基于改进 Hodrick-Prescott 分解模型的近红外自适应降噪方法

谢德红[1]，李俊锋[2]，刘 蔚[3]，万晓霞[4]，叶 艺[1]

1. 南京林业大学轻工与食品学院，江苏 南京 210037
2. 河南牧业经济学院包装与印刷工程学院，河南 郑州 450046
3. 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
4. 武汉大学湖北省文物颜色信息数字化与虚拟再现工程研究中心，湖北 武汉 430079

摘 要 在检测果蔬农药残留的近红外光谱采集过程中，往往会受噪声干扰获得低信噪比近红外光谱，且近红外光谱中表征农药和果蔬化学组分的谱峰微弱且重叠度高，因而此近红外光谱降噪普遍存在易平滑微弱的农药组分谱峰、或增加非测量物化学组分谱峰的危险，导致在后续以仅挖掘红外光谱谱峰特征为前提的分类和化学组分分析中，恶化近红外光谱的分类精度、影响农药残留成分的正确分析。针对抑制近红外光谱噪声与保持近红外光谱谱峰的矛盾，提出一种改进 Hodrick-Prescott 分解模型的自适应降噪方法。在该方法的 Hodrick-Prescott 分解模型中，以染噪光谱与复原光谱之间残差的 L2 范数为残差项，描述高斯噪声结构，以复原光谱信号二阶差分的 L2 范数为正则化项，惩罚复原光谱、迫使从染噪光谱中复原的光谱倾向于梯度减少的方向，以平滑噪声、保持原始谱峰信息。该方法同时结合 L-曲线方法，自适应地获取染噪光谱在 Hodrick-Prescott 优化方程中的正则化参数，并通过求解该曲线最大曲率点对应的参数获得最优正则化参数，确保能平衡 Hodrick-Prescott 分解模型中正则化项和残差项，以得到较为理想的光谱复原结果。实验以农药残留和未残留的上海青近红外光谱为基本数据、通过降噪前后信噪比、以及支持向量机分类模型的识别率，对比分析 bior6.8 小波分解方法、sym8 小波分解方法、互补集合模态分解方法的降噪效果。实验结果显示，该方法在处理 18.79 dB 信噪比染噪近红外光谱时获得了 33.35 dB 信噪比；在实施上海青农药残留检测中，处理训练集与测试集近红外光谱数据后，训练所得支持向量机分类模型的训练集识别率达 93.58%、测试集识别率达 71.18%，此识别率明显高于上述三种方法降噪后的结果，接近于原始未染噪声光谱数据。该方法在近红外光谱降噪方面具有明显的优势，能应用于农药残留近红外光谱检测的前期处理。

关键词 Hodrick-Prescott 分解；L 曲线；自适应；降噪；近红外光谱

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2020)05-1650-06

引 言

农药的高效防治功能，对促进农作物增产、缓解农产品供需矛盾有举足轻重的作用。然而，农药大量不合理使用，导致农药残留超过食品安全标准，引发诸多食品安全问题[1]。当前一些化学、生物等检测方法精度虽高，但存在耗时费力、破损被测物、污染环境等缺点，难以满足现代农业快速、无损、批量及实时检测的需求。近红外光谱因其无损、快速、无污染的特点，在农药残留定性和定量检测应用中优势明显。但是，近红外光谱是典型的弱信号，被测物化学组分的谱峰相对微弱且重叠度高。此外，由于近红外光谱采集仪器和采集环境等因素，实测近红外光谱常受噪声干扰。在后续近红外光谱数据分类中，无法有效区别此光谱中被测物化学组分信息和噪声，进而影响分类精度。有效地去除光谱数据中的噪声，对建立稳定性好、分类精度高的农药残留检测的分类模型十分重要。

小波分解[2-3]和经验模太分解(empirical mode decomposition，EMD)[4-5]可较好刻画光谱信号的特征，近些年常被用于近红外等各类光谱信号的降噪中，并在信噪比小时有一

收稿日期：2019-04-11，**修订日期**：2019-08-26

基金项目：国家自然科学基金项目(61275172，61575147)，江苏省高校优势学科建设工程项目(164030934)，江苏省制浆造纸科学与技术重点实验室开放基金项目(201526)，南京林业大学青年科技创新基金项目(CX2018024)，南京林业大学大学生创新创业训练计划项(2018NFUSPITP680)，食品安全大数据技术北京市重点实验室开放基金项目(BTBD-2019KF02)资助

作者简介：谢德红，1979年生，南京林业大学轻工与食品学院讲师 e-mail：dehong.xie@gmail.com

定抑制噪声的效果。小波分解方法是在小波分解的基础上通过阈值手段去除噪声，因此小波分解的基函数(即小波基)和阈值选择与染噪光谱特征的适应性对降噪至关重要。有研究[3]提出利用 Stein 无偏风险估计(Stein's unbiased estimate of risk，SURE)有效地解决了阈值自适应问题，但小波基与染噪光谱特征的匹配问题却鲜有研究关注。据文献[6]研究发现，当小波基与被分解信号特征不匹配时，重构信号易产生 Gibbs 震荡现象。此震荡波破坏了光谱的几何特征，导致在后续近红外光谱数据分类中易被误认为被测物化学组分的谱峰，恶化分类精度、影响微弱的农药残留成分的正确分析。EMD 克服了小波基与被分解信号相适应的要求，可将任意特征的光谱信号分解成不同的震荡成分，即本征模态函数(intrinsic mode functions，IMFs)，但在信号重构时易产生混叠模态问题。混叠模态是由经验模分解的过程中信号因局部极值在很短的时间(或很短的波段)间隔内发生多次跳变导致的。在光谱信号染噪后，混叠模态中的信号状态很难完全避免。为此，有学者提出了集合模态分解(ensemble empirical mode decomposition，EEMD)[4]。此方法是一种噪声协助分析方法，因而当分解出的 IMFs 数量比较有限时，各 IMFs 不可避免的存在噪声，导致重构信号的信噪比降低。互补集合模态分解(complete ensemble empirical mode decomposition，CEEMD)[5]可抑制 IMFs 里的残余噪声，但当参数选择不当时，会产生错误成分使最后获得的 IMFs 不能真正满足 IMFs 的定义，导致重构信号仍会存在一些小震荡，在近红外光谱中呈现为额外的谱峰。为了提高近红外光谱的信噪比、减少光谱信号的失真以及额外谱峰的产生或残留、满足果蔬农药残留定性检测的应用需求，本研究提出了 L 曲线与 Hodrick-Prescott 分解模型相结合的自适应降噪方法，复原被测物表面原始近红外光谱，为后续近红外光谱数据分类、农药残留检测、及农药残留化学组分信息分析和挖掘奠定基础。

1　改进 Hodrick-Prescott 降噪的原理与方法

Hodrick-Prescott 分解模型[7]是一种信号分离方法，广泛用于经济数据中经济趋势和各时间点震荡波的分离。在果蔬农药残留的近红外光谱中，用以表征被测物化学组分的谱峰相对较弱且时常发生重叠，因而其光谱波形随波长的变化相对缓慢，如同经济趋势的特征；通常，近红外光谱采集过程中所产生的噪声随波长变化相对剧烈，如同经济数据中随时间点变化的震荡波特征。因此将依据染噪近红外光谱的特征，建立基于 Hodrick-Prescott 分解模型的近红外光谱降噪的方法，并与 bior6.8 小波分解方法、sym8 小波分解方法、互补集合模态分解方法进行比较，验证本方法的有效性。

1.1　Hodrick-Prescott 降噪原理与方法

设被测物的原始近红外光谱(即未染噪声的近红外光谱)为 $\boldsymbol{f}=(f_i, \cdots, f_j)\in R^n$，其近红外光谱在采集过程中受广义高斯噪声的污染获得染噪近红外光谱为 $\boldsymbol{s}=(s_i, \cdots, s_j)^{\mathrm{T}}\in R^n$，它们与噪声 $\boldsymbol{e}=(e_i, \cdots, e_j)^{\mathrm{T}}\in R^n$，关系如下

$$\boldsymbol{s}=\boldsymbol{f}+\boldsymbol{e} \tag{1}$$

下标 i 和 j 表示波长，n 为近红外波长范围内的波长点的数量。

利用 Hodrick-Prescott 分解方法降噪是通过引入一定约束条件，将光谱信号复原为具有存在性、唯一性且适定的结果，且此结果为受噪声干扰小的过程。已知染噪近红外光谱信号 $\boldsymbol{s}$，求原始近红外光谱 $\boldsymbol{f}$ 的 Hodrick-Prescott 分解方法可转化为目标函数 $J(\boldsymbol{f})$的最小化问题

$$J(\boldsymbol{f})=(1/2)\|\boldsymbol{s}-\boldsymbol{f}\|_2+\kappa\|\boldsymbol{D}\boldsymbol{f}\|_2 \tag{2}$$

此目标函数由残差项和正则化项两部分构成。其中，$\|\boldsymbol{s}-\boldsymbol{f}\|_2$ 为残差项，保证复原信号的保真度；$\|\boldsymbol{D}\boldsymbol{f}\|_2$ 为正则化项，约束近红外光谱 $\boldsymbol{f}$ 的波形特征；$\boldsymbol{D}^{(n-2)\times n}$为二阶差分矩阵

$$\boldsymbol{D}=\begin{bmatrix}1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1\end{bmatrix} \tag{3}$$

用于惩罚复原光谱信号的震荡波动，迫使其趋向梯度减少的方向。$\|\cdot\|_2$ 表示 l_2 范数。$\kappa\geqslant 0$ 为正则化参数，用以调整目标函数两部分的权重。若 $\kappa=0$，正则化项的惩罚作用失效，求得的复原光谱信号 $\boldsymbol{f}$ 趋于染噪光谱信号；反之，若 $\kappa\to+\infty$，则正则化项惩罚起主导作用，求得的复原光谱信号 $\boldsymbol{f}$ 只会尽可能满足 $\|\boldsymbol{D}\boldsymbol{f}\|_2$ 最小，因而其复原光谱信号逼真程度比较差、偏离原始光谱信号，甚至无法呈现原始光谱信号的基本波形。由此可知，选择合适的正则化参数 κ，对获得最优降噪效果至关重要。

1.2　正则化参数自适应获取方法

依上所析，正则化参数 κ 越大、惩罚越严重，复原出的光谱信号越平滑、含震荡波动程度越弱。虽易去除噪声，但也存在平滑掉有利于数据分类的小谱峰的危险；反之，正则化参数 κ 越小，正则化项的惩罚作用越小，复原光谱信号则越易残留噪声。此外，原始 Hodrick-Prescott 方法中的正则化参数 κ 的选择是非自适应的，选择一个固定值，难以适应实测近红外光谱非稳定特征及实测近红外光谱信噪比有大有小的状况。基于此，为了使正则化参数与实测染噪近红外光谱能自适应，提出利用 L-曲线方法[8]获得 Hodrick-Prescott 方法中的自适应值。

L-曲线是指各正则化参数下对应复原光谱信号的残差项 $\|\boldsymbol{s}-\boldsymbol{f}^{\kappa}\|_2$ 与正则化项 $\|\boldsymbol{D}\boldsymbol{f}\|_2$ 所构成的 log-log 尺度曲线，其形状接近 L 形。

$$l:=\{(\log\|\boldsymbol{s}-\boldsymbol{f}^{\kappa}\|_2, \log\|\boldsymbol{D}\boldsymbol{f}^{\kappa}\|^2): \kappa\geqslant 0\} \tag{4}$$

其中，"l"表示 L-曲线，$\boldsymbol{f}^{\kappa}$ 为每个 κ 值下的复原光谱信号。由式可知，每个 κ 值均存在一个最优 $\boldsymbol{f}^{\kappa}$ 使目标函数 $J(\boldsymbol{f})$趋于最小。由此，L-曲线上每个点则表示每个 κ 值下的最小 $J(\boldsymbol{f})$值，其拐点表示所有最小 $J(\boldsymbol{f})$值中最小值，且拐点处对应的 κ 值则为与当前染噪光谱信号最优适应值。

2　结果与讨论

实验采集 341 个上海青(一种青菜)近红外光谱样本，并在采集的近红外光谱上人工添加广义高斯噪声作为降噪对

象。为验证本方法的有效性，选用 bior6.8 和 sym8 小波基的 SURE 阈值方法[3]及 CEEMD 方法[5]作对比，并采用信噪比(signal noise ratios，SNR)、均方根误差(root mean square error，RMSE)及降噪光谱数据训练所得分类模型识别率评价降噪效果。

2.1 实验数据

以无锡迅杰光远科技有限公司的 IAS-2000 近红外光谱仪为近红外光谱采集设备，光谱仪采集参数设置如下：平均采样次数为 30 次，采集波段为 900～1 700 nm，测量间隔为 1 nm。光谱采集环境：装有空调的恒温环境(23 ℃)内。采集的实验样品为无农药残留上海青和喷洒了乐果(农药，C5H12NO3PS2)的上海青。其中，无农药残留上海青购自农贸市场，并通过小苏打和食盐浸泡、自来冲、蒸馏水先后冲洗晾干；残留农药上海青则由无农药残留上海青喷晒不同浓度配比的乐果溶液并晾干所得。其中，乐果为河南省周口市德贝尔生物化学品工程有限生产的 40%乐果乳油；不同浓度配比的乐果溶液由蒸馏水稀释所得，且稀释的最小和最大倍数分别为 50 倍和 1 500 倍。实验分别采集残留农药样本 291 个、未残留农药样本 50 个。从农药残留样本和未残留农药样本中分别选取 146 个农药残留样本和 25 个未残留农药样本构成训练样本集，剩余为测试样本集。其中，146 个农药残留样本上的乐果浓度横跨最小与最大配比。

为了便于验证降噪效果，设定上述采集的近红外吸收光谱为原始纯光谱信号，模拟叠加不同程度高斯噪声后的光谱为实验待处理的染噪光谱。图 1 上排图为稀释 50 倍的农药残留样本，叠加噪声后的近红外光谱信噪比分别为 19.07 dB[如图 1(a)]和 18.79 dB[如图 1(b)]。

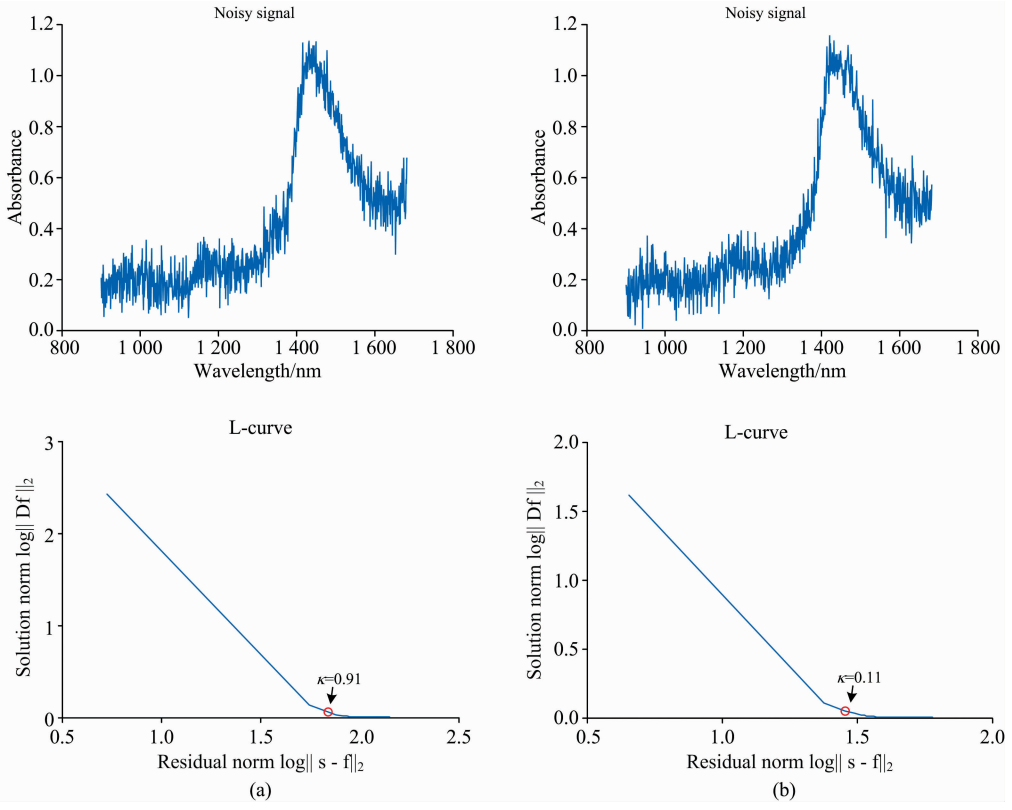

图 1 染噪近红外光谱和其 L-曲线

(a)：SNR=19.07 dB；(b)：SNR=18.79 dB

Fig. 1 Noise-contaminated NIR spectroscopy and their L-curve

(a)：SNR=19.07 dB；(b)：SNR=18.79 dB

2.2 降噪效果与分析

对任一组染噪近红外光谱，可利用 L-曲线方法获得此光谱的自适应 κ 值，其步骤执行如下：先设定 κ 的范围和步长，设定 $\kappa=0.001, 0.002, \cdots, 29.999, 30$；然后分别计算各 κ 值下残差项 $\| \boldsymbol{s}-\boldsymbol{f}^{\kappa} \|_2$ 与正则化项 $\| \boldsymbol{Df} \|_2$ 的值，在获得 L-曲线的基础上，获取曲线最大曲率点处的 κ 值；最后，在此 κ 值下，利用 Hodrick-Prescott 方法复原的近红外光谱 $\boldsymbol{f}^{\kappa}$ 即为降噪后的近红外光谱。依上述步骤，图 1 中两组染噪近红外光谱的自适应正则化参数分别为 0.91 和 0.11，如图 1(a,b)下排图所示。

为了衡量、比较降噪效果，首先采用了 SNR 和 RMSE 两个指标。其中，SNR 值大小与降噪效果成正比，RMSE 值大小则与降噪效果成反比。表 1 利用 SNR 和 RMSE 两个指标比较了 bior6.8 和 sym8 小波的 SURE 阈值方法(均为五层小波分解)[3]、CEEMD 方法[5]、以及本方法。首先，通过 SNR 和 RMSE 值的横向比较发现，本方法的 SNR 值最大、

RMSE 值最小，说明相对其他三种方法，本方法的降噪效果最优。其次，结合两组染噪近红外光谱的纵向数据比较发现，降噪方法优劣顺序未随噪声不同而变化，说明本文方法对不同噪声的降噪效果相对稳定。SNR 和 RMSE 均是从全局的角度考察降噪前后光谱相似程度，未能反映局部缺陷，而局部缺陷恰恰容易恶化后续数据的分类精度。例如，当降噪后光谱出现沿未染噪光谱上下波动的小波峰波谷，且波峰和波谷构成的增加和减少的面积能相互抵消时，SNR 和 RMSE 则不能从数据上体现额外小波峰波谷的局部缺陷存在，但此额外产生的小波峰波谷在后续数据分类中又极易误认为是物质化学组分的谱峰，影响分类的精确度。因此，本研究将继续从局部和后续数据分类精度进一步验证降噪效果。

当被测物含量有限时，其近红外光谱(或其对应化学组分的近红外光谱谱峰)也比较弱。例如，在本实验中，上海青本身的化学组分含量一般远高于残留农药的化学组分，因而农药残留上海青的近红外光谱中除较弱的、反映农药化学组分的谱峰外，主要为宽波段特征、且信号较强的上海青化学组分的近红外光谱。此特点使得残留样本的近红外光谱与未残留样本的近红外光谱识别度小。为了增强农药谱峰的可识别度，在近红外光谱数据分类、判断农药残留与否的过程中，常用导数处理方式抑制背景光谱信号(如上海青的近红外光谱)、增强目标识别物化学组分的谱峰(如农药化学组分的谱峰)。一阶导数的处理方式是双刃剑，会增强农药化学组分谱峰也会放大噪声等非光谱成份，但均有利于从局部判断降噪后光谱与原未染噪光谱的相似度。因此，利用光谱一阶导数谱图进一步评判各方法的降噪效果，如图 2 所示。

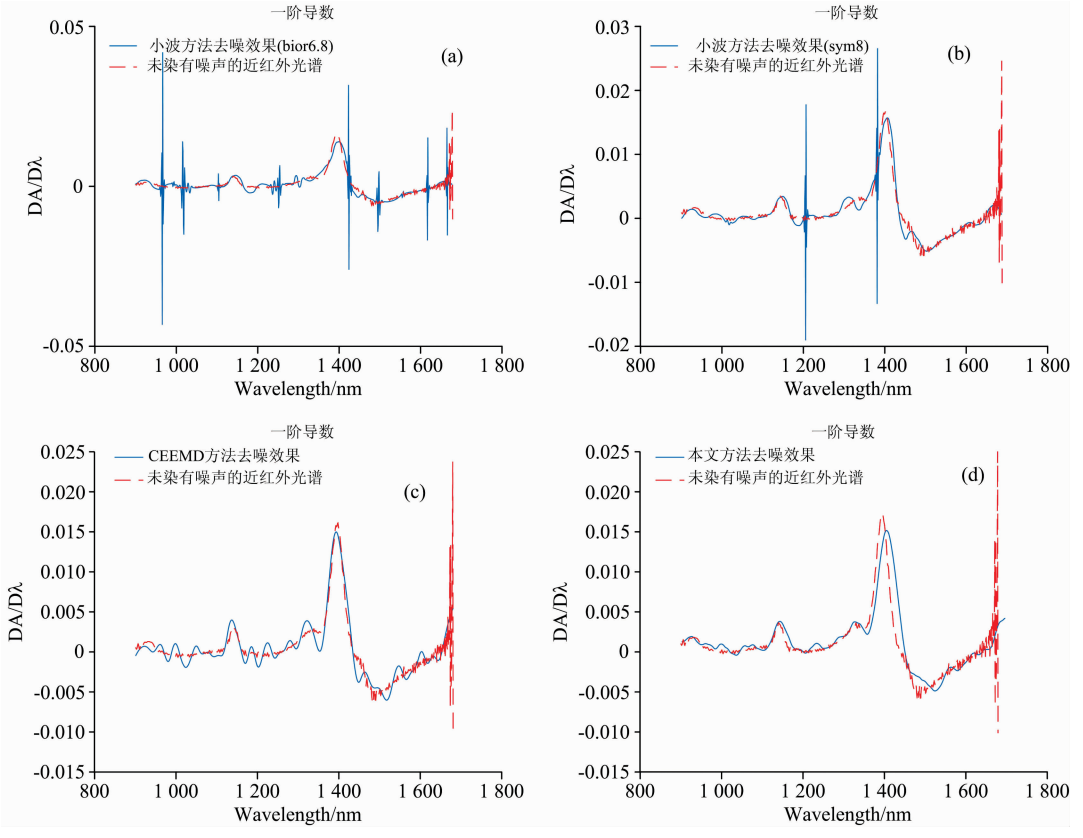

图 2　降噪后光谱和未染噪声光谱的一阶导数谱图

(a)：小波法(bior6.8 小波核)；(b)：小波法(sym8)；(c)：CEEMD 法；(d)：本方法

Fig. 2　First derivative spectrum of the denoised NIR spectroscopy and its original

(a)：Wavelet method (with bior6.8 wavelet kernel)；(b)：Wavelet method (with sym8 wavelet kernel)；(c)：CEEMD method；(d)：The proposed method

图 2 各分图均为图 1(a)对应染噪近红外光谱经过降噪和一阶求导后的结果。图中实线代表降噪后光谱的一阶求导结果，虚线则代表原未染噪声光谱的一阶求导结果。在图 2(a)中，实线图谱大约于 980，1 020，1 100，1 290，1 410，1 500和 1 620 nm 等波段处呈现明显的震荡波，而此震荡波在虚线的谱图上均不可见。由此推测，此额外震荡波为 bior6.8 小波方法降噪所产生、非被测物化学组分，会干扰后续光谱数据的分类、恶化分类精度；图 2(b)显示 sym8 小波方法也会产生较强的震荡波；图 2(c)实线图谱中虽无明显的震荡波，但相对于虚线所示谱图，仍然呈现额外的小波峰波谷，此小波峰波谷与某些物质化学组分的近红外二级和三级倍频波[9]难以区分，也极易被误当作上海青或农药的化学组分的小谱峰；图 2(d)中实线与虚线表示的谱图比较接近，且无明显额外的震荡波和小谱峰。由此分析可知，从局部特征

判断，相对上述其他三种降噪方法，本方法降噪后的效果优，也更有益于后续化学组分分析以及农药残留检测为目的的数据分类。

表 1 降噪前后近红外光谱信噪比与均方根误差对比

Table 1 Comparison of signal-to-noise ratio and the root mean square error of the denoised NIR spectroscopy

染噪光谱 SNR	评价指标	sym8	bior6.8	CEEMD	本方法
19.07	SNR	32.63	32.56	33.07	35.75
	RMSE	0.011 7	0.011 8	0.011 1	0.008 2
18.79	SNR	31.69	31.50	31.71	33.35
	RMSE	0.013 0	0.013 3	0.013 0	0.010 8

表 2 降噪前后近红外数据的 SVM 分类结果

Table 2 SVM classification results based different denoising methods

光谱	训练集		测试集	
	识别数量	识别率/%	识别数量	识别率/%
原始光谱	165/171	96.49	125/170	73.53
sym8	131/171	76.61	86/170	50.59
bior6.8	148/171	86.55	91/170	53.53
CEEMD	156/171	91.22	113/170	66.47
本方法	160/171	93.58	121/170	71.18

研究近红外光谱降噪的主要目的是减少噪声对后续农药残留检测中数据分类的干扰，从而提高训练所得分类模型的拟合能力和泛化能力。由此，对染噪训练样本数据和测试样本数据降噪处理后，再经过一阶导数预处理、主成分分析(PCA)选取十维特征的前提下，建立检测农药残留与否的支持向量机(SVM)[10]分类模型，并分别通过训练样本数据和测试样本数据获得此分类模型的识别率，以进一步评价降噪方法优劣。表 2 比较了原未染噪光谱及四种方法降噪后光谱的分类效果。由分类原理可知，上述额外震荡波和小谱峰的存在，会导致分类数据在特征空间的分布结构发生变化，不仅影响训练所得分类模型拟合数据的能力，更会恶化分类模型的泛化能力(即正确识别非训练近红外光谱数据所属类别的能力)。表 2 中的数据显示了与理论推导一致的结论。由表 2 中的分类模型的识别率还可发现，本方法降噪后的光谱数据训练所得的 SVM 分类模型，其训练集和测试集的识别率远高于其他三种方法的结果，接近原始无噪声近红外光谱数据的结果。此结果表明，本方法降噪效果优良，且对分类模型的准确率影响最优。

3 结 论

提出了一种改进 Hodrick-Prescott 分解模型的自适应近红外光谱降噪方法，充分利用 Hodrick-Prescott 分解模型中正则化项对复原光谱的惩罚作用，迫使复原光谱倾向于梯度减少的方向，并结合 L-曲线方法自适应地获取了 Hodrick-Prescott 分解模型中正则化参数，从而实现了染噪近红外光谱自适应降噪的目的。实验从信噪比、均方根误差及分类模型识别率数据证明了本降噪方法具有一定的优越性，可为近红外光谱定性检测提供可靠的噪声预处理方法。

References

[1] JIANG Meng-yun, GONG Wen-wen, LIU Qing-ju, et al(蒋梦云, 巩文雯, 刘庆菊, 等). Guangdong Agricultural Sciences(广东农业科学), 2018, 45(11): 60.

[2] ZHU Wen-bin, LEI Bing-shan, LEI Zhi-yong(朱文斌, 雷秉山, 雷志勇). Infrared Technology(红外技术), 2018, 40(11): 1047.

[3] Guo Q, Zhang C. Science Asia, 2012, 38: 207.

[4] Santhosh M, Venkaiah C, Vinod Kumar D M. Energy Conversion and Management, 2018, 168: 482.

[5] Elouaham S, Dliou A, Latif R, et al. International Journal of Computer Applications, 2016, 149(7): 39.

[6] Xia Y X, Ni Y Q. International Journal of Distributed Sensor Networks, 2018, 14(12): 1.

[7] Ouahilal M, Mohajir M E, Chahhou M, et al. Journal of Big Data, 2017, 4(1): 31.

[8] Chen M, Su H, Zhou Y, et al. Biomedical Optics Express, 2016, 7(12): 5021.

[9] Stuart B H. Infrared Spectroscopy: Fundamentals and Applications: John Wiley & Sons, Ltd, 2005.

[10] Ullah R, Khan S, Javaid S, et al. Biomedical Optics Express, 2018, 9(2): 844.

An Improved Hodrick-Prescott Decomposition Based Near-Infrared Adaptive Denoising Method

XIE De-hong[1], LI Jun-feng[2], LIU Di[3], WAN Xiao-xia[4], YE Yi[1]

1. School of Light Industry and Food, Nanjing Forestry University, Nanjing 210037, China
2. School of Packaging and Printing Engineering, Henan University of Animal Husbandry and Economy, Zhengzhou 450046, China
3. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China
4. Hubei Province Engineering Technical Center for Digitization and Virtual Reproduction of Color Information of Culture Relics, Wuhan University, Wuhan 430079, China

Abstract During the rapid detection of pesticide residue in fruits and vegetables by near-infrared (NIR) spectroscopy, NIR spectroscopy is often contaminated by noises. Meanwhile, the peaks in NIR spectroscopy of chemical components of pesticides and fruits and vegetables are weak and highly overlapped, so denoising the NIR spectroscopy has risks of smoothing weak peaks of pesticide components or generating peaks of non- chemical components. In the subsequent classification or chemical composition analysis, the above problems deteriorate the accuracy of classification of the NIR spectroscopy and influence analysis of chemical components of pesticide residue. In order to solve the conflict between noise suppression and peak maintenance of the NIR spectroscopy, an adaptive denoising method is proposed based on an improved Hodrick-Prescott decomposition model. In the model, L2 norm of the residual between the noisy near-infrared spectroscopy and its restored spectroscopy is used as the residual term to describe the Gaussian noise structure, and L2 norm of the second-order difference of the restored spectroscopy is used as the regularization term to penalize the restored spectroscopy. The penalty can force the restored spectroscopy to reduce its gradient, resulting in smoothing noises and keeping the original peaks. In order to acquire the regularization parameter in the optimization equation of the Hodrick-Prescott decomposition model adaptively, an L-curve method is combined into the method. So in the method, the optimal regularization parameters are obtained by solving the parameters corresponding to the maximum curvature point of the L-curve, which can balance the regularization term and the residual term in the Hodrick-Prescott decomposition model and finally obtain ideal restored spectroscopy. In order to compare wavelet decomposition methods with bior6.8 basis and sym8 basis and complete ensemble empirical mode decomposition method, signal-to-noise ratio (SNR) is computed, and a support vector machine (SVM) classification model is established by NIR spectroscopy of Shanghai Qing with pesticide and without pesticide. The experimental results show that the SNR value of the denoised NIR spectroscopy can reach to 33.35 dB when using the proposed method to deal with the noisy spectroscopy with 18.79 dB. Then the method is applied to denoising the NIR spectroscopy in the training set and the testing set, the recognition rate of vector machine classification model trained by the training set and testing set are 93.58% and 71.18% respectively. This recognition rate is significantly higher than the results using the above three denoising methods and is close to the results of the original uncontaminated spectroscopy. The results validate that the proposed method has better denoising effect than other methods mentioned, which can improve the stability of NIR classification model to pesticide residue detection.

Keywords Hodrick-Prescott decomposition; L-curve; Adaptive; Denoising; Near-infrared spectroscopy

(Received Apr. 11, 2019; accepted Aug. 26, 2019)