

小波分析的茶鲜叶全氮含量高光谱监测

王 凡^{1,2}, 陈龙跃^{2,3}, 段丹丹^{1,2,4*}, 曹 琼^{1,4}, 赵 钰¹, 蓝玩荣⁵

1. 国家农业信息化工程技术研究中心, 北京 100097
2. 清远市智慧农业农村研究院, 广东 清远 511500
3. 农芯科技(广州)有限责任公司, 广东 广州 510000
4. 湖南农业大学, 湖南 长沙 410125
5. 江门市农业技术服务中心, 广东 江门 529000

摘要 茶是世界上最受欢迎的饮料之一, 而氮素(N)是影响茶叶品质的主要成分之一, 因此快速准确地估算 N 素含量至关重要。由于测定 N 含量的化学方法繁琐耗时, 利用高光谱对茶鲜叶中 N 含量进行预测, 利用连续小波转换(CWT)提取的小波系数, 探究 CWT 不同分解层数对于 N 素含量的估测能力, 并讨论了不同波长选择算法所建模型的预测效果。首先, 采集广东省英德市茶园的 151 个茶鲜叶样品高光谱数据, 将获得的原始光谱通过卷积平滑(SG)、去趋势(Detrending)、一阶导数(1st)、多元散射校正(MSC)和标准正态变量变换(SNV)五种预处理方法进行预处理并作为参考。其次, 采用连续小波对原始光谱进行初步处理生成多尺度小波系数, 并进行相关性分析, 分别利用连续投影算法(SPA)、竞争性自适应加权采样法(CARS)和变量组合集分析(VCPA)方法进一步优化 CWT 变换后光谱数据的变量空间, 最后, 以特征变量为输入使用 PLSR 建立了 N 素定量监测模型, 并对比不同尺度不同方法估算 N 素的效果。结果表明, 连续小波分析方法可有效提升茶鲜叶光谱对 N 素含量的估测能力, 明显优于常规光谱处理方法。经连续小波分解后, 对茶鲜叶 N 素的预测能力随分解尺度的增加整体呈逐步降低的趋势, 其中在 1~6 尺度连续小波变换后的光谱与茶鲜叶 N 素存在良好的相关性, 表明小尺度的连续小波分解可有效应用于茶鲜叶 N 含量的监测。基于 CWT(1)-VCPA 方法建立的模型精度最高, 且变量数相比于全波段减少了 99.34%, 其建模与预测 R^2 达到 0.95 和 0.90, 相比于传统光谱处理方法, 精度提升了 11%, 证明 CWT-VCPA 可以有效降低光谱维度并大幅提升模型精度。实现了茶叶 N 素含量的高效量化预测, 为评估茶叶的其他成分提供了可靠技术参考。

关键词 茶鲜叶; 氮素; 连续小波变换; 高光谱; 变量组合集分析

中图分类号: S56 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3235-08

引 言

茶树是世界上广泛种植的作物, 因其独特的滋味以及强大的健康功效而受到人们的喜爱。众所周知, 氮(N)是植物光合作用和生长必需的元素, 实时进行氮素管理, 对于保证茶叶产量, 减少环境污染^[1]具有指导意义。

传统的作物氮素检测方法费时耗力, 对样品具有破坏性, 需具备较高的操作技能, 无法满足实时准确的测定需求^[2]。高光谱作为一种成熟的无损检测技术已被广泛应用于茶叶的营养诊断、种类识别、长势检测^[3-4]。光谱测量植生

化变量中有光谱信息冗余等问题, 易受到噪音和背景因素的影响。如何从复杂重叠的光谱信息中得到微弱的化学成分变化的有效信息, 提高检测精度, 是光谱分析的关键。研究表明, 偏最小二乘方法(PLSR), 通过使用主成分分析(PCA)将自变量投影到新空间以概括潜在变量, 可以处理大量的噪声和相关变量, 被证实同样适用于少量变量的情况^[5]。选择合适的特征提取方法可以降低模型的复杂程度并提高预测准确性和鲁棒性, 且少量变量即可提供与原始全光谱波段相媲美的信息。已有的研究报道了许多方法, 如 SPA, UVE, PCA, RF 以及 CARS 等^[6-7], 这些方法都是选择信息变量非常有效的策略, 然而其变量选择方法大多是利用蒙特卡洛采样法在

收稿日期: 2021-06-22, 修订日期: 2022-01-17

基金项目: 国家自然科学基金项目(21974012)资助

作者简介: 王 凡, 1994 年生, 国家农业信息化工程技术研究中心工程师 e-mail: wfnn9448@163.com

* 通讯作者 e-mail: duandd@nercita.org.cn

样本变量空间进行重复取样后一次性建模,却没有考虑到变量之间的组合效应。此外,在整个光谱区间进行变量搜索,这些方法仅消除了非信息性变量而未考虑干扰变量的影响,当存在干扰变量时,将无法获得令人满意的结果。连续小波变换(CWT)用于处理高光谱数据,通过缩放和移动小波函数将原始光谱分解为多个尺度,促进对细微特征信号的识别,从而构建对光谱敏感的生物和化学成分的预测模型,已广泛应用于植被参数评估。变量组合集群分析(VCPA)算法,是基于达尔文进化论的“适者生存”以及依据交叉验证均方根误差(RMSE)的一种变量选择方法^[8],从全光谱数据中选择光谱变量的最佳组合,利用变量的随机组合考虑变量间的相互作用,进行多次迭代的分析,直到剩下的变量没有无信息变量和干扰变量,在特征提取方面效果较好^[9-10]。连续投影算法(SPA)、竞争性自适应重加权算法(CARS)广泛用于高光谱信息筛选中以提高模型性能,而关于 VCPA 的应用研究较少;另外,基于小波分析的品质参数估算,在茶鲜叶监测上的应用鲜有报道,目前尚未有开发 CWT 与 VCPA 特征变量选择算法相结合的模型优化方法的应用实例。

以茶鲜叶为研究对象,利用小波分解对茶鲜叶光谱展开分析,确定了 CWT 在茶叶 N 素含量鉴定中的性能,结合实验室理化检测茶鲜叶 N 素含量,采用 SPA, CARS 和 VCPA 特征波段选择算法提取有益信息,建立茶鲜叶 N 素含量预测模型,并探究在小波算法不同分解尺度下,不同特征波段选择算法对茶鲜叶 N 素估算精度的影响。

1 实验部分

1.1 研究区概况

实验区位于广东省清远市英德市积庆里、创美、德高信、八百秀才和石门山 5 个茶园(图 1),该区域位于北纬 24.50°,东经 113.55°,平均海拔为 181 m,地处亚热带向中亚热带的过渡地区,属亚热带季风气候。茶鲜叶光谱数据采集在 2020 年 5 月、7 月和 9 月开展,覆盖 3 个品种(包括主栽品种英红九号,鸿雁十二号,金萱),采样点均匀分布于实验区内,共选取了 151 个样点,每个样点采集茶鲜叶样品约 200 g,进行化验。

1.2 数据采集

叶片的光谱反射率采用美国 ASD Fieldspec Pro3 光谱仪测量,其光谱覆盖范围为 350~2 500 nm,光谱重采样间隔为 1 nm。利用叶片夹持器(自带光源)选取叶片的中部较宽的部分进行光谱测定,手柄中的光源部分直接夹紧茶叶叶片所测部位,叶片夹持器测定使叶片处于密闭条件下,且所测面积相同,保证光谱数据获取的精确性和稳定性。取 10 片标准叶片在相同样点采集光谱数据,10 个光谱反射率的平均值作为该样点叶片的光谱反射率。

采集光谱数据后的茶鲜叶放于 120 °C 烘箱中杀青固样 30 min,然后置于 80~90 °C 下烘至足干。干燥后的样品磨碎后参照 GB5009.5—2016,通过凯氏定氮法测量其全氮量。

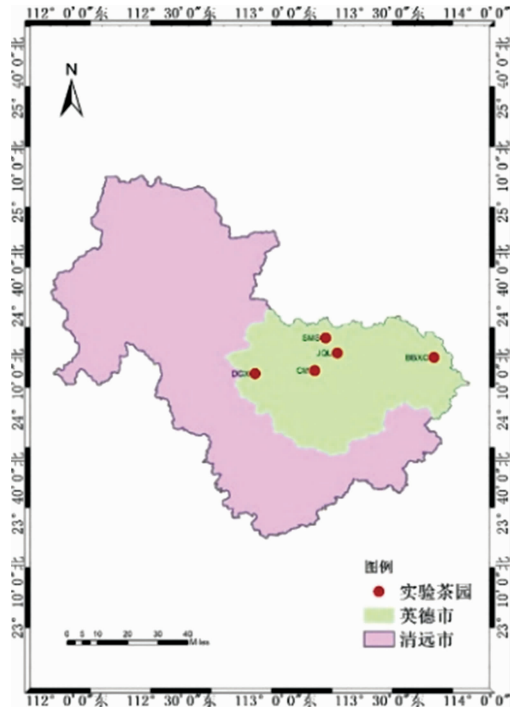


图 1 研究区域空间分布

Fig. 1 Spatial distribution of study area

1.3 数据分析

1.3.1 常规光谱预处理方法

采用卷积平滑(SG)、去趋势(Detrending)、一阶导数(1st)、多元散射校正(MSC)和标准正态变量变换(SNV)五种预处理方法,对原始光谱数据进行预处理作为参考。

1.3.2 小波变换

连续小波变换(CWT)起源于傅里叶变换信号处理技术,同时开展频率和时间分析,从复杂的原始光谱信号中分辨细微的变换,捕捉有益信息,表征光谱局部特征信息。小波变换分析广泛应用于高光谱数据的降维压缩、噪声去除、特征选择以及弱信号提取等方面。小波分析包括连续分解与离散分解两种类型的小波变换,Wang 等^[11]研究表明离散小波随着分解尺度的增加,光谱维度减半,不能对光谱数据进行较好的分析。连续小波变换是一种连续分解信号的方法,其输出光谱可以与原始光谱直接进行对比,比离散小波变换后获得的光谱更易于解释。通过以式(1)缩放和移动小波母函数

$$\Psi_{a,b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{\lambda-b}{a}\right) \quad (1)$$

式(1)中, λ 为茶叶叶片高光谱数据的波段数, a 和 b 分别代表伸缩因子(即宽度)和平移因子(即位置),这两个变量都为正实数。CWT 的输出公式如式(2)

$$W_f(a,b) = (f, \Psi_{a,b}) = \int_{-\infty}^{+\infty} f(\lambda) \Psi_{a,b}(\lambda) d\lambda \quad (2)$$

式(2)中, $f(\lambda)$ 为高光谱反射率,CWT 系数($W_f(a_i, b_j)$), $i=1,2,\dots,m$; $j=1,2,\dots,n$)包含两维(即 $m \times n$ 矩阵),其中一维为波长(1, 2, ..., n),另一维是分解尺度(1, 2, ..., m),因此,小波系数是一个行为尺度,列为波长的矩阵。

由于 Mexh 小波在高光谱分解方面表现出色^[12], 且光谱吸收特性与高斯函数相似, 因此选取 Mexh 作为小波母函数, 使用连续小波分析(continuous wavelet transformation, CWT)以执行尺度 $2^1 \sim 2^{10}$ 分解反射光谱为 10 个尺度, 并获取不同分解尺度下的不同波长的一系列小波系数, 进行了小波系数与茶叶 N 素含量的相关性分析, 并用于构建校准模型。

1.3.3 特征变量选择方法

连续投影(SPA)是一种前向变量选择方法, 使用简单的投影操作来获得具有最小共线性的变量子集, 并可以根据交叉验证的最小均方根误差(RMSE)选择最终的最佳波长。

竞争性自适应重加权采样(CARS)结合蒙特卡洛采样和 PLSR, 基于达尔文进化论, 遵循“适者生存”的原则, 根据每个变量的重要性, 以迭代和竞争的方式从 N 个蒙特卡洛采样顺序中选择 N 个变量子集。

变量组合集群分析(variable combination population analysis, VCPA)是一种新的有效的变量选择方法, 它由 Yun 等建立^[13]。VCPA 的实现过程分为两个基本步骤, 首先基于“适者生存”原理的指数递减函数(exponentially decreasing function, EDF)选择有用的变量并缩小变量空间, 其次通过二进制矩阵采样(binary matrix sampling, BMS)选择最佳光谱波长子集, 通过生成随机变量组合, 确保每个变量都有机会被选择并产生各种变量子集, 在每个 EDF 和 BMS 运行的周期中都采用模型集群分析(model population analysis, MPA)根据较低的 RMSECV 来确定最优变量子集。分别利用 SPA, CARS 和 VCPA 对连续小波变换处理后的光谱数据进行进一步的特征波段提取。

1.4 模型评价标准

经过连续小波变换后的光谱, 通过 SPA、CARS 和 VCPA 方法进行变量筛选, 利用偏最小二乘法(partial least squares regression, PLSR)构建茶鲜叶 N 含量的估测模型。采用 Kennard-Stone(KS)方法以近似 3:1 比例选取 113 个样本数据作为建模集, 用于构建估测模型, 剩余的 38 个样本数据作为验证集, 用于预测模型精度和稳定性的验证。模型精度采用模型评价参数决定系数(R^2)与均方根误差(RMSE)来综合评价模型表现, 其计算方法如式(3)和式(4)

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y'_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - Y'_i)^2}{n}} \quad (4)$$

式(3)和式(4)中, n 为样本个数, i 为样本, Y_i 和 Y'_i 分别是样品的实测与预测值, \bar{Y} 为实测样品平均值。

1.5 分析软件

使用 Matlab R2018b(The Math Works, Natick, USA)、Unscrambler X10.4(CAMO AS, Oslo, Norway)、SPSS 26 (IBM, USA)、Origin 2019b(OriginLab, USA)等软件。

2 结果与讨论

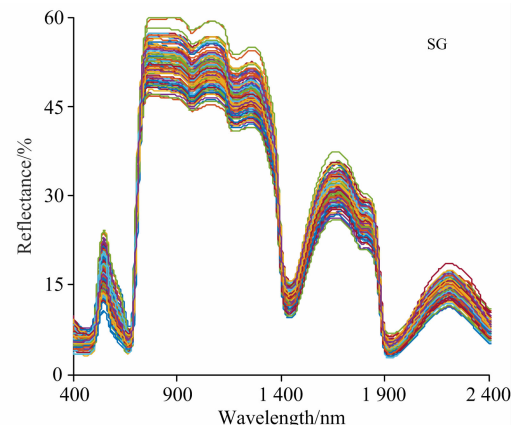
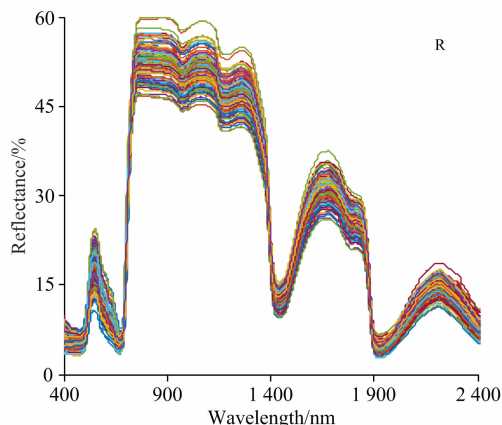
2.1 常规光谱预处理结果

图 2 展示了 5 种预处理后的高光谱。与原始光谱(R)相比, 预处理 1st 和 Detrending 方法明显改变了光谱特性, 在具体波段上凸显了光谱信息, 而 SG, MSC 和 SNV 预处理方法没有明显改变光谱吸收特性, 光谱特征变化幅度小, 与原始光谱相比更加平滑与集中。上述 5 种预处理算法有效地平滑光谱曲线, 且可以保留光谱信息的有用信息。

2.2 小波分析结果

2.2.1 基于小波变换的光谱分析

原始光谱经 CWT 处理后的结果如图 3 所示, D1—D10 分别代表了以 $2^1 \sim 2^{10}$ 为分解尺度经过 CWT 分解得到的 10 层小波能量系数模拟光谱分辨率。可以看出, 1—7 尺度上的光谱分辨率, 随着分解层数的增加, 光谱吸收、反射波谱特征均加强, 光谱波动范围加大, 8—10 尺度的光谱反射率起伏变化较大。总体来看, 随着 CWT 分解尺度的不断增加, 原始光谱反射率的波峰特征(761, 1 070, 1 217 和 1 645 nm)与波谷特征(670, 902, 1 447 和 1 931 nm)被不断放大, 并且均有一定的红移与蓝移现象。小尺度分解可以较好的保持原始光谱特性, 随着分解层数的增加, 光谱特征在不断凸显的过程中也不断被拉伸, 在大尺度分解中又可以凸显具体波段



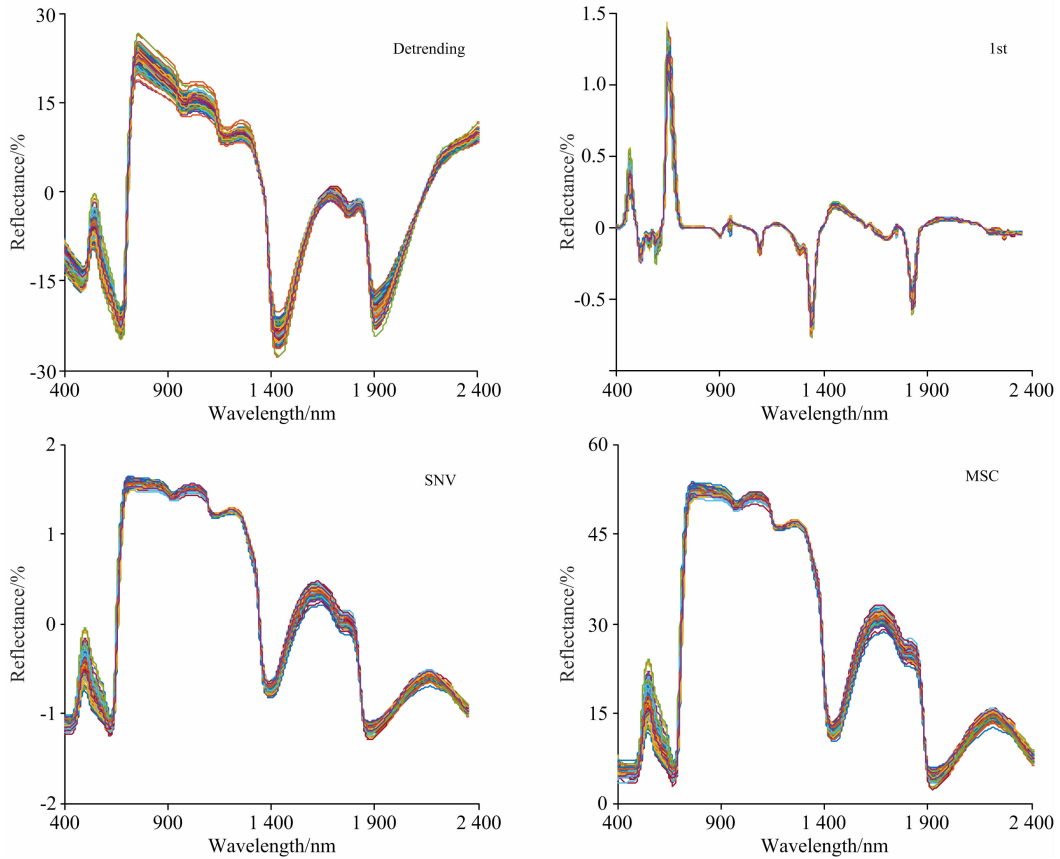


图 2 不同预处理方法下光谱反射率

Fig. 2 Reflectance spectra with different pre-processing methods

的细微特征信息。这是因为 CWT 是基于光谱曲线的形状特征，捕获光谱形状的差异并凸显特征^[14]，更有利于捕捉原始光谱反射率的微弱变化，有利于进行 N 含量对茶叶鲜光谱响应特征的分析。

异较大，且随着分解尺度的增加呈逐渐集中分布的趋势。由图可知，相关性较高的波段表现为更明显的红色区域，相关系数较大的波长主要与 N 素、叶绿素以及类胡萝卜素等物质的复杂性有关，对于选定的 493, 987, 1 193, 1 195 和 1 825 nm，这些波段大多为 N 素的重要光谱区域^[15]，另外可以看出，小波变换后的光谱在小尺度分解下与茶叶鲜 N 素存在良好的相关性，选取 1—6 尺度作进一步分析。

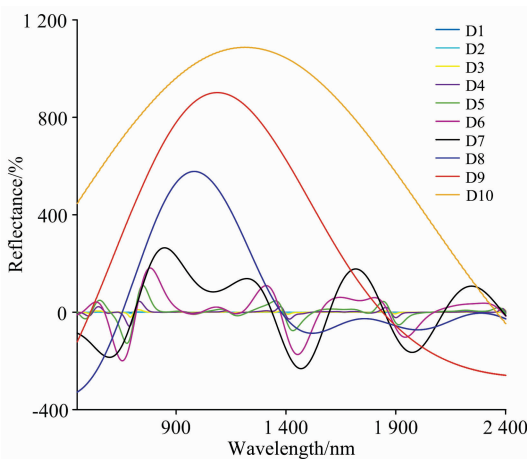


图 3 小波变换光谱特性

Fig. 3 Spectral characteristics of wavelet transform

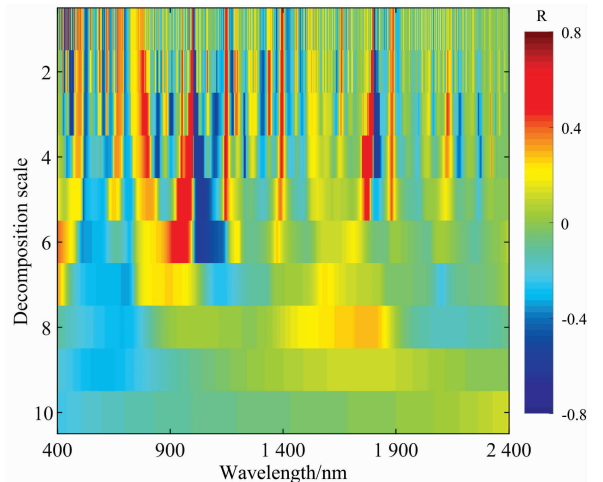


图 4 小波变换相关性分析

Fig. 4 Correlation analysis of wavelet transform

2.2.2 相关性分析

小波系数能量与 N 素实测数据进行相关性分析，如图 4 所示。可看出各尺度小波系数与茶叶鲜 N 含量的相关系数差

2.3 特征波长选择

为了比较不同变量筛选方法对茶鲜叶 N 含量高定量模型的影响，将经过 CWT 处理的光谱数据分别利用 SPA, CARS 和 VCPA 进行变量筛选。SPA 设置选择的最小和最大变量数的参数分别为 1 和 30，分别对 CWT 处理后的小波系数进行光谱特征变量的选取，结果如图 5(a)所示，光谱变量分别减少到 15, 16, 17, 11, 14 和 18 个。CARS 设置蒙特卡罗抽样运行次数为 50，十倍交叉验证用于评估每个子集的效用，分别对 CWT 处理后的小波系数进行光谱特征变量的选取。结果如图 5(b)所示，光谱变量分别减少到 16, 19, 17,

16, 15 和 17 个。VCPA 设置 EDF 剩余变量数为 100，迭代次数为 50，最优子模型的比例为 10%，BMS 运行次数为 1 000，VCPA 步骤中每个 BMS 采样的平均数设置为 20，并采用 5 折交叉验证分别对 CWT 处理后的小波系数进行光谱特征变量的选取，结果如图 5(c)所示。可以看出基于 CWT 小波分解后的光谱数据通过 VCPA 的五次迭代将全波段光谱变量分别减少到 12, 11, 10, 11, 8 和 8 个。相比 SPA 和 CARS, VCPA 选择的变量更少，这表明该方法在数据降维方面有着巨大的优势。

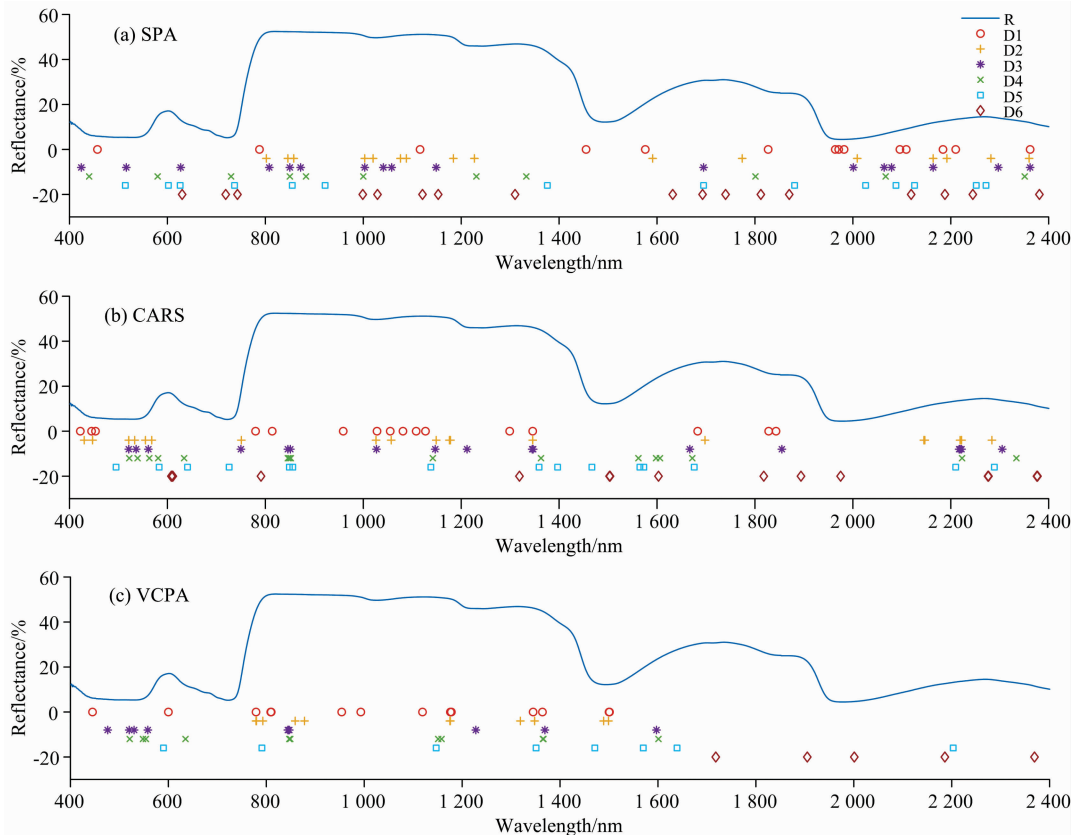


图 5 SPA, CARS 和 VCPA 筛选的特征变量

Fig. 5 The variables selected by SPA, CARS and VCPA

2.4 不同预处理方法结合特征波段筛选的 N 素含量预测

2.4.1 基于常规预处理方法和特征筛选的 N 素含量预测

采用卷积平滑 (SG)、去趋势 (Detrending)、一阶导数 (1st)、多元散射校正 (MSC) 和标准正态变量变换 (SNV) 5 种传统光谱处理技术分析茶鲜叶光谱特性，并采用偏最小二乘法建立茶鲜叶 N 素含量预测模型作为参考，结果如表 1 所示。

2.4.2 基于小波变换和特征筛选的 N 素含量预测

使用 CWT 处理后的光谱数据筛选后的特征波段作为自变量，以茶鲜叶 N 含量为因变量，采用偏最小二乘法构建茶鲜叶 N 含量的估测模型 (表 2)。由表 2 可知，经过 CWT 处理后光谱构建的模型均表现出较好的拟合精度，并且随着分解尺度的增加模型精度呈现下降趋势，说明小规模分解通常

能够改善 CWT 处理后的小波系数对分析物的敏感性，这与 Liu 等的研究结果一致^[16]。其中 SPA 选择变量估测茶鲜叶 N 含量的精度较低，1—6 尺度建模 R^2 范围在 0.65~0.83，RMSE 范围在 0.25~0.33，预测 R^2 范围在 0.65~0.83，RMSE 范围在 0.25~0.33；以 CARS 选择变量为因变量，1—6 尺度建模 R^2 范围在 0.73~0.91，RMSE 范围在 0.19~0.32，预测 R^2 范围在 0.64~0.88，RMSE 范围在 0.24~0.35；VCPA 选择变量估测模型表现最好，1—6 尺度建模 R^2 范围在 0.73~0.95，RMSE 范围在 0.16~0.29，预测 R^2 范围在 0.73~0.90，RMSE 范围在 0.23~0.35，其中在 1 尺度下的模型拟合效果最高。

图 6 所示为不同变量选择方法下最优估测模型实验值与预测值 1:1 对比图，可以看出 CWT-VCPA 模型表现最优，

表 1 基于不同预处理方法和变量选择方法的模型

Table 1 Models based on different preprocessing methods and variables selection methods

模型	光谱处理	主成分	建模集		预测集	
			R^2_C	RMSE _C	R^2_P	RMSE _P
SPA-PLSR	SG	8	0.50	0.42	0.37	0.48
	Detrending	9	0.59	0.35	0.58	0.38
	1st	7	0.63	0.37	0.61	0.35
	SNV	10	0.71	0.31	0.60	0.39
	MSC	8	0.62	0.38	0.68	0.29
CARS-PLSR	SG	5	0.62	0.35	0.55	0.41
	Detrending	10	0.75	0.30	0.71	0.29
	1st	11	0.79	0.26	0.73	0.32
	SNV	11	0.71	0.33	0.68	0.29
	MSC	11	0.73	0.29	0.62	0.39
VCPA-PLSR	SG	6	0.81	0.24	0.65	0.37
	Detrending	10	0.83	0.25	0.73	0.31
	1st	9	0.86	0.18	0.84	0.25
	SNV	7	0.81	0.25	0.67	0.35
	MSC	6	0.82	0.24	0.72	0.32

训练集和验证集的 R^2 分别为 0.95 和 0.90, RMSE 分别为 0.16 和 0.23, 二者差异较小, 相比于 CWT-SPA 和 CWT-CARS, 建模及预测集的预测精度最高。对比各模型可知, CWT 方法结合特征变量选择算法明显提升了模型的精度,

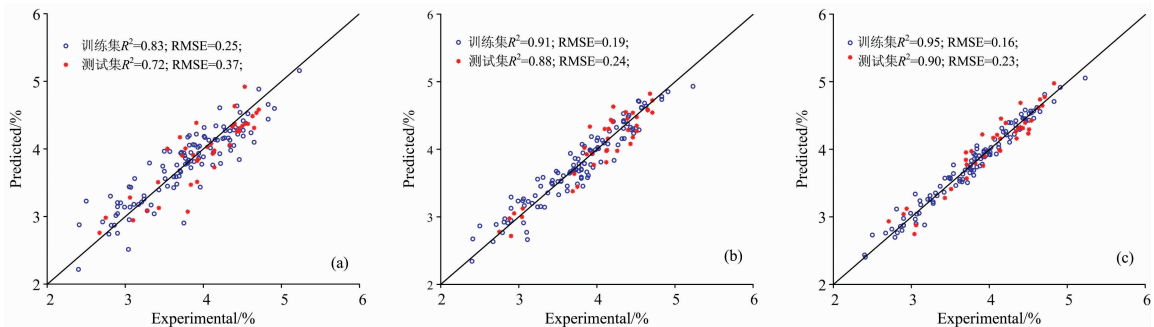


图 6 基于 (a)CWT(1 尺度)-SPA-PLSR, (b)CWT(1 尺度)-CARS-PLSR 和 (c)CWT(1 尺度)-VCPA-PLSR 的氮含量最佳模型的拟合 1:1 拟合图

Fig. 6 Fitting diagrams of the optimal models of tea nitrogen based on (a) CWT (1 scale)-SPA-PLS, (b) CWT (1 scale)-CARS-PLS and (c) CWT (1 scale)-VCPA-PLS

以茶鲜叶为研究对象, 构建并对比分析了常用 5 种光谱处理与连续小波分解方法的 N 含量估测模型。研究结果表明, 与常规光谱处理方法相比, 小波变换方法可明显提升茶鲜叶光谱对 N 含量的响应, 且模型精度与稳定性均较高。这是由于在常规的光谱预处理方法中, SNV, MSC 和 Detrending 通常用于校正低频散射信号、1st 可以消除信号的基线漂移, 而 CWT 可以同时消除光谱信号中的高频和低频噪声, 并且连续小波分解本质上是一种基于相关性的方法, 可以捕获对茶多酚含量最敏感的小波特征, 这种频谱相关性的方法在凸显光谱细微变化特征方面效果显著, 更有利于后续对光谱特征波段的提取。随着分解尺度的增加, 模型预测精度逐步降低, 主要原因是光谱中含有较多的噪声, 造成了光谱的

且 CWT-VCPA 筛选变量估算 N 含量效果最好。

表 2 基于不同分解层数和变量选择方法的模型

Table 2 Models based on different decomposition scales and variables selection methods

模型	分解尺度	主成分	建模集		预测集	
			R^2_C	RMSE _C	R^2_P	RMSE _P
SPA-PLSR	1	5	0.83	0.25	0.72	0.37
	2	7	0.72	0.30	0.68	0.36
	3	10	0.74	0.31	0.69	0.33
	4	11	0.78	0.28	0.67	0.35
	5	11	0.79	0.26	0.60	0.41
	6	11	0.65	0.33	0.60	0.42
CARS-PLSR	1	7	0.91	0.19	0.88	0.24
	2	8	0.81	0.24	0.76	0.32
	3	12	0.81	0.25	0.73	0.31
	4	8	0.82	0.26	0.69	0.29
	5	10	0.78	0.28	0.64	0.35
	6	11	0.73	0.32	0.68	0.32
VCPA-PLSR	1	7	0.95	0.16	0.90	0.23
	2	7	0.83	0.23	0.82	0.27
	3	10	0.81	0.26	0.79	0.26
	4	8	0.82	0.24	0.73	0.30
	5	8	0.81	0.25	0.77	0.30
	6	7	0.73	0.29	0.74	0.35

有益信息与噪声的分布波段区间有较高的重叠性, 且随着 CWT 分解尺度的增加, 其内含的有益信息与无益噪声信息重新分配, 导致引入的噪声信息较多从而降低模型的预测精度。在 1—6 尺度模型精度达到最低时相较于传统光谱变换仍然有一定程度的提升, 表明 CWT 能提取光谱特征中的有益信息, 并一定程度上分离无用噪声, 提升光谱信息信噪比。基于三种不同波段筛选方法 (SPA, CARS 和 VCPA) 都可以在减少变量的同时提升模型预测效果, 且相比于 SPA 和 CARS, VCPA 选择的变量更少, 预测效果更好。这归因于 EDF 对变量的连续收缩, 在 EDF 的每次运行中, 都会消除一些贡献较小的变量。此外, 还与 BMS 和 MPA 的使用有关, BMS 是一种采样策略, 它为所有变量提供了相同的机会

进行采样,使用不同的变量组合用于子模型的构建。MPA 用于提取子模型中包含的最佳变量组合,比较每个子模型的 RMSECV,保留最佳 10%子模型中具有最高频率的变量^[17]。因此,VCPA 通过最少的选择变量,保证了光谱数据集更好的可预测性。当将最佳波长与全光谱进行比较时,基于 CWT 分解光谱结合 VCPA 筛选特征变量的 PLSR 模型比其他光谱变换的 PLSR 模型具有更高的性能($R^2=0.95$),并且减少了全光谱数据变量的 99.34%,极大的简化了建模过程。综上可知,CWT-VCPA 方法在实现对茶鲜叶 N 含量的定量监测方面具有强大的潜力。

3 结 论

采用传统光谱处理技术和连续小波技术对光谱分别开展处理和分析,并通过 SPA, CARS 和 VCPA 方法进一步筛选特征波段,建立预测茶鲜叶 N 素含量的模型。旨在研究基于连续小波技术结合特征变量选择方法预测茶鲜叶 N 素含量

的最优模型,结论如下:

(1)相比于传统光谱处理方法,经过 CWT 分解光谱结合变量选择方法构建的茶鲜叶 N 含量估算模型均表现出令人满意的结果,且随着分解层数的增加,模型的精度整体有逐步下降的规律。表明 CWT 处理可有效提升光谱对 N 的敏感性,且小规模分解通常能够改善这种关系,提高模型性能。

(2)以 SPA, CARS 和 VCPA 筛选出的特征变量为模型输入因子均在 CWT 分解尺度为 1 时达到最佳估算效果,CWT-SPA-PLSR 建模 R^2 为 0.83, RMSE 为 0.25; CWT-CARS-PLSR 建模 R^2 为 0.91, RMSE 为 0.19; CWT-VCPA-PLSR 建模 R^2 为 0.95, RMSE 为 0.16。

(3)经过 1 尺度 CWT 分解结合 VCPA 所构建的模型表现最佳($R_C^2=0.95$, $RMSE_C=0.16$; $R_P^2=0.90$, $RMSE_P=0.23$),与传统光谱处理方法相比,变量数减少了 99.34%, R_C^2 提升了 11%, R_P^2 提升了 7%。表明 CWT-VCPA 方法可以筛除高光谱中大量无用或多余的信息,从而可以有效地降低定量检测模型的复杂度,实现对茶鲜叶 N 含量的定量监测。

References

- [1] WANG Fan, ZHAO Chun-jiang, XU Bo, et al(王 凡,赵春江,徐 波,等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2020, 36(24): 273.
- [2] Chen Q S, Chen M, Liu Y, et al. Journal of Food Science and Technology, 2018, 55: 4363.
- [3] Yamashita H, Sonobe R, Hirono Y, et al. Scientific Reports, 2020, (10): 17360.
- [4] Hazarika A K, Chanda S, Sabhapondit S, et al. Journal of Food Science and Technology, 2018, 55: 4867.
- [5] Thiele-Bruhn S, Emmerling C, Harbich M, et al. Journal of Near Infrared Spectroscopy, 2016, 24(3): 255.
- [6] Sun J, Zhou X, Hu Y, et al. Computers and Electronics in Agriculture, 2019, 160: 153.
- [7] Hui C, Zan L, Chao T, et al. Vibrational Spectroscopy, 2018, 99: 178.
- [8] Li H, Zhu J, Jiao T, et al. Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy, 2020, 243: 118765.
- [9] Al-Kaf H A G, Alduais N A M, Saad A M H Y, et al. IEEE Access, 2020, 8: 1.
- [10] Wang Y J, Jin S S, Li M H, et al. Computers and Electronics in Agriculture, 2020, 175: 105538.
- [11] Wang G, Fang Q, Teng Y, et al. International Journal of Applied Earth Observation and Geoinformation, 2016, 53: 48.
- [12] Thiele-Bruhn S, Emmerling C, Harbich M, et al. Journal of Near Infrared Spectroscopy, 2016, 24(3): 255.
- [13] Yun Y H, Wang W T, Deng B C, et al. Analytica Chimica Acta, 2015, 862: 14.
- [14] Qi X, Jiang J, Cui X, et al. Food Analytical Methods, 2019, 13(2): 1.
- [15] XUE Li-hong, CAO Wei-xing, LUO Wei-hong, et al(薛利红,曹卫星,罗卫红,等). Chinese Journal of Plant Ecology(植物生态学报), 2004, (2): 172.
- [16] Liu W W, Li M J, Zhang M Y, et al. Ecosystem Health and Sustainability, 2020, 6(1): 1726211.
- [17] AN Si-yu, ZHANG Lei, SHANG Xian-zhao, et al(安思宇,张 磊,尚献召,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2021, 41(1): 206.

Estimation of Total Nitrogen Content in Fresh Tea Leaves Based on Wavelet Analysis

WANG Fan^{1,2}, CHEN Long-yue^{2,3}, DUAN Dan-dan^{1,2,4*}, CAO Qiong^{1,4}, ZHAO Yu¹, LAN Wan-rong⁵

1. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

2. Qingyuan Academy of Smart Agriculture, Qingyuan 511500, China

3. Nongxin Technology (Guangzhou) limited Liability Company, Guangzhou 510000, China

4. Hunan Agricultural University, Changsha 410125, China

5. Jiangmen Agricultural Technology Service Center, Jiangmen 529000, China

Abstract Tea is one of the most popular beverages globally, which is greatly affected by the content of nitrogen (N) in quality. Due to the complicated and time-consuming method for determining N content in fresh tea leaves by traditional chemical analysis, this paper proposes a means for N content prediction by hyperspectral technique. The wavelet coefficients extracted by continuous wavelet transform (CWT) technology are used to estimate N content by different decomposition layers of CWT. Moreover, the predictive effects of models built by different wavelength selection algorithms are also discussed. Several 151 hyperspectral data of tea samples were collected from tea gardens in the Yingde City of Guangdong Province. The original spectra data are preprocessed by smoothing (SG), detrending (Detrending), first derivative (1st), multiple scattering correction (MSC), and standard normal variable transformation (SNV) while comparing with CWT. Then, continuous wavelet multi-scale analysis is applied to process the original spectrum for generating wavelet coefficients, and Pearson correlation analysis was also performed. Next, three kinds of methods, including successive projections algorithm (SPA), competitive adaptive weighted sampling (CARS) and variable combination population analysis (VCPA), are adopted to optimize the variable space of the spectral data after CWT transformation. At last, quantitative models of N content prediction are established and compared by PLSR with characteristic variables selected by the three above mentioned methods as input. The overall results show that the continuous wavelet analysis algorithm can improve the model's efficiency for estimating the N content of the fresh tea leaves by hyperspectral data. Furthermore, it has better performance than other conventional spectral preprocessing methods significantly. With continuous wavelet decomposition, the precision of the model for N content prediction gradually decreases with the increase of the decomposition scale. There is a good correlation between the spectrum after the continuous wavelet transforms on the scale of 1~6 and the N in fresh tea leaves, which shows that the small-scale continuous wavelet algorithm can be well applied to monitor N content in fresh tea leaves. The model established by CWT (1scale)-VCPA method has the best performance, and the number of variables is reduced by 99.34% compared to the full band. The R^2 of the calibration model and prediction model respectively, are 0.95 and 0.90. Compared with the traditional spectral processing method, the accuracy is improved by 11%. It is proved that the combination of CWT-VCPA can obviously reduce the spectral dimension and improve the accuracy of the model. This research achieves an efficient way for N content prediction of tea, which provides a technical basis and reliable reference for other components evaluation of tea.

Keywords Fresh tea leaves; Nitrogen; Continuous wavelet transform; Hyperspectral; Variable combination population analysis

(Received Jun. 22, 2021; accepted Jan. 17, 2022)

* Corresponding author