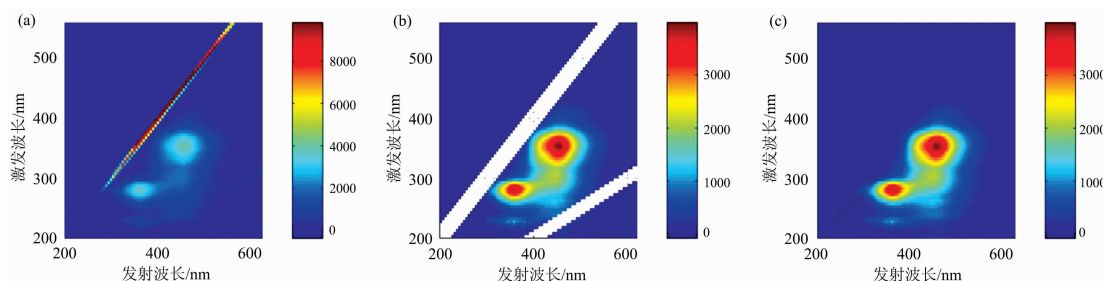


图 3 插值法消除 EEM 数据中瑞利散射

(a): 具有可见瑞利散射的原始数据; (b): 扣除瑞利散射后的缺陷数据; (c): 使用插值法修复后的数据

Fig. 3 Rayleigh scattering eliminated from EEM data by interpolation

(a): Raw data with visible Rayleigh scattering; (b): Defect data after deducting Rayleigh scattering;

(c): Data repaired by interpolation

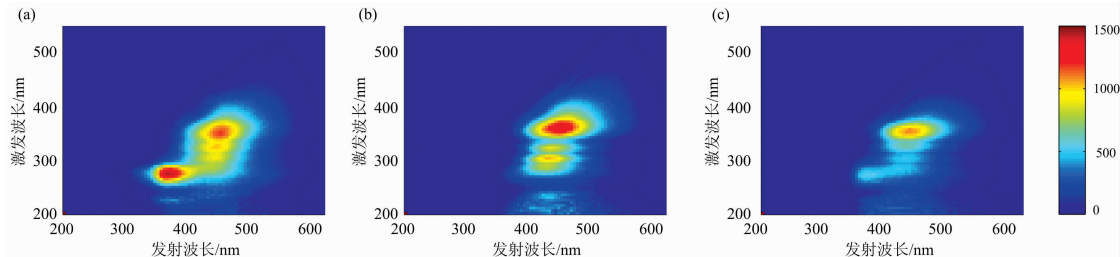


图 4 扣除散射后不同地区白术样品典型 EEM 等高线图

(a): 安徽白术样本; (b): 湖南白术样本; (c): 浙江白术样本

Fig. 4 Typical EEM contour map of AM samples in different provinces after deducting scattering

(a): Anhui AM; (b): Hunan AM; (c): Zhejiang AM

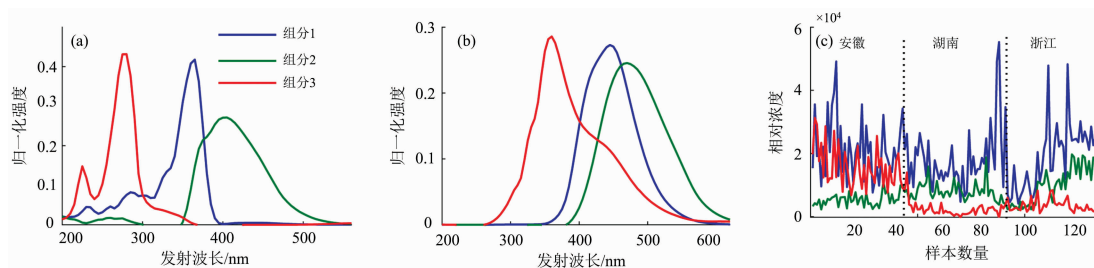


图 5 $N=3$ 时 ATLD 解析三维矩阵获得的归一化激发光谱图(a)、归一化发射光谱图(b)和相对浓度图(c)

Fig. 5 The normalized excitation spectra profiles (a), the normalized emission spectra profiles (b) and the relative concentration spectra profiles (c) obtained by ATLD analysis of three-way matrix with $N=3$

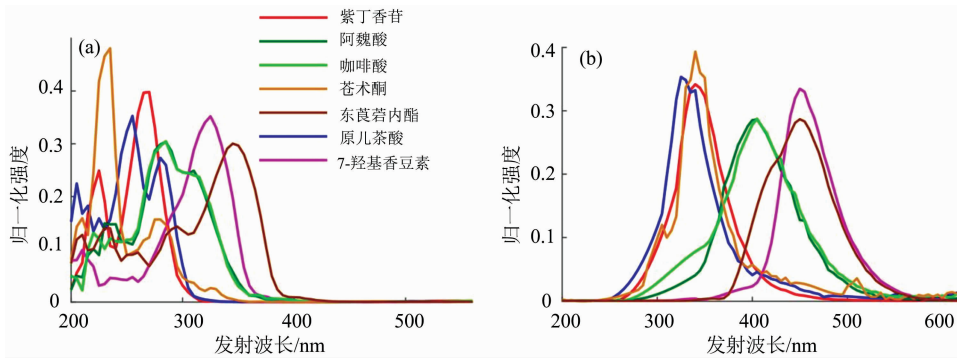


图 6 白术中可能含有的 7 种具有荧光信号的标准物质的归一化激发光谱图(a)和归一化发射光谱图(b)

Fig. 6 The normalized excitation spectra profiles (a) and normalized emission spectra profiles (b) of 7 standard substances with fluorescence signals that may be contained in AM

2.3 白术产地分类

使用 PLS-DA 和 *k*NN 两种模式识别方法对不同产地的白术进行分类。首先通过交叉验证来优化 PLS-DA 的 LVs 以及 *k*NN 的 *k* 值。根据交叉验证的结果，最终确定 PLS-DA 的 LVs 为 10，*k*NN 的 *k* 值为 2(见表 3)。表 3 列出交叉验证模型中两种方法获得的分类准确率(CCR%)，其中 PLS-DA 的 CCR 为 91%，*k*NN 的 CCR 为 88%。为了更好地评估分类性能，表 3 给出了特异性和灵敏度两种品质因子参数，特异性表示模型拒绝所有其他类别样本的能力，灵敏度表示模型正确识别该类样本的能力。PLS-DA 在交叉验证中对于各类别获得的特异性、灵敏度分别在 92%~98%和 81%~97%范围内，而 *k*NN 分别为 93%~95%和 85%~91%，这意味着两个分类模型均能比较有效地识别该类样本并拒绝他类样本。同时，使用以上两种方法验证了训练集样本的类别，所得分类结果如表 3 所示。其中 PLS-DA 方法的 CCR 为 100%，且其特异性、灵敏度都为 100%，能完全准确地对训

练集的样本归类。而 *k*NN 方法的 CCR 为 88%，稍逊于 PLS-DA 的结果。可以明显看出，无论是特异性还是灵敏度，安徽白术的检测结果都略大于湖南白术和浙江白术。另外，可以通过 PLS-DA 解析训练集时前 3 个潜变量的得分值所作出的三维散点图可更直观地比较分类结果，如图 7 所示。从图 7 可以看出，安徽、湖南和浙江白术在三维空间中被大致的区分开来，其中安徽白术明显与湖南和浙江白术区分开，而湖南白术和浙江白术有少量重叠，很好地解释了安徽白术的分类结果略大于湖南和浙江白术的原因，也说明了为什么在表 4(支持信息)混淆矩阵中，湖南白术和浙江白术都有少量被相互误分。用上述已验证的白术分类模型来对测试集样本进行预测，结果如表 3 所示。其中对于测试集，PLS-DA 的 CCR 为 87%，而 *k*NN 的 CCR 仅为 73%。从整体来看，两种分类方法均获得较高的分类准确率，其中 PLS-DA 最令人满意，交叉验证、训练集和测试集获得 CCR 都比 *k*NN 高，且结果比较稳定。

表 3 PLS-DA 和 *k*NN 两种分类模型的复杂度参数(LVs/*k*)，交叉验证、训练集、测试集和预测集获得的分类准确率、特异性和灵敏度

Table 3 The complexity parameters (LVs/*k*) of the PLS-DA and *k*NN classification models, the classification accurate rate, specificity and sensitivity obtained by cross-validation, training set, test set and prediction set

模型	LVs/ <i>k</i>	CCR/%	特异性 ^a /%			灵敏度 ^b /%			
			安徽	湖南	浙江	安徽	湖南	浙江	
PLS-DA	10	交叉验证	91	98	92	96	97	91	81
		训练集	100	100	100	100	100	100	100
		测试集	87	95	89	95	100	91	63
<i>k</i> NN	2	交叉验证	88	95	93	94	91	88	85
		训练集	88	95	93	94	91	88	85
		测试集	73	89	89	82	100	55	63

注：^a 特异性定义为 $TN/(TN+FP)$ ，TN：真阴性，FP：假阳性；^b 灵敏度定义为 $TP/(TP+FN)$ ，TP：真阳性，FN：假阴性
 Note: ^a The specificity is defined by $TN/(TN+FP)$, TN: True Negatives, FP: False Positives; ^b The sensitivity is defined by $TP/(TP+FN)$, TP: True Positives, FN: False Negatives

表 4 列出了训练集和测试集中两种分类方法所获得的混淆矩阵，从表 4 可以非常直观地看出两种模型对样本的分类结果。其中，PLS-DA 分类方法在训练集和测试集都能 100% 正确地对安徽白术进行分类，而对湖南和浙江白术，

无论是 PLS-DA 还是 *k*NN，都不能对其进行 100% 正确的归类，而绝大多数样本都被正确地归类。总之由于白术内源荧光物质的多样性和体系的复杂性，再加之荧光光谱有着较大的重叠部分，因此获得以上分类结果还是非常理想的。

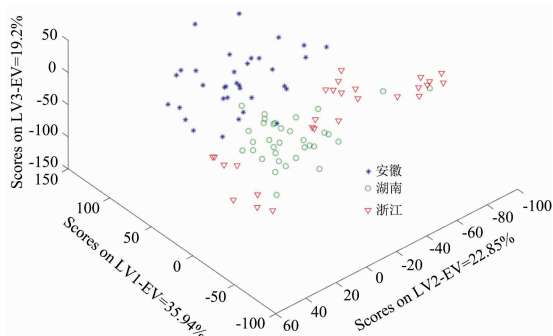


图 7 根据 PLS-DA 前 3 个潜变量的得分值所作的训练样本三维散点图

Fig. 7 Three-way scatter diagram of training samples based on the scores of the first three latent variables of PLS-DA

表 4 白术分类模型中 PLS-DA 和 k NN 分类方法获得的训练集、测试集和预测集的混淆矩阵

Table 4 Confusion matrix of training set, test set and prediction set obtained by PLS-DA and k NN classification methods in AM classification model

模型	预测	真实					
		PLS-DA			k NN		
		安徽	湖南	浙江	安徽	湖南	浙江
训练集	安徽	34	0	0	31	1	2
	湖南	0	34	0	2	30	2
	浙江	0	0	27	1	3	23
测试集	安徽	11	0	1	11	1	1
	湖南	0	10	2	0	6	2
	浙江	0	1	5	0	4	5
预测集	安徽	0	0	1	0	0	0
	湖南	0	0	1	0	0	1
	浙江	0	0	8	0	0	9

References

- [1] LIU He-hua(刘荷花). Capital Food Medicine(首都食品与医药), 2019, 26(14): 115.
- [2] LIAO Bao-sheng, SONG Jing-yuan, XIE Cai-xiang, et al(廖保生, 宋经文, 谢彩香, 等). China Journal of Chinese Materia Medica(中国中药杂志), 2014, 39(20): 3881.
- [3] Song H P, Li R L, Zhou C, et al. Journal of Ethnopharmacology, 2014, 159: 23.
- [4] Liu H, Zhu Y, Zhang T, et al. Molecules, 2013, 18(11): 13357.
- [5] Zhou J, Qu F, Barry J A, et al. Journal of Clinical and Experimental Medicine, 2016, 9(2): 2758.
- [6] Xu C L, Zhao Y F, Shang X Y, et al. Journal of Animal and Feed Sciences, 2012, 21(2): 302.
- [7] Zhu B, Zhang Q L, Hua J W, et al. Journal of Ethnopharmacology, 2018, 226: 143.
- [8] ZHANG Long-kai, QIN Jie-yuan, ZHAO Wei-bo, et al(张龙开, 覃婕媛, 赵维波, 等). Journal of Anhui Agricultural Sciences(安徽农业科学), 2019, 47(23): 203.
- [9] Hu L, Chen X F, Yang J, et al. Rapid Commun. Mass. Spectrom., 2019, 33: 1703.
- [10] Wang T, Wu H L, Long W J, et al. Food Chemistry, 2019, 293: 348.
- [11] Hu Y, Wu H L, Yin X L, et al. Analytica Chimica Acta, 2018, 1021: 28.
- [12] Ballabio D, Consonni V. Analytical Methods, 2013, 5(16): 3790.
- [13] WU Hai-long, LONG Wan-jun, GU Hui-wen, et al(吴海龙, 龙婉君, 谷惠文, 等). Journal of Yangtze University • Natural Science Edition(长江大学学报 • 自然科学版), 2021, 18(3): 74.

2.4 新样本测试

为了更好地检验两种模型对白术样本的正确分类能力,证明本模型能对白术道地药材进行产地溯源,从浙江省金华市磐安县和台州市天台县共采集 10 个新的浙白术药材样本,并使用所建立的两个模型对其进行预测,预测结果如表 3 和表 4 所示。其中 PLS-DA 的 CCR 为 80%,只有两个样本被错误分类, k NN 的 CCR 为 90%,只有一个样本被错误分类到湖南。结果表明,两种模型即使在面对未知的白术样本时,仍能有效地识别该样本并正确对其产地进行预测。进一步验证了本工作所提方法的可行性与可靠性,证明本方法能有效地对白术进行产地溯源。

3 结论

采用三维荧光与化学计量学相结合对不同产地白术样本进行荧光光谱表征和产地溯源。首先优化并选择 40% (v/v) 的乙醇水溶液为萃取剂,获得白术样本的 EEM 谱图,并使用交替三线性分解(ATLD)对白术三维荧光光谱进行表征,再结合两种化学模式识别方法(PLS-DA 和 k NN)对获得的荧光光谱图建立分类模型,最后用已建立的分类模型对测试集和预测集进行预测分类。结果显示,PLS-DA 在交叉验证和预测集获得的 CCR 分别为 91% 和 80%, k NN 在交叉验证和预测集获得的 CCR 分别为 88% 和 90%,两种分类方法均得到了满意的结果。上述结果表明,采用 EEM 荧光光谱与化学计量学策略相结合对不同产地白术进行分类,与传统的产地溯源方法相比,具有灵敏度高、试剂消耗量小、检测速度快和仪器便携等优点。本工作为白术的产地溯源提供了一个新颖有效的思路,可快速判断白术样本是否属于道地产区,能有效地杜绝以次充优现象,有利于维护白术市场秩序和消费者权益,在药材市场品质检测方面有着广阔的应用前景。

- [14] Wu H L, Shibukawa M, Oguma K. *Journal of Chemometrics*, 1998, 12(1): 1.
- [15] Bahram M, Bro R, Stedmon C, et al. *Journal of Chemometrics*, 2006, 20(3-4): 99.
- [16] Bro Rand Kiers H A L. *Journal of Chemometrics*, 2003, 17(5): 274.

Origin Traceability of *Atractylodes Macrocephala* Koidz. by Using Three-Way Fluorescence Coupled With Chemometrics

CHEN Wei¹, WU Hai-long^{1*}, WANG Tong^{1*}, CHANG Yue-yue¹, CHEN Yao², YANG Jian³, FU Hai-yan⁴, YANG Xiao-long⁴, LI Xu-fu⁵, YU Ru-qin¹

1. State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China
2. Hunan Key Lab of Biomedical Materials and Devices, College of Life Sciences and Chemistry, Hunan University of Technology, Zhuzhou 412008, China
3. National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, State Key Laboratory Breeding Base of Dao-di Herbs, Beijing 100700, China
4. The Modernization Engineering Technology Research Center of Ethnic Minority Medicine of Hubei Province, School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan 430074, China
5. Beijing Tongrentang Pingjiang *Atractylodes Macrocephala* Koidz. Co., Ltd., Pingjiang 414500, China

Abstract *Atractylodes macrocephala* Koidz. (AM) is a famous traditional Chinese medicine, the quality of AM is directly related to the producing area, so it is urgent to find a method to quickly identify the quality of AM and trace its origin. AM contains various endogenous fluorescent substances, which can provide a basis for identifying the origin of AM by three-way fluorescence technology. In this paper, the three-way fluorescence spectrum of AM was characterized by alternating trilinear decomposition (ATLD), and then three-way fluorescence was combined with partial least squares discriminant analysis (PLS-DA) and k -Nearest Neighbors (k NN) to trace the origin of AM samples. The results show that the classification models established by PLS-DA and k NN could effectively distinguish samples from the three major AM production areas (Anhui, Hunan, and Zhejiang), and the classification accurate rate (CCR) of Zhejiang AM, a Dao-di AM herbs, is as high as 80% and 90% in the prediction set, respectively. The proposed method provides a novel and rapid solution for the origin traceability of AM, which is conducive to maintaining the order of AM market and its industrial development.

Keywords *Atractylodes macrocephala* Koidz.; Chemometrics; Three-way fluorescence; Alternating trilinear decomposition; Origin traceability

(Received May 25, 2021; accepted Aug. 24, 2021)

* Corresponding authors