

玉米秸秆组分近红外漫反射光谱 (NIRS) 测定方法的建立

刘丽英, 陈洪章*

中国科学院过程工程研究所, 生化工程国家重点实验室, 北京 100080

摘要 玉米秸秆是我国产量最大的秸秆生物质资源,但目前还没有快速高效的组分分析方法,本研究利用傅里叶变换近红外漫反射光谱(NIRS)技术,采用偏最小二乘法(PLS),在国内首次建立了NIRS测定玉米秸秆中灰分、半纤维素、纤维素、Klason木质素、酸不溶灰分和水分含量的校正模型,该模型稳定,适合不同地区、不同品种的玉米秸秆及其不同部位。实验结果表明,采用一阶导数+Karl Norris滤波预处理,谱区在 $4\ 100\sim 7\ 500\text{ cm}^{-1}$,能得到理想的预测模型。该模型对玉米秸秆各组分的交叉验证均方差(RMSECV)范围为 $0.090\ 3\sim 1.45$,预测误差(RMSEP)范围为 $0.256\ 9\%\sim 2.581\ 9\%$,预测相关系数 $0.871\ 1$ 。该研究对加速我国秸秆生物质的工业转化具有重要意义。

关键词 生物质;玉米秸秆组分;近红外漫反射光谱(NIRS);模型

中图分类号: O657.3 **文献标识码**: A **文章编号**: 1000-0593(2007)02-0275-04

引言

在我国各个发展领域执行可持续发展战略(Sustainable development strategy)是国民经济持续、快速、健康发展的关键和保证。1973年第一次石油危机冲击后,开发利用生物质能源的技术得到世界各国的重视和发展。我国是农业大国,秸秆是生物质资源中高值化应用潜力最大的资源之一。而玉米秸秆是我国三大秸秆资源之一,约占我国秸秆总产量的36.18%。目前玉米秸秆由于受到自身特点和转化技术的制约,还没有得到高效全利用,主要用作反刍动物饲料和农村生活燃料,少量用作工业原料或者还田、废弃燃烧。随着生物质利用研究的进展,玉米秸秆在工业中的应用将会越来越多。而玉米秸秆无论是作为动物饲料,还是工业原料,都需要对其中的主要成分纤维素、半纤维素和木质素等进行测定,为应用提供基础数据。

传统方法对玉米秸秆这几大成分的分析,不仅测定速度慢、成本高,而且使用试剂有害,不适合大批样品的测定。近红外反射光谱(NIRS)具有测定速度快、样品制备简单、不耗费化学试剂、操作简便、信息丰富等优点^[1],非常适合多组分测定。在纤维素原料方面,NIRS技术最早用于青贮饲料的分析,而且主要集中于中性洗涤纤维(NDF)与酸性洗涤纤维(ADF)含量、消化率和其他营养成分的研究与评价上,以指导反刍家畜生产^[2-4]。国内也在这方面做了大量工

作,如中国农业大学利用近红外光谱技术对玉米秸秆的NDF和ADF含量,对绵羊日粮的消化率等进行了测定,建立了测定模型^[5]。随着对生物质研发力度的加大,国外研究者将NIRS技术应用到生物质组分的测定上,并建立了组分数据库,为生物质的高效转化提供了有效分析手段^[6,7],而国内目前还没有建立针对生物质工业应用的NIRS快速高效分析方法。因此研究利用NIRS快速、准确测定玉米秸秆的纤维素、半纤维素、木质素、灰分的方法,建立玉米秸秆的光谱数据库,将为我国玉米秸秆的利用提供高效分析平台和信息平台。

1 实验

1.1 实验材料

供试样品采自2003~2004年,山东、河北、山西、辽宁、河南5省种植的不同品种的玉米秸秆。以上样品共计380份,用植物粉碎机粉碎,过40目筛,样品装入密封袋中,常温下避光贮存。

1.2 仪器与样品光谱测定

实验仪器为Nicolet公司(美国)生产的Nexus傅里叶变换近红外光谱仪。仪器工作参数为:谱区范围 $4\ 000\sim 10\ 000\text{ cm}^{-1}$,扫描次数64次,分辨率 8 cm^{-1} 。取适量的玉米秸秆粉盛入旋转样品池,用压样制样器在样品杯内制样。每个样品重复装样3次,计算其平均光谱,存入计算机中。其中5

收稿日期:2005-12-02,修订日期:2006-04-17

基金项目:国家重点基础研究发展计划“973”项目(2004CB719700)资助

作者简介:刘丽英,女,1976年生,中国科学院过程工程研究所博士研究生

*通讯联系人

e-mail: hzchen@home.ipe.ac.cn

个样品的近红外漫反射光谱如图 1 所示, 近红外光谱图有多处吸收峰, 可以作为定量分析的依据。

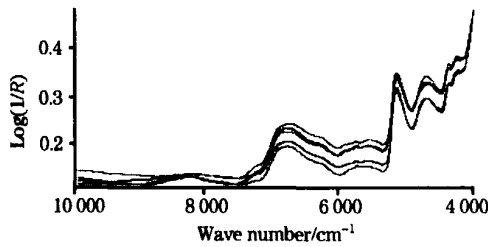


Fig 1 The five random spectra of maize stover

1.3 样品中各组分的含量测定

样品的水分含量按照国标 GB2677.2—81 测定, 灰分按照国标 GB2677.3—81 测定, 纤维素、半纤维素、Klason 木素和酸不溶灰分的测定按照 Van Soest^[8] 方法的原理, 结合王玉万改进方法^[9]和 Contreras Lara 等^[10]改进的尼龙袋法进行。

1.4 建立 NIRS 数学模型的方法

偏最小二乘 (PLS) 是一种相对而言较新的多元数据处理技术, 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。因此, 试验中采用 TQ Analyst V6.2 商用光谱定量分析软件中的 PLS 算法, 将采集的玉米秸秆不同部位的样品的近红外光谱与其相应的水分、灰分、半纤维素、纤维素、Klason 木素和酸不溶灰分 (AUA) 基础测定数据分别进行拟合建模。以校正集样品的交叉验证均方差 (RMSEVC) 为指标优化光谱预处理方法和模型参数, 以对未知样品的预测均方差 (RMSEP) 考察模型的预测准确度。RMSEVC 和 RMSEP 的计算公式如下

$$\text{RMSEVC} = \sqrt{\frac{(\hat{c}_i - c_i)^2}{n-1}}$$

$$\text{RMSEP} = \sqrt{\frac{(\hat{c}_i - c_i)^2}{m}}$$

式中: \hat{c}_i 是标准化学方法测得的值, c_i 是 NIR 预测值, n 是校正集样品数, m 是验证集样品数。

2 实验结果与分析

2.1 玉米秸秆组分的化学分析结果

实验样品中灰分、水分、半纤维素、纤维素、木质素和酸不溶灰分 (AUA) 的测试标准误差分别为 2.76, 0.87, 3.38, 5.30, 2.52 和 0.30 (见表 1, 表中 SD 为标准差)。样品中各组分的含量变幅很大, 说明样品来源范围广, 其含量变化范围基本上可以覆盖当前生产上推广玉米品种和青贮品种中可能出现的纤维含量, 可以很好地用于建立玉米秸秆不同组分近红外光谱校正模型。

2.2 光谱数据的预处理

对采集到的原始光谱, 进行必要的平滑和求导等数学预处理, 其目的是过滤噪声、提高信噪比, 消除基线漂移的干扰。从建模的效果看, Karl Norris 导数平滑滤波^[11] (Karl

norris derivative filter) 优于 Savitzky-Golay 平滑滤波^[12]: 应用一阶导数与 Karl Norris 滤波结合处理生成的导数光谱建模, 可获得理想效果。但采用 Karl Norris 导数滤波时, 应慎重选择段长 (Segment length) 和段间距 (Gap between segments)。段长不易大于 13, 较为理想的段间距在 4~6 之间比较合适, 本试验中采用了 5。

Table 1 Components of the samples

Components	Samples	Range/ %	Average/ %	SD/ %
Ash	68	2.34 ~ 14.26	6.40	2.76
Water	56	3.15 ~ 6.75	5.00	0.87
Hemicellulose	68	18.33 ~ 37.41	31.09	3.38
Cellulose	68	22.17 ~ 45.32	31.22	5.30
Klason lignin	68	6.68 ~ 17.26	11.57	2.52
AUA	68	0.25 ~ 1.41	0.67	0.30

2.3 校正模型的优化及交叉验证结果

当模型的雏形建立后, 必须应用交互验证方法, 对模型逐步优化, 其目的是剔除奇异点 (Outlier) 和强影响点, 确定最优主因子, 直到获得最佳模型。

判断和剔除奇异点和强影响点, 一是采用 Mahalanobis 距离^[13] (马氏距离), 即 $MD_i = (t_i - \bar{t}) S_k^{-1} (t_i - \bar{t})^T$ 。设定显著性水平 (0.05), 通过 Dixon 检验来剔除。二是通过杠杆值来考察对模型的影响^[14], 即采用“剔一”交互验证的方法, 结合杠杆值, 即 $h_i = \frac{1}{m} + t_i^T (T_k^T - T_k)^{-1} t_i$, 与学生化残差值 (Studentized residual) 等指标联合诊断^[15], 逐步剔除这些强影响点。

光谱范围和 PLS 主因子数是影响模型质量的两个主要参数。其中光谱范围的选择过程实际就是样品中各组分光谱信息的筛选过程, 它直接影响到光谱信息的利用效果, 继而影响模型的准确性。PLS 主因子数则关系到模型的稳定性, 太小会出现“欠拟合”, 太大会出现“过拟合”。

根据以上模型建立的规则, 经 TQ Analyzer V6.2 软件优化得到的各组分的光谱范围、PLS 主因子数和校正集样品交叉验证结果如表 2。表 2 显示, 样品各组分建模光谱范围很宽, 在 4100~7500 cm^{-1} 波谱范围内都能得到较好的交叉验证结果。灰分、半纤维素、纤维素、Klason 木素、酸不溶灰分 (AUA) 和水分的模型的相关系数分别为 0.934 1, 0.931 7, 0.956 9, 0.964 7, 0.962 1 和 0.860 9, 交叉验证均方差 (RMSEVC) 分别为 0.991 %, 1.27 %, 1.45 %, 0.599 %, 0.090 % 和 0.547 %, 说明模型预测值和化学测定值接近, 完全可满足工业生产中中对组分测定精度的要求。

2.4 模型对玉米秸秆各组分的预测

校正模型建立后, 采用外部验证的方法对所建模型实际预测效果进行评价检验。预测样品各组分的预测值和化学测定值进行成对数据双尾 t 检验 (见表 3), 各组分化学测定值与预测值间的 t 检验值均未达到显著水平, 表明预测值与化学值之间的差异不显著。灰分、半纤维素、纤维素、Klason 木质素、酸不溶灰分 (AUA) 和水分的预测误差 (RMSEP) 分别为 0.774 6 %, 1.807 2 %, 0.256 9 %, 2.581 9 %, 0.256 9 %, 1.807 2 %。

0.515 8 % 和 1.032 5 % ; 预测值与化学测定值间的相关系数分别为 0.946 1, 0.931 2, 0.959 2, 0.922 8, 0.871 1 和

0.931 7 (见图 2)。上述结果表明预测值与化学值比较接近, 可用来对未知样品进行实际预测。

Table 2 Results of optimal wave number range and PLS factor number and cross validation results

	Ash	Hemicellulose	Cellulose	Klason lignin	AUA	Water
Spectrum range/cm ⁻¹	4 080 ~ 7 578	4 100 ~ 7 770	4 100 ~ 7 770	4 083 ~ 7 773	4 083 ~ 7 773	4 067 ~ 7 773
PLS factor	10	13	13	11	12	10
RMSECV	0.991	1.27	1.45	0.599	0.090	0.547
r ²	0.934 1	0.931 7	0.956 9	0.964 7	0.962 1	0.860 9

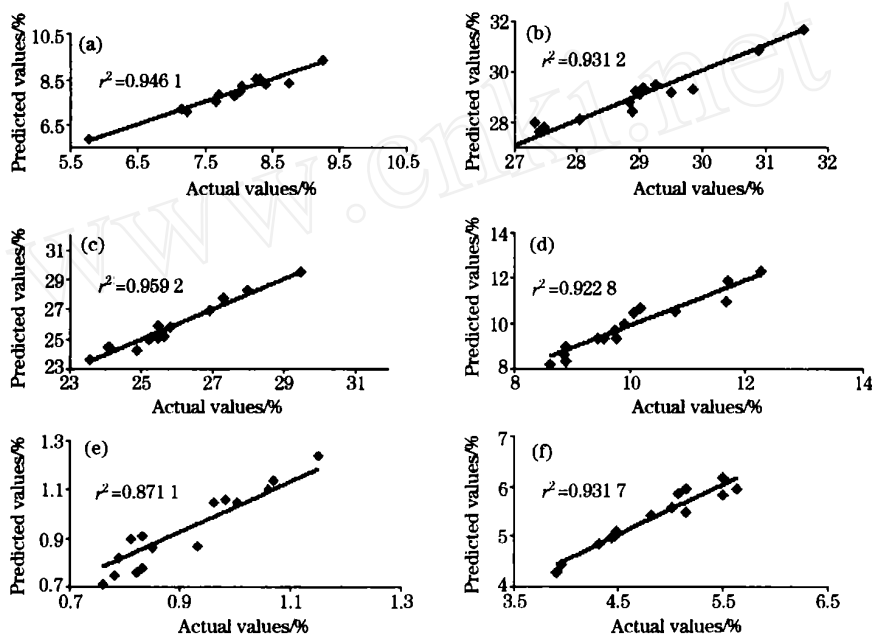


Fig 2 The correlations between measured and predicted values of corn stover components

(a) : ash ; (b) : Hemicellulose ; (c) : Cellulose ; (d) : Klason lignin ; (e) : AUA ; (f) : Water

Table 3 Evaluation of the prediction models with validation sets

Components	Samples	Actual averages/ %	Predicted averages/ %	r ²	tValues	t _{0.05}	RMSEP/ %
Ash	15	7.87	7.90	0.946 1	0.159 0	1.753 1	0.774 6
Water	15	4.80	4.84	0.931 7	0.279 6	1.753 1	1.032 5
Hemicellulose	15	28.87	28.94	0.931 2	0.207 4	1.753 1	1.807 2
Cellulose	15	25.79	25.78	0.959 2	0.024 6	1.753 1	0.256 9
Klason lignin	15	10.02	9.92	0.922 8	0.346 4	1.753 1	2.581 9
AUA	15	0.91	0.93	0.871 1	0.634 9	1.753 1	0.515 8

3 讨论

目前国内还没有建立不同地区玉米秸秆的 NIRS 快速测定方法, 以用于生物质可再生能源的组分预测。利用 NIRS 技术进行玉米秸秆组分快速测定也仅限于中性洗涤纤维 (NDF) 和酸性洗涤纤维 (ADF), 主要用于粗饲料测定和育种材料的选择^[5, 10]。而美国能源部和国家可再生能源实验室已经利用玉米秸秆为模型材料, 建立了 NIRS 测定方法, 为利用玉米秸秆生产燃料乙醇提供了有效分析手段^[6, 7]。由于不

同地区和不同品种间玉米秸秆组分存在很大差异, 需要建立针对我国玉米秸秆的 NIRS 分析方法。美国能源部和国家可再生能源实验室建立的模型主要采用的组分指标为葡萄糖、木糖、阿拉伯糖、木质素和蛋白质, 主要是针对秸秆酶解产糖的组分测定, 而本实验采用的组分测定指标为水分、半纤维素、纤维素、Klason 木质素、总灰分和酸不溶灰分, 这也是我国在秸秆造纸、产糖发酵、加工材料等生产中普遍采用的几大指标, 而且本实验采用的样品范围广, 包括不同省区和不同品种, 预测精度较高, 因此更适于我国的实际需要, 对加速我国的可再生能源研发进程有重要意义。

中国农业大学也曾用近红外光谱法研究过玉米的品种^[16]。

参 考 文 献

- [1] YAN Yan-lu, ZHAO Long-lian, LI Jun-hui, et al (严衍禄, 赵龙莲, 李军会, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(6): 777.
- [2] Zimmer E, Gurrath P A, Buxton D R. Euphytica. , 1990, 48: 73.
- [3] Jung H G, Mertens D R, et al. Crop Sci. , 1998, 38: 205.
- [4] Schwab E C, Shaver R D, Lauer J G, et al. Animal Feed Science and Technology, 2003, 109: 1.
- [5] BAI Qi-lin, CHEN Shao-jiang, DONG Xiao-ling, et al(白琪林, 陈绍江, 董晓玲, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析) 2004, 11(4): 1345.
- [6] Matt A. Sanderson, Foster Agblevor, et al. Biomass and Bioenergy, 1996, 11(5): 365.
- [7] Hames B R, Thomas S R, Sluiter A D, et al. Applied Biochemistry and Biotechnology, 2003, 105: 5.
- [8] Goering H K, Van Soest P J. Agric. Handbook 379. ARS. USDA. Washington D. C. 1970, 1.
- [9] WANG Yr-wan, XU Wen-yu(王玉万, 徐文玉). Chinese Bulletin of Microbiology(微生物学通报), 1987, 14(2): 81.
- [10] Contreras Lara, Guti érez Ch ávez D L, Valdivia Macedo I, et al. Arch. Zootec. , 1999, 48: 351.
- [11] Thermo Nicolet Corporation. TQ Analyst User 's Guid. Masison: Thermo Nicolet Corporation, 2003. 153.
- [12] Savitzky A, Golay M J E. Anal. Chem. , 1964, 36: 1627.
- [13] Infometrix Inc. Multivariate Data Analysis. Woodinville: Infometrix Inc. , 2003. 5.
- [14] Weisberg S. Applied Linear Regression. New York: John Wiley & Sons, 1985, Chapter 5.
- [15] Cook R D, Weisberg S. Residuals and Influence in Regression. New York: Chapman and Hall, 1982, Chapter 2.
- [16] WEI Liang-ming, JIANG Hai-ying, LI Jun-hui, et al(魏良明, 姜海鹰, 李军会, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(9): 1404.

Prediction of Maize Stover Components with Near Infrared Reflectance Spectroscopy

LIU Li-ying, CHEN Hong-zhang *

State Key Laboratory of Biochemical Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100080, China

Abstract The components concentrations in maize stover were analyzed with 67 samples selected from 380 samples of different provinces and varieties in order to serve the biomass utilization of our country. The technique of near infrared reflectance spectroscopy (NIRS) and partial least square (PLS) regression were used to establish the models. The results showed that the calibration models developed by the spectral data pretreatment of the first derivative + Karl Norris derivative filter were the best for ash, hemicellulose, cellulose, Klason lignin, acid unsolvable ash, and water in the spectral region of 4 100-7 500 cm^{-1} . The root mean square error of cross validation (RMSECV) for the above six components was 0.991, 1.27, 1.44, 0.599, 0.0903 and 0.547, respectively; the root mean square error of prediction (RMSEP) was 0.7746%, 1.8072%, 0.2569%, 2.5819%, 0.5158% and 1.0325%, respectively. The models can be used to measure various samples in biomass transformation industry.

Keywords Biomass; Components of maize stover; Near infrared reflectance spectrum(NIRS); Models

(Received Dec. 2, 2005; accepted Apr. 17, 2006)

* Corresponding author