doi:10.6041/j.issn.1000-1298.2025.12.042

基于近红外光谱特征信号的秸秆粒径校正方法研究

朱礼强　郭建涛　杨增玲　刘　贤　韩鲁佳
(中国农业大学工学院, 北京 100083)

摘要: 近红外光谱技术已被广泛用于秸秆理化特性的快速分析研究,具有检测速度快、效率高、无损等优点。然而实际应用时,秸秆的粒径不同会降低模型的检测精度及其稳定性。针对几何粒径水稻秸秆样本,分别获取其在 10 000 ~ 4 000 cm^{-1} 波段的近红外光谱。采用多项式拟合(Polynomial fitting, PF)消除秸秆样本光谱信号中的背景信号。使用单因素方差分析对多项式拟合的几何粒径秸秆样本光谱特征信号进行统计学检验($P<0.05$)。运用主成分分析(Principal component analysis, PCA)及 Biplot 解析光谱特征信号对几何粒径秸秆样本光谱差异的贡献度。分别利用 Pearson 与 Spearman 相关性分析筛选关键光谱特征信号。根据二次多项式拟合数学模型,假设吸光度由粒径相关光谱信息(Particle size related spectrum information, PSRSI)与非粒径相关光谱信息(Non-particle-size-related spectrum information, nPSRSI)构成,提出了一种基于关键光谱特征信号吸光度与粒径的粒径回归校正方法(Particle size regression correction method, PSRCM)。结果表明:几何粒径秸秆样本 5 180 cm^{-1} 光谱特征信号存在极显著差异性($P\leqslant0.05$,$P\leqslant0.01$,$P\leqslant0.001$)且与粒径呈现极显著相关性(Pearson's r 为 0.96)。基于二次多项式数学模型的表现最优,决定系数 R^2 与均方根误差(RMSE)的平均值分别为 0.53 和 0.002 1。PSRCM 可有效校正 6 824、5 360、5 180 cm^{-1} 等特征信号引起的几何粒径秸秆样本光谱差异,校正效果明显优于标准正态变量变换(Standard normal variate transformation, SNV)与乘性散射校正(Multiplicative scatter correction, MSC)。本研究为秸秆近红外光谱无标样模型传递校正方法提供了理论依据。

关键词: 水稻秸秆; 几何粒径; 近红外光谱; 光谱特征信号; 光谱差异; 校正方法
中图分类号: O433.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)12-0460-10　　**OSID**:

Straw Particle Size Correction Method Based on Near-infrared Spectral Characteristic Signals

ZHU Liqiang　GUO Jiantao　YANG Zengling　LIU Xian　HAN Lujia
(*College of Engineering*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract: Near-infrared spectroscopy has been widely used for the rapid analysis of the physical and chemical properties of straw, offering advantages such as fast detection speed, high efficiency, and non-destructive testing. However, in practical applications, differences in straw particle size can reduce the detection accuracy and stability of the model. For rice straw samples with varying geometric particle sizes, near-infrared spectra were obtained in the 10 000 ~ 4 000 cm^{-1} wavelength range. Polynomial fitting (PF) was employed to remove background signals from the spectral data of the straw samples. A one-way analysis of variance (ANOVA) was used to perform statistical testing ($P<0.05$) on the spectral feature signals of the geometrically sized straw samples after polynomial fitting. Principal component analysis (PCA) and Biplot analysis were employed to assess the contribution of spectral feature signals to the spectral differences among straw samples with different geometric particle sizes. Pearson and Spearman correlation analyses were used to screen for key spectral feature signals. Based on the quadratic polynomial fitting mathematical model, it was assumed that absorbance was composed of particle size-related spectral information (PSRSI) and non-particle size-related spectral information (nPSRSI). A particle size regression correction method (PSRCM) was proposed based on the absorbance and particle

收稿日期: 2025-07-02　修回日期: 2025-07-23
基金项目: 国家重点研发计划项目(2022YFD2002101)
作者简介: 朱礼强(2001—),男,博士生,主要从事农业信息化研究,E-mail: BS20253070905@cau.edu.cn
通信作者: 刘贤(1979—),女,教授,博士生导师,主要从事生物质资源与利用研究,E-mail: lx@cau.edu.cn

size of key spectral feature signals. The results indicated that the spectral feature signals at 5 180 cm^{-1} of the geometric particle size straw samples exhibited extremely significant differences ($P \leqslant 0.05$, $P \leqslant 0.01$, $P \leqslant 0.001$) and showed a highly significant correlation with particle size (Pearson's r was 0.96). The quadratic polynomial mathematical model performed optimally, with an average coefficient of determination R^2 and root mean square error (RMSE) of 0.53 and 0.002 1, respectively. PSRCM effectively corrected the spectral differences caused by characteristic signals at 6 824 cm^{-1}, 5 360 cm^{-1}, and 5 180 cm^{-1}, with correction effects significantly superior to those of standard normal variate transformation (SNV) and multiplicative scatter correction (MSC). The research result can provide a theoretical basis for the transfer correction method of straw near-infrared spectroscopy models, which can be applied without the need for standard samples.

Key words: rice straw; geometric particle size; near-infrared spectroscopy; spectral characteristic signal; spectral difference; correction method

0　引言

近红外光谱具有分析速度快[1]、效率高[2]、绿色环保[3]等优势，广泛应用于秸秆理化特性的定量分析研究[4-5]。采用漫反射近红外光谱测量不规则样品时，近红外光与样本内部分子相互作用时，会发生多次反射、吸收、透射、衍射、折射及散射，当散射系数(S)保持不变时，Kubelka－Munk 方程与样品的浓度呈线性关系[6]。然而，实际情况下，固体颗粒的粒度、形状、分布等多种物理因素均会影响光的散射[7]。样品的粒径和均匀度显著影响光谱信号的稳定性和模型的预测性能[8-11]。因此，采用合适的校正方法消除异质性秸秆分子光谱信号中的粒径信息具有重要意义。

常用的散射校正方法包含标准正态变量变换(Standard normal variate transformation, SNV)[12]、乘性散射校正(Multiplicative scatter correction, MSC)[13]、最小－最大归一化(Min－max normalization, MMN)[14]等，已被广泛应用于农业[15]、制药[16]、烟草[17]等领域。然而，这些校正方法仅适用于光谱散射信息的校正且不能应用于化学组分发生复杂变化与特定已知干扰组分的光谱中。扩展乘性散射校正(Extended multiplicative scatter correction, EMSC)[18]是一种在 MSC 算法基础上拓展的算法，可以消除复杂基质与指定干扰组分引起的光谱信号变化，广泛应用于材料[19]、食品[20]领域。然而，EMSC 需要已知化学组分与干扰组分的参考光谱才能消除背景信号的干扰。实际场景下，原料状态及基质构成具有复杂、动态变化的特点，使得 EMSC 的应用范围被限制。文献[21－22]研究提出的粒径/直径校正方法效果优异，但是在验证算法有效性的过程中，通常需要测定化学值，存在耗时、费力、成本高昂的缺点。因此，通过分析光谱信号的变化规律开发校正方法具有节约算力、降本增效的优点。李阳等[23]通过探究参考球漫反射光谱在 600、700、800、900 nm 光谱信号处的强度随探头检测距离的变化规律，提出一种距离校正方法(Distance correction, DC)，消除了样本物理属性差异导致探头到样本表面的距离变化对苹果光谱影响。MI 等[24]测定大米、糯米和西米 3 种不同样品的近红外光谱，通过随机选择的 2 个光谱信号在不同粒度之间进行拟合方程，进一步基于吸光度和粒度研究了一种新的粒度校正方法。结果表明：在判别分析中，相比 SNV 与 MSC，新校正方法处理的 3 种不同样品的光谱可以通过主成分分析(Principal component analysis, PCA)直接准确地区分。然而，相关研究未探索光谱信号与粒径的相关关系，并且此校正方法只适用于颗粒形状规则的原料。目前关于秸秆粒径校正方法的相关研究尚未见报道。

本文拟以水稻秸秆为研究对象，获取几何粒径秸秆样本在波段 10 000～4 000 cm^{-1} 的光谱信号。通过化学计量学与数理统计学分析方法系统挖掘粒径相关的光谱特征信号，基于光谱信号的变化规律构建一种粒径校正方法。研究在实际应用场景下，利用无标样算法直接校正秸秆样本几何粒径对光谱产生的影响，以提高定标模型的检测精度与检测效率。

1　材料与方法

1.1　水稻秸秆样本收集与制备

水稻秸秆样本于 2024 年 10 月收集于中国农业大学上庄实验站。收集的秸秆置于实验室自然风干，去除分蘖、表皮、叶片、根部，获取水稻秸秆主茎秆，以最大程度确保粉碎样本内部组分的均匀性。分别采用切短与粉碎 2 种处理方式获取几何粒径秸秆。取等质量的 5 份水稻秸秆样本，利用铡刀将其切短，分别获取粒径为 50、30、20、10、5 mm 的秸秆。取等质量的 2 份水稻秸秆样本，利用分别放置 1、0.45 mm 筛片的锤片式粉碎机粉碎秸秆，进一步将

它们分别放置在自动振筛机上,设定振筛时间为 20 min,过40 目和 18 目筛,获取粒径为 1、0.45 mm 的秸秆。将几何粒径秸秆分别收集至管盖为聚乙烯(Polyethylene, PE)、管体为聚丙烯(Polypropylene, PP)材质的塑料离心管,标记为 50、30、20、10、5、1、0.45 mm 水稻秸秆样本,每种水稻秸秆样本包含 3 个平行样本,放置于 4℃贮藏室存放。光谱扫描前,样本放置室温 24 h。

1.2 光谱信号采集

采用 Bruker MPA II 型傅里叶变换近红外光谱仪(布鲁克,德国)积分球附件在波段 10 000 ~ 4 000 cm^{-1}采集几何粒径水稻秸秆样本的近红外光谱信号,分辨率设定为 8 cm^{-1},背景扫描 32 次,样本扫描 32 次。每个样本扫描 3 次,每次间隔 15 min 扫描一次背景光谱,防止环境因素对光谱信号产生影响。共 7 组样本,每组样本的光谱信号为 3 个平行样本光谱信号的平均值,每个平行样本的光谱信号是光谱仪扫描每个样本 6 次光谱信号的平均值。以每组样本的光谱信号作为最终结果进行分析。

1.3 统计分析方法

本研究采用软件 IBM SPSS Statistics 27(IBM,美国)单因素方差分析(one-way ANOVA)[25]中的 Duncans 多重比较对几何粒径水稻秸秆样本光谱特征信号进行统计学检验($P<0.05$),置信区间设定为 95%。采用多项式拟合(Polynomial fitting, PF)[26]消除仪器背景、光散射、基线漂移、噪声等低频干扰噪声对水稻秸秆样本光谱信号产生的影响,多项式的阶数设置为 2,该算法及 PSRCM、SNV、MSC 算法代码均通过 Matlab R2021a(MathWorks,美国)编写。运用 Matlab 中 PLS_Tool box 8.2(Eigenvector Research, Inc.,美国)[27]的主成分分析及 Biplot[28]衡量光谱特征信号对几何粒径水稻秸秆样本光谱差异的贡献度。在此基础上,利用 Origin 2021(Origin Lab,美国)分别进行 Person 和 Spearman 相关性分析[29]以进一步筛选出关键光谱特征信号($P\leqslant0.05$,$P\leqslant0.01$,$P\leqslant0.001$)并绘制图形。

1.4 粒径校正模型与评价

本研究在文献[24]所提出校正方法思路的基础上进行改进,提出了一种基于光谱特征信号吸光度的粒径校正方法,将其命名为粒径回归校正方法(Particle size regression correction method, PSRCM)。

首先,采用统计分析方法筛选出 1 个光谱特征信号,它与粒径呈现极强相关性。其次,在坐标图内,分析该信号点的分布情况。最后,分别建立 3 种基于粒径值与该信号点处吸光度的数学模型,分别为四参数逻辑斯蒂数学模型、二次多项式数学模型、指数衰减数学模型

$$y_{p,j}=\frac{A_1-A_2}{1+e^{(x_p-x_0)/dx}}+A_2\quad(p=1,2,\cdots,m)\tag{1}$$

$$y_{p,j}=a+bx_p+cx_p^2\tag{2}$$

$$y_{p,j}=s-qr^{x_p}\tag{3}$$

式中 x_p——几何粒径

x_0——使吸光度介于A_1与A_2之间一半的位置,代表信号变化最敏感的几何粒径位置

A_1——几何粒径非常小时,吸光度的理论最大值

A_2——几何粒径非常大时,吸光度的理论最小值

dx——斜率系数,控制光谱曲线在x_0附近的陡峭程度

a——几何粒径趋近 0 时的理论吸光度值

b——一次项系数,控制光谱曲线的线性变化趋势

c——二次项系数,控制光谱曲线的弯曲程度

s——几何粒径非常大时,光散射趋于稳定后吸光度的最终值

q——初始吸光度与极限吸光度的差值

r——吸光度随几何粒径变化的速率

m——光谱条数

$y_{p,j}$——x_p下的光谱在某一波数点 j 处的吸光度

将几何粒径x_p和几何粒径样本光谱信号从 10 000 cm^{-1}至 4 000 cm^{-1}的吸光度$y_{p,j}$分别放置于粒径矩阵 $\boldsymbol{x}_p=[x_1\quad x_2\quad\cdots\quad x_m]$与吸光度矩阵 $\boldsymbol{y}_{p,j}=[y_{1,10\,000}\quad y_{2,10\,000}\quad\cdots\quad y_{m,j}]$,然后将它们分别代入 3 种数学模型(式(1)、(2)、(3)),采用最小二乘法进行逐波数点拟合,通过决定系数(Determination coefficients, R^2)和均方根误差(Root mean square error, RMSE)的最大值、平均值以及最小值作为评价指标以确定最优的数学模型。

基于本研究分析优化获取的二次多项式拟合数学模型的基础上,假设样本光谱信号的吸光度 y 被分为两部分,a^*代表非粒径相关光谱信息(Non-particle-size-related spectrum information, nPSRSI),x代表粒径,$b^*x+c^*x^2$代表粒径相关光谱信息(Particle size related spectrum information, PSRSI),b^*、c^*表示 PSRSI 系数。

为了计算 nPSRSI,采集不同粒径样本的 m 条光谱信号(列向量 $\boldsymbol{Y}_i$,其中 $i=1,2,\cdots,m$)和粒径X_j。几何粒径与样本光谱信号的吸光度分别被放置于粒

径向量$\boldsymbol{X}_j=(X_1,X_2,\cdots,X_m)$与光谱矩阵 $\boldsymbol{Y}=[Y_1\ Y_2\ \cdots\ Y_m]$。根据公式(2)，创建回归矩阵 $\boldsymbol{R}=[1\ x\ x^2]$。因此，模型可以被写作

$$\boldsymbol{Y}=\boldsymbol{AR}\quad(\boldsymbol{A}\in\mathbf{R}^{u\times 3}) \tag{4}$$

式中　$\boldsymbol{A}$——由 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$、$\boldsymbol{\gamma}$ 向量组成的光谱系数矩阵

u——矩阵 $\boldsymbol{A}$ 的行向量数量，取 1 501

采用最小二乘法求解 $\boldsymbol{A}$，即

$$\boldsymbol{A}=\boldsymbol{YR}^{\mathrm{T}}(\boldsymbol{RR}^{\mathrm{T}})^{-1} \tag{5}$$

矩阵 $\boldsymbol{A}$ 中的第 1 列向量 $\boldsymbol{\alpha}$ 为几何粒径样本的 nPSRSI，第 2 列的向量 $\boldsymbol{\beta}$ 与第 3 列的向量 $\boldsymbol{\gamma}$ 均用于计算 PSRSI，即$b^*x+c^*x^2$。

因此，样本的近红外光谱信号可以表达为

$$z_i=a^*+b_i^*x+c_i^*x^2+\delta\quad(i=1,2,\cdots,m) \tag{6}$$

式中　z_i——每个样本的光谱信号

b_i^*、c_i^*——每个样本的 PSRSI 系数

δ——其它低频干扰信号

如果理论上已知公式(6)的系数a^*、b_i^*、c_i^*，此时通过 PSRCM 校正的样本光谱信号 z_{i_c} 为

$$z_{i_c}=z_i-b_i^*x-c_i^*x^2-\delta \tag{7}$$

假设此研究严格控制了其它变量对样本光谱信号的干扰，所以 $\delta\approx 0$。

此时，设定参考光谱的粒径后，即可得到公式

$$e_r^*=b^*x_\lambda+c^*x_\lambda^2 \tag{8}$$

式中　x_λ——参考光谱的粒径

e_r^*——参考光谱的 PSRSI

然后，在 PSRCM 校正的几何粒径样本的光谱信号的基础上，添加参考光谱的 PSRSI，最终得到公式

$$z_{j_c}=z_{i_c}+e_r^* \tag{9}$$

式中　z_{j_c}——PSRCM 校正的 j 条样本光谱信号

最终，假设z_{j_c}为参考光谱的平行光谱。

残差[30]被用于观察实际值与拟合值之间的差，样本的残差越小，表明拟合效果越好。设定 0.45 mm 秸秆样本光谱信号为参考光谱，对 1、5、10、20、30 mm 秸秆样本光谱信号进行了相同的残差分析，结果相似。变异系数(Coefficient of variation，CV)[31]是衡量样本离散程度的一个归一化量度，其定义为标准差与平均值的比值，变异系数越小，数据离散程度越小。当 Pearson 相关性分析中的 $P>0.05$ 时，说明光谱差异校正成功，同样进行了 Spearman 相关性分析，结果类似。

以几何粒径水稻秸秆样本在波段 10 000 ~ 4 000 cm^{-1} 内未校正的光谱信号为基准，利用 PSRCM 校正粒径引起的光谱差异，并以 SNV、MSC 作为对比算法，分别采用残差、变异系数平均值和相关性分析作为校正效果评价指标。

2　结果与讨论

2.1　几何粒径秸秆样本光谱背景信号的剔除

2.1.1　几何粒径秸秆样本原始光谱信号

通过傅里叶变换近红外光谱仪扫描几何粒径秸秆样本，获取其平均光谱信号曲线。图 1 为几何粒径水稻秸秆样本原始光谱信号。

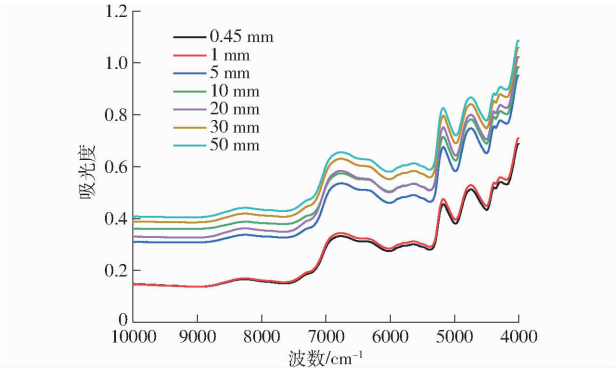


图 1　几何粒径水稻秸秆样本原始光谱信号

Fig. 1　Geometric particle size of rice straw sample raw spectral signal

由图 1 可知，随着粒径的增加，秸秆样本近红外光谱信号的吸光度整体呈现逐渐上升的趋势。其中，0.45 mm 与 1 mm 秸秆样本光谱信号曲线分布紧密，而 5、10、20、30、50 mm 秸秆样本光谱信号分布较近。当秸秆样本的粒径为 0.45 mm 时，其平均吸光度最小为 0.27；粒径为 50 mm 时，平均吸光度最大为 0.57。上述结果表明，相比于粗糙秸秆，0.45 mm 秸秆样本的比表面积大幅度增加，它的均匀度更好，近红外光照射在小粒径秸秆体系内部的散射光程路径增加。这与王海东等[32]得出样品的散射系数 S 与粒度密切相关，S 随着粒度的增加而减少，根据 Kubellka－Munk 函数，吸光度会增加，整个样品的光谱向上移的结论类似。

2.1.2　基于多项式拟合的几何粒径秸秆光谱信号

多项式拟合法[33]是一种基线校正方法，它可以消除激光波长、仪器伪影、荧光效应、光散射、基线漂移、噪声等各种因素造成的背景信号，将真正的光谱信号与基本基线分离，从而保留感兴趣的光谱特征，增强光谱特征，更精确地分析分子振动。

采用多项式拟合法消除几何粒径秸秆样本近红外光谱信号的基线。粒径对近红外光谱的二级倍频区(Second overtone region，2nd Overtone) 10 000 ~ 7 100 cm^{-1}、一级倍频区(First overtone region，1st Overtone) 7 100 ~ 5 000 cm^{-1}、合频区(Combination band region，CR) 5 000 ~ 4 000 cm^{-1} 均有不同程度的影响[34]。PASIKATAN 等[35]研究表明：散射作用存

在于近红外光谱全波长。因为光谱特征信号对于近红外光谱定标模型的构建和预测精度的提升具有重要作用,所以,所有后续分析都基于多项式拟合的秸秆样本光谱特征信号。图 2 为多项式拟合的几何粒径秸秆样本光谱图。

如图 2 可知,采用多项式拟合法消除低频干扰信号对几何粒径秸秆样本近红外光谱信号在 10 000 ~ 4 000 cm^{-1}产生的影响,增强了 15 个代表性秸秆样本光谱特征信号(8 880 ~ 4 176 cm^{-1})的响应。从理论上分析,这些代表性的光谱特征信号中均含有不同程度的粒径相关光谱信息,因此,这些光谱特征信号对开发针对秸秆样本粒径的校正方法具有潜力。所以,需要采用统计学与化学计量学分析方法进一步优化上述粒径相关的光谱特征信号。

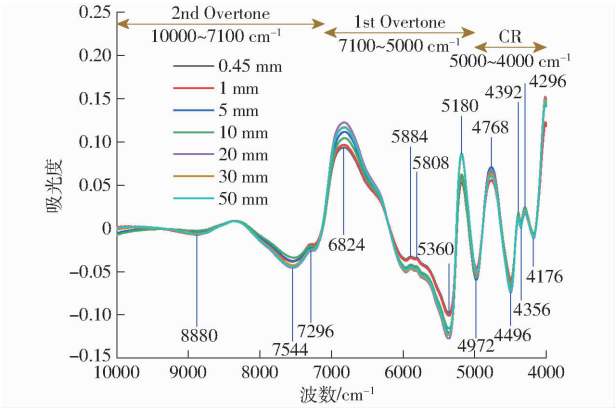


图 2 基于多项式拟合的几何粒径秸秆样本光谱信号

Fig. 2 Spectral signals of straw samples of different grain sizes based on polynomial fitting

2.2 粒径相关的光谱信息挖掘

采用单因素方差分析对几何粒径秸秆样本光谱特征信号的吸光度进行显著性检验($P < 0.05$)。表 1 为基于基线校正的几何粒径秸秆样本光谱特征信号的方差分析结果。

表 1 方差分析结果

Tab. 1 Analysis of variance results

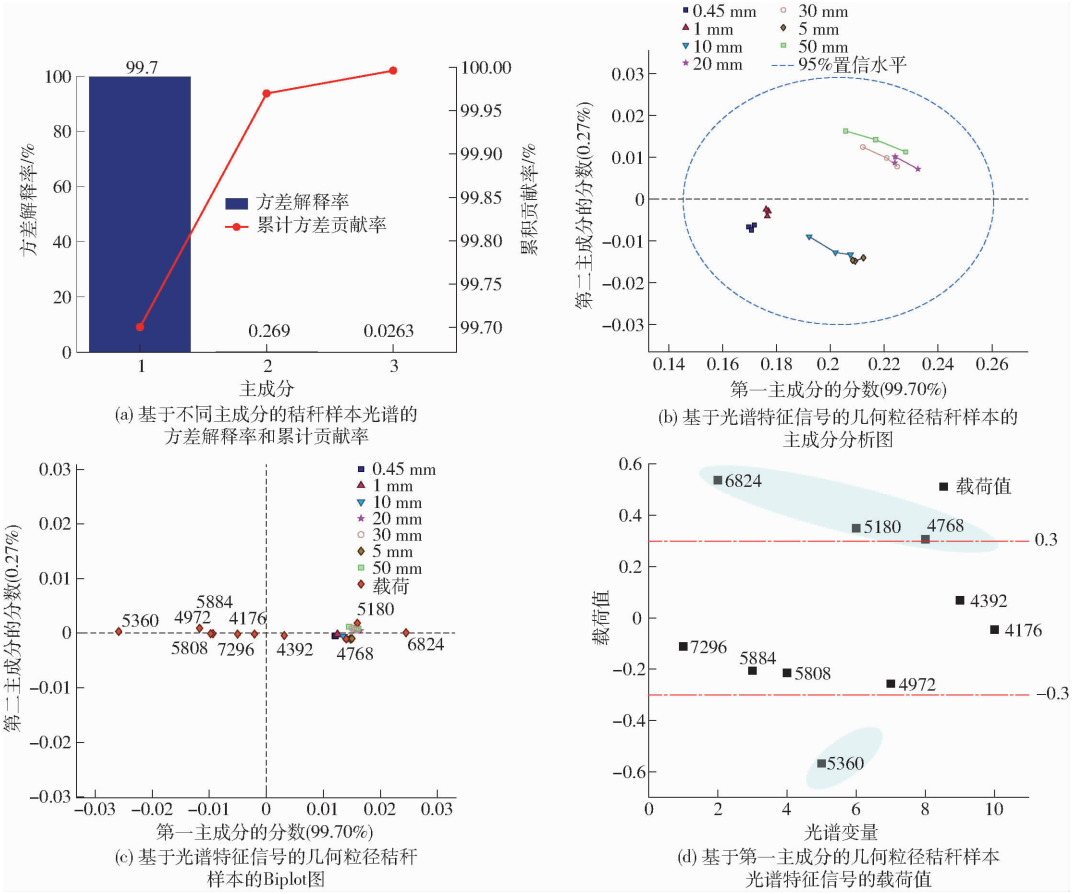
光谱信号/cm^{-1}	粒径/mm						
	0.45	1	5	10	20	30	50
8 880	$(-6.6 \pm 0.1)^{c}$	$(-7.3 \pm 0.9)^{c}$	$(-4.1 \pm 0.6)^{ab}$	$(-2.7 \pm 0.7)^{a}$	$(-6.7 \pm 1.0)^{c}$	$(-6.0 \pm 1.5)^{c}$	$(-6.8 \pm 1.9)^{c}$
7 544	$(-37.7 \pm 0.2)^{ab}$	$(-38.6 \pm 1.2)^{ab}$	$(-37.7 \pm 1.8)^{ab}$	$(-33.8 \pm 2.4)^{a}$	$(-45.6 \pm 2.5)^{c}$	$(-42.6 \pm 3.7)^{bc}$	$(-44.5 \pm 4.7)^{c}$
7 296	$(-18.5 \pm 0.2)^{a}$	$(-19.1 \pm 0.2)^{a}$	$(-22.5 \pm 0.7)^{bc}$	$(-20.7 \pm 1.4)^{ab}$	$(-26.1 \pm 1.2)^{d}$	$(-24.7 \pm 1.6)^{cd}$	$(-25.2 \pm 2.0)^{d}$
6 824	$(93.6 \pm 0.5)^{e}$	$(96.8 \pm 0.7)^{de}$	$(111.8 \pm 2.2)^{bc}$	$(104.5 \pm 5.0)^{cd}$	$(122.6 \pm 3.9)^{a}$	$(117.2 \pm 5.5)^{ab}$	$(116.8 \pm 8.3)^{ab}$
5 884	$(-32.7 \pm 0.2)^{a}$	$(-34.3 \pm 0.3)^{b}$	$(-43.6 \pm 0.1)^{d}$	$(-42.0 \pm 1.3)^{c}$	$(-47.2 \pm 0.6)^{e}$	$(-46.2 \pm 0.1)^{e}$	$(-44.7 \pm 1.4)^{d}$
5 808	$(-33.9 \pm 0.2)^{a}$	$(-35.7 \pm 0.3)^{b}$	$(-45.4 \pm 0.1)^{d}$	$(-43.8 \pm 1.4)^{c}$	$(-49.2 \pm 0.6)^{e}$	$(-48.3 \pm 0.5)^{e}$	$(-46.6 \pm 1.7)^{d}$
5 360	$(-97.8 \pm 0.6)^{a}$	$(-100.8 \pm 0.4)^{a}$	$(-120.8 \pm 0.9)^{bc}$	$(-115.7 \pm 4.5)^{b}$	$(-127.2 \pm 2.8)^{d}$	$(-123.2 \pm 3.7)^{cd}$	$(-121.5 \pm 6.0)^{bcd}$
5 180	$(54.0 \pm 0.7)^{e}$	$(58.7 \pm 0.4)^{d}$	$(62.1 \pm 0.6)^{c}$	$(62.3 \pm 1.0)^{c}$	$(85.8 \pm 0.7)^{ab}$	$(84.6 \pm 0.6)^{b}$	$(86.8 \pm 0.9)^{a}$
4 972	$(-46.3 \pm 0.3)^{a}$	$(-46.0 \pm 0.5)^{a}$	$(-59.2 \pm 0.4)^{e}$	$(-56.8 \pm 2.4)^{de}$	$(-54.6 \pm 0.9)^{cd}$	$(-52.5 \pm 1.5)^{bc}$	$(-50.6 \pm 3.4)^{b}$
4 768	$(55.4 \pm 0.3)^{d}$	$(55.8 \pm 0.6)^{d}$	$(70.6 \pm 0.6)^{a}$	$(67.0 \pm 3.4)^{ab}$	$(65.8 \pm 2.3)^{b}$	$(63.0 \pm 3.1)^{bc}$	$(60.2 \pm 3.7)^{c}$
4 496	$(-59.4 \pm 0.4)^{a}$	$(-61.5 \pm 0.2)^{a}$	$(-67.5 \pm 0.4)^{b}$	$(-66.4 \pm 1.7)^{b}$	$(-74.1 \pm 1.1)^{c}$	$(-72.3 \pm 1.7)^{c}$	$(-72.4 \pm 2.3)^{c}$
4 392	$(11.0 \pm 0.1)^{d}$	$(10.8 \pm 0.6)^{d}$	$(18.0 \pm 0.6)^{a}$	$(17.5 \pm 1.0)^{a}$	$(14.3 \pm 0.3)^{b}$	$(13.6 \pm 0.2)^{bc}$	$(12.8 \pm 0.7)^{c}$
4 356	$(2.8 \pm 0.2)^{b}$	$(2.7 \pm 0.4)^{b}$	$(6.1 \pm 0.5)^{a}$	$(5.5 \pm 0.6)^{a}$	$(1.1 \pm 0.5)^{c}$	$(1.2 \pm 0.2)^{c}$	$(1.0 \pm 0.3)^{c}$
4 296	$(17.7 \pm 0.1)^{c}$	$(17.9 \pm 0.6)^{c}$	$(24.5 \pm 0.7)^{a}$	$(24.1 \pm 0.7)^{a}$	$(21.1 \pm 0.4)^{b}$	$(21.3 \pm 0.5)^{b}$	$(20.5 \pm 0.4)^{b}$
4 176	$(-8.3 \pm 0.1)^{ab}$	$(-8.8 \pm 0.4)^{bc}$	$(-8.2 \pm 0.6)^{ab}$	$(-7.1 \pm 0.8)^{a}$	$(-11.2 \pm 0.9)^{d}$	$(-10.1 \pm 1.5)^{cd}$	$(-10.9 \pm 1.0)^{d}$

注:同行中不同小写字母表示几何粒径秸秆样本组别间的吸光度存在显著性差异($P < 0.05$)。

如表 1 所示,经过多项式拟合的 15 个秸秆样本的光谱特征信号(8 264 ~ 4 016 cm^{-1})在几何粒径组别之间均存在显著差异($P < 0.05$)。其中,进一步通过 Duncans 多重检验分析可知,几何粒径秸秆样本在波数 7 296、6 824、5 884、5 808、5 360、5 180、4 972、4 768、4 392、4 176 cm^{-1}处光谱信号差异相对较大,说明此 10 个秸秆样本的光谱特征信号可能与粒径高度相关。

以上述 10 个光谱特征信号为变量进行主成分分析,得到 Biplot 图后,进一步分析 10 个光谱特征信号对秸秆样本光谱差异的贡献度,以载荷值的绝对值 0.3 为分界线,筛选贡献度较大的光谱特征信号。

由图 3 可知,PC1 的方差解释率高达 99.70%,表明在该主成分下,秸秆样本光谱特征信号中包含了绝大部分的粒径相关光谱信息。图 3 中,PC1 与 PC2 分别占总方差解释率的 99.70% 与 0.27%,50、30、20 mm 秸秆样本分布在第二主成分正半轴,10、5、1、0.45 mm 秸秆样本分布在第二主成分的负半轴,所有样本分布在第一主成分的正半轴。PARK 等[36]认为较长的变量对 PCA 模型变量的贡献具有关键作用。因此,图 3 中,波数 6 824、5 360、5 180、4 768 cm^{-1}相比于其它光谱特征信号对几何粒径秸秆样本光谱差异的贡献较大。通过图 3 可知,以第



(a) 基于不同主成分的秸秆样本光谱的方差解释率和累计贡献率

(b) 基于光谱特征信号的几何粒径秸秆样本的主成分分析图

(c) 基于光谱特征信号的几何粒径秸秆样本的Biplot图

(d) 基于第一主成分的几何粒径秸秆样本光谱特征信号的载荷值

图 3　主成分分析结果

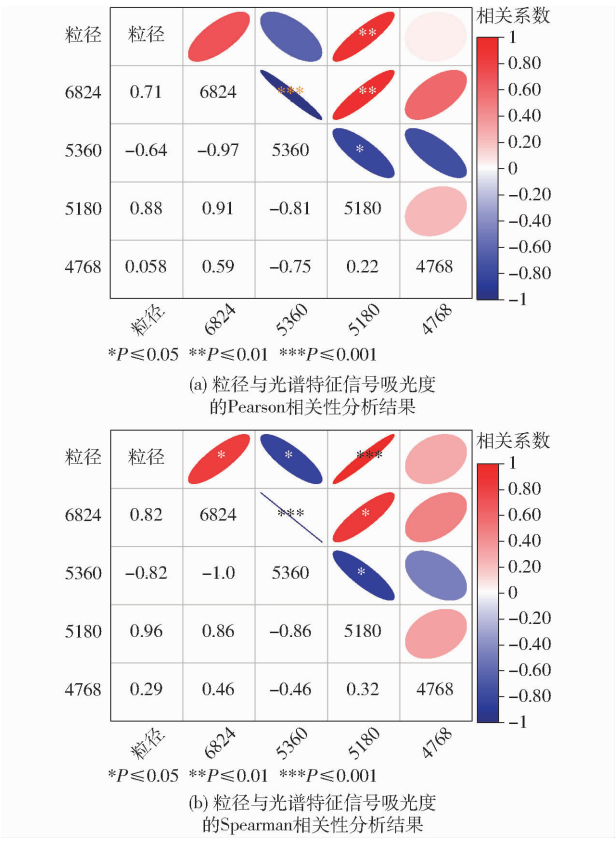
Fig. 3　Principal component analysis results

一主成分载荷值的绝对值 0.3 为分界线，进一步筛选粒径相关的光谱特征信号，它们分别为 6 824、5 360、5 180、4 768 cm^{-1}。上述结果表明，PCA 及 Biplot 方法筛选出的光谱特征信号含有绝大部分的粒径信息。

采用几何粒径与 PCA 及 Biplot 筛选的贡献度较大的光谱特征信号的吸光度值分别进行 Pearson 与 Spearman 相关性分析。

通过图 4 可知，光谱特征信号在波数 5 180 cm^{-1} 处与秸秆样本的粒径存在极显著差异（$P \leqslant 0.05$，$P \leqslant 0.01$），且Pearson 相关系数达到了 0.88。图 4 中光谱特征信号在波数 6 824、5 360 cm^{-1} 处与粒径存在显著差异（$P \leqslant 0.05$），Spearman 相关系数的绝对值均为 0.82，光谱特征信号 5 180 cm^{-1} 与粒径存在极显著差异（$P \leqslant 0.05$，$P \leqslant 0.01$，$P \leqslant 0.001$），Spearman 相关系数高达 0.96。上述结果表明，通过 Pearson 与 Spearman 相关性分析筛选了关键光谱特征信号 6 824、5 180、5 360 cm^{-1}。Pearson 与 Spearman 相关性分析结果均表明光谱特征信号 5 180 cm^{-1} 与粒径之间呈现极强相关关系（相关系数的绝对值为 0.8 ~ 1）。

5 180 cm^{-1} 位于近红外光谱 1st Overtone，靠近



(a) 粒径与光谱特征信号吸光度的Pearson相关性分析结果

(b) 粒径与光谱特征信号吸光度的Spearman相关性分析结果

图 4　相关性分析结果

Fig. 4　Correlation analysis results

CR,含有丰富的秸秆样本化学结构信息。然而,相比于长波光谱区域,短波光谱区域受粒径的影响程度较强。WILSON 等[37]观察到散射系数的波长依赖性遵循从可见近红外(Visible and near-infrared, VIS - NIR)到短波近红外(Short-wave infrared spectroscopy, SWIR)区域呈现幂衰减。本研究结果与相关研究的分析结果相悖,这可能是本研究采用的相关性分析方法的作用机制是基于吸光度与粒度之间的线性或非线性单调作用关系,然而,相比于长波近红外区域,米氏散射[38]和几何光学在短波近红外区域导致光谱响应与粒度呈现高度复杂的非线性与非单调关系,相关性分析不能捕捉到这种响应关系。然而,尽管粒径对短波区域的散射作用较大,但 1st Overtone 与 CR 的吸光度与粒径之间表现出强烈的统计相关性规律,这对建立粒径校正模型具有潜在价值。

2.3 校正方法研究

图 5 分别为粒径与关键光谱特征信号 5 180 cm^{-1} 处的吸光度进行非线性拟合后的 3 种数学模型的结果。

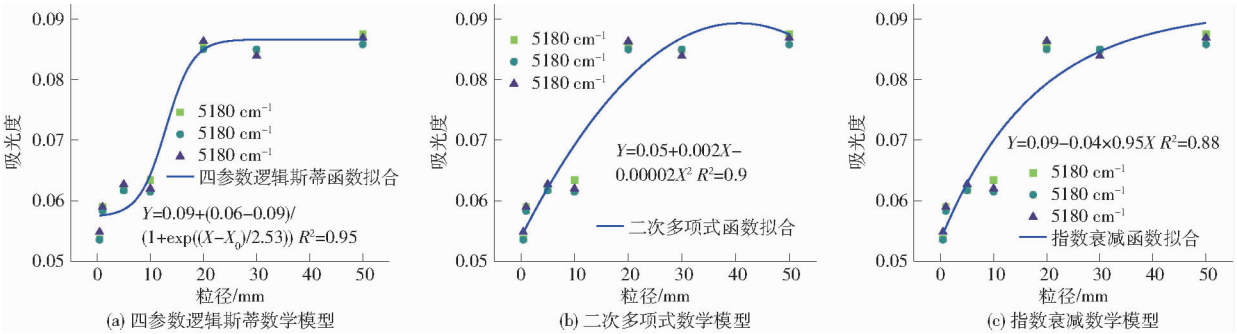


图 5 非线性拟合结果

Fig. 5 Non-linear fitting results

图 5 中,3 种数学模型的 R^2 分别为 0.95、0.9、0.88。上述结果表示,3 种数学模型均良好地拟合了几何粒径秸秆样本在光谱信号 5 180 cm^{-1} 处的曲线。图 6 为 3 种数学模型的 R^2 和 RMSE。

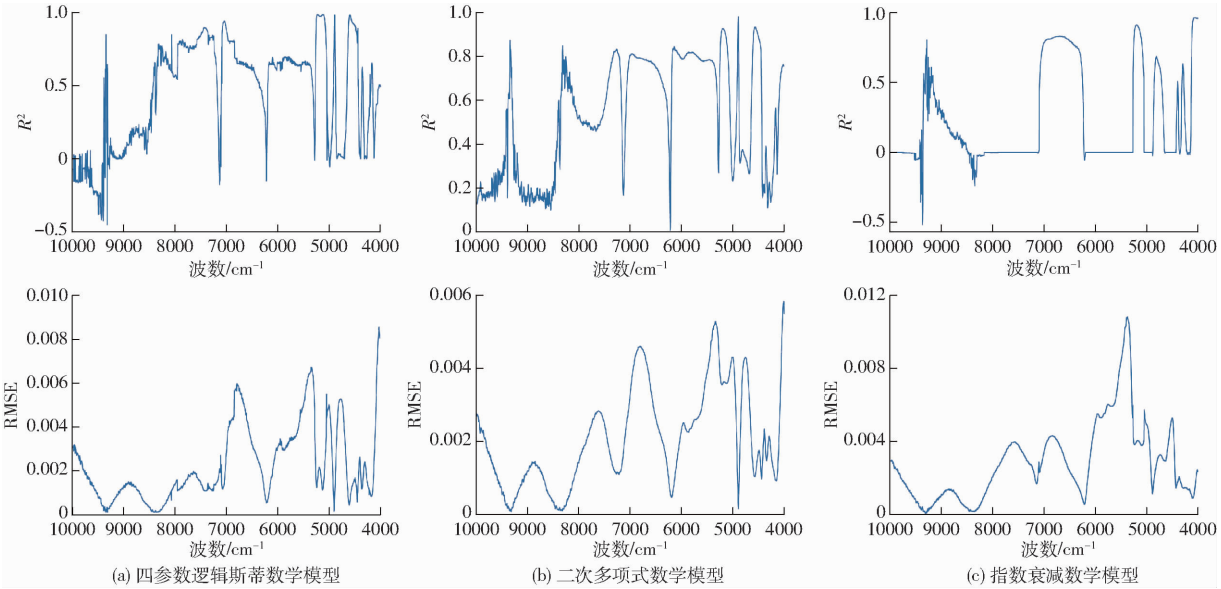


图 6 不同数学模型的评价结果

Fig. 6 Evaluation results of different mathematical models

由图 6 可看出,相比于四参数逻辑斯蒂与指数衰减数学模型的 R^2,二次多项式数学模型的 R^2 在 10 000 ~ 4 000 cm^{-1} 内均大于 0。二次多项式数学模型的 RMSE 相对于四参数逻辑斯蒂与指数衰减数学模型,它的整体变异性较小。进一步计算不同数学模型评价指标的最大值、平均值、最小值,如表 2 所示。

表 2 不同数学模型的评价结果

Tab. 2 Evaluation results of different mathematical models

模型	R^2			RMSE		
	最大值	平均值	最小值	最大值	平均值	最小值
四参数逻辑斯蒂	0.99	0.45	-0.45	0.008 6	0.002 3	0.000 1
二次多项式	0.98	0.53	0.007 2	0.005 9	0.002 1	0.000 1
指数衰减	0.96	0.22	-0.52	0.010 8	0.002 9	0.000 1

由表 2 可知，二次多项式数学模型在 10 000 ~ 4 000 cm^{-1}范围内，R^2平均值为 0.53，RMSE 的平均值为 0.002 1，而四参数逻辑斯蒂数学模型与指数衰减数学模型的 R^2平均值分别为 0.45 与 0.22、RMSE 的平均值分别为 0.002 3 与 0.002 9。分析上述结果可知，二次多项式数学模型的效果最优。因此，将其作为假设模型进行校正方法的研究。

2.4　校正方法的应用与效果评价

根据 1.4 节所建立的 PSRCM 来校正粒径引起的秸秆样本光谱差异。图 7 为几何粒径秸秆样本光谱信号中的 nPSRSI 与 PSRSI 结果。

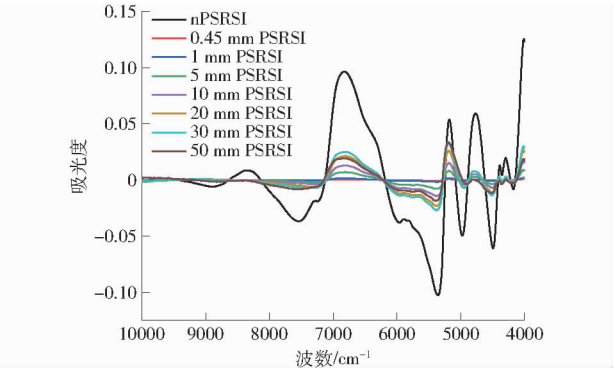
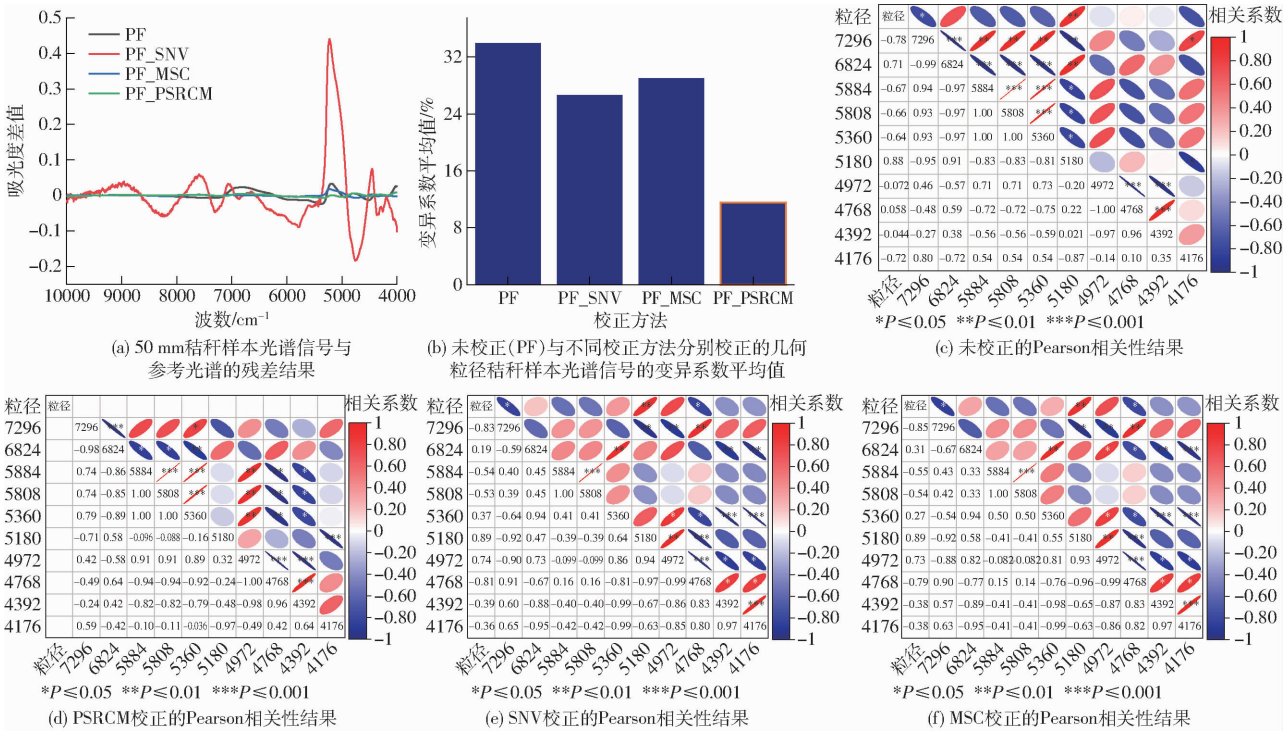


图 7　混合光谱信号的分离与解析结果

Fig. 7　Separation and analysis results of mixed spectral signals

由图 7 可知，PSRCM 分离了秸秆样本光谱信号中的 PSRSI 与 nPSRSI。在波段 10 000 ~ 4 000 cm^{-1}，随着粒径的增加，0.45、1、5、10、20、30、50 mm 秸秆样本光谱信号中的 PSRSI 逐渐增大，其中 0.45 mm 秸秆的 PSRSI 最小，50 mm 秸秆样本的 PSRSI 最大。

由图 8 可知，相比未校正，SNV 校正的光谱信号残差波动较大，尤其在波段 6 500 ~ 5 500 cm^{-1}、5 360 ~ 4 920 cm^{-1}，MSC 与 PSRCM 校正的几何粒径秸秆样本光谱信号残差整体变化幅度较为平稳，靠近水平线；以未校正为基准，PSRCM 校正的几何粒径秸秆样本光谱信号的变异系数平均值明显低于 SNV 与 MSC；PSRCM 校正了全部粒径相关的光谱特征信号，尤其是关键光谱特征信号5 180 cm^{-1}，相比未校正光谱，SNV 校正和 MSC 校正未能有效校正部分粒径相关的光谱特征信号。上述结果表明，PSRCM 整体的校正效果明显优于 SNV 与 MSC。这可能是因为 SNV 是对每一条秸秆样本的光谱信号分别求取各自的平均值和标准差，然而，粒径使秸秆样本光谱信号的吸光度在不同光谱波段处存在较大差异。尽管 MSC 采用最小二乘拟合机制将每一条秸秆样本光谱与参考光谱（通常为平均光谱）拟合，但是，MSC 保留了全部的秸秆样本光谱信息。相比之下，PSRCM 消除了秸秆样本光谱信号中的 PSRSI。



(a) 50 mm秸秆样本光谱信号与参考光谱的残差结果　(b) 未校正(PF)与不同校正方法分别校正的几何粒径秸秆样本光谱信号的变异系数平均值　(c) 未校正的Pearson相关性结果

(d) PSRCM校正的Pearson相关性结果　(e) SNV校正的Pearson相关性结果　(f) MSC校正的Pearson相关性结果

图 8　光谱差异校正结果

Fig. 8　Spectral difference correction results

3　结论

（1）本研究基于近红外光谱特征信号吸光度与粒径，提出了一种用于校正秸秆粒径的新方法。采用多项式拟合消除低频干扰信号。随后，利用单因素方差分析、主成分分析以及相关性分析方法进行

粒径相关光谱特征信号的深度挖掘,发现几何粒径的秸秆样本光谱信号 5 180 cm^{-1} 处表现极显著差异($P \leqslant 0.05$,$P \leqslant 0.01$,$P \leqslant 0.001$),光谱信息吸光度与粒度的相关系数高达 0.96。

(2)基于秸秆样本信号 5 180 cm^{-1} 处的吸光度与粒径,分别利用四参数逻辑斯蒂数学模型、二次多项式数学模型、指数衰减数学模型,以最小二乘法进行逐波数点拟合,结果表明,相比其它 2 种数学模型,二次多项式数学模型效果最优,其 R^2 与 RMSE 的平均值分别为 0.53 与 0.002 1。基于优化获取的二次多项式数学模型,假设吸光度由 nPSRSI 与 PSRSI 组成,通过一系列的矩阵变换与数学推理,建立了 PSRCM,其以残差、变异系数、相关性分析作为评价指标的校正效果明显优于 SNV 和 MSC。

参 考 文 献

[1] 刘翠玲, 徐金阳, 孙晓荣, 等. 基于无参数高效算法的近红外光谱模型传递研究[J]. 农业机械学报, 2023, 54(2): 396 - 402.
LIU Cuiling, XU Jinyang, SUN Xiaorong, et al. Near infrared spectroscopy calibration transfer based on parameter-free and efficient calibration enhancement algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 396 - 402. (in Chinese)

[2] ATANASKOVA E, ANGELOVSKA V, CHACHOROVSKA M, et al. Development of novel portable NIR spectroscopy process analytical technology (PAT) tool for monitoring the transition of ibuprofen to ibuprofen sodium during wet granulation process [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2024, 317: 124369.

[3] STANIMIROVA I, DASZYKOWSKI M, HOPKE P K. The role of chemometrics in improving clinical data analysis and diagnostics[J]. Trends Anal. Chem., 2024, 173: 117642.

[4] LI J, SUN L, LI R, et al. Application of siSVR - Vis/NIR to the nondestructive determination of acid detergent fiber content in corn straw[J]. Optik, 2020, 202: 163717.

[5] 李钰, 李欣荣, 刘保仓, 等. 近红外光谱技术分析茴香秸秆的营养组成[J]. 草业科学, 2025, 42(2): 433 - 442.
LI Yu, LI Xinrong, LIU Baocang, et al. Analysis of the nutritional composition of fennel straw by near-infrared spectroscopy [J]. Pratacultural Science, 2025, 42(2): 433 - 442. (in Chinese)

[6] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.

[7] 王动民, 纪俊敏, 高洪智. 多元散射校正预处理波段对近红外光谱定标模型的影响[J]. 光谱学与光谱分析, 2014, 34(9): 2387 - 2390.
WANG Dongmin, JI Junmin, GAO Hongzhi. The effect of msc spectral pretreatment regions on near infrared spectroscopy calibration results[J]. Spectroscopy and Spectral Analysis, 2014, 34(9): 2387 - 2390. (in Chinese)

[8] 杨丹, 刘新, 王川丕, 等. 绿茶样品粒度对近红外光谱图和模型的影响[J]. 食品科技, 2012, 37(11): 278 - 281.
YANG Dan, LIU Xin, WANG Chuanpi, et al. Effect of green tea particle size on near infrared spectrogram and nitrogen content model[J]. Food Science and Technology, 2012, 37(11): 278 - 281. (in Chinese)

[9] 贺城, 杨增玲, 黄光群, 等. 秸秆和煤混燃物中秸秆含量近红外光谱测定[J]. 农业机械学报, 2011, 42(10): 125 - 128,4.
HE Cheng, YANG Zengling, HUANG Guangqun, et al. Quantitative analysis of straw content in co-firing biomass-coal blends by near infrared spectroscopy[J]. Transactions of the Chinese Society of Agricultural Machinery, 2011, 42(10): 125 - 128,4. (in Chinese)

[10] 贺城, 韩鲁佳, 杨增玲, 等. 近红外光谱法在线检测秸秆和煤混燃物中秸秆含量[C]//中国农业工程学会 2011 年学术年会论文集, 2011: 1905 - 1908.

[11] 薛俊杰. 玉米、小麦秸秆原料、热解过程及固体产物特性 NIRS 快速分析研究[D]. 北京: 中国农业大学, 2016.
XUE Junjie. Study on rapid analysis of characteristics of raw material, process and solid product of pyrolysis of corn and wheat straw using NIRS[D]. Beijing: China Agricultural University, 2016. (in Chinese)

[12] WU R, TIAN Z, ZHANG C, et al. Uniformity evaluation of stem distribution in cut tobacco and single cigarette by near infrared spectroscopy[J]. Vibrational Spectroscopy, 2022, 121: 103401.

[13] SILALAHI D D, MIDI H, ARASAN J, et al. Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data[J]. Vibrational Spectroscopy, 2018, 97: 55 - 65.

[14] PHOOMWARIN K, SAENGPRACHATANARUG K, POSOM J, et al. Comparison of sample preparation methods for higher heating values in various sugarcane varieties using near-infrared spectroscopy[J]. Smart Agricultural Technology, 2025, 11: 101053.

[15] 黄梦曦. 近红外光谱(NIRs)在测定森林土壤碳氮磷含量中的应用研究[D]. 长沙: 中南林业科技大学, 2017.
HUANG Mengxi. Application of near infrared spectroscopy (NIRs) in determination of soil nitrogen, phosphorus and nitrogen in forest soil[D]. Changsha: Central South University of Forestry & Technology, 2017. (in Chinese)

[16] WANG K, WANG Z, XU H, et al. Experimental study on online detection of near-infrared spectroscopy suitable for continuous drug production[J]. Journal of Drug Delivery Science and Technology, 2025, 104: 106528.

[17] XIE C, QIAO M, YANG L, et al. Establishment of a general prediction model for protein content in various varieties and colors of peas using visible-near-infrared spectroscopy[J]. Journal of Food Composition and Analysis, 2024, 127: 105965.

[18] JOCHEMSEN A, ALFREDSEN G, MARTENS H, et al. Exploring the use of extended multiplicative scattering correction for near infrared spectra of wood with fungal decay[J]. Chemometrics and Intelligent Laboratory Systems, 2024, 252: 105187.

[19] AFSETH N K, KOHLER A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial[J]. Chemometrics and Intelligent Laboratory Systems, 2012, 117: 92 – 99.
[20] FREITAG S, ANLANGER M, LIPPL M, et al. Simplifying wheat quality assessment: using near-infrared spectroscopy and analysis of variance simultaneous component analysis to study regional and annual effects[J]. ACS Meas. Sci. Au, 2024, 4(6): 695 – 701.
[21] LI H, ZAN J, ZHANG L, et al. Improving the accuracy of NIR detection of moldy core in apples using different diameter correction methods[J]. Postharvest Biology and Technology, 2025, 219: 113279.
[22] YULIA M, SUHANDY D. The feasibility of using explicit method for linear correction of the particle size variation using NIR Spectroscopy combined with PLS2 regression method[J]. IOP Conference Series: Materials Science and Engineering, 2018, 334(1): 012062.
[23] 李阳, 彭彦昆, 李永玉. 基于距离校正和数据融合的苹果可溶性固形物含量预测模型优化[J]. 农业机械学报, 2024, 55(增刊): 336 – 345.
LI Yang, PENG Yankun, LI Yongyu. Optimization of apple soluble solids content prediction models based on distance correction and data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(Supp.): 336 – 345. (in Chinese)
[24] MI J, ZHANG L, ZHAO L, et al. Particle size regression correction for NIR spectrum based on the relationship between absorbance and particle size[J]. Frontiers of Optoelectronics, 2013, 6(2): 216 – 223.
[25] 高冰, 王梦妍, 姚玉梅, 等. 基于组成特性的肉骨粉种属鉴别标志性变量挖掘[J]. 农业工程学报, 2020, 36(18): 275 – 282.
GAO Bing, WANG Mengyan, YAO Yumei, et al. Markers mining for species discrimination based on component characteristics of meat and bone meal[J]. Transactions of the CSAE, 2020, 36(18): 275 – 282. (in Chinese)
[26] YAN C. A review on spectral data preprocessing techniques for machine learning and quantitative analysis[J]. iScience, 2025, 28(7): 112759.
[27] SHI Z, REN Z, YANG Z, et al. Deployment strategy of multiple miniaturized near-infrared spectrometers based on spectral transfer for characterizing soil organic matter and nitrogen[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2024, 320: 124620.
[28] GAO B, QIN Q, XU X, et al. Enhancement of species-specific analysis for meat and bone meal by matrix fragments-related spectral fusion[J]. Vibrational Spectroscopy, 2024, 130: 103644.
[29] 谷玲霄, 方涛, 杜林丹, 等. 基于 CA/SPACARS 算法的小麦条锈病特征波段优选与监测模型构建[J]. 农业机械学报, 2025, 56(6): 487 – 498.
GU Lingxiao, FANG Tao, DU Lindan, et al. Feature band selection and construction of monitoring model of wheat stripe rust based on CA/SPACARS algorithm[J]. Transactions of the Chinese Society of Agricultural Machinery, 2025, 56(6): 487 – 498. (in Chinese)
[30] 王晶晶. 基于近红外光谱的大米水分及蛋白质含量检测方法研究[D]. 哈尔滨: 东北农业大学, 2017.
WANG Jingjing. The research of moisture and protein content of the detection method in rice based on near infrared spectroscopy[D]. Harbin: Northeast Agricultural University, 2017. (in Chinese)
[31] 秦玉华. 烟叶通用近红外定量模型稳健性研究[D]. 青岛: 中国海洋大学, 2014.
QIN Yuhua. Research on robustness of near infrared tobacco quantitative model[D]. Qingdao: Ocean University of China, 2014. (in Chinese)
[32] 王海东,秦玉昌,吕小文,等. 近红外光谱技术(NIRS)测定玉米粉碎粒度的研究[J]. 饲料工业,2008,29(19):41 – 44.
[33] 林润瑞. 拉曼光谱技术结合机器学习算法在乳腺癌筛查中的应用[D]. 成都: 电子科技大学, 2024.
LIN Runrui. Application of raman spectroscopy combined with machine learning algorithms in breast cancer screening[D]. Chengdu: University of Electronic Science and Technology of China, 2024. (in Chinese)
[34] 耿姝. 适宜中药材体系的近红外分析方法影响因素研究[D]. 杭州: 浙江大学, 2016.
GENG Shu. Studies on influential factors of near-infrared spectroscopy for chinese medicinal materials system[D]. Hangzhou: Zhejiang University, 2016. (in Chinese)
[35] PASIKATAN M C, STEELE J L, SPILLMAN C K, et al. Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials[J]. J. Near Infrared Spectrosc., 2001, 9(3): 153 – 164.
[36] PARK M, YU J Y, KO J A, et al. Application of UV – Vis – NIR and FTIR spectroscopy coupled with chemometrics for quality prediction of katsuobushi based on the number of smoking treatments[J]. Food Chemistry, 2024, 442: 138604.
[37] WILSON R H, NADEAU K P, JAWORSKI F B, et al. Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization[J]. J. Biomed Opt., 2015, 20(3): 030901.
[38] SANG V Q, 冯鹏, 汤斌, 等. 基于米氏散射理论的水中悬浮颗粒物散射特性计算[J]. 激光与光电子学进展, 2015, 52(1): 231 – 238.
SANG V Q, FENG Peng, TANG Bin, et al. Study on properties of light scattering based on Mie scattering theory for suspended particles in water[J]. Laser & Optoelectronics Progress, 2015, 52(1): 231 – 238. (in Chinese)