

基于太赫兹光谱数据融合实现多组分橡胶添加剂的定量检测

殷贤华^{1,2}, 陈慧聪^{1,2}, 张活^{1,2*}

¹桂林电子科技大学电子工程与自动化学院, 广西 桂林 541004;

²广西自动检测技术与仪器重点实验室, 广西 桂林 541004

摘要 橡胶添加剂含量是判断橡胶质量好坏的一个重要标志。现有的检测方法不能满足橡胶中添加剂快速、准确、无损的检测需求,而且当检测对象为多组分混合物时,混合物的吸收光谱会出现重叠和失真现象,从而导致预测结果不准确。针对此问题,本文提出了一种基于太赫兹时域光谱结合化学计量学方法、数据融合的定量分析方法。以丁腈橡胶、白炭黑、氧化锌、防老剂 H 和防老剂 MB 制成五组分混合物并将其作为实验样品,利用太赫兹时域光谱系统获取并计算了五组分混合物在 0.3~1.6 THz 范围内的吸收光谱,然后通过求一阶导数的方式获得样品的导数光谱数据。低层数据融合直接将吸收光谱数据与导数光谱融合;中层数据融合将蒙特卡罗无信息变量消除法和连续投影算法进行特征提取后的变量进行融合;高层数据融合使用多元线性回归法进行融合。基于蒙特卡罗无信息变量消除法的中层数据融合的预测精度高于单一光谱的预测精度,而且预测性能最好。研究结果表明,太赫兹光谱结合支持向量回归、数据融合的方法可以实现多组分混合物中防老剂 MB 的快速、准确、无损检测,对于促进橡胶工业的快速发展具有十分重要的意义。

关键词 光谱学; 橡胶及添加剂; 多组分混合物; 太赫兹光谱; 定量检测; 数据融合

中图分类号 O436

文献标志码 A

DOI: 10.3788/CJL230807

1 引言

我国是轮胎橡胶生产大国,其产量在一定程度上影响着国民经济的发展。丁腈橡胶(NBR)具有耐油、耐磨、耐热、黏性强、气密性好、抗静电等性能,已被广泛应用于军工、汽车、航空等领域。氧化锌(ZnO)是轮胎制造传统配方中不可缺少的一种添加剂,但其过量使用会对人体、环境以及水中对锌敏感的动植物产生危害^[1]。白炭黑也是轮胎生产中的重要添加剂之一,其含量影响着橡胶的性能。橡胶是一种高分子材料,会因长期使用而老化。为减缓橡胶的老化速度,厂家会在橡胶中添加防老剂。随着人们保护环境意识的不断提高,防老剂 H 逐渐被无毒无污染的防老剂 MB 替代。目前,防老剂 MB 含量的检测方法主要有燃烧检验法、化学分析法、色谱法以及红外光谱法等^[2],但这些检测方法都存在前期处理过程复杂、检出速度慢、不能准确反映添加剂的实际含量等缺点。太赫兹时域光谱技术是一种新兴的光谱技术,具有操作简单、分析时间相对较短、结果稳定等独特优点,已得到了广泛应用。迄今为止,国内外已在橡胶^[3]、食品^[4-7]、药品^[8-9]检

测等领域对该技术进行了相应的研究与应用。有人利用太赫兹时域光谱技术对多组分混合物进行了定量检测,如:Komatsu等^[10]利用太赫兹光谱技术对三组分乙丙橡胶中滑石粉与氢氧化镁的含量进行了检测;Ma等^[11]利用太赫兹光谱技术和反向神经网络对小麦粉中三元农药混合物进行了定量检测;Sun等^[12]采集了4种双酚类混合物的太赫兹光谱,分析了它们的吸收特性,并结合支持向量回归对其进行了定量分析;殷贤华等^[13]利用太赫兹时域光谱技术测定了四组分橡胶添加剂混合物的太赫兹光谱,然后采用Zernike矩方法对混合物中目标成分的特征信息进行提取,并结合偏最小二乘法和支持向量回归对四组分混合物中的目标成分进行了定量检测。目前,在利用太赫兹光谱技术与化学计量方法相结合对混合物中目标物质含量进行分析时,会面临混合物中较多组分导致的光谱重叠和失真问题,采用常规方法进行建模分析很难得到满意的结果^[14]。鉴于此,笔者利用太赫兹时域光谱技术、数据融合方法及支持向量回归模型对五组分混合物中的防老剂 MB 进行定量分析,以期对橡胶及助剂的多组分混合物中目标成分含量的检测与分析提供一种新方法。

收稿日期: 2023-05-08; 修回日期: 2023-07-03; 录用日期: 2023-07-25; 网络首发日期: 2023-08-07

基金项目: 国家自然科学基金(62161005, 62261012)、广西自动检测技术与仪器重点实验室基金项目(YQ22107)

通信作者: *cpxu_ck@163.com

2 基本原理

2.1 实验设备

本次实验样品的太赫兹时域光谱数据是由美国 Zomega 公司的 Z-3 透射式太赫兹时域光谱系统采集的^[15], 图 1 为 Z-3 系统实物图。实验设备主要由超快飞秒光纤激光器和 Z-3 太赫兹时域光谱系统组成, 系统实验原理如图 2 所示。该系统的工作原理为: 激光光源产生激光, 激光经过分束器后被分为两束, 其中一束

用于产生泵浦光脉冲, 另一束用于产生功能探测光脉冲。泵浦光经过调制后入射到发射晶体碲化镓上, 产生太赫兹脉冲。经过传输后, 太赫兹脉冲聚焦透射过待测样品, 携带着样品信息的太赫兹脉冲继续传输, 与探测光一同会聚并入射到探测电光晶体碲化锌上。在太赫兹脉冲作用下, 碲化锌发生光电效应, 最后经过探测器的解调和采集, 获得样品的太赫兹时域光谱信号。水汽对太赫兹的吸收很大, 因此需要将样品仓充满氮气, 以排除水汽的干扰。

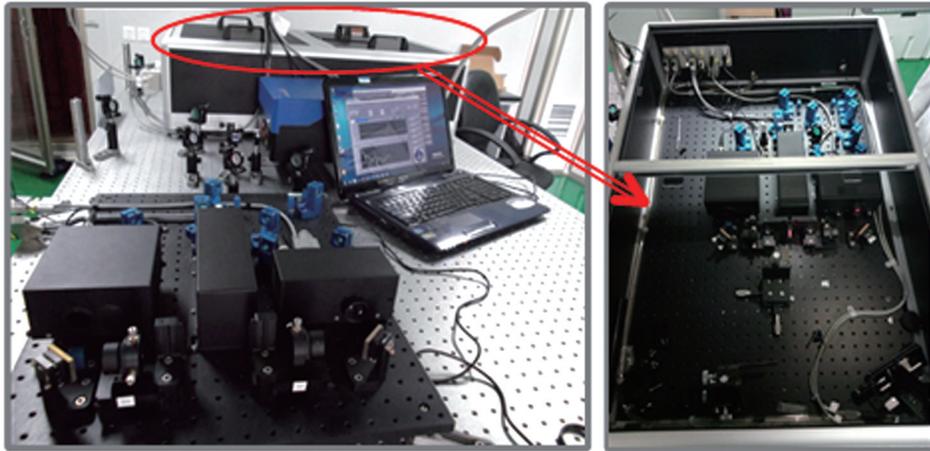


图 1 Z-3 系统实物图

Fig. 1 Physical photo of the Z-3 system

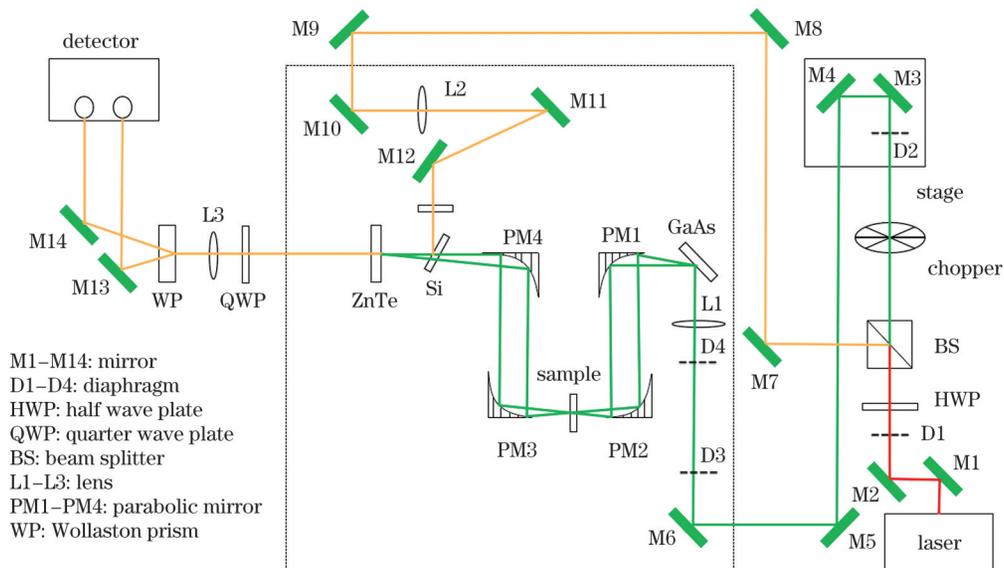


图 2 太赫兹系统原理图

Fig. 2 Schematic diagram of THz system

2.2 样品制备

实验对象是由丁腈橡胶、白炭黑和氧化锌为主要成分, 并混合不同比例的防老剂 H 和防老剂 MB 制成的五组分混合物。丁腈橡胶是纯度为 93.75% 的白色粉末, 购于天津明基金泰橡塑制品加工有限公司; 防老剂 H 的纯度为 99%, 防老剂 MB 的纯度为 85%, 均购于东莞市聚高化工科技有限公司; 白炭黑和氧化锌的纯度均大于 99.5%。所有实验样品均为固体粉末。实

验样品的制备采用压片法, 在样本制备前需要把原料放入真空恒温干燥箱中在 50 °C 下干燥 2 h 左右, 目的是减少样本中水分的干扰。然后按照比例把 5 种原料分别称量好, 在研钵中进行充分研磨后混合倒入模具中, 使用压片机在“6.1 t”压力参数下进行压片, 将混合粉末压成厚度约为 1 mm、直径约为 12 mm、质量为 200 mg 的上下表面平行且光滑的圆片, 最后把压好后的样本放置于恒温干燥箱中。另外还制作了 5 种单一

物质的实验样品。五组分混合物中各成分的比例如表 1 所示。总共制作了 9 种不同含量的混合物样品,并且每种比例下制作了 36 个样品。

表 1 五组分混合物的配制比例

Table 1 Five-component mixture content ratio

Number	Mass fraction / %				
	NBR	Silica	ZnO	Antioxidant	
				H	MB
1	60	10	6	0	24
2	60	10	6	3	21
3	60	10	6	6	18
4	60	10	6	9	15
5	60	10	6	12	12
6	60	10	6	15	9
7	60	10	6	18	6
8	60	10	6	21	3
9	60	10	6	24	0

2.3 光学参数提取

使用 Z-3 太赫兹时域光谱技术对物质进行检测时,常用折射率、吸收系数、吸光度等光学参数表征物质的性质。其中吸光度表示太赫兹波被材料吸收的程度,其为无量纲的相对量。使用吸光度描述样品的吸收特性,可以减小样本厚度不均匀对光谱分析产生的影响^[16]。为了计算样品的太赫兹吸光度,还需要对于干燥空气进行测量,将其太赫兹时域光谱作为参考信号。通过太赫兹时域光谱系统实验获取时域的参考信号 $E_{\text{ref}}(t)$ 和时域的样品信号 $E_{\text{sam}}(t)$,将参考信号和样品信号进行快速傅里叶变换转换成频域信号 $E_{\text{ref}}(\omega)$ 和 $E_{\text{sam}}(\omega)$ 。依据 Beer-Lambert 定律,吸光度 A 的计算公式为

$$A = -\log_{10} \left| \frac{E_{\text{sam}}(\omega)}{E_{\text{ref}}(\omega)} \right|^2. \quad (1)$$

2.4 数据融合

数据融合是将不同的互补信息进行结合的方法,以便优化所获得的信息,得到比单一信息更好的推论^[17-18]。导数光谱是由太赫兹吸收光谱进行一阶求导得到的,具有提高分辨率和信噪比、消除或降低噪声干扰的优势^[19]。太赫兹吸收光谱和导数光谱都具有独特的优势,将它们相互融合可以更加准确地描述物质的特征。可以在多个层面上对多种信息进行融合处理,依据数据抽象的层次可以将数据融合分成低层数据融合、中层数据融合、高层数据融合三种类型^[20-22]。低层数据融合又名数据层融合。在数据层融合中,吸收光谱数据和导数光谱数据直接连接在一起,形成一个新的数据矩阵,然后在此基础上建立定量分析模型。需要注意的是,在融合之前需要将吸收光谱数据和导数光谱数据归一化到 $[-1, 1]$ 范围内,以便在融合前获得类似的数值权重^[23]。中层数据融合又名特征层融合。

在特征层融合中,使用 SPA 和 MCUVE 算法对吸收光谱数据和导数光谱数据进行特征提取,然后利用特征变量建立定量分析模型。高层数据融合又名决策层融合。在决策层融合中,吸收光谱和导数光谱都分别通过定量分析模型进行处理,然后将每个模型的处理结果整合在一起,形成一个“集成决策”^[24]。笔者使用多元线性回归法^[25]处理高层数据融合问题。利用多元线性回归公式以及吸收光谱和导数光谱训练集的预测结果可以得到回归公式的权重(权重是依据不同光谱的贡献度来确定的),然后根据两种光谱训练集和预测集的预测结果以及权重来获得融合的预测值。多元线性回归公式为

$$y_{i, \text{act-cal}} = b_0 + b_1 x_{\text{THz, cal}} + b_2 x_{\text{DE, cal}}, \quad (2)$$

$$y_{i, \text{h-cal}} = b_0 + b_1 x_{\text{THz, cal}} + b_2 x_{\text{DE, cal}}, \quad (3)$$

$$y_{i, \text{h-val}} = b_0 + b_1 x_{\text{THz, val}} + b_2 x_{\text{DE, val}}, \quad (4)$$

式中: $y_{i, \text{act-cal}}$ 是训练集的真实值; $x_{\text{THz, cal}}$ 是吸收光谱训练集的预测值; $x_{\text{DE, cal}}$ 是导数光谱训练集的预测值; $x_{\text{THz, val}}$ 是吸收光谱预测集的预测值; $x_{\text{DE, val}}$ 是导数光谱预测集的预测值; $y_{i, \text{h-cal}}$ 是高层数据融合训练集的预测值; $y_{i, \text{h-val}}$ 是高层数据融合预测集的预测值; b_1 是吸收光谱的权重; b_2 是导数光谱的权重; b_0 是多元线性回归的截距。

2.5 样本集划分

Kennard-Stone(KS)算法是一种常用的样本集划分方法,该算法可以有效地对样本集进行划分。它的基本原理是计算两样本间的欧氏距离,选择最远距离的样品置于训练集中,然后计算剩余样本与被选出的全部样本的欧氏距离,选出欧氏距离最大和最小的样本放入训练集中。重复上述计算,直至选到指定的校准集数目为止。KS 算法可以在一定程度上避免校准集样本分布的不均匀性,这是由于它将具有代表性光谱的样本选入训练集,同时将余下的光谱差异相近的样本选入预测集^[26]。计算样本集欧氏距离的公式为

$$d_x(m, n) = \sqrt{\sum_{j=1}^i [x_m(j) - x_n(j)]^2}, m, n \in [1, N]. \quad (5)$$

2.6 支持向量回归

支持向量机(SVM)是 Vapnik 团队^[27]于 20 世纪 90 年代提出的一种有监督的学习模型。支持向量回归(SVR)将支持向量机的思想推广到回归问题中,在解决非线性问题时具有更显著的优势。其主要思想是通过核函数将低维非线性数据点映射到高维空间,并将数据点拟合到合适的超平面上,从而使数据点与超平面之间的误差距离最小,实现对数据点和超平面之间的非线性关系的描述和预测。SVR 模型的惩罚参数和径向基核函数参数对于预测的准确率至关重要,为了得到最佳的预测精度,本文选用网格搜索算法优化惩罚参数和核参数。

2.7 特征提取

连续投影(SPA)算法对向量之间的投影进行分析,不断地进行循环选择,选取彼此之间最小共线性以及具有最小冗余度的变量进行组合,并且尽可能保留原始光谱数据信息,从而提高模型的预测性能和处理速度^[28]。其工作原理是随机选取一个波长作为起始点,循环计算其在其他变量上的投影,然后将投影向量最大的变量放入组合中,以保证新选取的变量与上一个变量之间的线性关系最弱^[29]。SPA算法克服了变量之间存在的共线性问题,在降低光谱复杂度的同时,加快了模型的处理速度。

蒙特卡罗无信息变量消除法(MCUVE)是在无信息变量消除法(UVE)的基础上进行改进得到的。UVE对全光谱的特征波段进行选择,MCUVE利用蒙特卡罗算法代替留一法将样本中的训练集随机划分成多个子数据集,并且在子数据集中运行UXE-PLS模型^[30]。该算法通过不断分割训练集来降低运行速度,可以更加有效地利用样本间的内在相关性,更方便地评定光谱数据中波长的贡献度,利用贡献度来选择波长,从而实现光谱数据中无用信息波长的剔除。

2.8 评价指标

使用训练集相关系数(R_c)和训练集均方根误差(RMSEC)来评估模型的相关性能,同时使用预测集相关系数(R_p)和预测集均方根误差(RMSEP)来评估模型的预测性能。相关系数(R)是一个不大于1的正数,其值越接近1,表示相关性越大^[31]。均方根误差(RMSE,在公式中记为 e_{RMSE})是预测值的偏差平方之和与预测量 n 之比的平方根。均方根误差和相关系数的计算公式分别为

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{pre}} - y_{\text{act}})^2}, \quad (6)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_{\text{pre}} - y_{\text{act}})^2}{\sum_{i=1}^n (y_{\text{act}} - y_{\text{mean}})^2}}, \quad (7)$$

式中: n 是模型使用的样本数量; y_{act} 是样本集的真实值; y_{pre} 是样本集的预测值; y_{mean} 是平均值。

3 分析与讨论

3.1 光谱分析

使用太赫兹时域光谱系统分别对丁腈橡胶、白炭黑、氧化锌、防老剂H和防老剂MB等5种物质及五组分混合物进行测量,得到的5种纯物质的吸收光谱如图3所示。可以看出,5种纯物质在分析频段范围内均存在明显的吸收峰。丁腈橡胶在1.11 THz和1.38 THz处有吸收峰,氧化锌在1.33 THz处有吸收峰,白炭黑在1.32 THz和1.50 THz处有强度较为明显的特征吸收峰,防老剂H在1.32 THz处有一个特征吸收峰,防

老剂MB在1.48 THz处有一个特征吸收峰。可以看出,在上述5种物质中,氧化锌和防老剂H的特征吸收峰位置比较接近,而其他物质根据各自的吸收峰和吸收光的强度就能进行类别的区分。

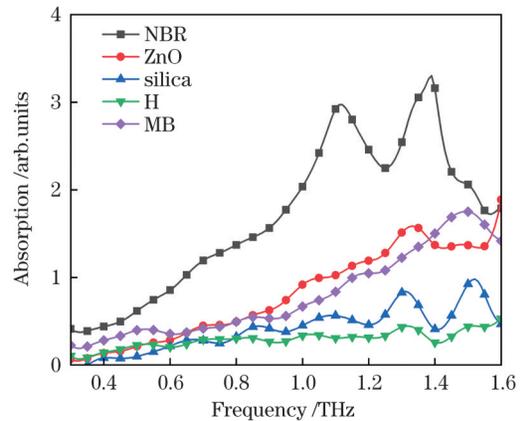


图3 5种纯物质的吸收光谱

Fig. 3 Absorption spectra of five kinds of pure substances

图4是不同比例的五组分混合物分别取平均后的吸收光谱。随着混合物中防老剂MB质量分数的增加,吸光度呈现递增的趋势,说明混合物的吸收光谱与防老剂MB在混合物中的质量分数有一定的线性关系。此外,不同比例的混合物的吸收峰强度在1.26 THz和1.52 THz处可以较好地分辨:防老剂MB的质量分数为15%~24%的混合物的吸收峰和吸收强度都存在较为明显的区别,防老剂MB的质量分数为0%~12%的混合物的吸收峰和吸收强度的差异较小,且前者(MB的质量分数为15%~24%)的吸收强度较后者(MB的质量分数为0%~12%)的大。

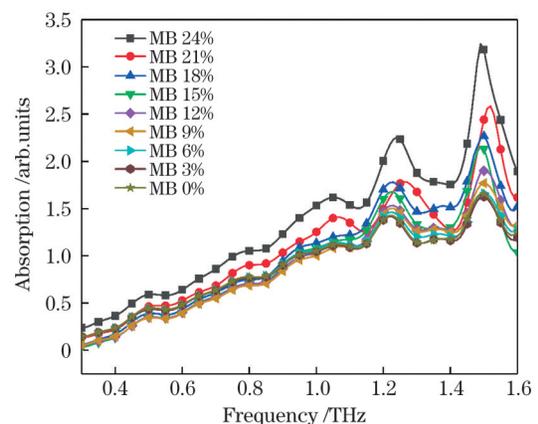


图4 五组分混合物的平均吸收光谱

Fig. 4 Average absorption spectra for five-component mixtures

图5是五组分混合物的全部吸收光谱,其中丁腈橡胶、氧化锌和白炭黑在混合物中所含比例不变,分别占60%、6%和10%,防老剂H与防老剂MB有9种比例,总含量不变,共占24%。混合物中不同组分之间相互干扰,使得实验光谱更加复杂。与纯组分光谱相比,混合物的吸收峰数量、位置、强度都发生了不同程

度的非线性变化,存在着重叠以及一定程度的失真现象,必须借助定量分析模型进行各组分的含量检测。

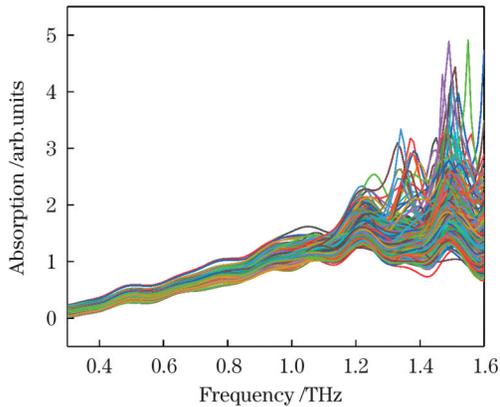


图5 五组分混合物的吸收光谱

Fig. 5 Absorption spectra of five-component mixtures

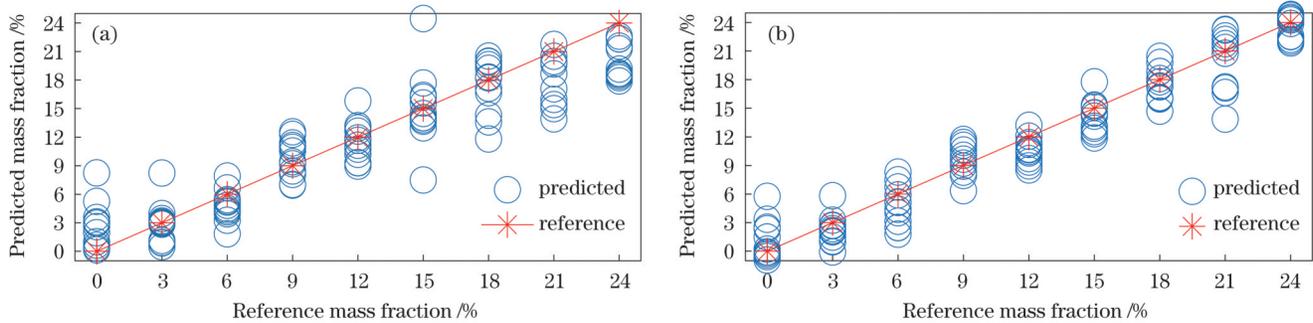


图6 不同模型对防老剂MB质量分数的预测结果。(a)PLSR模型;(b)SVR模型

Fig.6 Predicted mass fraction of antioxidant MB with different models. (a) PLSR model; (b) SVR model

为了更加直观地定量比较两种模型对五组分混合物中防老剂MB含量的预测结果,计算了相关系数和均方根误差(R_c 、RMSEC、 R_p 、RMSEP)。表2为五组分混合物的定量分析结果。可知:相较于PLSR模型,SVR模型的 R_c 从0.9269提高到0.9760,提升了0.0491, R_p 从0.9260提高到0.9653,提升了0.0393。这一结果说明SVR模型的预测精度更高,模型更加稳定,亦说明五组分混合物中的防老剂MB的质量分数与吸光度之间存在非线性关系。

表2 PLSR模型和SVR模型对五组分混合物的建模结果
Table 2 PLSR and SVR modeling results of five-component mixtures

Model	Training set		Prediction set	
	R_c	RMSEC / %	R_p	RMSEP / %
PLSR	0.9269	2.9108	0.9260	2.9947
SVR	0.9760	1.6899	0.9653	2.0219

通过上述模型对比可以发现SVR模型更好。因此,在SVR模型的基础上,结合数据融合方法对五组分混合物中的防老剂MB进行定量分析,分析结果如图7所示。基于吸收光谱的防老剂MB质量分数的分析结果整体波动更小,且预测值相比基于导数光谱的预测值更接近真实值,说明基于吸收光谱的预测精度更高。

3.2 多组分混合物的定量分析

首先用KS法选出216个数据作为训练集,余下的108个作为预测集。采用偏最小二乘回归(PLSR)、支持向量回归(SVR)和数据融合方法建立多组分混合物中防老剂MB的定量分析模型。用训练集和预测集的相关系数 R_c 和 R_p 以及均方根误差RMSEC和RMSEP来评价模型的优劣。

预测集中防老剂MB含量的预测值与参考值的关系如图6(a)、(b)所示,其中米字点表示参考值,圆圈表示预测值,斜线表示零误差线,预测值越接近零误差线,预测结果就越接近真实值,预测效果越好;反之,预测效果越差。由图6可以明显看出,与PLSR模型相比,SVR模型对五组分混合物中防老剂MB含量的预测值更接近真实值。这说明SVR模型的预测性能更好。

表3给出了混合物中防老剂MB的SVR模型的评价指标。可以看出:太赫兹吸收光谱的相关系数 R_c 和 R_p 分别为0.9760和0.9653,均方根误差RMSEC和RMSEP分别为1.6899%和2.0219%;导数光谱的相关系数 R_c 和 R_p 分别为0.9778和0.9528,均方根误差RMSEC和RMSEP分别为1.6249%和2.3523%。通过对比可以发现吸收光谱预测集的相关系数高于导数光谱,且吸收光谱的均方根误差RMSEP更小,这说明吸收光谱的预测性能优于导数光谱,泛化能力更强。

数据融合预测集中防老剂MB含量的预测值和参考值的关系如图8所示。图8(a)为低层数据模型的预测结果,与吸收光谱相比,其预测值到零误差线的距离更小。从表4可以看出,相比于吸收光谱,低层数据融合的预测集相关系数 R_p 从0.9653提升为0.9701,均方根误差RMSEP从2.0219%下降为1.8791%。说明低层数据融合可以提高模型的预测精度。但对比MCUVE中层数据融合和高层数据融合的建模结果可知,低层数据融合的预测精度较低。这是由于低层数据融合保留了所有的光谱信息,其中有一些冗余信息,降低了模型的预测能力。

图8(b)、(c)分别给出了SPA中层数据融合和MCUVE中层数据融合的分析结果,可以发现两者的

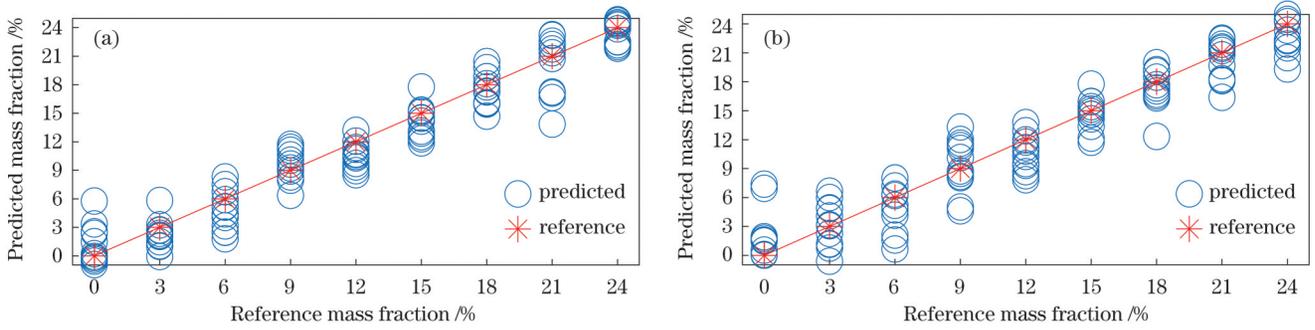


图 7 防老剂 MB 的预测结果。(a)基于吸收光谱的预测结果;(b)基于导数光谱的预测结果

Fig.7 Predicted results of antioxidant MB. (a) Predicted results based on absorption spectra; (b) predicted results based on derivative spectra

表 3 五组分混合物的 SVR 建模结果

Table 3 SVR modeling results of five-component mixtures

Spectrum	Training set		Prediction set	
	R_c	RMSEC / %	R_p	RMSEP / %
Absorption	0.9760	1.6899	0.9653	2.0219
Derivative	0.9778	1.6249	0.9528	2.3523

预测结果不同,说明特征提取方法与中层数据融合之间有着密切的关系。表 4 显示:SPA 中层数据融合的相关系数 R_c 和 R_p 分别 0.9841 和 0.9693,相较于吸收光谱, R_c 提升了 0.0081, R_p 提升了 0.0040;MCUVE 中层数据融合相关系数 R_c 和 R_p 分别为 0.9900 和 0.9758,相较于吸收光谱, R_c 提升了 0.0140, R_p 提升了 0.0105。说明两种中层数据融合都可以改善模型的预测性能,这是由于中层数据融合不仅去除了噪声,也挑选出了带有被测物特征信息的变量。

高层数据融合建立在全光谱的基础上,融合各个模型的结果。吸收光谱模型的预测值和导数光谱模型

的预测值与多组分混合物中防老剂 MB 的实际含量的多元线性拟合方程为 $y=0.5076x_{THz}+0.5288x_{DE}-0.4572$ 。由线性回归方程中吸收光谱和导数光谱的权重系数可知,导数光谱数据对于模型性能的贡献度高于吸收光谱数据。图 8(d) 显示,高层数据融合的预测值相比吸收光谱更接近实际参考值,并且从表 4 可以看出高层数据融合的 R_c 和 R_p 分别为 0.9872 和 0.9732,均方根误差 RMSEC 和 RMSEP 分别为 1.2367% 和 1.7798%,相较于吸收光谱, R_p 提升了 0.0079, RMSEP 下降了 0.2421 个百分点,说明高层数据融合提高了模型的预测精度,增强了模型的稳定性。这是由于高层数据融合以线性回归的方式对每一个预测结果都加以利用,避免有用信息丢弃的同时也降低了噪声和冗余信息对建模的不利影响。

本研究同大多数研究所得结果一样,数据融合模型的预测效果明显优于单一光谱模型^[32],这表明数据融合能够有效地改善模型的预测性能。低层数据融合

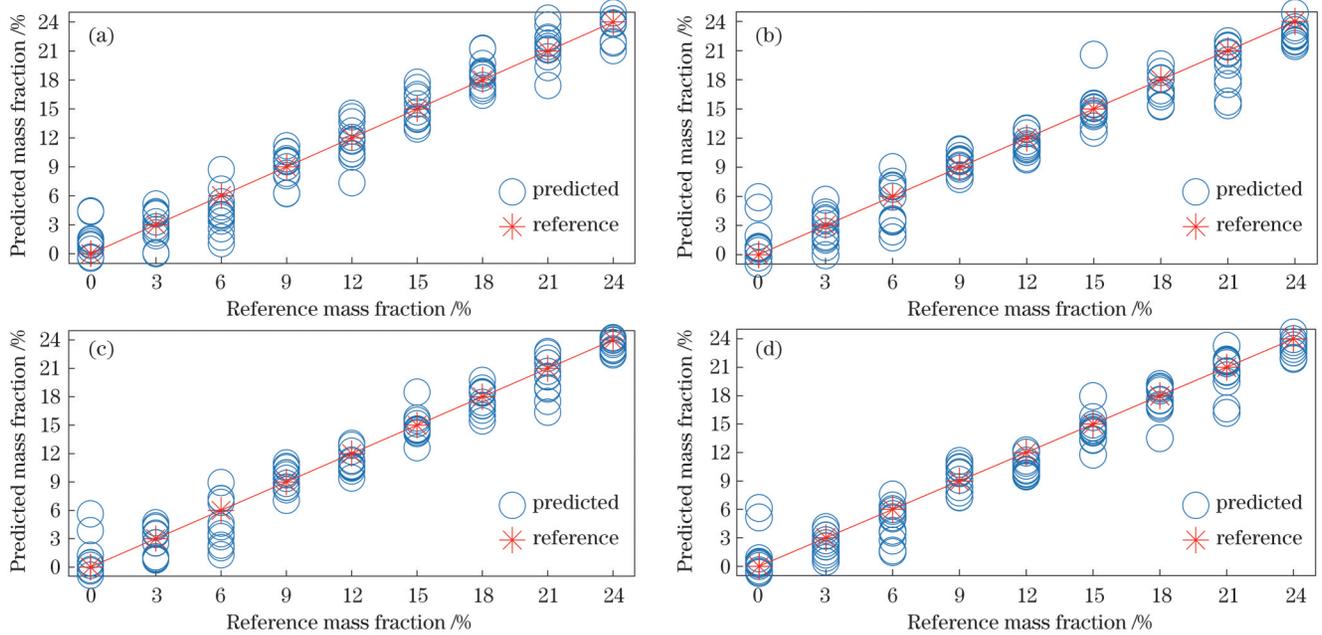


图 8 防老剂 MB 的预测结果。(a)低层数据融合;(b)SPA 中层数据融合;(c)MCUVE 中层数据融合

Fig.8 Predicted results of antioxidant MB. (a) Low-level data fusion; (b) mid-level data fusion (SPA); (c) mid-level data fusion (MCUVE); (d) high-level data fusion

表 4 五组分混合物的数据融合建模结果
Table 4 Data fusion modeling results of five-component mixtures

Model	Training set		Testing set	
	R_c	RMSEC / %	R_p	RMSEP / %
Low-level data fusion	0.9908	1.0476	0.9701	1.8791
Mid-level data fusion(SPA)	0.9841	1.3759	0.9693	1.9040
Mid-level data fusion(MCUVE)	0.9900	1.0947	0.9758	1.6941
High-level data fusion	0.9872	1.2367	0.9732	1.7798

的预测精度低于 MCUVE 中层数据融合和高层数据融合的预测精度,因为低层数据融合保留了所有的光谱信息,其中包含了与被测物无关的冗余信息,这些信息会给定量分析带来一定的负面影响。SPA 中层数据融合的预测性能比其他几种数据融合的预测性能差,这是由于 SPA 中层数据融合丢失了有效信息,从而阻碍了模型的预测能力。高层数据融合相比单一光谱、低层数据融合、SPA 中层数据融合具有更好的预测效果,因为它通过不同的加权系数对每个光谱的预测结果进行融合。因此,检测多组分混合物中防老剂 MB 含量的模型按预测效果从高到低排序为 MCUVE 中层数据融合、高层数据融合、低层数据融合、SPA 中层数据融合、吸收光谱、导数光谱。

4 结 论

笔者利用太赫兹时域光谱技术、MCUVE 中层数据融合和 SVR 研究了一种快速检测多组分混合物中防老剂 MB 含量的新方法,该方法可以解决现有橡胶及添加剂检测方法的不足以及由多组分混合物吸收光谱重叠和失真导致的定量结果准确性不高的问题。通过分析五组分混合物的吸收光谱发现混合物的吸光度随着防老剂 MB 含量增加呈现递增的趋势,特别是在吸收峰处,递增的趋势更加明显,说明混合物的吸光度与防老剂 MB 含量之间存在一定的线性关系。为了进一步分析混合物中防老剂 MB 的含量,采用 PLSR 和 SVR 方法建立定量分析模型,结果显示 SVR 的预测性能更好,说明混合物的吸光度与防老剂 MB 质量分数之间存在非线性关系。在 SVR 模型的基础上结合数据融合方法进行定量分析,结果表明,相较于单一光谱,4 种数据融合方法的预测准确性和稳定性均得到了明显提升,且 MCUVE 中层数据融合的预测性能最好。

综上所述,可以将数据融合与 SVR 结合起来应用到太赫兹光谱对橡胶多组分混合物的定量分析中,从而为橡胶质量监控和检测提供了一种有效方法。此外,本文所用数据融合方法将吸收光谱和导数光谱进行融合,这两种光谱本质上都属于太赫兹光谱,因此,拟在下一步工作中将其他光谱(比如红外光谱和拉曼光谱)与太赫兹光谱进行融合,以便获得待测物质的更多信息,提高模型的预测性能。

参 考 文 献

- [1] 陈慧. 有机锌代替间接法氧化锌在轮胎胎面胶中的应用[J]. 中国橡胶, 2016, 32(10): 45-48.
Chen H. Application of organic zinc instead of indirect zinc oxide in tire tread compound[J]. China Rubber, 2016, 32(10): 45-48.
- [2] 李淑娟. 橡胶助剂结构和组分分析[J]. 橡胶科技, 2014, 12(10): 47-50.
Li S J. Structure and composition analysis of rubber additives[J]. Rubber Science and Technology, 2014, 12(10): 47-50.
- [3] Yin X H, Feng M L, Jiang Y, et al. Quantitative analysis of the 2-mercaptobenzothiazole based on terahertz time-domain spectroscopy and an improved support vector regression[J]. Infrared Physics & Technology, 2021, 119: 103953.
- [4] 马卿效, 李春, 李天莹, 等. 基于太赫兹光谱和机器学习算法的二元及三元混合物定量分析[J]. 激光与光电子学进展, 2022, 59(19): 1930003.
Ma Q X, Li C, Li T Y, et al. Quantitative analysis of binary and ternary mixtures based on terahertz spectroscopy and machine learning algorithm[J]. Laser & Optoelectronics Progress, 2022, 59(19): 1930003.
- [5] 胡军, 刘燕德, 孙旭东, 等. 基于 BP 神经网络的太赫兹时域光谱对面粉中苯甲酸的定量检测研究[J]. 激光与光电子学进展, 2020, 57(7): 073002.
Hu J, Liu Y D, Sun X D, et al. Quantitative determination of benzoic acid in flour based on terahertz time-domain spectroscopy and BPNN model[J]. Laser & Optoelectronics Progress, 2020, 57(7): 073002.
- [6] Shao Y N, Liu J L, Zhu Z, et al. Quantitative detection on metabolites of *Haematococcus pluvialis* by terahertz spectroscopy [J]. Computers and Electronics in Agriculture, 2021, 186: 106223.
- [7] 刘燕德, 李茂鹏, 胡军, 等. 基于太赫兹检测技术的咖啡豆品种鉴别[J]. 激光与光电子学进展, 2021, 58(16): 1630002.
Liu Y D, Li M P, Hu J, et al. Identification of coffee-bean varieties using terahertz detection technology[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1630002.
- [8] 饶近秋, 陈李乙, 白鹏鹏, 等. 基于太赫兹时域光谱技术的四种基原郁金的鉴别研究[J]. 激光与光电子学进展, 2021, 58(22): 2200002.
Rao J Q, Chen L Y, Bai P P, et al. Identification of four origins of curcuma based on terahertz time-domain spectroscopy[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2200002.
- [9] 曹瑶瑶, 李霞, 白军朋, 等. 基于 THz-TDS 技术的培氟沙星和氟罗沙星抗生素定性定量检测研究[J]. 光谱学与光谱分析, 2022, 42(6): 1798-1803.
Cao Y Y, Li X, Bai J P, et al. Qualitative and quantitative detection of pefloxacin and fleroxacin antibiotics based on THz-TDS technology[J]. Spectroscopy and Spectral Analysis, 2022, 42(6): 1798-1803.
- [10] Komatsu M, Izutsu T, Ohki Y, et al. Terahertz spectroscopic analysis of ethylene-propylene-diene copolymer[C]//Proceedings of 2014 International Symposium on Electrical Insulating Materials, June 1-5, 2014, Niigata, Japan. New York: IEEE Press, 2014: 338-341.
- [11] Ma Q X, Teng Y, Li C, et al. Simultaneous quantitative determination of low-concentration ternary pesticide mixtures in

- wheat flour based on terahertz spectroscopy and BPNN[J]. Food Chemistry, 2022, 377: 132030.
- [12] Sun Y W, Huang J L, Shan L X, et al. Quantitative analysis of bisphenol analogue mixtures by terahertz spectroscopy using machine learning method[J]. Food Chemistry, 2021, 352: 129313.
- [13] 殷贤华, 郭超, 李安, 等. Zernike 矩在橡胶添加剂太赫兹光谱定量分析中的应用研究[J]. 光谱学与光谱分析, 2019, 39(9): 2725-2731.
- Yin X H, Guo C, Li A, et al. Application of Zernike moment in terahertz spectrum quantitative analysis of rubber additives[J]. Spectroscopy and Spectral Analysis, 2019, 39(9): 2725-2731.
- [14] 陈涛, 李智, 莫玮, 等. 太赫兹时域光谱的药物多组分同时定量测定[J]. 光谱学与光谱分析, 2013, 33(5): 1220-1225.
- Chen T, Li Z, Mo W, et al. Simultaneous quantitative determination of multicomponents in tablets based on terahertz time-domain spectroscopy[J]. Spectroscopy and Spectral Analysis, 2013, 33(5): 1220-1225.
- [15] 陈涛, 蔡治华, 胡放荣, 等. 结构相似单糖和二糖分子的太赫兹时域光谱研究[J]. 光谱学与光谱分析, 2019, 39(3): 686-692.
- Chen T, Cai Z H, Hu F R, et al. A study of terahertz spectra of monosaccharides and disaccharides with structural similarities[J]. Spectroscopy and Spectral Analysis, 2019, 39(3): 686-692.
- [16] 王启超, 汪家春, 王泉, 等. 太赫兹波探测光子晶体涂层覆盖目标的可行性[J]. 发光学报, 2017, 38(2): 248-253.
- Wang Q C, Wang J C, Wang X, et al. Feasibility of applying terahertz wave to detect target covered with photonic crystal coating [J]. Chinese Journal of Luminescence, 2017, 38(2): 248-253.
- [17] Yang X H, Li Y P, Wang L, et al. Determination of 10-HDA in royal jelly by ATR-FTMIR and NIR spectral combining with data fusion strategy[J]. Optik, 2020, 203: 164052.
- [18] Sanaeifar A, Li X L, He Y, et al. A data fusion approach on confocal Raman microspectroscopy and electronic nose for quantitative evaluation of pesticide residue in tea[J]. Biosystems Engineering, 2021, 210: 206-222.
- [19] 卫辰洁, 王继芬, 范琳媛, 等. 基于光谱数据融合和人工神经网络的汽车灯罩鉴别[J]. 中国塑料, 2020, 34(12): 59-64.
- Wei C J, Wang J F, Fan L Y, et al. Identification of automobile lampshade based on spectral data fusion technology and artificial neural network[J]. China Plastics, 2020, 34(12): 59-64.
- [20] 周昆鹏. 基于多光谱信息融合的水质 COD 检测模型与实验研究 [D]. 秦皇岛: 燕山大学, 2017: 61-23.
- Zhou K P. Model and experimental study on water quality COD detection based on multi-spectral information fusion[D]. Qinhuangdao: Yanshan University, 2017: 61-23.
- [21] 陈达, 骆文欣, 黄志轩, 等. 基于多光谱融合的奶粉掺假诊断方法[J]. 纳米技术与精密工程, 2017, 15(5): 384-388.
- Chen D, Luo W X, Huang Z X, et al. Adulterated milk powder diagnosis method based on multi-spectra fusion[J]. Nanotechnology and Precision Engineering, 2017, 15(5): 384-388.
- [22] Wu X M, Zhang Q Z, Wang Y Z. Traceability of wild *Paris polyphylla* Smith var. *yunnanensis* based on data fusion strategy of FT-MIR and UV-Vis combined with SVM and random forest[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 205: 479-488.
- [23] Yin X H, Chen H C, Li A, et al. Quantitative analysis of textile delusterant based on terahertz spectral and data fusion strategies[J]. Infrared Physics & Technology, 2022, 125: 104293.
- [24] Zhu J Y, Fan X, Han L, et al. Quantitative analysis of caprolactam in sauce-based food using infrared spectroscopy combined with data fusion strategies[J]. Journal of Food Composition and Analysis, 2021, 104: 104130.
- [25] Li Y, Huang Y, Xia J J, et al. Quantitative analysis of honey adulteration by spectrum analysis combined with several high-level data fusion strategies[J]. Vibrational Spectroscopy, 2020, 108: 103060.
- [26] 金航峰. 基于光谱和高光谱图像技术的蚕茧品质无损检测研究 [D]. 杭州: 浙江大学, 2013: 20-21.
- Jin H F. Study on nondestructive testing of cocoon quality based on spectrum and hyperspectral image technology[D]. Hangzhou: Zhejiang University, 2013: 20-21.
- [27] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [28] 王松静. 基于光谱分析与图像处理的模式识别研究[D]. 杭州: 浙江大学, 2013: 20-21.
- Wang S J. Research on pattern recognition based on spectral analysis and image processing[D]. Hangzhou: Zhejiang University, 2013: 20-21.
- [29] 章海亮, 叶青, 罗微, 等. 基于 LS-SVM 高光谱成像鱼新鲜度鉴别[J]. 华东交通大学学报, 2017, 34(2): 105-111.
- Zhang H L, Ye Q, Luo W, et al. Classification of fish freshness based on LS-SVM and hyperspectral imaging technology[J]. Journal of East China Jiaotong University, 2017, 34(2): 105-111.
- [30] 李鸿博. 基于近红外光谱的红松子品质检测模型研究[D]. 哈尔滨: 东北林业大学, 2021: 74-77.
- Li H B. Study on quality detection model of Korean pine nuts based on near infrared spectroscopy[D]. Harbin: Northeast Forestry University, 2021: 74-77.
- [31] 张活. 基于太赫兹时域光谱技术的中药检测方法研究[D]. 西安: 西安电子科技大学, 2018: 75-76.
- Zhang H. Study on detection method of traditional Chinese medicine based on terahertz time domain spectroscopy[D]. Xi'an: Xidian University, 2018: 75-76.
- [32] Zhou L, Zhang C, Qiu Z J, et al. Information fusion of emerging non-destructive analytical techniques for food quality authentication: a survey[J]. TrAC Trends in Analytical Chemistry, 2020, 127: 115901.

Quantitative Detection of Multi-Component Rubber Additives Based on Terahertz Spectral Data Fusion

Yin Xianhua^{1,2}, Chen Huicong^{1,2}, Zhang Huo^{1,2*}

¹School of Electronic Engineering and Automation, Guilin University of Electronic Science and Technology, Guilin 541004, Guangxi, China;

²Guangxi Key Laboratory of Automatic Detection Technology and Instrument, Guilin 541004, Guangxi, China

Abstract

Objective The content of rubber additives is an important determinant of rubber quality. Current testing methods for these additives include combustion testing, chemical analysis, chromatography, and infrared spectroscopy. However, these detection techniques present challenges such as intricate pre-processing, time-intensive operations, laborious procedures, and potential

inaccuracies in reflecting the genuine additive content. These types of limitations hinder their ability to cater to the growing demand for swift, precise, and non-destructive detection in rubber. This is a significant challenge for the advancement of the rubber industry in China. Furthermore, when analyzing multi-component mixtures, the absorption spectra can overlap and become distorted, leading to unreliable results. In this study, we leverage terahertz time-domain spectroscopy, data fusion, and chemometrics to quantitatively assess additives in five-component mixtures. This offers an innovative approach for detecting and analyzing the content of target components in multi-component mixtures of rubber and its auxiliaries.

Methods In this study, a five-component mixture composed of NBR, silica, zinc oxide, antioxidant H, and antioxidant MB was used as an experimental sample. The terahertz time-domain spectroscopy system was utilized to capture and compute the absorption spectra of the five-component mixture within the range of 0.3–1.6 THz and to analyze its spectral characteristics. The derivative spectral data of the sample were derived by taking the first-order derivatives. Initially, the KS algorithm was employed to segment the sample set data, which was then quantitatively analyzed using partial least squares regression and support vector machine regression models. Subsequently, three data fusion methods were employed to process the data. Specifically, the low-level data fusion directly combined the absorption spectrum data with the derivative spectrum; the mid-level data fusion merged variables after feature extraction via the Monte Carlo uninformative variable elimination and successive projections algorithm; and the high-level data fusion was executed using multiple linear regression. Finally, the predictive accuracy of the models was assessed based on the correlation coefficient and root mean square error.

Results and Discussions Through the absorption spectra of five pure substances — NBR, silica, zinc oxide, antioxidant H, and antioxidant MB — it is evident that there are noticeable absorption peaks within the range of the analyzed frequency band for all five pure substances (Fig. 3). The absorption spectra of the five-component mixtures are averaged individually for each proportion. It is observable that as the content of antioxidant MB in the mixtures increases, the absorbance also rises, suggesting a linear relationship between the absorption spectra of the mixtures and content of antioxidant MB (Fig. 4). The full absorption spectra of the five-component mixtures reveal complexity in the mixtures, with overlapping and some distortion (Fig. 5). The comparison between the predicted and reference values of the antioxidant MB content in the prediction set reveals that SVR aligns more closely with the actual value than PLSR does when predicting the antioxidant MB content in the five-component mixtures. This indicates that the SVR model predicts more effectively (Fig. 6). Both the correlation coefficient and root mean square error demonstrate that SVR predicts with superior accuracy, suggesting a non-linear relationship between the content of antioxidant MB in the five-component mixture and absorbance (Table 2). Based on the SVR model, when comparing the prediction results of absorption spectra to derivative spectra, it is found that the analytical results for the content of antioxidants MB from absorption spectra fluctuate less (Fig. 7). In comparing the correlation coefficients and root mean square errors of absorption and derivative spectra using the SVR model, the prediction accuracy for antioxidant MB content from absorption spectra is higher, indicating a superior predictive capability of absorption spectra (Table 3). The comparison between the predicted and reference values of antioxidant MB content for the data fusion prediction set demonstrates that the data fusion model predicts significantly better than a single spectrum, suggesting that the data fusion method enhances the model's predictive performance (Fig. 8). The predictive accuracy of the Monte Carlo-based uninformative variable elimination method for mid-level data fusion surpasses the accuracy of the single spectrum and other data fusions (Table 4).

Conclusions In the current study, a new method for rapid detection of antioxidant MB content in rubber multi-component mixtures is investigated using terahertz time-domain spectroscopy, MCVUE mid-level data fusion, and SVR. Analysis of the absorption spectra of the five-component mixtures and quantitative analytical models reveals linear and non-linear relationships between the absorbance of the mixtures and antioxidant MB content. Results from quantitative analyses, which combine data fusion methods based on SVR, indicate that prediction accuracy and stability of all four data fusion methods significantly surpass that of a single spectrum. Specifically, the prediction performance of MCVUE mid-level data fusion is the best. In conclusion, the combination of terahertz time-domain spectroscopy, data fusion methods, and SVR modeling addresses the shortcomings of existing rubber and additive detection methods and the accuracy challenges posed by overlapping and distortion phenomena in the absorption spectra of multi-component mixtures. This approach holds significant scientific value and promises substantial market application potential.

Key words spectroscopy; rubber and additives; multi-component mixtures; terahertz spectroscopy; quantitative detection; data fusion