基于 THz-ATR 光谱的玉米种子水分定量模型优化方法研究

吴静珠[1]，李晓琪[1]，孙丽娟[2]，刘翠玲[1]，孙晓荣[1]，孙 梅[1]，余 乐[1]

1. 北京工商大学食品安全大数据技术北京市重点实验室，北京 100048
2. 中国农业科学院作物科学研究所，北京 100081

摘 要 应用太赫兹时域光谱技术结合区间偏最小二乘法筛选玉米种子水分 THz 特征波段，并采用支持向量机构建基于特征谱区的抗非线性干扰的种子水分快速定量分析模型。实验以郑单 958 玉米种子为例，制备含水量范围 9.58%～12.71%的种子粉末样本 40 组(每组取样 3 份)，采用衰减全反射(ATR)附件扫描得到 120 份样本太赫兹时域光谱，根据 SPXY(光谱-理化值共生距离算法)法划分得到训练集样本 90 份，测试集样本 30 份。种子水分对太赫兹波具有强烈吸收，首先采用基于偏最小二乘线性回归的移动区间(mwPLS)、独立区间(iPLS)、后向区间(biPLS)和联合区间(siPLS)方法筛选最优特征谱区组合；鉴于环境水分、种子其他成分及系统噪声对种子水分太赫兹光谱存在不可避免的非线性干扰，在上述光谱特征区间进一步采用基于 RBF 核函数的支持向量机和网格搜索法构建得到预测性能最优的种子水分快速定量分析非线性模型，训练集均方根误差为 0.021 2，预测集均方根误差为 0.069 7，相对分析误差为 12.345 7，相较于传统偏最小二乘线性回归模型，模型性能得到提升。种子水分含量是影响种子贮藏安全和种子活力的重要因素，实验结果表明：太赫兹时域光谱结合化学计量学可以有效筛选种子水分特征吸收谱区，建立抗干扰、高精度的种子水分快速定量分析模型，有望成为未来种子质量快速测定领域一项极具应用潜力的补充技术。

关键词 太赫兹时域光谱；衰减全反射；种子水分；谱区筛选；支持向量机
中图分类号：O433.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2021)07-2005-07

引 言

种子水分是我国《农作物种子质量标准》(GB4404.1—2008)规定的四大必检项目之一。水分含量与种子生理变化过程，如种子萌发、成熟及贮藏等，都有着密切的关系[1]。水分含量过高容易加快种子的劣变死亡；较低的水分含量则有利于保持种子活力的稳定性。因此实时、精准地快速获取种子水分含量对于农业生产领域的选种、育种、增产增收具有非常重要的意义[2]。

传统的种子水分测定方法如烘干法(GB/T 3543.6—1995)具有较高的检测精度，但检测时间较长、操作过程繁杂，难以满足现代农业发展提出的快速检测需求。近年来蓬勃兴起的光谱技术以其快速、绿色、便捷、多组分分析等特点已然成为种子质量检测领域的研究热点，其中太赫兹(Terahertz，THz)时域光谱技术以其丰富的指纹谱性、安全性和穿透性，有望成为该领域继近红外、X 光技术之后的一项极具应用潜力的补充技术[3-5]。

有报道应用太赫兹技术结合化学计量学方法在玉米、水稻和棉花种子转基因鉴别[6-8]、向日葵种子饱满度判别[9]、储粮品质判别等[10]方面均取得了极具应用前景的研究结论。水对太赫兹波具有强烈吸收，原因在于水是极性分子，且水分子通过四个氢键和其附近的水分子连接，形成一个局域四面体结构。当太赫兹波穿过水分子时，水网络结构中的氢键受激产生共振，水分子偶极发生旋转取向，并经弛豫形成新的氢键网络，水分子间在皮秒量级时间内可以发生多种相互作用所致[11]。有报道采用太赫兹时域光谱最小值成像建立绿萝叶片水分含量多元回归预测模型，预测集相关系数达 0.989 1，预测均方根误差为 0.024 4。步正延等[12]采用自适应阈值分割法对 THz 特征波段下的大豆叶片光谱图像进行分割，并分别建立了基于多元线性回归(MLR)、反向传播神经网络(BP-ANN)和最小二乘支持向量机(LS-SVM)的大豆叶片水分含量预测模型，实验结果表明，基于叶肉特征的 LS-SVM 模型具有较高的预测精度，为作物叶片水分含量测定提供了一种行之有效的检测手段。

应用太赫兹时域光谱技术检测种子水分含量具有较好的

收稿日期：2021-02-20，**修订日期**：2021-05-09
基金项目：国家自然科学基金项目(61807001)资助
作者简介：吴静珠，女，1979 年生，北京工商大学食品安全大数据技术北京市重点实验室教授 e-mail：pubwu@163.com

理论基础和可行性，但是由于环境水分对太赫兹波存在强烈吸收，且种子主要成分如蛋白、脂肪、淀粉等均存在 THz 吸收[13-14]，此外系统自身噪声[15]等多重因素，都给种子水分 THz 光谱表征和准确预测造成了严重困扰，因此应用 THz 技术直接进行种子水分检测目前鲜有报道。本文以玉米种子为例，重点探索采用太赫兹衰减全反射(attenuated total reflection，ATR)技术结合化学计量学方法建立抗干扰性强、预测性能优的玉米种子水分定量分析模型，以期为基于 THz 技术的种子成分精确测定提供理论基础和技术支撑。

1 实验部分

1.1 样本制备

实验选取郑单 958 玉米种子。为制备不同含水率样本，将种子样本置于 40 ℃，100%相对湿度环境中，每隔 2 h 得到一个批次样本，将该批次样本晾晒 0，2，4，6 和 8 h 后分别取样，共计制备 8 批次 40 组样本，每份样本 100 g。

应用 FW-200 高速万能粉碎机将种子样本粉碎，采用梅特勒-托利多 HB43-S 卤素水分测定仪对样本进行快速水分测定。40 组样本含水率统计信息如表 1 所示。

表 1 样本含水率统计信息

Table 1 Statistical information of moisture content of samples

样本容量	最小值/%	最大值/%	平均值/%	标准差	极差/%	变异系数/%
40	9.58	12.71	11.04	0.82	3.13	7.42

1.2 THz 光谱采集

相较于压片透射方式，ATR 方式无需制样前处理，所需样本量少，采集过程更为方便，快捷，本实验采用 TeraPulse 4000 型太赫兹时域光谱系统及 ATR 附件采集样本光谱。光谱仪参数设定：光谱范围为 0.2～359.94 cm^{-1}，分辨率为 0.94 cm^{-1}，单个样本扫描次数为 900 次取平均。实验环境温度为 22 ℃，为研究环境水分对模型的影响，实验过程中不采用任何吹扫系统来去除环境水分。40 组样本每组分别取样 3 次，共采集得到 40×3 份样本光谱曲线。以第一个批次样本分别晾晒 0，2，4，6 和 8 h 后的样本光谱的时域和频域谱为例，如图 1 所示。

根据图 1(a)：样本时域信号峰值约在－6 ps 处相对参考信号产生明显右移，说明 THz 波通过一定厚度的样本产生了时延；且样本时域信号峰值明显低于参考信号峰值，表明样本对 THz 波有强烈吸收，但是在不同时延处，样本信号表现有所差异，如在－8 ps 处，信号值大致随着晾晒时长降低，但是在－5 ps 处，信号值又呈现出反向的变化趋势。将时域信号经快速傅里叶变换(fast Fourier transform，FFT)运算后计算得到功率谱如图 1(b)所示：在整个谱区范围内参考信号略高于样本信号；不同的玉米样本所含成分相同，只是成分含量各有差异，因此样本光谱曲线相似；与未经过任何样本的参考信号相比，光谱曲线存在差异，尤其是在小于 20 cm^{-1}的范围内。通过太赫兹光学参数计算得到第一个批次不同含水率样本的吸光度谱，如图 2 所示。

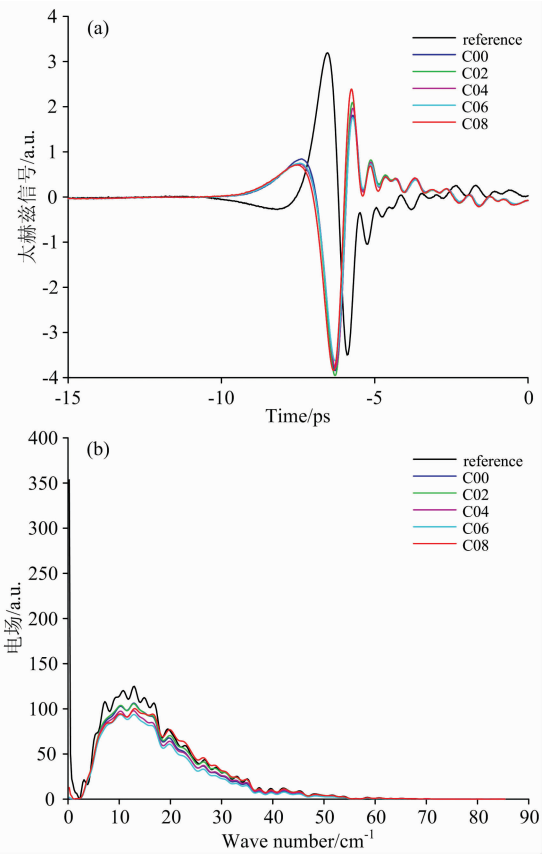

图 1 第一批次不同含水率样本时域光谱及频域光谱

(a)：时域谱；(b)：频域谱

Fig. 1 Time domain spectra and frequency domain spectra of the first batch of samples with different water contents

(a)：Time domain spectra；

(b)：Frequency domain spectra

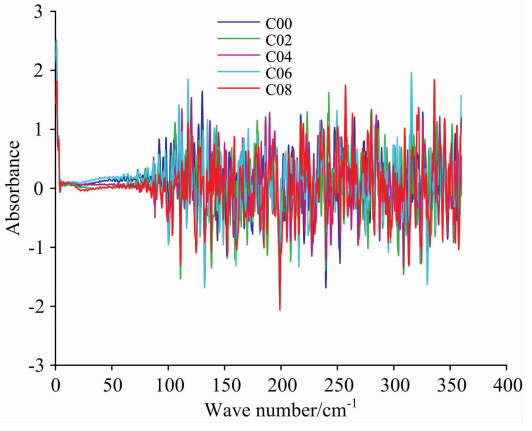

图 2 第一批次不同含水率样本吸光度谱

Fig. 2 Absorbance spectra of the first batch of samples with different water content

从图 2 中可以看出在 0.2～85.37 cm^{-1}范围内，随着波

数的增加，样品吸光度谱线变化较为一致且平缓，但谱线上并无肉眼可观的明显吸收峰；在 85.37～359.94 cm^{-1} 范围内，谱线突然出现极为剧烈地、杂乱无章的变化，难以分析光谱的规律性变化。综上光谱特点，实验通过光谱信号处理、筛选有效光谱区间来消除仪器噪声，净化谱图，提升光谱信噪比，以期为构建种子水分定量分析模型提供高质量的基础数据。

1.3　数据处理方法

1.3.1　SPXY 数据集划分

SPXY (sample set partitioning based on joint x-y distance)方法[16] 是在 KS(Kennard-Stone)法的基础上提出的，在样品间距离计算时将 x 变量和 y 变量同时考虑在内，以保证最大程度表征样本分布，有效地覆盖多维向量空间，增加样本间的差异性和代表性，提高模型稳定性。其距离公式如式(1)—式(3)

$$d_x(p,q)=\sqrt{\sum_{j=1}^{N}[x_p(j)-x_q(j)]^2};\ p,q\in[1,N] \quad (1)$$

$$d_y(p,q)=\sqrt{(y_p-y_q)^2}=|y_p-y_q|;\ p,q\in[1,N] \quad (2)$$

$$d_{xy}(p,q)=\frac{d_x(p,q)}{\max\limits_{p,q\in[1,N]}d_x(p,q)}+\frac{d_y(p,q)}{\max\limits_{p,q\in[1,N]}d_y(p,q)};\ p,q\in[1,N] \quad (3)$$

式中，x 表示吸光度谱数据，y 表示含水率的真实值。p，q 为样本编号，N 表示样本的光谱点数量。SPXY 是用 $d_{xy}(p,q)$ 代替了 $d_x(p,q)$，同时为了确保样本在 x 和 y 空间具有相同权重，因此将 $d_x(p,q)$ 和 $d_y(p,q)$ 分别除以他们在数据集中的最大值得到标准化后的距离公式 $d_{xy}(p,q)$。

本实验采用 SPXY 法对原始样本集按照 3∶1 比例划分训练集和测试集，即训练集样本数为 90，测试集样本数为 30。

1.3.2　光谱预处理

由于 THz 光谱中存在较为明显的噪声，首先采用移动窗口平滑(Smooth)、多元散射校正(multiplicative scatter correction，MSC)、卷积平滑(savitzky-golay，SG)以及标准正态矢量变换(standard normal variate transform，SNV)进行数据预处理，筛选出适合于本实验数据的预处理方法。

1.3.3　偏最小二乘回归

偏最小二乘回归法(partial least squares regression，PLSR)是目前分子光谱化学计量学方法中最为经典的线性回归建模算法。由于种子水分对太赫兹波有强烈的正相关吸收，采用 PLSR 来建立种子水分定量预测线性模型，通过 iPLS，biPLS，siPLS 和 mwPLS 四种方法筛选表征种子水分最佳 THz 特征谱区。

1.3.4　支持向量机回归

支持向量机的核心思想是以结构风险最小化为原则在特征空间建立最优分类面完成模型的训练，具有良好的泛化性能和强非线性逼近能力，非常适合于小样本，维度高且非线性的样本分类。支持向量回归(support vector regression，SVR)将原始样本数据投射到高维的特征空间，再通过线性回归在高维空间中寻求最优的回归超平面，使得预测值与真实值偏差尽可能小，从而实现预测和数据分析的目的[17]。

为消除环境水分，种子其他成分以及系统自身噪声等引起的种子水分 THz 特征谱区上的非线性干扰，采用基于 RBF 核函数的支持向量机和网格搜索法优化建立种子水分 THz 特征谱区的非线性模型，进一步提高种子水分模型的预测准确性和抗干扰性。

1.3.5　模型评价指标

采用相关系数(r)，训练集均方根误差(root mean square error of calibration，RMSEC)，预测集均方根误差(root mean square error of prediction，RMSEP)和相对分析误差(residual predictive deviation，RPD)评价模型的性能[18]。相对分析误差计算公式如式(4)所示

$$\mathrm{RPD}=\frac{\mathrm{SD}}{\mathrm{RMSEP}} \quad (4)$$

SD 表示测量值的标准差，RMSEP 为分析样品的预测均方根误差。在样本待测组分标准差一致的前提下，RPD 值越大，说明模型的准确性越高。通常认为，若 RPD＜1.4，表明所建模型预测结果不可靠；若 1.4＜RPD＜2.0，表明模型的预测结果可以接受；若 RPD＞2.0，则表明模型的预测准确性很高，能够用于模型分析。

2　结果与讨论

2.1　基于 PLSR 的种子水分定量线性模型的建立与分析

根据图 2，样本吸光度谱在 85.37 cm^{-1} 附近出现了明显变化，＞85.37 cm^{-1} 的谱线变化剧烈，无规律性。因此实验分别以 0.2～359.94 和 0.2～85.37 cm^{-1} 谱区信息作为输入，采用不同方法对光谱进行预处理后比较建模结果，如表 2 所示。

从表 2 中可以看出：(1)在不同光谱预处理方式下，基于 0.2～359.94 cm^{-1} 谱区建立的种子水分 PLSR 模型自预测性能较好，但是泛化性能很差；(2)基于 0.2～359.94 cm^{-1} 建立的所有模型性能均劣于基于 0.2～85.37 cm^{-1} 谱区模型，说明＞85.37 cm^{-1} 的谱区包含噪声较大，导致模型预测性能较差；(3)在 0.2～85.37 cm^{-1} 谱区内，所有模型的 RPD＞3，说明模型具有较高的实用性和可靠性，其中经 Smooth 光谱预处理后建立的模型相较于其他预处理方法具有更好的预测性能，其相关系数 r 为 0.996 9，RMSEP 为 0.198 6，RPD 为 4.323 4。

2.2　基于区间 PLSR 的种子水分 THz 特征谱区筛选

由于在 0.2～85.37 cm^{-1} 范围内，经 Smooth 预处理所建模型性能较优，因此进一步考虑在该范围内分别采用 iPLS，biPLS，siPLS 和 mwPLS 四种方法筛选表征种子水分 THz 特征谱区，结果如图 3 所示。

iPLS 将光谱分别划分为 10～30 个子区间的情况下建立 iPLS 谱区筛选模型，综合考虑 r 和 RMSECV，最终确定在光谱划分为 10 个子区间，并采用第 2 个子区间进行建模时，模型性能最佳。THz 特征波段筛选结果如图 3(a)所示，区间对应的波数范围为 8.63～17.02 cm^{-1}。

表 2 种子水分 PLSR 模型预测结果

Table 2 Results of seed moisture PLSR model predictions

波数范围/cm^{-1}	变量个数	预处理	主成分数	r	RMSEC	RMSEP	RPD
1.2～359.94	1 502	—	10	0.863 4	0.002 2	0.703 8	1.298 8
		Smooth	10	0.805 2	0.013 5	0.821 0	1.127 7
		MSC	10	0.859 9	0.005 8	0.818 3	1.016 4
		SG	10	0.844 1	0.004 0	0.740 9	1.252 7
		SNV	10	0.864 4	0.003 8	0.713 3	1.221 9
0.2～85.37	357	—	4	0.996 5	0.158 6	0.209 9	4.239 8
		Smooth	4	0.996 9	0.179 4	0.198 6	4.323 4
		MSC	3	0.995 7	0.183 3	0.232 6	3.888 2
		SG	4	0.995 4	0.151 8	0.239 3	3.829 9
		SNV	2	0.995 9	0.201 2	0.226 3	3.866 7

biPLS 将光谱分别划分为 10～30 个子区间的情况下建立 biPLS 谱区筛选模型，综合考虑 r 和 RMSECV 的值，biPLS 谱区筛选模型在将光谱划分为 27 个子区间，2 个子区间(1，8)联合所建模型最优。THz 特征波段筛选结果如图 3(b)所示，区间对应的波数范围为 0.2～3.11 和 23.26～26.13 cm^{-1}。

siPLS 将光谱分别划分为 10～30 个子区间，在相同区间划分数下，又分别尝试联合其中 2 个，3 个和 4 个子区间建立模型。最终在将光谱区域划分 15 个子区间，采用 4 个子区间联合(2，3，4，10)建模性能最佳。THz 特征波段筛选结果如图 3(c)所示，区间对应的波数范围为：5.75～11.27，11.51～17.02，17.26～22.78，51.79～57.31cm^{-1}。

mwPLS 以窗口宽度为 11 作为初始窗口宽度，设窗口滑动步长为 1，窗口宽度增加步长为 10，逐步将窗口宽度增加至 121，分别建立不同窗口宽度下的谱区筛选模型，当窗口宽度为 51 时，RMSECV 值达最小。THz 特征波段筛选结果如图 3(d)所示，区间对应的波数范围为 11.75～18.94 cm^{-1}。

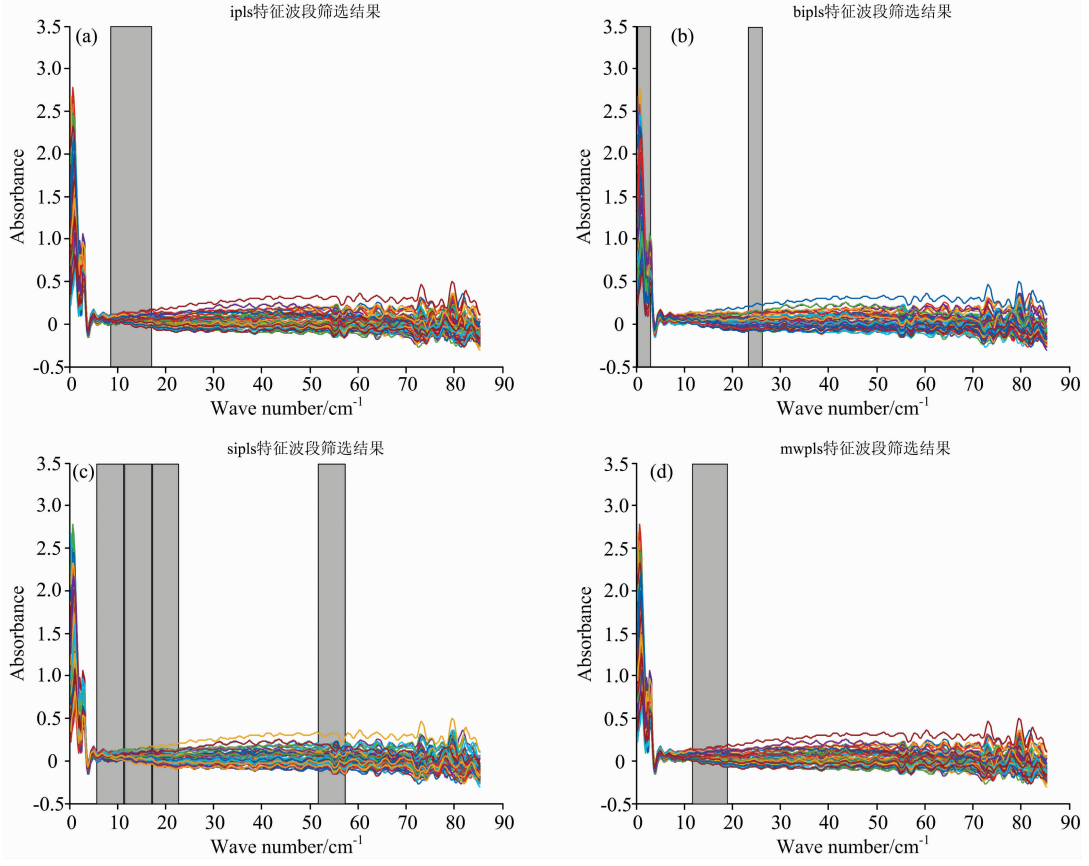

图 3 基于区间 PLSR 的种子水分 THz 特征谱区

(a)：iPLS；(b)：biPLS；(c)：siPLS；(d)：mwPLS

Fig. 3 THz characteristic spectral region of seed moisture based on interval PLSR

(a)：iPLS；(b)：biPLS；(c)：siPLS；(d)：mwPLS

特征波段筛选结果如表 3 所示：四种方法筛选得到共同 THz 谱区在 10～20 cm^{-1}范围内，siPLS 方法建立的模型相关系数 r 最高，交互验证均方根误差最小，其筛选得到的部分谱区 51.79～57.31 cm^{-1}与理论模拟所预测的水在太赫兹波段的弱吸收峰(60 cm^{-1}附近)接近[19]。

2.3　基于 SVR 和特征谱区的种子水分定量非线性模型的建立与优化

采用上述筛选的特征谱区建立 PLSR 模型结果如表 4 所示。与表 2 中基于 0.2～85.37 cm^{-1}所建 PLSR 模型相比，输入数据变量个数减少，RPD 均大于 2，但模型的预测误差却有所增大。

考虑到环境水分影响、种子中非目标成分的干扰以及系统噪声等，实验在上述四种方法已筛选得到与种子水分密切相关的 THz 特征谱段上，分别采用 SVR 非线性模型进行建模分析。SVR 核函数为 RBF，惩罚参数 c 和核参数 g 采用网格搜索法进行寻优，设定惩罚参数 c 和核参数 g 的范围为 $[2^{-10}, 2^{10}]$，步长为 0.5。模型预测结果如表 5 所示。

表 3　特征波段筛选结果

Table 3　Characteristic band screening results

方法	主因子数	r	RMSECV	特征波段/cm^{-1}
iPLS	5	0.658 1	0.631 0	8.63～17.02
biPLS	10	0.662 9	0.606 9	0.2～3.11，23.26～26.13
siPLS	9	0.762 0	0.526 2	5.75～11.27，11.51～17.02，17.26～22.78，51.79～57.31
mwPLS	5	0.644 9	0.585 4	11.75～18.94

表 4　基于特征区间的 PLSR 线性模型预测结果

Table 4　Prediction results of PLSR linear model based on feature intervals

建模方法	特征变量筛选	变量个数	主因子数	r	RMSEC	RMSEP	RPD
PLSR	iPLS	36	3	0.996 5	0.202 0	0.209 8	4.088 9
	biPLS	27	2	0.995 7	0.215 4	0.234 2	3.861 3
	siPLS	96	3	0.995 6	0.215 6	0.234 1	3.675 5
	mwPLS	31	2	0.994 4	0.243 4	0.266 5	3.368 1

表 5　基于特征区间的 SVR 非线性模型预测结果

Table 5　Prediction results of SVR nonlinear model based on characteristic intervals

建模方法	特征变量筛选	变量个数	c	g	r	RMSEC	RMSEP	RPD
SVR	iPLS	36	4	2	0.984 5	0.025 3	0.103 4	8.294 9
	biPLS	27	90.509 7	0.176 8	0.870 3	0.056 8	0.340 9	2.652 6
	siPLS	96	2.828 4	0.353 6	0.993 0	0.021 2	0.069 7	12.345 7
	mwPLS	31	16	2.828 4	0.967 8	0.018 5	0.145 6	6.162 7

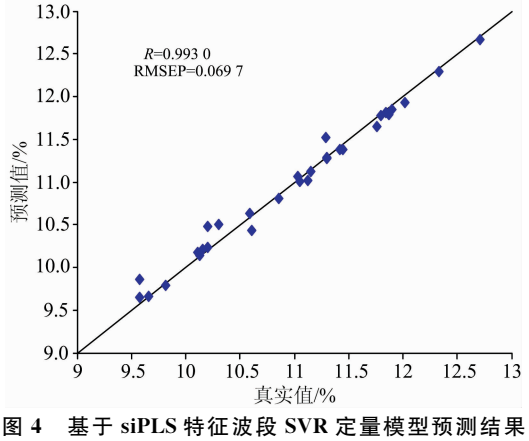

图 4　基于 siPLS 特征波段 SVR 定量模型预测结果

Fig. 4　Quantitative model prediction results based on siPLS feature band SVR

从表 5 可得，根据 iPLS，biPLS，siPLS 和 mwPLS 筛选的特征谱区建立的 SVR 模型，相较于之前表 2、表 4 建立的 PLSR 模型，预测性能得到了明显提升，该模型不但具有较少变量输入个数，而且具有更高的相关系数值及更低的预测均方根误差。其中基于 siPLS 筛选波段所建立的 SVR 模型性能最优，其测试集的真值和预测值相关性如图 4 所示，RMSEP 为 0.069 7，r 为 0.993 0。理论研究表明水在太赫兹波段的吸收峰在 60 cm^{-1}附近[24-25]，确实与优选的波长对应的吸收峰相近。因此在种子水分的 THz 检测中，种子水分对 THz 波的强烈吸收占据了 THz 光谱的主要分量，而环境水分、种子非目标成分以及系统噪声等非线性干扰在 THz 光谱中占据了次要分量，因此首先采用线性模型筛选特征谱区可以最大程度地保留水分的 THz 特征，然后再结合非线性建模方法有效地降低非目标信息的非线性干扰，从而有效提升模型性能。

3　结　论

种子水分对太赫兹波有敏感吸收。采用 THz 时域光谱系统及其衰减全反射附件测得 120 组不同含水率样本的吸光度谱，在 0.2～85.37 cm^{-1}范围内，经 Smooth 光谱预处理后建立种子水分的 PLSR 线性模型具有较好的预测性能；在此基础上分别应用 iPLS，biPLS，siPLS 和 mwPLS 特征谱区筛

选表征玉米种子水分 THz 特征谱区组合；研究构建基于特征谱区的 SVR 非线性模型能有效降低环境水分、种子非目标成分以及系统噪声的干扰，大幅度提高模型预测性能，可实现对玉米种子水分含量快速、精准定量检测，为采用太赫兹时域衰减全反射技术结合化学计量学算法在种子质量检测领域提供了思路与方法。

References

[1] LI Zhen-hua, WANG Jian-hua(李振华, 王建华). Sciences Agricultural Sinica(中国农业科学), 2015, 48(4): 646.

[2] WANG Jian-hua(王建华). Maize Seed Quality Evaluation Manual(玉米种子质量评价手册). Beijing: China Agricultural University Press(北京: 中国农业大学出版社), 2015.

[3] YAO Jian-quan(姚建铨). Journal of Chongqing University of Posts and Telecommunications · Natural Science Edition(重庆邮电大学学报 · 自然科学版), 2010, 22(6): 703.

[4] ZHANG Cun-lin, MU Kai-jun (张存林, 牧凯军). Progress in Laser and Optoelectronics(激光与光电子学进展), 2010, 47(2): 1.

[5] Li Bin, Hua Kai, Shen Yin. IEEE Access, 2020, (8): 56092.

[6] Lian Feiyu, Xu Degang, Fu Maixia, et al. IEEE Transactions on Terahertz Science and Technology, 2017, 7(4): 378.

[7] Liu Wei, Liu Changhong, Hu Xiaohua, et al. Food Chemistry, 2016, 210: 415.

[8] Liu Jianjun, Li Zhi, Hu Fangrong, et al. Opt. Quant. Electron., 2015, 47: 685.

[9] Sun X, Liu J. Journal of Infrared, Millimeter, and Teraherz Waves, 2020, 41: 307.

[10] Ge Hongyi, Jiang Yuying, Xu Zhaohui, et al. Optics Express, 2014, 22(10): 12533.

[11] LIU Li-ping, WANG Yu-fei, ZHANG Ya-zhou, et al(刘丽萍, 王煜斐, 张亚洲, 等). Advances in Analytical Chemistry(分析化学进展), 2018, 8(1): 1.

[12] BU Zheng-yan, LI Zhen-feng, SONG Fei-hu, et al(步正延, 李臻峰, 宋飞虎, 等). Acta Agriculturae Zhejiangensis(浙江农业学报), 2018, 30(8): 1420.

[13] LIANG Chuan, QI Shu-ye, LI Xi-ran, et al(梁 川, 戚淑叶, 李曦染, 等). Food Safety and Quality Detection Technology(食品安全质量检测学报), 2014, 5(3): 730.

[14] SHEN Xiao-chen, LI Bin, LI Xia, et al(沈晓晨, 李 斌, 李 霞, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(S1): 288.

[15] ZHANG Ji-yang, REN Jiao-jiao, CHEN Si-hong, et al(张霁旸, 任娇娇, 陈思宏, 等). Chinese Journal of Lasers(中国激光), 2020, 47(1): 326.

[16] Roberto K H G, Mario C U A, Gledson E J, et al. Talanta, 2005, 67(4): 736.

[17] Yun Xue, Bin Zou, Yimin Wen, et al. Sustainability, 2020, 12(11): 4441.

[18] Mohammed Kamruzzaman, Yoshio Makino, Seiichi Oshita. LWT-Food Science and Technology, 2016, 66: 685.

[19] FAN Shu-ting, MA Ying-yu, SHU Guo-xiang, et al(范姝婷, 马莹玉, 舒国响, 等). Journal of Shenzhen University · Science and Engineering(深圳大学学报 · 理工版), 2019, 36(2): 200.

Study on the Optimization Method of Maize Seed Moisture Quantification Model Based on THz-ATR Spectroscopy

WU Jing-zhu[1], LI Xiao-qi[1], SUN Li-juan[2], LIU Cui-ling[1], SUN Xiao-rong[1], SUN Mei[1], YU Le[1]

1. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China
2. Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract Characteristic Terahertz (THz) bands of maize seed moisture were screened using the Terahertz time-domain spectroscopy technique combined with the interval partial least squares method. The support vector machine was used to construct a rapid quantitative analysis model of seed moisture based onthe characteristic spectral region against nonlinear interference. Take Zhengdan 958(Corn variety), for example, in this experiment, 40 sets of seed powder samples (3 samples from each set) with moisture content ranging from 9.58% to 12.71% were prepared. Terahertz time-domain spectra of 120 samples were collected by Terapluse 4 000 terahertz time-domain system with Attenuated Total Reflection (ATR) module. According to the SPXY method, 90 training set samples and 30 test set samples were obtained. Given the strong absorption of terahertz waves by seed moisture, the moving interval (mwPLS), independent interval (iPLS), backward interval (biPLS) and synergy interval (siPLS) methods based on partial least squares linear regression were firstly used to screen the optimal combination of the characteristic spectral regions. In view of the inevitable nonlinear interference of environmental moisture, other seed components and systematic noise on the terahertz spectrum of seed moisture, a nonlinear model for rapid quantitative analysis of seed moisture with optimal prediction performance was further constructed using support vector machine and grid search method based on RBF kernel function on the above spectral feature intervals. The optimal SVR model was obtained with a lower root mean square error of the training set (RMSEC) of 0.021 2, a lower root mean square error of the prediction (RMSEP) of 0.069 7 and a higher residual predictive deviation (RPD) of 12.345 7. The model performance was significantly improved compared with the traditional partial least squares linear regression model. Seed moisture content is an important factor in seed storage safety and seed vigour. The experimental results show that THz time-domain spectroscopy combined with the chemometric method can effectively be used to screen the characteristic absorption spectral region of seed moisture and establish an interference-resistant and high-precision model for rapid quantitative analysis of seed moisture, which is expected to be a Promising complementary technology in the field of rapid seed quality determination.

Keywords Terahertz time-domain spectral; Attenuated total reflection; Seed moisture; Spectral region screening; Support vector machine

(Received Feb. 20, 2021; accepted May 9, 2021)