

## 基于近红外光谱技术的茶园土壤酸碱状况快速判别

盛孟鸽<sup>1</sup>, 王凯<sup>2</sup>, 王玉洁<sup>1</sup>, 许姗姗<sup>1</sup>, 胡欣<sup>1</sup>, 侯智炜<sup>1</sup>, 张正竹<sup>1</sup>, 谷勋刚<sup>3</sup>, 宁井铭<sup>1\*</sup>

(1. 安徽农业大学茶树生物学与资源利用国家重点实验室, 合肥 230036; 2. 安徽农业大学理学院, 合肥 230036; 3. 安徽农业大学资源与环境学院, 合肥 230036)

**摘要:** 为了实现对茶园土壤酸碱状况量化判别, 以7个省份茶园313份土壤为材料, 以酸碱度(pH表示) 值在4.5~5.5的范围为最适宜茶树生长区间, 将pH值划分为<4.5, 4.5~5.5和>5.5 3个范围, 提出了将近红外光谱信息与贝叶斯(Bayes)判别相结合进行定性判别酸碱状况是否适合茶树正常生长。在此基础上, 采用多元线性回归(multiple linear regression, MLR)定量预测pH值。通过一阶导数(first derivative, 1stDer)对光谱预处理, 通过逐步判别分析(stepwise discriminant analysis)优选20条特征光谱, 基于特征光谱数据结合Bayes判别构建定性判别模型, 再通过MLR构建pH值的定量预测模型。结果表明, 采用本研究的方法和构建的模型对茶园土壤酸碱状况总体准确判别率达83.54%, pH值预测相关系数均在0.928 6以上, 预测精度较高。证明运用该方法能实现对茶园土壤酸碱状况快速预测。

**关键词:** 一阶导数; 贝叶斯判别; 逐步判别分析; 多元线性回归; 茶园土壤; 酸碱

中图分类号: S571.105

文献标识码: A

文章编号: 1672-352X(2018)04-0580-08

### Rapid identification soil acid-base of tea garden based on near infrared spectroscopy

SHENG Mengge<sup>1</sup>, WANG Kai<sup>2</sup>, WANG Yujie<sup>1</sup>, XU Shanshan<sup>1</sup>, HU Xin<sup>1</sup>,  
HOU Zhiwei<sup>1</sup>, ZHANG Zhengzhu<sup>1</sup>, GU Xungang<sup>3</sup>, NING Jingming<sup>1</sup>

(1. State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036; 2. School of Science, Anhui Agricultural University, Hefei 230036; 3. School of Resources and Environment, Anhui Agricultural University, Hefei 230036)

**Abstract:** In order to develop a digitized discriminant on soil base conditions in tea garden, 313 soil samples were collected from tea gardens of seven different provinces. Usually, the pH value of soil between 4.5 and 5.5 is the most suitable range for the growth of tea plant. The soil pH values were divided into three ranges of <4.5, 4.5-5.5 and >5.5, respectively. Near infrared spectroscopy (NIRS) information combined with Bayesian discriminant was used to judge whether the base conditions are suitable for the normal growth of tea plant or not. On this basis, a multiple linear regression (MLR) model was established for the quantitative prediction of the pH values. The first derivative (1<sup>st</sup>Der) was used to preprocess NIRS data for removing redundant information. Twenty characteristic spectra were optimized and selected by stepwise discriminant analysis. According to the characteristic spectral data combined with Bayesian discriminant, the qualitative prediction model was established, and the quantitative prediction model of pH value was also established according to MLR. The results indicated that stepwise discriminant analysis combined with Bayesian discriminant model was a promising tool for the discrimination of soil base conditions, and the total accurate recognition rate of 83.54% was achieved. In addition, the result also showed that the correlation coefficients of the three ranges were all above 0.928 6 by MLR model, and the prediction accuracy was extremely high. The results of this study proven that rapidly and scientifically discrimination soil base conditions in tea garden can be achieved via the method developed in this paper.

**Key words:** first derivative; Bayesian discriminant; stepwise discriminant analysis; multiple linear regression; tea garden soil; acid-base condition

收稿日期: 2018-01-23

基金项目: 国家重点研发计划(2016YFD0200900)和中国现代农业(茶叶)专项体系(CARS-19)共同资助。

作者简介: 盛孟鸽, 硕士研究生。E-mail: 1330409483@qq.com

\* 通信作者: 宁井铭, 博士, 教授。E-mail: ningjm@ahau.edu.cn

茶树的正常生长易受到土壤化学环境的影响, 其中土壤酸碱度、无机养分对其影响最大。土壤酸碱度(通常用 pH 值表示)是土壤酸碱状况的综合体现。茶树是多年生, 喜酸植物, 其生长受土壤酸碱度影响较大, 茶树生长的适宜 pH 值为 4.5~6.5 之间, 最适 pH 值为 4.5~5.5<sup>[1]</sup>。当土壤过酸或偏碱时, 叶片发黄不利于光合作用, 茶树的生长受阻, 从而影响茶叶的产量和品质。土壤 pH 的改变能影响矿质盐分溶解度, 从而影响土壤养分的有效性<sup>[2]</sup>。当 pH<4.5 或 >6.5 时, 茶树对氮、磷、钾的吸收能力显著下降<sup>[3]</sup>。因此, 准确掌握土壤 pH 值的变化不仅能维持茶树旺盛的生长势态, 而且可以提高茶叶产量和品质。传统实验室检测方法操作步骤繁琐、成本高, 不利于大规模、高效地实时掌握土壤养分信息<sup>[4]</sup>。近红外光谱技术(near infrared spectrometry, NIRS)具有操作简便、准确高效、不破坏样品、多组分同时测定等优点, 其在精准农业上的应用已受到国内外学者的高度关注<sup>[5-6]</sup>。

NIRS 在土壤学方面的研究始于 1965 年, Bower 和 Hanks<sup>[7]</sup>首次运用 NIRS 检测土壤, 发现有机质影响着太阳辐射能对土壤的反射和吸收。随后在土壤肥力因子方面的研究日益增多。2003 年 Choe 等<sup>[8]</sup>从佛罗里达 3 个不同类型 15 个取样点收集 270 份土壤样品, 基于偏最小二乘法(partial least squares, PLS)建立预测模型预测样品的 pH 和营养(P, K, Ca 和 Mg 等)含量, 实验证明构建的模型可以用来

开发土壤养分传感器, 用于现场的作物管理。Marin-González 等<sup>[9]</sup>收集欧洲(捷克共和国, 丹麦, 荷兰和英国)不同领域的 4 个农场 160 个土壤样品, 用同样的方法建模并加以验证实现极好的预测。而 Kissel 等<sup>[10]</sup>则认为采用一阶导数(first derivative, 1<sup>st</sup>Der)处理后, 对波长在为 950~1 650nm 范围的光谱进行分析, 用改进的偏最小二乘回归(partial least squares regression, PLR)来开发校准模型是有前途的。Li 等<sup>[11]</sup>、李颖等<sup>[12]</sup>、董一威等<sup>[13]</sup>都基于 NIRS 对土壤的 pH 进行预测, 研究结果大多都认为运用 PLS 预测效果最好, 但很少有人针对样品的最优制备条件进行研究。有的研究直接采用风干土进行<sup>[14]</sup>, 有的则风干研磨后进行检测, 对土壤粒度的选取也不统一<sup>[12-13]</sup>。另外, 基于 NIRS 对土壤 pH 预测时特征光谱的选取上很少有人去关注, 构建的模型仅适应当地的土壤预测, 具有一定局限性<sup>[10,15]</sup>。

本研究以中国 7 个省份的茶园 313 份土壤为研究对象, 采用近红外光谱仪获取土壤在 800~2 500 nm 波长范围的光谱信息, 通过光谱预处理消除外界因素的干扰后筛选特征波长下的光谱信息, 结合化学计量学方法构建茶园土壤酸碱状况的快速判别模型, 并尝试在定性判别的基础上采用多元线性回归模型(multiple linear regression, MLR)对茶园土壤 pH 定量预测, 从而实现从定性到定量的快速判别(主要技术路线如图 1), 为田间实时便携式传感器 pH 测定仪的开发奠定理论依据。

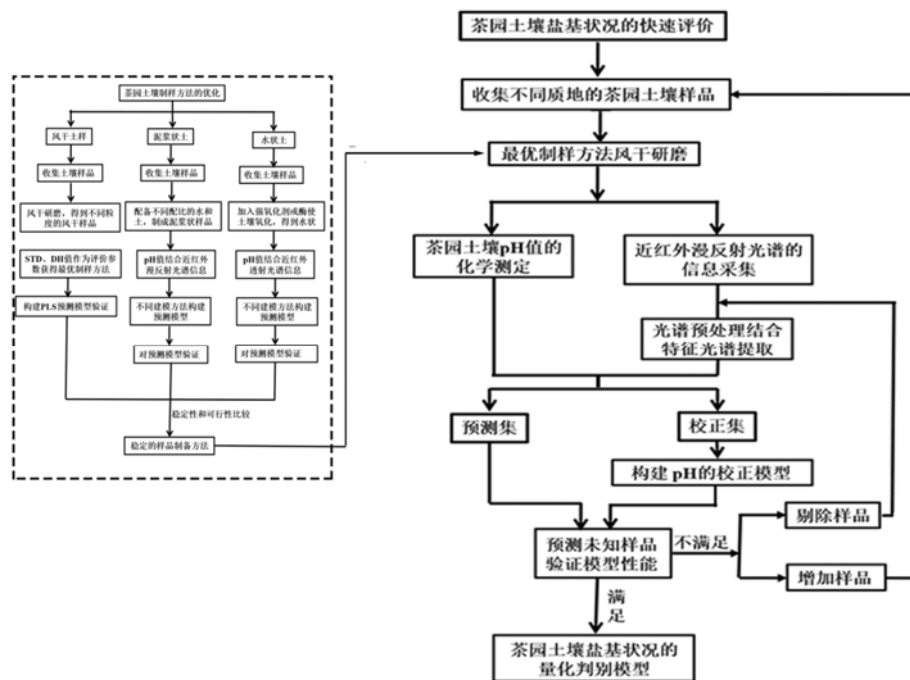


图 1 研究技术路线

Figure 1 Key steps of the main technique route

## 1 材料与方法

### 1.1 材料

**1.1.1 制样条件优化** 于2016年7月23日从安徽省黄山市祁门县祥源茶厂采集的90份茶园土壤样品(如图2a),用来优化样品的制备条件。将风干土样研磨后分别制成泥浆状土、水状土、风干土(如图3)3种状态。通过实验,确定最优的土壤样品制样方法。

**1.1.2 茶园土壤酸碱状况的量化判别** 本实验共收集的313个茶园土壤样本,除了1.1.1中用于优化的90个土壤样本外,还包括福建、云南、湖北、四川、山东、贵州和安徽7个省份不同土壤类型的产茶区223份样本(图2b)。采用5点采样法取样。每个土壤样品重约1 kg,装于透明塑料袋内,并贴好标签带回实验室,剔除杂质,放在通风阴凉处风干。根据最优样品制备条件,将样品研磨、统一过筛后,分为2部分。一部分用来采集近红外光谱信息,另一部分用来进行土壤pH化学测定。数据分析前,将313个样本以3:1的比例随机划分为校正集和预测集(如表1所示),其中校正集样本用来构建校正模型,预测集样本用于验证模型预测性能,验证模

型的稳健性和适应性。

### 1.2 光谱信息的采集和处理

本研究采用傅里叶变换MPA型近红外光谱仪(带有PbS检测器)进行茶园土壤样本近红外光谱数据的采集。为了获得高质量的近红外光谱数据信息,开始操作前需要对仪器预热30 min左右,并进行系统适应性测试检验仪器性能。同时,对实验参数光谱分辨率、扫描范围等进行调节设置,在漫反射模式下获得土壤样品光谱数据。光谱扫描波数范围约为:12 493~39 99  $\text{cm}^{-1}$ (波长范围约为800~2 500 nm),采样间隔为3.86  $\text{cm}^{-1}$ ,光谱分辨率为8  $\text{cm}^{-1}$ ,样本和背景都被一个积分球扫描32次。为了减少实验误差,在整个实验过程要保持恒定室温( $25\pm 1$ ) $^{\circ}\text{C}$ 和湿度( $45\pm 1$ )%。进行空白校正后,风干土壤样本(约20 g)装入样品杯(直径97 mm)中,经压样器充分压实使近红外光无法穿透样品后放入光谱仪漫反射探头上,通过每次120 $^{\circ}$ 手动转换样品杯的位置3次旋转对土壤样本扫描3次,进行光谱信息采集。

### 1.3 土壤pH值的检测

茶园土壤pH值的测定按照国际标准ISO 10390:2005执行<sup>[16]</sup>。

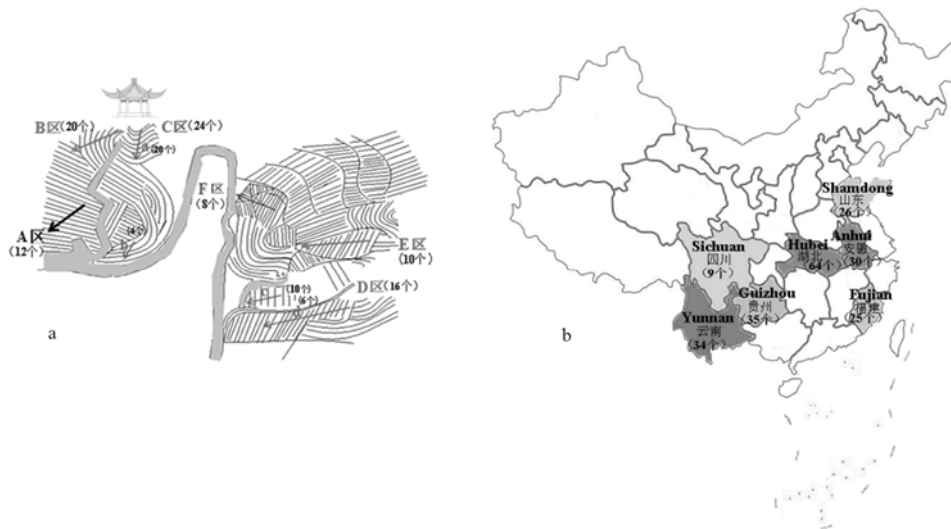


图2 研究区域样点分布和每个区域样品数

Figure 2 The study area of all soil samples and the number of samples

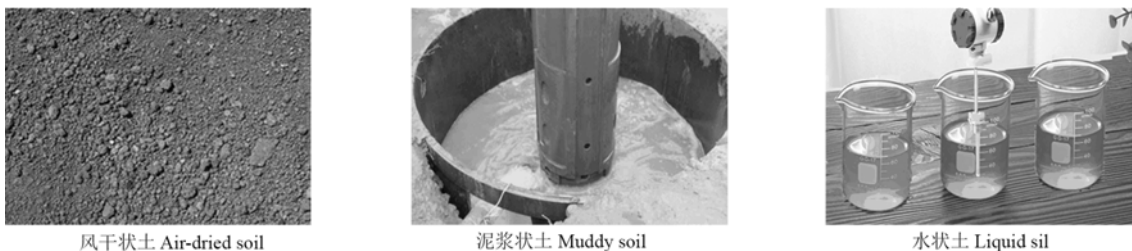


图3 不同状态下的茶园土壤样品

Figure 3 Tea garden soil samples in different states

表 1 土壤样品集的划分  
Table 1 The division of soil sample sets

不同范围的 pH 值 pH value of different ranges	每个范围样本数 Sample number	校正集 Calibration set	预测集 Prediction set
总样本 Total	313	234	79
<4.5	123	92	31
4.5~5.5	135	101	34
>5.5	55	41	14

## 1.4 数据分析与模型构建

**1.4.1 光谱数据预处理** 近红外光谱检测仪在采集光谱信息时,容易受杂散光干扰。仪器性能的变化,也可能使光谱产生很大的噪声。在构建模型之前,对光谱预处理可以尽可能多地减少无关因素对光谱的影响,使构建的模型更加准确、可靠。本研究采用 1<sup>st</sup>Der 预处理光谱,将重叠的光谱进行分解,消除与波长无关的光谱基线偏移、漂移,减少可能由于背景干扰引起的系统误差<sup>[17]</sup>,增加样品间的光谱特征差异,使原始光谱特征峰呈现的更加明显。具体计算方法为:假设原光谱数据的 X 轴(波长)是等间距的,则一阶导数为:

$$\Delta x' = \frac{(x_{j+1}^i - x_j^i)}{x_j^i} \quad (1)$$

式(1)中  $x_j^i$  表示第  $i$  个样本,第  $j$  条光谱。

**1.4.2 判别指标的优化筛选** 全光谱变量较多,相邻波段间光谱矩阵的相关性可能存在信息冗余,影响构建模型的预测精度和性能。因此,剔除不敏感的波段,优选有效特征变量,不仅降低模型的复杂度、提高运算速度,而且有助于增强模型的鲁棒性。常用的特征波长提取方法有主成分分析(PCA)<sup>[18]</sup>、竞争性自适应权重系数法(CARS)<sup>[19]</sup>、连续投影算法(SPA)<sup>[20]</sup>等。经过反复实验用 PCA、CARS 和 SPA 来提取土壤 pH 特征变量光谱进行建模取得较好的预测效果是相当困难的,因此本研究确定采用逐步判别分析选取特征波长优选变量。逐步判别分析是利用一些检验规则,根据筛选变量的重要性逐步引入判别式因子的方法<sup>[21]</sup>。也就是说,通过对所有变量进行筛选,从这些变量中优选最具影响力的变量,这样逐步依次引入。在引入过程中,每引入一个新变量,再对原有变量的判别能力进行检验,筛除判别能力差的变量。如此循环,直至没有新的变量可以继续引入,也没有已选变量可以筛除时逐步筛选结束,然后利用已筛选的最优变量建立判别函数<sup>[22]</sup>。其中,变量的重要性是根据构建的统计量来判断的。这种根据重要性的大小筛选特征变量的方法,更加准确客观。

**1.4.3 定性判别模型选取** 采用近红外光谱进行建模和分类识别常见的方法有支持向量机(SVM)<sup>[23]</sup>、线性判别分析法(LDA)<sup>[24]</sup>、极限学习机(extreme learning machine, ELM)<sup>[25]</sup>等通过预测分类识别率来确定最佳的分类判别模型。本研究使用的 Bayes 判别是一种基于先验概率,根据新的信息,获得后验概率的方法,误判引起的损失更小<sup>[26]</sup>。在本研究中,将 313 个总样本按 3:1 划分为训练集和测试集,利用测试集样本,采用逐步判别筛选的指标变量,使用 Bayes 判别方法构造判别函数,计算后验概率<sup>[27-28]</sup>,以判断样本属于哪一个 pH 范围内的土壤,并将校正集样本代入该判别函数进行检验。二次判别函数表示为式(2)和式(3):

$$Z_j = -\frac{1}{2} D_j(X) = -\frac{1}{2} [-2 \ln q_j + \ln |S_j| + d_j^2(X)], j = (1, 2, 3) \quad (2)$$

$$P\{X \in G_j | X\} = \frac{e^{z_j(X)}}{\sum_{j=1}^k e^{z_j(X)}} \quad (3)$$

式(2)中  $q_j$  是先验概率,  $S_j$  是协方差矩阵,  $d_j^2$  是样本  $X$  与第  $j$  类之间马氏距离,  $P$  是后验概率。

**1.4.4 定量预测模型的构建** 在定性分类的基础上,根据优化筛选的指标组合,采用原样本集的划分方法,再基于定性判别模型结果,采用多元线性回归(multiple linear regression, MLR)进行多元定量预测。MLR 是由两个或两个以上自变量共同来预测或估计因变量的回归分析<sup>[29]</sup>。多元线性回归的一般形式:  $Y = a_0 + a_1 X_1 + a_2 X_2 \cdots + a_p X_p$  ( $P$  为解释变量的数目,  $a_1, a_2, \dots, a_p$  为偏回归系数,  $\varepsilon$  为随机误差项)。运算形式为:逐步判别优选的原指标,用偏最小二乘法进行回归估计,得到回归方程;对回归方程假设检验。

**1.4.5 数据处理和模型评价** 采用 Matlab 2014a (The Mathworks Inc., Massachusetts, USA)、OPUS6.5 (MPA150, Bruker Optik GmbH, Ettlingen, Germany)、Origin 9.0 软件对数据进行处理。

定性判别模型的评价指标为构建的模型在校正集和预测集的准确判别率,即每类判别正确的样品总和与总样本数的百分比越接近 100%;定量预测

模型的评价指标为校正集和预测集相关系数越接近1, 预测集分布范围越广, 模型精度越高。

## 2 结果与分析

### 2.1 样品制样条件的优化

泥浆状土样虽然可以模拟土壤环境, 但是由于在实际操作过程中加入的量特别少, 水分易挥发, 容易搅拌不均匀会使实验结果误差较大, 故此方法不可取。水状土样, 目前还没有找到某种酶或者强氧化剂就可以使土壤完全分解, 因此无法用近红外透射光谱进行透射分析。风干土壤样品是土壤养分检测中最常用的, 通过实验取得较好的预测效果。

本研究采用风干土壤样品, 将其研磨成60、80、100和120目4个不同粒度, 以光谱吸光度标准偏差值(standard deviation, STD)和差谱值(differential spectrum height, DH)作为评价参数, 来确定最优的样品制备条件。实验表明: 样品的粒度为100目时, 光谱重复性最好, 并用PLS模型进行对单因素实验加以验证。因此, 样品的最优制备条件为100目, 后续实验采用100目的风干样品进行研究。

### 2.2 不同地区茶园土壤样品pH值的分析

来自我国主要产茶省的不同土壤类型茶园土壤

样品pH值的统计结果如表2所示。从表2可以看出, 茶树土壤pH值在4.5~5.5范围内, 有135个样本, 所占的比例最大; 其次在pH值<4.5的范围内, 有123个样本; pH值从3.04~7.45, 标准偏差为0.68, pH范围覆盖的广泛。表明选择的样本合理, 具有较好的代表性, 有利于进一步分析建模。

### 2.3 光谱预处理方法的选取

计算3种不同pH范围土壤样本的平均值绘制的光谱曲线如图4a所示。不同范围内样本pH的光谱曲线趋势相似, 大约在1415 nm、1920 nm和2207 nm处出现明显的特征峰。当pH值介于4.5~5.5之间时, 图谱的吸光值在大约800~1156 nm范围内迅速降低, 在大约1157~2203 nm范围整体达到最小; pH值>5.5时整条光谱的吸光值都相对偏高。由于总体样本数较多, 3个范围样本pH值差异较小存在谱峰重叠。通过对2203个光谱数据进行1<sup>st</sup>Der预处理后最终得到2202个光谱值, 3个范围样本pH光谱曲线如图4b所示, 从图中可以看出预处理后样品间的光谱特征差异增大, 光谱特征峰更明显。这与Kissel等<sup>[10]</sup>分析的粒度为2 mm田间风干土壤的处理相近。

表2 茶园土壤样本的pH值统计结果

Table 2 The statistical results of pH in tea garden soil sample

取样地区 Location	样本集 Subsets	最大值 Maximum	最小值 Minimum	平均值 Mean	标准差 Standard deviation
安徽 Anhui	120	7.45	3.91	4.43	0.66
云南 Yunnan	35	6.33	3.78	5.25	0.64
福建 Fujian	25	6.28	3.04	4.56	0.77
贵州 Guizhou	34	6.54	3.72	4.84	0.70
四川 Sichuan	9	6.58	4.43	5.54	0.76
山东 Shandong	26	6.87	3.98	5.15	0.74
湖北 Hubei	64	6.87	3.95	4.94	0.57
pH<4.5	123	4.49	3.04	4.21	0.22
4.5≤pH≤5.5	135	5.50	4.51	4.90	0.26
pH>5.5	55	7.45	5.51	5.98	0.41
总样品 Total sample	313	7.45	3.04	4.82	0.68

### 2.4 光谱特征波长的优选

经过1<sup>st</sup>Der预处理后的对应2202个光谱数据进行逐步判别分析。具体的筛选步骤如下: (1) 根据总离差阵与分组离差阵和构建统计量。选取判别贡献率最大的, 再对其构建F统计量。利用F检验检验其显著性。(2) 若F检验结果显著, 则引入该新指标, 然后对指标集中原引入的指标构建F统计量检验。若不显著, 则剔除原引入指标(3)如此循

环, 直到没有新的指标可以继续引入, 没有旧的指标可以剔除。根据指标选取越少越好的原则, 本研究循环共进行30步, 最终筛选30个特征变量。对这30个变量进行判别组合分析, 最终优选879.45、897.72、944.49、952.12、975.77、979.09、1001.40、1013.93、1031.69、1039.13、1047.53、1128.21、1163.66、1172.60、1191.46、1393.14、1496.90、1516.16、1701.20和2052.75 nm这前20条特征光

谱, 用于后期盐基状况判别分析。

## 2.5 盐基状况定性判别模型的构建

光谱信息能表征茶园土壤内部特性, 为了更好地体现土壤 pH 的变化, 本研究将优选的前 20 个特征变量对应的特征光谱数据作为 Bayes 判别模型的输入值, 建立盐基状况的判别模型, 结果 (表 3) 显示, 通过逐步判别分析优选变量结合模式识别进行 Bayes 判别可以得到较高的准确判别率, 校正集、预测集的总体判别率分别达到 87.18%, 83.54%; 在预测集中有 5 个 pH<4.5 的样本被误判为 pH 值介于 4.5~5.5, 6 个 pH 值介于 4.5~5.5 的样本 4 个被

误判为 pH <4.5, 2 个被误判为 pH>5.5。2 个 pH >5.5 的样本, 2 个被误判为 pH<4.5。通过比较可知, pH<4.5 和介于 4.5~5.5 产生误判的相对较多, 这可能是因为光谱数据综合了 7 个省份不同类型茶园土壤的内部信息, 而介于 pH 值 4.5~5.5 的样品来源范围相对较小为建立高精度的模型增加一定难度。但是, 通过 1<sup>st</sup>Der 预处理, 采用逐步判别分析优选特征光谱, 结合 Bayes 判别进行模式识别, 这种方法为茶园土壤酸碱状况的判别提供了一个可靠的新思路。

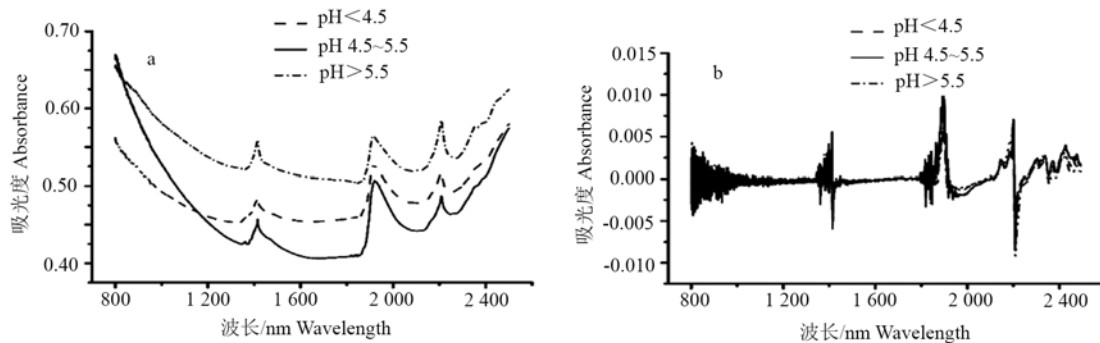


图 4 不同 pH 范围样本 (a) 原始、(b) 一阶导数预处理后的光谱曲线

Figure 4 The spectral curves of all samples with three different ranges: (a) Original spectra; (b) 1st derivative preprocessing spectra

表 3 特征波长下茶园土壤酸碱状况的 Bayes 判别模型识别结果

Table 3 The predicted results of Bayes discrimination model based on characteristic wavelength in calibration and prediction set.

样本集 Sample sets	pH 值 pH value	样本(n) Sample number	正确判别结果 Correct discrimination results			总体识别率/% Total discrimination correct rate
			< 4.5	4.5~5.5	> 5.5	
校正集 Calibration set	< 4.5	92	78	7	7	84.78
	4.5~5.5	101	9	86	6	85.15
	> 5.5	41	0	1	40	97.56
预测集 Prediction set	< 4.5	31	26	5	0	83.87
	4.5~5.5	34	4	28	2	82.35
	> 5.5	14	2	0	12	85.71

## 2.6 酸碱状况定量预测模型的构建

在茶园生产管理过程中, 土壤 pH 值随着自然因素 (生物、气候、母质等) 和人为活动 (施肥、灌溉等) 的影响一直处于动态变化。通过定性判别可以监测到土壤相对稳定地分布在一定 pH 范围内, 判定茶树是否正常生长。但是, 要采取科学合理的手段改善土壤酸碱度 (缓冲性、合理施肥等), 定量检测 pH 值来预测酸碱状况是十分必要的。为了更加准确地量化判别茶园土壤酸碱状况, 基于定性判别的结果, 对样本的 pH 值进行定量预测。采用优选的 20 个特征变量对应的光谱数据构建 MLR 模型对 pH 值定量预测, 依据评价指标对模型精度

进行判断。我国茶区分布范围较广泛, 不同茶区样本土壤类型多样, 土壤内部特性差异很大, 加上 3 个 pH 范围的取样相对集中, 因此建模难度较大。然而, 从表 4 中看出,  $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$ ..... $\alpha_{21}$  为 3 个 MLR 模型的回归系数。校正集的预测相关系数都达到 0.99 以上。为验证模型的有效性和适应性, 对每个 pH 范围未参与建模的样品进行预测, 预测相关系数均在 0.928 6 以上。将预测值与实际测定值进行方差分析, 两种方法测的结果差异不显著。这表示优选出的特征光谱用来预测 pH 值是可行的。可见, 采用 MLR 构建的模型预测土壤 pH 值效果极佳。

表 4 采用多元线性回归模型对茶园土壤 pH 值定量预测结果

Table 4 The predict results of soil pH value in tea garden using multivariate linear regression model

回归系数 Regression coefficient	pH<4.5	4.5≤pH≤5.5	pH>5.5
$\alpha_1$	4.44	4.74	6.79
$\alpha_2$	12.04	-168.71	363.85
$\alpha_3$	61.14	-27.11	-79.10
$\alpha_4$	-129.05	-843.36	-2 604.07
$\alpha_5$	103.70	3.34	-183.20
$\alpha_6$	-231.63	-278.37	-24.88
$\alpha_7$	1226.91	180.05	1 885.25
$\alpha_8$	-109.16	171.78	64.89
$\alpha_9$	17.40	-1.44	-150.66
$\alpha_{10}$	128.38	-1.41	446.58
$\alpha_{11}$	-251.93	46.20	-738.90
$\alpha_{12}$	-50.45	134.50	427.39
$\alpha_{13}$	-17.52	-4.60	3.08
$\alpha_{14}$	16.52	-109.01	478.80
$\alpha_{15}$	-29.35	0.52	-25.27
$\alpha_{16}$	-1.35	88.90	117.02
$\alpha_{17}$	-389.85	-356.14	-1 107.18
$\alpha_{18}$	190.72	55.26	432.26
$\alpha_{19}$	-377.48	-160.28	-652.82
$\alpha_{20}$	203.35	-172.58	1 120.81
$\alpha_{21}$	71.70	125.82	-6.52
校正集 Calibration set	1	0.990 1	1
预测集 Prediction set	1	1	0.928 6

### 3 结论

本研究基于近红外光谱技术结合模式识别成功实现了茶园土壤酸碱状况的量化判别。通过 1<sup>st</sup>Der 对光谱数据进行预处理,采用逐步判别分析优选 20 个特征变量及对应特征光谱值,结合 Bayes 判别构建茶园土壤酸碱状况的快速判别模型,模型的判别率分别达到 83.87%、82.35%和 85.71%。在定性判别的基础上,仅用优选 20 个特征变量对应的光谱数据结合 MLR 模型建立 pH 值的定量预测模型,预测相关系数在 0.928 6 以上,预测效果较好。因此,快速判别模型的成功构建,为以后开发便携式传感器应用于茶园管理提供理论依据。

### 参考文献:

- [1] 骆耀平. 茶树栽培学 [M]. 5 版.北京:中国农业出版社, 2015.
- [2] 赵仁全, 黄艳红, 李琳. 土壤 pH 值对茶树生长和茶叶品质影响的研究进展[J]. 农技服务, 2012(3):295-296.
- [3] FUNG K F, WONG M H. Effects of soil pH on the uptake of Al, F and other elements by tea plants[J]. J Sci Food Agr, 2002, 82(1):146-152.
- [4] 陆婉珍. 现代近红外光谱分析技术 [M]. 2 版.北京:中国石化出版社, 2007:9-11.
- [5] 褚小立, 袁洪福. 近红外光谱分析技术发展和应用现状[J]. 现代仪器与医疗, 2011(5):1-4.
- [6] CROFT H, KUHN N J, ANDERSON K. On the use of remote sensing techniques for monitoring spatiotemporal soil organic carbon dynamics in agricultural systems[J]. Catena, 2012, 94(9):64-74.
- [7] BOWERS S, HANKS R. Reflection of radiant energy from soil[J]. Soil Sci, 1965, 100(2):130-138.
- [8] CHOE J S, SANCHEZ J F, RAO S M, et al. Estimating chemical properties of Florida soils using spectral reflectance[J]. T Asabe, 2003, 46(227):1443-1453.
- [9] MARÍN-GONZÁLEZ O, KUANG B, QURAIISHI M Z, et al. On-line measurement of soil properties without direct spectral response in near infrared spectral range[J]. Soil Till Res, 2013, 132(4):21-29.
- [10] KISSEL D, CHRISTY C, SHAABAN S, et al. Measurement of pH, pH buffering capacity, and other soil properties with NIR reflectance spectroscopy[C]//The 18th world congress of soil science. Philadelphia, 2006.
- [11] LI M Z, SASAO A, SHIBUSAWA S, et al. Soil parameters estimation with NIR spectroscopy[J]. J Jpn Soc Agr Mach, 2000, 62(3): 111-120.
- [12] 李颀, 张小超, 苑严伟, 等. 北京典型耕作土壤养分的

- 近红外光谱分析研究[J]. 农机化研究, 2012, 34(4): 115-118.
- [13] 董一威, 黄金丽, 孙宝利, 等. 我国苹果栽培区土壤参数的近红外光谱检测研究[J]. 光谱学与光谱分析, 2009, 29(8): 2075-2078.
- [14] 朱登胜, 吴迪, 宋海燕, 等. 应用近红外光谱法测定土壤的有机质和 pH 值[J]. 农业工程学报, 2008, 24(6): 196-199.
- [15] 王昶. 近红外光谱快速评估土壤和有机肥质量研究[D]. 南京: 南京农业大学, 2014.
- [16] 李强, 赵秀兰, 胡彩荣. ISO10390:2005 土壤质量 pH 的测定[J]. 污染防治技术, 2006, 19(1): 53-55.
- [17] 邓小蕾, 李民赞, 郑立华, 等. 基于反射光谱预处理的苹果叶片叶绿素含量预测[J]. 农业工程学报, 2014, 30(14): 140-147.
- [18] NING J M, SUN J J, LI S H, et al. Classification of five Chinese tea categories with different fermentation degrees using visible and near-infrared hyperspectral imaging[J]. *Int J Food Prop*, 2017, 20(sup2): 1515-1522.
- [19] 刘燕德, 姜小刚, 熊松盛, 等. 基于 CARS 算法的脐橙果园土壤有机质近红外检测[J]. 中国农机化学报, 2016, 37(2): 240-244.
- [20] 何东健, 陈熙. 土壤有机质含量田间实时测定方法[J]. 农业机械学报, 2015, 46(1): 127-132.
- [21] 孟凡奇, 李广杰, 李明, 等. 逐步判别分析法在筛选泥石流评价因子中的应用[J]. 岩土力学, 2010, 31(9): 2925-2929.
- [22] 卢文喜, 李俊, 于福荣, 等. 逐步判别分析法在筛选水质评价因子中的应用[J]. 吉林大学学报 (地球科学版), 2009, 39(1): 126-130.
- [23] ZHANG S, JIE D, ZHANG H. NIR spectroscopy identification of persimmon varieties based on PCA-SVM[C]// *Computer and computing technologies in agriculture IV- Ifip Tc 12 Conference, Ccta 2010, Nanchang, China, October 22-25, 2010, Selected Papers. DBLP*, 2017: 118-123.
- [24] WEI X, LIU F, QIU Z J, et al. Ripeness classification of astringent persimmon using hyperspectral imaging technique[J]. *Food Bioprocess Tech*, 2014, 7(5): 1371-1380.
- [25] HUANG G, HUANG G B, SONG S J, et al. Trends in extreme learning machines: A review[J]. *Neural Networks*, 2015, 61: 32-48.
- [26] ZHANG J F, WANG S R, CHEN L F, et al. Multiple Bayesian discriminant functions for high-dimensional massive data classification[J]. *Data Min Knowl Disc*, 2017, 31(2): 465-501.
- [27] GONG F Q, LI X B, ZUO Y J. A Bayes discriminant analysis method for predicting the hazard classification of rock burst and its application[C]// *International symposium on risk control and management of design, construction and operation in underground engineering. Dalian, 2009*.
- [28] HU Y X, LI X B. Bayes discriminant analysis method to identify risky of complicated goaf in mines and its application[J]. *T Nonferr Metal Soc*, 2012, 22(2): 425-431.
- [29] PREACHER K J, CURRAN P J, BAUER D J. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis[J]. *J Educ Behav Stat*, 2006, 31(4): 437-448.