

基于 CARS 和 K-S 的马拉硫磷农药浓度吸收光谱预测模型研究

甄欢仪, 马瑞峻*, 陈瑜*, 孙小鹏, 马创立

华南农业大学工程学院, 广东 广州 510642

摘要 利用马拉硫磷在紫外/可见光波长范围内的不同浓度梯度的吸光度光谱数据, 建立其快速有效的定量预测分析模型。在预测模型的建立过程中, 参与建模的波长变量和校正集样本的优劣对定量分析模型的预测能力有着决定性作用。首先对实验样本是否存在异常样本进行检查, 然后将 200.08~750.04 nm 波长范围的光谱数据采用不同预处理方法并建立 PLS 模型, 进而将预处理结果最优(均值中心化)的光谱数据采用竞争性自适应重加权采样(CARS)算法和蒙特卡洛无消息变量消除法(MC-UVE)分别筛选出关键波长变量并建立相应的 PLS 预测模型, 模型结果表明, CARS 算法在关键变量筛选的性能上优于 MC-UVE 算法; 再将 CARS 算法筛选出的 18 个波长变量(为原来变量数的 1.137 8%)结合 Kennard-Stone (K-S)算法和蒙特卡洛交叉验证(MCCV)分别优选出的 44 个建模样本(原来样本数的 88%)建立 CARS-K-S_s-PLS 和 CARS-CCV_s-PLS 定量预测模型, R_p^2 分别为 0.998 2 和 0.998 9, RMSEP 分别为 0.863 4 和 1.026 2, RPD 分别为 24.163 5 和 20.330 1, CARS-K-S_s-PLS 模型略优于 CARS-CCV_s-PLS 模型。结果表明, CARS 算法能够淘汰与样本浓度相关性较弱的变量, 同时有效剔除无关光谱信息, K-S 算法能帮助选择更优的建模样本集, 马拉硫磷农药的紫外/可见光吸收光谱结合 CARS 算法和 K-S 算法所建立的 CARS-K-S_s-PLS 模型能够用来定量预测马拉硫磷农药浓度。研究工作为利用光谱技术快速检测有机磷农药浓度提供一定的理论依据和实验基础, 在有机磷农药快速检测领域具有良好的应用前景。

关键词 马拉硫磷; 紫外/可见吸收光谱; 关键变量筛选; 样本优选; 定量预测

中图分类号: O657.39 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)05-1601-06

引言

马拉硫磷是一种低毒高效、残效期短的有机磷杀虫剂, 不仅被广泛应用于农业生产, 而且其原药及其混合剂是谷物常用的储粮保护剂, 但不合理地使用马拉硫磷农药会对人体和环境造成不可逆转的影响。有机磷农药的速测方法有酶联免疫法、酶抑制法和生物传感器法等。其中酶抑制法因其成本低、检测时间短和操作简单等特点, 成为蔬菜农药残留快速检测最为常用的方法, 但受不同的前处理方法、基质干扰方法和最低检测限高的影响, 酶抑制检测的样品部分结果存在假阳性和假阴性风险^[1-2]。寻找一种成本低、灵敏度高、操作简单方便和能够实现现场准确快速检测有机磷农药的方法具有重要的现实意义和实际应用价值。

光谱技术是一种快速、无损检测技术, 但直接使用光谱技术检测有机磷农药, 存在光谱背景噪声干扰明显和吸收峰

偏移不稳定、特征不明显、检测精度难以满足定量要求等问题。然而将光谱技术和化学计量学分析方法结合, 对光谱数据进行去噪声预处理、提取关键变量和特征样本以及建立预测模型来压缩有用信息, 可有效地解决光谱检测中存在的问题, 并被不同领域的研究人员广泛应用于实际检测中^[3-4]。目前采用光谱技术快速有效地定量检测水体中有机磷农药的研究鲜有报道。

本工作以马拉硫磷农药为研究对象, 采用紫外/可见吸光度光谱仪获取不同浓度实验样品的的光谱数据, 结合竞争性自适应重加权采样(competitive adaptive reweighted sampling, CARS)变量选择算法和经典的 Kennard-Stone(K-S)算法样本选择方法, 建立偏最小二乘(PLS)定量分析预测模型。为验证 CARS 算法的变量选择上的性能, 将其与常用有效的变量筛选方法蒙特卡洛无消息变量消除法(Monte Carlo-uninformative variable elimination, MC-UVE)进行系统的比较。并将 K-S 算法与蒙特卡洛交叉验证(Monte Carlo

收稿日期: 2019-08-13, 修订日期: 2019-12-18

基金项目: 国家重点研发计划项目(2016YFD0800901)资助

作者简介: 甄欢仪, 女, 1995年生, 华南农业大学工程学院硕士研究生 e-mail: 370124342@qq.com

* 通讯联系人 e-mail: maruijun_mrj@163.com; chenyu219@126.com

cross-validation, MCCV)在样本优选性能进行比较。期望通过对关键变量和更优样本的筛选得到鲁棒性好且简单可靠的定量预测模型,实现对马拉硫磷农药浓度的快速有效的定量检测。

1 实验部分

1.1 样本

马拉硫磷实验样本配制:用 1/10 000 电子天平称取 98% 马拉硫磷标准样本 0.204 1($\pm 0.000 2$) g,用少量甲醇超声溶解,然后加入甲醇准确定容至 400 mL,摇匀,得到浓度为 $500 \text{ mg} \cdot \text{L}^{-1}$ 的马拉硫磷标准液。使用纯净水作为稀释剂配制不同浓度的实验样本,浓度为 $0.1 \sim 1.00 \text{ mg} \cdot \text{L}^{-1}$ (浓度梯度为 $0.05 \text{ mg} \cdot \text{L}^{-1}$),共 19 个样本; $1.50 \sim 10 \text{ mg} \cdot \text{L}^{-1}$ (浓度梯度为 $0.5 \text{ mg} \cdot \text{L}^{-1}$),共 18 个样本; $12 \sim 50 \text{ mg} \cdot \text{L}^{-1}$ (浓度梯度为 $2 \text{ mg} \cdot \text{L}^{-1}$),共 20 个样本; $55 \sim 100 \text{ mg} \cdot \text{L}^{-1}$ (浓度梯度为 $5 \text{ mg} \cdot \text{L}^{-1}$),共 10 个样本。合计 67 个浓度样本。

划分样本集前,参照文献[5]先进行异常样本剔除检查,结果表明并无异常样本,然后进行下一步样本划分。采用基于变量之间欧式(欧几里得)距离的 K-S 算法,按 3:1 的比例在特征空间中划分出校正集(50 个样本)和预测集(17 个样本)。

1.2 紫外/可见光谱数据采集

使用美国 Ocean Optics 海阳光学公司的便携式吸光度测量仪(型号为 FLAME-CHEM-UV-VIS)采集马拉硫磷的每个样本浓度的光谱数据。光谱仪的光学分辨率为 1.5 nm,波长范围为 200~850 nm,其间隔为 0.39 nm,得到波长变量总数为 1 914 个。光谱数据采集软件设置积分时间为 25 ms,平滑度为 2,每个样本扫描 20 次,取平均值为最终光谱。

1.3 关键变量筛选与样本优选的方法

1.3.1 关键变量筛选方法

CARS 算法基于变量回归系数绝对值 $|b_i|$,遵循“适者生存”原则,使用指数衰减函数(exponentially decreasing function, EDF)和自适应重加权采样技术(adaptive reweighted sampling, ARS)去除回归系数 $|b_i|$ 值相对小的波长变量,留下回归系数 $|b_i|$ 值权重值较大且共线性小的变量组成新的变量子集。算法的具体执行步骤可参考文献[8]。

MC-UVE 算法^[6]基于变量回归系数 b_i 的稳定性值(回归系数 b_i 均值和标准差比值的绝对值),可得到由回归系数较稳定且含有有效信息较多的变量组成最优变量子集。该方法在许多光谱波长变量筛选研究中已被证明能有效剔除无信息变量。

1.3.2 样本优选方法

建模样本的优劣关乎整个模型的质量,影响模型预测能力。蒙特卡罗交叉验证 MCCV 通过多次的蒙特卡罗(MC)采样,随机将原来校正集样本划分分校正子集和预测子集,建立相应的 PLS 模型并得到模型的预测子集浓度预测偏差(RMSEP),最后通过设定样本的预测残差平均值和预测残差标准差的阈值选择更优样本集。K-S 法基于光谱特征选取样本。

1.4 模型评价

对 PLS 模型的评价采用决定系数 R^2 、均方根误差 RMSE(包括校正集均方根误差 RMSE 和预测集均方根误差 RMSEP)和预测集的 RPD。一个好的模型应具有高的 R^2 和低的 RMSE。RPD 是预测集的标准偏差与预测均方根误差的比值,反映了模型的分辨能力和稳健性,当 $\text{RPD} \geq 3$ 时,模型可应用于定量分析和实际检测。本研究中的样本划分、数据预处理、PLS 建模、关键波长变量选择、样本优选都基于 The Unscrambler X10.4 和 MATLAB R2017b 软件平台进行。

2 结果与讨论

2.1 光谱数据预处理

马拉硫磷不同浓度梯度下的原始吸收光谱(200~850 nm)如图 1 所示。

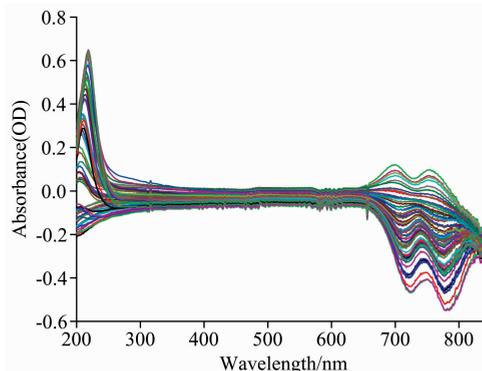


图 1 马拉硫磷原始吸收光谱图

Fig. 1 Original absorption spectrum of malathion

由图 1 可知,由于光谱仪在其适用的波长范围两端光源能量较低,光谱基线两端出现漂移下滑现象,同时全光谱存在较大的背景噪声干扰,检测时,在 200~250 nm 波长范围内,光谱出现吸收峰向左偏移不稳定和低浓度与高浓度光谱曲线之间空缺的现象,另外超过 750 nm 光谱基线下滑现象逐渐严重,影响农药浓度在可见光部分有效的吸光度检测。研究中将 200.08~750.04 nm 的光谱数据分别使用基线校正、基线补偿、不同点的 S-G 平滑、一阶导数、二阶导数、标准化和均值中心化等数据预处理方法进行处理,进一步获取有效的分析信号以提高模型预测性能。通过对不同预处理后 PLS 模型结果对比分析,表明均值中心化预处理效果最好。后续的变量筛选、样本优选和模型建立研究均采用均值中心化处理后的光谱数据,如图 2 所示。

2.2 关键变量的选取与模型结果

由图 1 和图 2 可知,马拉硫磷在紫外波段范围内的吸收光谱的吸收峰在 203.52~217.6 nm 之间偏移不稳定,最大吸收波长点难以确定。而传统的标准曲线定量分析方法只能针对有明显吸收峰出现且其相应的波长点位置稳定的农药样品进行分析,使用传统的标准曲线定量分析法建立的马拉硫磷定量预测模型稳定性会降低,所以需要采用化学计量学方法进行处理并建立定量预测分析模型。

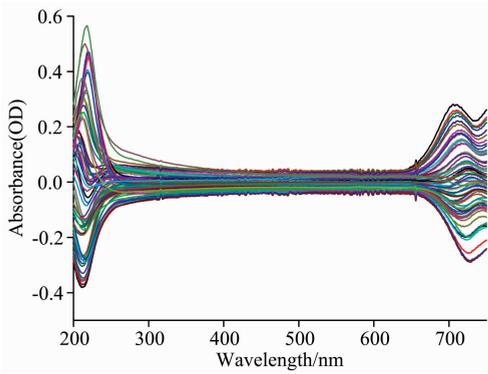


图 2 均值中心化后的马拉硫磷吸收光谱图

Fig. 2 Mean centering absorption spectrum of malathion

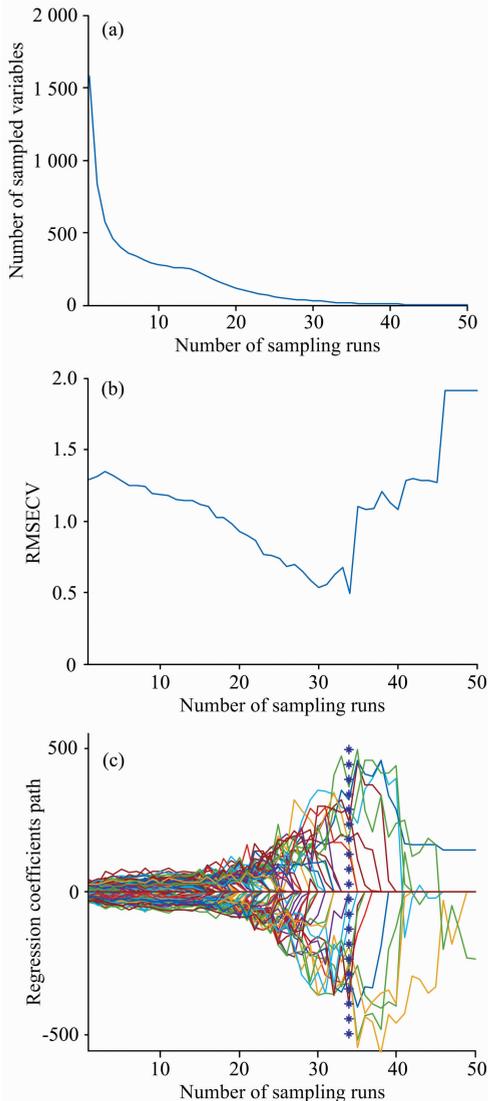


图 3 竞争性自适应重加权算法选出的关键变量

(a): 采样变量数的变化趋势; (b): RMSECV 值的变化趋势; (c): 每个变量的回归系数路径变化趋势

Fig. 3 Key variables selected by CARS

(a): The changing trend of the number of sampled variables; (b): The changing trend of RMSECV value; (c): The changing trend of regression coefficients of each variables

马拉硫磷吸收光谱的每条光谱线含 1 582 个波长变量, 波长变量数目非常多, 需要在不削弱原来模型预测性能的情况下, 对波长变量和校正样本集进行优选, 简化模型, 建立一个简单和预测能力强的马拉硫磷浓度定量预测模型。

光谱数据在执行一次 CARS 算法(设置 MC 采样为 50 次)后, 随采样次数增加, 波长变量保留数变化情况、50 个不同的变量子集模型交叉验证均方根误差 RMSECV 值、波长变量回归系数变化路径分别如图 3(a), (b)和(c)所示。

如图 3(a)所示, 随着采样次数增加, 由于指数衰减函数 EDF 的作用, 波长变量的保留数目变化先快速后缓慢, 在采样 10 次之前先大幅减少, 经历了一个“粗选”过程之后再慢慢“精选”; 如图 3(b)所示, RMSECV 先减小, 说明多重共线性或含无关信息的波长变量被剔除, 模型预测能力提高, RMSECV 达最小值[对应图 3(c)星号垂直标示的第 34 次]之后逐渐增大, 说明一些含有有关马拉硫磷浓度值重要信息的波长变量被剔除, 导致模型的预测性能下降。因此由 CARS 算法在第 34 次采样时有最小 RMSECV, 获得了 18 个关键波长变量。

执行一次 MC-UVE 算法后的波长变量筛选结果如图 4 所示。

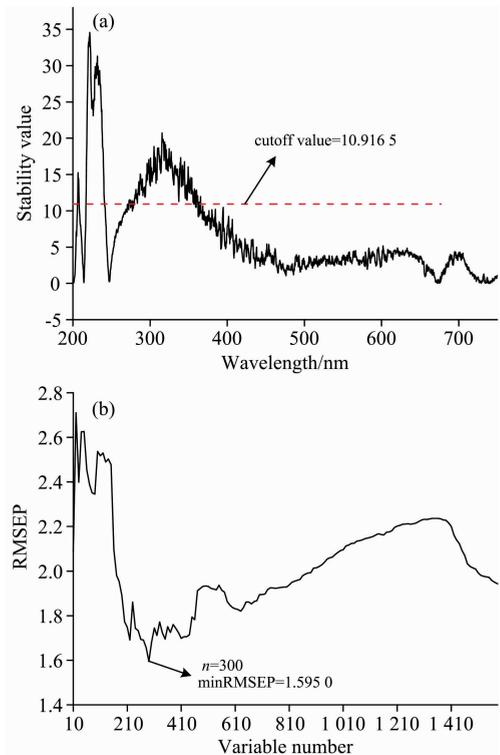


图 4 蒙特卡洛无消息变量消除法选出的关键变量

(a): 波长变量回归系数稳定性值;

(b): 不同变量子集的 PLS 模型的 RMSEP

Fig. 4 Key variables selected by MC-UVE

(a): Stability value of regression coefficient of wavelength variable; (b): The RMSEP value corresponding to PLS model obtained by different subset of variables

1 582 个波长变量的回归系数稳定性值变化情况如图 4

(a)所示,图中的水平虚线为稳定性阈值,该阈值是 RMSEP 最小的变量子集中波长变量回归系数的最小稳定性值,低于该值的波长变量被剔除;将样正集和预测集的波长变量按稳定性值从大到小排序,每次加入 10 个波长变量进行 PLS 建模和预测,直到加入所有的波长变量,得到不同变量子集 PLS 模型 RMSEP 的结果如图 4(b)所示,当波长变量数的保留数目为 300 时, RMSEP=1.595 0 为最小值,获得最优变量子集并得到稳定性阈值,之后保留变量数目增多, RMSEP 增大,说明无关信息增加,影响了预测的结果。

潜变量(latent variables, LVs)是一个非常重要的模型参数,直接影响到模型的预测性能。当潜变量选择少时,所建立的模型可能丢失较多有用的光谱信息,出现欠拟合现象,而当潜变量选择多时,模型中可能包含过多噪声,出现过拟合现象^[7]。因此合理选择建模 LVs 是充分利用光谱信息和降低噪声的有效方法之一。通过不同潜变量 PLS 模型预测 RMSEP 合理选择各模型的潜变量,采用 Full-PLS 和 CARS-PLS 模型潜变量选择过程如图 5 所示。

采用全波段变量以及由上述 CARS 算法和 MC-UVE 算法各自得到的波长变量子集,并分别选择合理的 LVs 建立相应的马拉硫磷浓度 PLS 定量预测模型,综合对比分析两种变量筛选方法在关键波长变量上的选取性能,分别建立的模型结果见表 1。

由图 4 和表 1 可知,运行一次 MC-UVE 后,得到 LVs=5 时,MC-UVE-PLS 的 minRMSEP=1.595 0。同时由图 5 可知,当 LVs=5 时,得到 Full-PLS 和 CARS-PLS 的 minRMSEP 分别为 1.226 8 和 0.923 0,三个模型的 LVs 相同。对于 MC-UVE-PLS 模型,其变量数虽然从全波段 1 582 个变量减少到 300 个,但其 RMSEP 高于以及 RPD 明显低于 Full-PLS 模型的值,可能是因为一些回归系数较稳定但含有无关信息的变量被保留了下来,导致模型的预测性能没有提高。而

CARS-PLS 模型在全面提高模型的预测性能上,仅用了 18 个变量,占全波段变量数的 1.13%,其 RMSEP 就比 Full-PLS 模型的值降低了 24.8%,且 RPD 值有所增加。

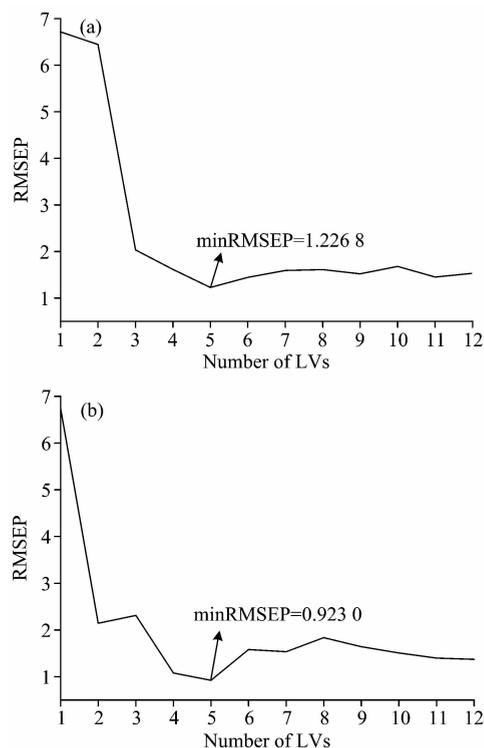


图 5 不同潜变量的 RMSEP 的变化情况

(a): 不同潜变量 Full-PLS 的 RMSEP 变化情况;

(b): 不同潜变量 CARS-PLS 的 RMSEP 变化情况

Fig. 5 Variation of RMSEP with LVs

(a): Variation of RMSEP of Full-PLS model with LVs;

(b): Variation of RMSEP of CARS-PLS model with LVs

表 1 不同波长变量集的马拉硫磷浓度 PLS 模型预测结果

Table 1 The PLS prediction model results of malathion concentration by using the different variables set

Model	Number of variables	LVs	Calibration set		Prediction set		
			R_c^2	RMSEC	R_p^2	RMSEP	RPD
Full-PLS	1 582	5	0.998 4	1.144 9	0.996 3	1.226 8	17.005 8
MC-UVE-PLS	300	5	0.998 5	1.082 1	0.993 5	1.595 0	13.080 1
CARS-PLS	18	5	0.998 5	1.105 2	0.997 9	0.923 0	22.603 2

通过对上述模型结果对比分析,表明 CARS 算法在马拉硫磷浓度吸收光谱关键变量的筛选性能上优于 MC-UVE 算法,说明 CARS 算法不但不会削弱模型的预测能力,还能进行关键变量的有效筛选,简化模型,提高模型稳健性。由此可知,模型的预测能力还与建模变量数和质量有关,通过提取关键变量和消除变量之间的多重共线性能够增强模型的预测能力。

2.3 样本优选与模型结果

对于建模集样本,除了要消除光谱变量之间存在较为严重的共线性之外,样本与样本之间差异性的大小也会影响模型的预测性能。合理地消除或者降低这种差异性,也有助于提高定量预测模型的预测性能。采用 MCCV 法对 50 个

样本运行一次(设置 MC 采样 2 500 次),以预测集(样本浓度预测)的残差均值为横坐标,残差标准差为纵坐标,得到 50 个样本的分布如图 6 所示。

将图 6 中预测残差平均值大于 3 和预测残差标准差大于 0.8 的 6 个样本剔除,被剔除的奇异样本序号分别为 24, 45, 47, 48, 49 和 50。最后得到优选的 44 个校正集样本。

为了验证 K-S 算法样本优选方法的性能,采用 MCCV 作为对照方法与其进行合理有效的比较,用 K-S 算法也筛选 44 个建模样本。由 MCCV 法和 K-S 算法分别得到的 44 个建模样本与 CARS 算法筛选的 18 个关键变量组成新的建模集,两模型 LVs 的选择过程分别如图 7(a, b)所示,然后分别建立相应的 PLS 定量预测模型,结果见表 2。其中马拉硫磷浓

度 CARS-CCV_s-PLS 模型预测结果如图 8 所示。

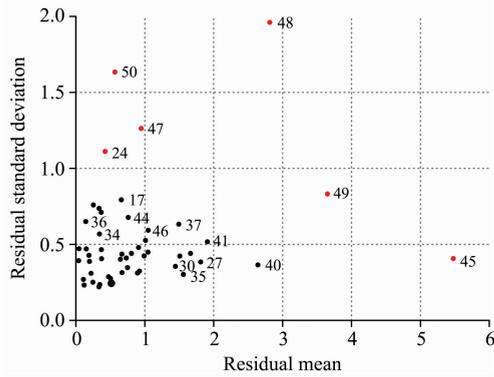


图 6 MCCV 法 50 个样本分布图

Fig. 6 The distribution diagram of 50 samples obtained from MCCV

由图 7(a, b)可知, 当 LVs 为 5 时, CARS-K-S_s-PLS 和 CARS-MCCV_s-PLS 的 minRMSEP 分别为 0.863 4 和 1.026 2。

由表 2 和图 8 可知, CARS-K-S_s-PLS 预测模型的 RMSEC 与 RMSEP 的值更相近, 即 RMSE 值更稳定, 且 CARS-K-S_s-PLS 的 RPD 高于 CARS-MCCV_s-PLS 的值, 说明 CARS-K-S_s-PLS 模型的预测能力更好, K-S 法能够对校正集样本进行优选, 且能够提高预测模型预测能力。MCCV 法通过样本残差均值和残差标准偏差这两个统计参数来剔除样本浓度预测结果情况较差的样本, 而 K-S 算法在特征空间中采用样本光谱为其特征变量, 基于变量之间的欧式距离来均匀选取样本, 充分考虑了样本光谱信息的影响。

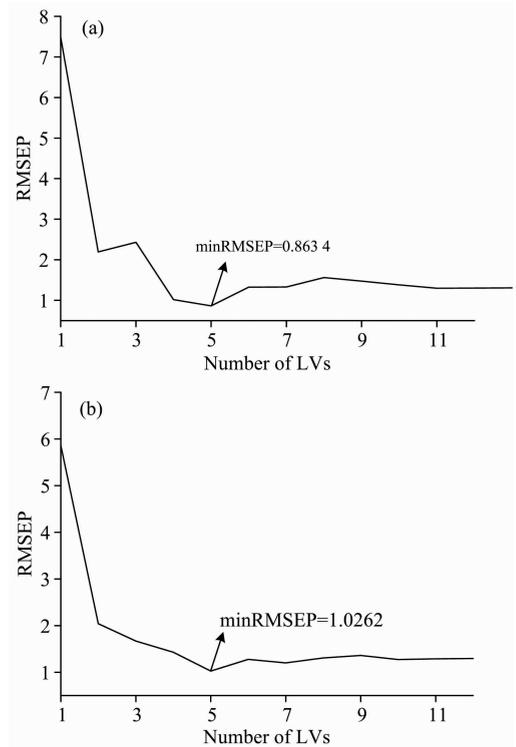


图 7 不同潜变量的 RMSEP 的变化情况

(a): 不同潜变量 CARS-K-S_s-PLS 的 RMSEP 变化情况; (b): 不同潜变量 CARS-MCCV_s-PLS 的 RMSEP 变化情况

Fig. 7 Variation of RMSEP with LVs

(a): Variation of RMSEP of CARS-K-S_s-PLS model with LVs; (b): Variation of RMSEP of CARS-MCCV_s-PLS model with LVs

表 2 基于关键变量和优选样本的 PLS 模型预测结果

Table 2 PLS model prediction results based on key variables and selected effective samples

Model	Number of variables	Number of calibration set samples	LVs	Calibration set		Prediction set		
				R ² _c	RMSEC	R ² _p	RMSEP	RPD
CARS-K-S _s -PLS	18	44	5	0.999 1	0.845 6	0.998 2	0.863 4	24.163 5
CARS-MCCV _s -PLS	18	44	5	0.998 8	0.702 3	0.998 9	1.026 2	20.330 1

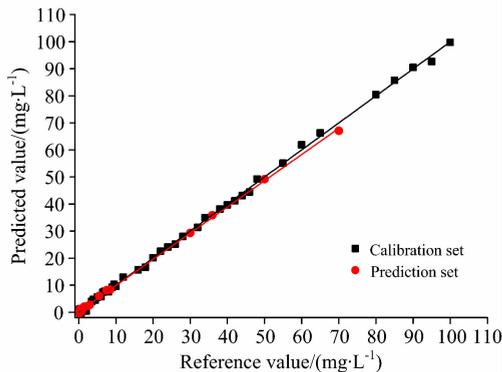


图 8 CARS-K-S_s-PLS 模型预测结果

Fig. 8 Prediction result of CARS-K-S_s-PLS model

3 结 论

(1)对马拉硫磷不同浓度梯度吸收光谱原始数据采用不同的预处理方法并建模, 结果表明均值中心化预处理结果最优。

(2)采用 CARS 算法和 MC-UVE 算法分别对全波段变量进行筛选并建模, 结果表明 CARS-PLS 模型预测结果最优, 建模变量数最少, 选出的 18 个关键波长变量比全波段的 1 582 个变量减少了 98.87%, 简化模型的同时提高了模型的预测能力。

(3)采用 CARS 算法筛选的 18 个波长变量, 分别结合 MCCV 法和 K-S 算法各自优选的 44 个样本建模, 结果表明 K-S 法在样本优选性能上略优, 可对校正集样本进行优选, 提高模型预测性能。

(4)通过对比分析各种算法建立的马拉硫磷浓度定量预测模型,结果表明采用 CARS 算法结合 K-S 法建立的 CARS-K-S_s-PLS 模型最优,可用于马拉硫磷浓度快速准确

的定量检测。

(5)该研究为采用紫外/可见吸光度光谱法快速实时检测水体中马拉硫磷农药浓度提供了理论依据。

References

- [1] YE Xue-zhu, WANG Qiang, ZHAO Shou-ping, et al(叶雪珠, 王强, 赵首萍, 等). China Vegetables(中国蔬菜), 2013, (4): 76.
- [2] JIANG Lu, YE Lin, YANG Xue, et al(姜露, 叶麟, 杨雪, 等). Food and Fermentation Industries(食品与发酵工业), 2016, 42(1): 200.
- [3] QIN Xiu-yuan, ZHAI Yuan-yuan, LI Xiao-yun, et al(秦修远, 翟媛媛, 李晓云, 等). Chinese Journal of Analysis Laboratory(分析试验室), 2017, 36(7): 790.
- [4] LI Shui-fang, ZHANG Xin, LI Jiao-juan, et al(李水芳, 张欣, 李娟娟, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2014, 30(6): 249.
- [5] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011. 89.
- [6] Cai Wensheng, Li Yankun, Shao Xueguang. Chemometrics and Intelligent Laboratory Systems, 2008, 90(2): 188.
- [7] Bian Xinhui, Li Shujuan, Shao Xueguang, et al. Chemometrics and Intelligent Laboratory Systems, 2016, 158: 174.
- [8] Li Hongdong, Liang Yizeng, Xu Qingsong, et al. Analytica Chimica Acta, 2009, 648(1): 77.

Study on Prediction Model of Malathion Pesticide Concentration Absorption Spectra Based on CARS and K-S

ZHEN Huan-yi, MA Rui-jun*, CHEN Yu*, SUN Xiao-peng, MA Chuang-li
College of Engineering, South China Agricultural University, Guangzhou 510642, China

Abstract In this study, the fast and effective quantitative prediction analysis model was established by using the absorption spectrum data of different concentration gradients of malathion in the ultraviolet/visible wavelength range. In the process of establishing a prediction model, the quality of the calibration set samples and wavelength variables involved in the modeling plays a decisive role in the predictive ability of the quantitative analysis model. Therefore, firstly checked whether there were abnormal samples in the experimental samples, then used the different preprocessing methods for the spectral data in the wavelength range of 200.08 to 750.04 nm and then established corresponding PLS model. Further based on the spectral data of the optimal preprocessing result (mean centering), competitive adaptive weighted algorithm (CARS) and Monte Carlo-uninformative variable elimination method (MC-UVE) were used to select the key wavelength variables respectively and established corresponding PLS prediction model. Model results indicated that CARS algorithm was superior to MC-UVE algorithm in the performance of key variable screening; then 18 wavelength variables (1.137 8% of the original variable number) selected by CARS algorithm combined with the 44 modeled samples (88% of the original sample number) respectively obtained from Kennard-Stone (K-S) algorithm method and Monte Carlo cross-validation method (MCCV) to establish CARS-K-S_s-PLS and CARS-CCV_s-PLS quantitative prediction model, which R_p^2 were 0.998 2 and 0.998 9, RMSEP were 0.863 4 and 1.026 2, and RPD were 24.163 5 and 20.330 1, as a result the CARS-K-S_s-PLS model was slightly better CARS-CCV_s-PLS model. The experimental results showed that the CARS algorithm could eliminate variables with weak correlation with sample concentration and effectively eliminate irrelevant spectral information. The K-S algorithm can help to select a better modeling sample set. UV-Vis absorption spectrum of malathion pesticides combined with the CARS-K-S_s-PLS model established by the CARS algorithm and K-S algorithm can predict malathion pesticide concentration. This study provides a certain of the important theoretical basis and experimental basis for the rapid detection of organophosphorus pesticide concentration by spectroscopy technology, and has a good application prospect in the field of rapid detection of organophosphorus pesticide.

Keywords Malathion; UV-Vis absorption spectrum; Key variable screening; Sample selection; Quantitative prediction

* Corresponding authors

(Received Aug. 13, 2019; accepted Dec. 18, 2019)