

基于 PLS 算法的生物质秸秆元素分析 NIRS 快速检测

李晓金 朱 凯 牛智有 程旭云

华中农业大学工学院, 武汉 430070

摘要 为探讨利用近红外光谱技术快速检测生物质秸秆中 N、C、H、S 和 O 元素的可行性, 采集并制备水稻、小麦、油菜和玉米秸秆样本 199 个, 采用近红外光谱(NIRS) 分析技术, 结合偏最小二乘(PLS)化学计量学算法, 在 $7\ 400\sim 5\ 550\ \text{cm}^{-1}$ 波段范围内, 比较不同光谱预处理方法的定标效果, 建立最优的生物质秸秆中 N、C、H、S 和 O 元素的定量分析模型, 并用独立的验证集样本对模型进行验证。验证结果表明: 所建立的 N 元素的定量分析模型可用于实际检测; O 元素的定量分析模型可进行实际估测; 采用近红外技术用于 C 元素定量分析是可行的, 但模型需要进一步优化; H、S 元素采用 NIRS 技术无法进行定量分析。

关键词 近红外光谱; 偏最小二乘算法; 生物质秸秆; 元素分析

中图分类号 S 216.2 **文献标识码** A **文章编号** 1000-2421(2015)02-0131-05

生物质秸秆是一种重要的可再生能源, 其元素的组成对燃烧和热解特性有直接关系。农作物秸秆是生物质的一个重要组成部分, 是当今世界上仅次于煤炭、石油和天然气的第四大能源, 在能源系统中占有重要地位^[1]。中国农业生产以种植为主, 秸秆资源丰富。农作物秸秆资源化利用有助于缓解能源短缺, 减轻化石燃料燃烧造成的环境污染和温室效应。成型燃料、直燃发电、秸秆气化和热裂解制生物油是秸秆资源化利用的重要途径。由于生物质组成元素的性质及其含量与燃烧性能密切相关, 因此, 实现生物质秸秆基本元素组成含量的快速检测, 对研究生物质秸秆资源化利用及其燃烧性能意义重大。

近红外光谱主要反映样品中有机分子含氢基团的特征信息, 是一种快速、高效、无损的检测方法, 已广泛应用于农业、食品、医药和石化等领域, 在生物质秸秆检测方面也有大量研究^[2-5]。相关研究结果表明, 近红外光谱技术可用于农产品化学元素的快速预测^[6-7]。笔者利用近红外光谱(near-infrared spectroscopy, NIRS)分析技术, 运用偏最小二乘(partial least squares, PLS)化学计量算法, 建立生物质秸秆元素组成含量的定量分析模型, 旨在探讨利用近红外分析技术测定生物质秸秆中化学元素组成的可行性。

1 材料与方法

1.1 样品收集与制备

分别从湖北、湖南、四川、重庆、贵州和云南等地区采集秸秆样品共 199 份, 其中水稻 100 份、小麦 39 份、油菜 39 份和玉米 21 份, 根据作物种植面积确定样品采集数量, 涵盖不同区域、不同气候和不同品种。将采集的秸秆样品置于室外摊开晾晒至干燥, 用粉碎机对秸秆进行粉碎后置于 $45\ ^\circ\text{C}(\pm 5\ ^\circ\text{C})$ 烘箱中烘约 8 h, 水分控制在 $5.5\%(\pm 1\%)$, 然后用锤式旋风磨进行研磨, 使其全部通过孔径 0.35 mm 筛。将研磨好的样品装入封口袋后贴上标签并编号, 置于室温(约 $25\ ^\circ\text{C}$)下保存, 供光谱扫描和 N、C、H、O、S 元素含量的测定。

1.2 近红外光谱采集

测试仪器为 Nicolet 公司(美国)生产的 Nicolet Antaris II 傅立叶变换近红外(FT-NIR)光谱仪, 钨卤灯光源, InGaAs 检测器, 石英样品池, 采用积分球固体采样模块采集光谱。仪器工作参数: 谱区范围 $10\ 000\sim 4\ 000\ \text{cm}^{-1}$, 分辨率 $8\ \text{cm}^{-1}$, 扫描 32 次, 背景扫描 1 次/h, 数据以 $\text{Log}(1/R)$ 形式存储, 在室温(约 $25\ ^\circ\text{C}$)下进行扫描。每个样本装填 3 次, 分别扫描后取平均光谱作为样本的原始光谱。

收稿日期: 2014-03-19

基金项目: 国家公益性行业(农业)科研专项(201003063-04)

李晓金, 硕士研究生。研究方向: 农产品加工技术与装备。E-mail: lixiaojin@webmail.hzau.edu.cn

通信作者: 牛智有, 教授。研究方向: 农产品加工技术与装备。E-mail: nzhy@mail.hzau.edu.cn

1.3 元素含量测定

样品的 N、C、H、S、O 元素含量(质量分数,%)按照美国试验材料学会协会(ASTM)标准测定^[8]。采用的仪器为 EA 3000 型元素分析仪(意大利 EuroVector 公司)。每个样品处理 3 个重复,取其平均值为样品元素的最终含量。

1.4 NIRS 定量分析模型的建立与分析

偏最小二乘(PLS)方法是多元数据处理的经典方法之一,它在考虑全谱区各波长点的光谱参数的同时,还兼顾到被分析样品内部各成分之间的关系。采用 TQ Analyst 8.3 软件中的 PLS 化学计量算法建立秸秆元素分析组成的定量分析模型。

使用漫反射方法采集粉状样品光谱时,由于受样品颗粒尺寸、均匀性等影响,光程无法保持恒定,为了消除这些因素的干扰,采用多元散射校正(MSC)、变量标准化(SNV)进行光程校正,然后结合 S-G 平滑、一阶导数、二阶导数方法及其组合对原始光谱进行预处理。

为了考察定量分析模型的可行性,用独立的验证集进行外部验证。校正集和验证集的选择依据样品化学元素含量的分布和光谱数据主成分得分空间分布。比较模型的统计参数包括校正集决定系数(R_c^2)、均方根误差(RMSEC)、交互验证决定系数(R_{cv}^2)和交互验证均方根误差(RMSECV)。验证集的统计参数包括决定系数(R_p^2)、均方根误差(RMSEP)、平均偏差(Bias)(验证集实测值与预测值之差的平均值)和相对分析误差(RPD)。Bias 是模型系统误差的体现,其值越小说明模型精度越高^[9]; RPD 大于 2.5 时表明模型可以用于定量分析,大于 2 小于 2.5 时表明模型可以进行定量分析,但模型需要进一步优化^[10]。

2 结果与分析

2.1 元素测定结果及样品集划分

样品中 N、C、H、S 和 O 含量(质量分数)的测试范围、均值、标准差和变异系数统计结果见表 1。

表 1 元素分析组成统计结果

Table 1 Statistical result of elemental analysis components %

元素 Elements	范围 Range	均值 Mean	标准差 SD	变异系数 CV
N	0.30~2.79	1.18	0.60	4.98
C	40.21~49.53	43.96	2.18	4.96
H	4.45~7.95	6.16	0.75	12.18
S	0.11~0.34	0.21	0.05	23.81
O	32.79~40.69	37.20	1.83	4.92

由表 1 可知,样品中元素 N、C 和 O 的变化范围较广,元素 H 和 S 变异系数较大,说明样品来源和状态范围广,基本可覆盖秸秆样品 N、C、H、S 和 O 元素含量的可能范围,其值在其他文献报道的范围内^[11-12],具有很好的代表性。结合浓度梯度法和主成分分析方法进行样品分集,选取 149 个样品为校正集,50 个样品为验证集。

2.2 样品近红外光谱

对 199 个秸秆样品进行光谱扫描,得到近红外漫反射光谱如图 1 所示。

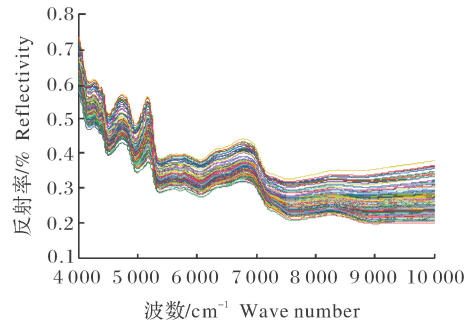


图 1 样品近红外反射光谱

Fig. 1 NIR spectra of samples

由图 1 可知,在全波长范围内,光谱存在多个吸收峰,可以反映样品的组成信息,为样品的 N、C、H、S 和 O 的定量分析提供了丰富的信息。在 7 400~5 550 cm^{-1} 波段范围内,光谱存在明显吸收峰,在此波段范围内主要反映了秸秆中含氢组分的信息^[14],能够筛选出反映秸秆样品物化特性的有效光谱信息,因此,选取光谱波段为 7 400~5 550 cm^{-1} 建立预测模型。

2.3 建模结果

采用多元散射校正(MSC)、变量标准化(SNV)进行光程校正,结合 SG 平滑、一阶导数、二阶导数方法及其组合对原始光谱进行预处理,通过 PLS 化学计量算法,分别建立秸秆样品的 N、C、H、S 和 O 的 NIRS 定量分析模型。通过综合比较分析决定系数(R_c^2)、均方根误差(RMSEC)、交互验证决定系数(R_{cv}^2)和交互验证均方根误差(RMSECV)值的大小,确定最优的 N、C、H、S、O 定量分析模型。校正集最优定标结果见表 2。

由表 2 可知,对 N 元素而言,采用变量标准化 SNV、一阶导数和 S-G 平滑处理组合后,建模的主因子数为 7,模型的决定系数(R_c^2)为 0.949,均方根误差(RMSEC)为 0.131%,交互验证决定系数(R_{cv}^2)为 0.863,交互验证均方根误差(RMSECV)

表 2 元素组成校正集定标模型

Table 2 The calibration models of calibration set of elements

元素 Elements	光谱预处理 Pretreatment	主因子数 Factors	决定系数 R_c^2	均方根误差 RMSEC/%	交互验证决定系数 R_{cv}^2	交互验证均方根误差 RMSECV/%
N	SNV+一阶导数+S-G 平滑 SNV+ First derivative+ S-G smooth	7	0.949	0.131	0.863	0.215
C	MSC+二阶导数+S-G 平滑 MSC+ Second derivative+ S-G smooth	8	0.878	0.684	0.803	0.872
H	MSC+二阶导数+S-G 平滑 MSC+ Second derivative+ S-G smooth	8	0.616	0.439	0.490	0.509
S	SNV+一阶导数+S-G 平滑 SNV+ First derivative+ S-G smooth	8	0.601	0.031	0.407	0.039
O	SNV+一阶导数+S-G 平滑 SNV+ First derivative+ S-G smooth	10	0.890	0.677	0.792	1.053

为 0.215%，说明模型的定标结果很好，预测值接近实测值；对 C、O 元素而言，主因子数分别为 8 和 10，决定系数(R_c^2)分别为 0.878 和 0.890，均方根误差(RMSEC)分别为 0.684% 和 0.677%，交互验证决定系数(R_{cv}^2)分别为 0.803 和 0.792，交互验证均方根误差 (RMSECV) 分别为 0.872% 和 1.053%，说明模型的定标效果结果较好，预测值比较接近实测值；对 H、S 元素而言，其决定系数(R_c^2)仅为 0.6 左右，均方根误差(RMSEC)分别为 0.439% 和 0.031%，交互验证决定系数(R_{cv}^2)不及 0.5，交互验证均方根误差 (RMSECV) 分别为 0.509% 和 0.039%，说明模型的定标效果很差，预测值与实测值有很大误差。

2.4 模型验证

采用独立的验证集 50 个样品对所建基于 PLS 化学计量算法的 NIRS 定量分析模型的性能进行验证，考察模型对 N、C、H、S 和 O 元素含量的预测精度。以验证集决定系数(R_p^2)、均方根误差 (RMSEP)、平均偏差(Bias)和相对分析误差(RPD)为评价指标，其验证评价指标统计结果见表 3。

表 3 样品 N、C、H、S、O 元素 NIRS 定标模型的验证指标统计结果

Table 3 The verification indexes statistics of NIRS calibration model for elements of N, C, H, S and O

元素 Elements	决定系数 R_p^2	均方根误差/% RMSEP	相对分析误差 RPD	平均偏差/% Bias
N	0.895	0.186	3.125	0.010
C	0.781	1.090	1.497	0.141
H	0.440	0.487	1.252	0.110
S	0.385	0.087	1.288	0.002
O	0.810	0.923	2.185	0.273

从表 3 可知，N 元素模型的决定系数(R_p^2)为 0.895，均方根误差(RMSEP)为 0.186%，Bias 仅为 0.01%，RPD 为 3.125，RPD 大于 2.5，说明模型的预测值接近实测值，所建立的定量分析模型可用于定量分析；O 元素模型的决定系数(R_p^2)为 0.810，均方根误差(RMSEP)为 0.923%，Bias 为 0.273%，RPD 为 2.185，RPD 大于 2 而小于 2.5，模型的预测值与实测值有一定的误差，说明模型 NIRS 技术用于 O 元素的定量分析是可行的，但所建立的定量分析模型只能用于 O 元素的估测，模型精度需进一步提高；C 元素模型的决定系数(R_p^2)为 0.781，均方根误差(RMSEP)为 1.090%，Bias 为 0.141%，RPD 为 1.497，模型的预测值与实测值有较大的误差，说明 NIRS 技术用于 C 元素的定量分析具有一定潜力，但需要对模型精度进一步优化；对 H、S 元素而言，模型的决定系数(R_p^2)分别为 0.440 和 0.385，RPD 均很小，所建立的 H、S 定量分析模型的预测值与实测值有很大误差。

为进一步说明 N、C、H、S 和 O 元素 NIRS 定量分析模型的精度，分别将模型预测值与实测值进行回归分析，得到样品 N、C、H、S、O 的实测值与 NIRS 预测值关系如图 2 所示。

从图 2 可知，N 元素模型预测值散点在回归线附近的集中程度较高，其预测值与实测值之间回归方程的斜率(slope)为 0.89，说明预测值与实测值有较好的拟合精度；C 元素模型的预测值与实测值之间回归方程的斜率为 0.93，接近于 1，说明预测值与实测值有很好的拟合精度；O 元素模型的预测值与实测值之间回归方程的斜率为 0.88，说明预测值与实测值有较好的拟合精度；H、S 元素模型的预测

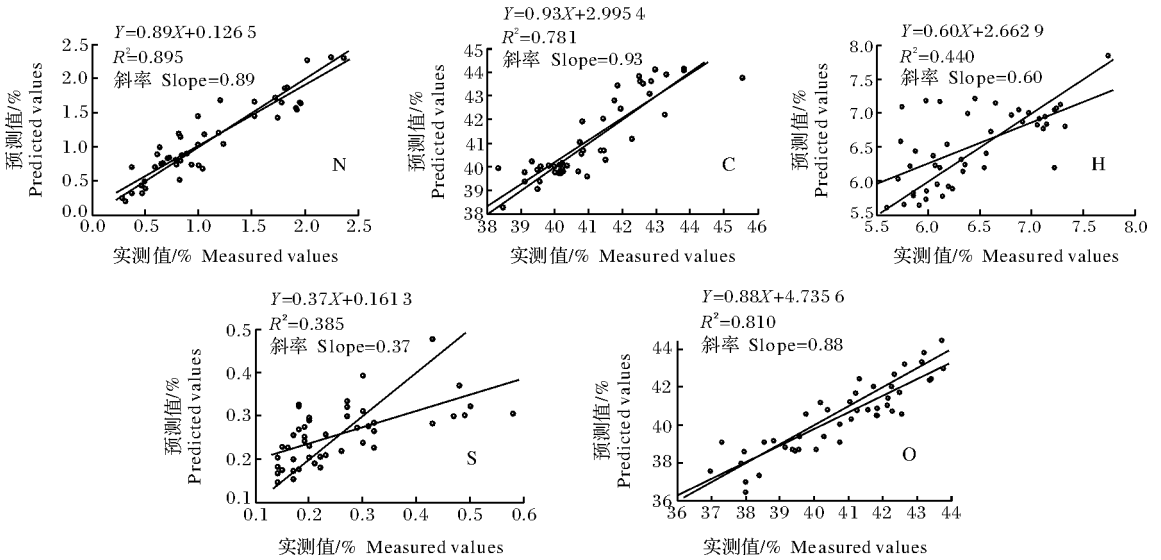


图 2 秸秆样品 N、C、H、S、O 的实测值与 NIRS 预测值的关系

Fig. 2 Correlations of measured and NIRS predicted values for N,C,H,S,O of straw samples

值散点在回归线附近的离散程度很高,预测值与实测值之间回归方程的斜率分别为 0.60 和 0.37,远远小于 1,说明预测值与实测值的拟合精度很差。

3 讨论

应用近红外光谱(NIRS)技术,结合偏最小二乘(PLS)化学计量算法,建立生物质秸秆样品元素分析组成的定量分析模型,探讨利用近红外光谱技术定量分析生物质秸秆中 N、C、H、S 和 O 元素的可行性。在以往的研究中,尚未见对多种秸秆的元素进行同步快速检测的报道。

本试验通过 PLS 化学计量算法,分别建立了秸秆样品的 N、C、H、S 和 O 的 NIRS 定量分析模型,并比较模型的性能参数,获得最优的定标模型。结果表明: N 元素定标模型的定标结果很好,预测值接近实测值; C、O 元素模型的定标效果结果较好,预测值比较接近实测值; H、S 元素模型的定标效果很差,预测值与实测值有很大误差。

本试验采用独立验证集样本对所建立 NIRS 定量分析模型的预测精度进行了验证。结果表明: N 元素定量分析模型的预测值接近实测值; O 元素定量分析模型的预测值与实测值有一定的误差,说明 NIRS 技术用于 O 元素的定量分析是可行的,所建立的定量分析模型可用于 O 元素的估测,模型精度需进一步提高; C 元素定量分析模型的预测值与实测值有一定的误差,需要对模型精度进一步优化;

H、S 元素定量分析模型的预测值与实测值有很大误差。

根据对秸秆元素 N、C、H、S 和 O 定标模型进行整体评价,结果表明 N 元素的定标结果最理想,其次为 O 元素和 C 元素,但 C 元素的预测精度有待提高,而 H、S 元素的预测效果均不太理想,不能进行 NIRS 定量分析。

模型预测精度不高的主要原因:首先,不同元素在秸秆中的存在形态不同,即是否均为化合物形态,如在秸秆样品中 S 元素由于含量低,其测定精度难于保证;其次,秸秆样品在保存过程受到外界环境影响,如湿度、温度等,由于秸秆粉体在环境中容易产生吸湿或解吸,最终造成秸秆元素分析组成发生一定的变化,导致对测试结果有影响,特别是对 H 元素的影响较大。

另外,将水稻、小麦、油菜和玉米 4 种秸秆一起作为试验样品,对其整体进行建模分析,由于 4 种秸秆的理化性质有较大差异,获取的近红外光谱信息也有差异。这些因素均会对秸秆元素分析组成的定标模型产生一定的影响,导致 NIRS 定标模型预测精度受到影响。因此,在规范试验过程、确保测试仪器稳定和有效控制实验室环境的条件下,如何有效地利用 4 种秸秆的光谱信息,分别建立水稻、小麦、油菜和玉米秸秆 NIRS 定标模型值得进一步研究探讨,同时可以尝试采用其他化学计量算法建立定量分析模型。

参 考 文 献

- [1] 娄玥芸, 张惠芳. 秸秆生物质能源的应用现状与前景[J]. 化学与生物工程, 2010, 27(9): 73-75.
- [2] 谢丽娟, 应义斌, 于海燕, 等. 近红外光谱分析技术在蔬菜品质无损检测中的应用研究进展[J]. 光谱学与光谱分析, 2007, 27(6): 1131-1135.
- [3] CHUNG H. Applications of near-infrared spectroscopy in refineries and important issues to address [J]. Applied Spectroscopy Reviews, 2007, 42(3): 251-285.
- [4] 皇才进, 韩鲁佳, 刘贤, 等. 基于近红外光谱技术的秸秆工业分析[J]. 光谱学与光谱分析, 2009, 29(4): 960-963.
- [5] 皇才进, 韩鲁佳, 刘贤, 等. 基于 LOCAL 算法的秸秆水分和热值的近红外光谱快速分析[J]. 红外与毫米波学报, 2009, 28(3): 184-187.
- [6] MORON A, COZZOLINO D. Determination of macro elements in alfalfa and white clover by near-infrared reflectance spectroscopy [J]. Journal of Agricultural Science, 2002, 139(4): 413-423.
- [7] COZZOLINO D, MORON A. Exploring the of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes [J]. Animal Feed Science and Technology, 2004, 111: 161-173.
- [8] ASTM. Standard test method for ash in biomass: U S, E1755-01 [S]. 2007.
- [9] MAUREL V B, AHUMADA E F, PALAGOS B, et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy [J]. Trends in Analytical Chemistry, 2010, 29(9): 1074-1081.
- [10] LOMBORG C J, HOLM-NIELSEN J B, POPIEL P O, et al. Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage [J]. Bioresource Technology, 2009, 100: 1711-1719.
- [11] 杨增玲, 薛俊杰, 贺城. 不同光谱采集条件下麦秸低位热值 NIRS 快速检测[J]. 农业机械学报, 2013, 44(8): 139-142.
- [12] 窦沙沙, 何芳, 王丽红, 等. 玉米秸秆热解挥发分元素含量分析[J]. 可再生能源, 2006, 129(5): 22-24.
- [13] 姚红宇, 唐光木, 葛春辉, 等. 炭化温度和时间与棉秆炭特性及元素组成的相关关系[J]. 农业工程学报, 2013, 29(7): 199-206.
- [14] BURNS D A, CIURCZAK E W. Handbook of near-infrared analysis [M]. 2nd ed. New York: Marcel Dekker, 2001: 431-433.

NIRS rapid detection of elemental analysis for straw biomass based on PLS algorithm

LI Xiao-jin ZHU Kai NIU Zhi-you CHENG Xu-yun

College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

Abstract To investigate the feasibility of fast detection of the elements of N, C, H, S and O of straw biomass by using the near-infrared spectroscopy (NIRS) technology, 199 straw samples have been collected and prepared, including the straw of rice, wheat, canola and corn. Near-infrared diffuse reflectance spectroscopy combined with PLS chemometric algorithms has been used to compare the calibration effect with different spectral pretreatment methods in 7 400-5 550 cm^{-1} wavelength range, and the optimal calibration analysis models for N, C, H, S and O element of straw biomass have been established, then the independent samples of validation set were used to validate the model. The validation results show that the established quantitative analysis model for N element can be used in practical detecting; the established quantitative analysis model for O element can be used in practical estimation; the quantitative analysis model of C element by using near-infrared technology is feasible, but the model needs to be further optimized; H and S element can not be quantitatively analyzed by using NIRS technology.

Key words near-infrared spectroscopy (NIRS); partial least squares (PLS) algorithm; straw biomass; elemental analysis