

A Novel Adaptive Robust NIR Modeling Method Based on Sparse Bayesian Learning

Yuqiang Li , Wenli Du , Senior Member, IEEE, Xinjie Wang , and Huijing Yu 

Abstract—The near-infrared (NIR) method has shown great potential in estimating key parameters in various industrial processes. Selecting characteristic wavelengths from high-dimensional spectra is crucial in building a prediction model with the satisfactory performance. However, existing wavelength selection methods based on the wavelength or waveband importance estimation are time-consuming and unstable. To improve the stability and generalization of the established model, a novel adaptive robust method is proposed for NIR modeling in this work, in which the pattern-coupled sparse model has been developed to estimate the spectral peak wavebands adaptively. Furthermore, a robust method based on global-local shrinkage is developed for characteristic wavelength selection. Compared with the state-of-the-art techniques, the proposed method is more accurate and robust for NIR modeling on both benchmark and real datasets.

Index Terms—Characteristic wavelength selection, robust modeling, sparse Bayesian learning (SBL), waveband estimation.

I. INTRODUCTION

NEAR-INFRARED (NIR) has been widely adopted to characterize the physiochemical properties of petrochemical [1], pharmaceutical [2], biomedical [3], and agricultural [4] products due to its simple, noninvasive, and nondestructive characteristics. As an indirect analysis method, building the

Manuscript received 20 August 2023; revised 19 December 2023; accepted 7 February 2024. Date of publication 18 March 2024; date of current version 5 June 2024. This work was supported in part by the National Key Research and Development Program-Intergovernmental International Science and Technology Innovation Cooperation Project under Grant 2021YFE0112800, in part by the National Natural Science Foundation of China under Grant 62136003, in part by the National Natural Science Foundation of China under Grant 62303186 and Grant 62373154, and in part by the Fundamental Research Funds for the Central Universities under Grant 222202417006. Paper no. TII-23-3186. (Corresponding author: Wenli Du.)

Yuqiang Li and Wenli Du are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: yuqiangli@mail.ecust.edu.cn; wldu@ecust.edu.cn).

Xinjie Wang is with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: wangxijie0621@foxmail.com).

Huijing Yu is with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China, and also with the China Aviation Lithium Battery (Shenzhen) Company Ltd., Shenzhen 518000, China (e-mail: y20170094@mail.ecust.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2024.3367007>.

Digital Object Identifier 10.1109/TII.2024.3367007

prediction model that reveals the relationship between the acquired spectra and the properties of interest is the basis of NIR analysis and its application. However, most existing methods have challenges in establishing the calibration models with generalization and stability [5]. On the one hand, there are redundant information and interferences caused by instrument vibration and fluorescence interference in the measured spectra, which inevitably affects the attraction of useful information and subsequent modeling analysis. On the other hand, with the widespread application of high-resolution spectrometers and the limitation of sampling conditions, the number of samples collected from the industrial process is generally much smaller than the spectral dimensionality, making it more challenging to extra the characteristic wavelengths through existing methods [6].

Over the past decades, various methods have been developed for spectral wavelength selection. According to the continuous characteristics of the analyzed spectra, wavelength selection approaches can be generally categorized as wavelength point selection (WPS) and wavelength interval selection (WIS) [7]. The WPS methods regard a single wavelength point or wavelength combinations as a unit and estimate their contribution to model performance, which can effectively eliminate redundant variables and improve model accuracy [8]. However, the computational cost of the WPS method increases dramatically as the wavelength dimension increases. In contrast, the WIS methods consider each waveband containing a certain number of continuous wavelengths as a unit for interval selection, which can effectively reduce the spectral dimensionality and provide a better interpretation of the selected intervals associated with the functional groups (C-H, O-H, N-H, S-H, etc.) of interest. Nevertheless, most existing WIS methods divide the entire waveband into multiple equal-length intervals, which inevitably divides some continuous characteristic wavelengths or introduce irrelevant information into the peak intervals [9]. To balance computational cost and model performance, the hybrid wavelength selection methods combined with the merits of WIS and WPS have been applied to wavelength selection, where WIS is used for rough interval searching, and WPS is used for fine wavelength selection [10]. Although various hybrid strategies have been developed, the equal-length interval division adopted by existing methods cannot accurately describe the actual peak waveband characteristics [11]. Specifically, there are differences in the peak widths of different spectral peaks, so it is difficult to accurately divide the peak wavebands through equal-length interval division [12].

In fact, the wavelengths that characterize the functional groups and attribute information are mainly located in the corresponding spectral peak wavebands. Compared with large-scale wavelengths, the number of spectral peaks is relatively small, that is, the analyzed NIR generally exhibits sparsity [13]. Therefore, the sparse coefficients obtained from representation learning can be used to identify the peak and no-peak wavebands of the analyzed spectra, thereby reducing the impact of unreasonable interval division [14]. Furthermore, sparse signals, including NIR, audio, and Electroencephalogram, usually have a cluster structure where neighboring coefficients are coupled to each other, and the nonzero coefficients appear in clusters [15]. Recent studies point out that using sparse Bayesian learning (SBL) to explore these inherent properties of sparse signals cannot only effectively reduce the computational cost but also help to improve the accuracy of spectral sparse reconstruction significantly [15], [16]. However, traditional SBL framework adopts variable-independent hierarchical models that do not consider variable coupling properties for sparse learning, which is prone to inaccurate representation in the presence of interference [16]. Therefore, it is necessary to explore the feasibility of combining spectral sparsity and wavelength coupling properties for peak waveband estimation.

In general, the spectral wavelengths within peak wavebands contain comprehensive information on the physicochemical properties of process objects. However, we usually focus on some significant attributes in the actual industrial processes modeling. Therefore, it should also be reasonably believed that only a few wavelengths within peak wavebands play a significant role in constructing the prediction model of the target attributes. With the help of regularization methods, such as least absolute shrinkage and selection operator (Lasso) and Elastic net, we can determine the useful variables by setting appropriate penalty terms. However, the L_1 method has no analytical solution for variable selection, and the L_2 method cannot reduce the spectral dimensionality through nonzero sparsification [17]. In recent years, an alternative learning method under the Bayesian inference (BI) framework has been extensively studied [18]. Compared with regularization methods, BI can introduce robustness to the analytical models due to their unique ability to marginalize out uncertainties associated with parameter estimates [19]. In the past few years, various probability distributions have been used for sparse coefficient learning, such as Gaussian [18], Laplacian [20], Student's-t [17], and hierarchical distributions [21]. Nevertheless, a limitation of these methods is that the prior shrinkage effect is uniform for all coefficients, which makes it challenging to promote high levels of sparsification while avoiding the overshrinking of significant variables [22]. Therefore, it is worthwhile to develop new wavelength selection methods for NIR analysis by incorporating appropriate sparsity priors.

To improve the accuracy and stability of the NIR model, the adaptive waveband estimation and robust characteristic wavelength selection (AWE-RCWS) method is proposed in this work, which can simultaneously improve the generalization and robustness by sparse representation and inference. The main contributions of this work are summarized as follows.

- 1) To achieve accurate peak waveband estimation of the analyzed spectra, the adaptive waveband estimation method based on the pattern-coupled sparse learning is proposed, in which the hierarchical sparse model is used to characterize spectral sparsity and wavelength coupling properties simultaneously.
- 2) To select the characteristic wavelengths from the estimated peak wavebands, the robust characteristic wavelength selection method based on the global-local shrinkage is developed, which promotes a high level of global sparsity while simultaneously retaining locally significant explanatory variables. In addition, the posterior distributions of the model parameters are derived.
- 3) Comparison experiments conducted on benchmark and real NIR datasets indicate that the AWE-RCWS method achieves better prediction accuracy and robustness than the existing methods.

The rest of this article is organized as follows. Section II briefly introduces the sparse learning for NIR analysis and modeling. The proposed method for waveband estimation and characteristic wavelength selection is given in Section III. Section IV describes the measured NIR datasets, comparative methods, and experimental settings. Comparison results are given in Section V. Finally, Section VI concludes this article.

II. PRELIMINARIES

In this section, we briefly introduce the SBL framework for spectra sparse representation. In addition, the BI has been discussed for NIR modeling and wavelength selection.

A. Sparse Representation of NIR

For the original spectral vector $\mathbf{x} \in \mathbb{R}^M$, it generally consists of the spectrum vector $\mathbf{s} \in \mathbb{R}^M$ and random noise vector $\mathbf{n} \in \mathbb{R}^M$, and they satisfy

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \quad (1)$$

where \mathbf{n} is the Gaussian noise with zero mean and variance σ^2 . \mathbf{s} can be approximated by a finite linear combination of learning dictionaries as follows [23]:

$$p(\mathbf{x} | \mathbf{A}, \mathbf{w}, \sigma^2) = \mathcal{N}(\mathbf{x} | \mathbf{Aw}, \sigma^2) \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{M \times M}$ is the Gaussian lineshape dictionary and $\mathbf{w} \in \mathbb{R}^M$ is the weight vector. In SBL framework, the weight \mathbf{w} is considered to be relatively sparse and assigned a Gaussian prior distribution

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{j=1}^M p(w_j | \alpha_j) \quad (3)$$

where $p(w_j | \alpha_j) = \mathcal{N}(w_j | 0, \alpha_j^{-1})$ and $\boldsymbol{\alpha} \triangleq \{\alpha_j\}_{j=1}^M$ are non-negative hyperparameters controlling the sparsity of the spectral signal \mathbf{s} . Clearly, for wavelength x_j located in the nonpeak regions, the corresponding coefficient w_j will tend to zero when α_j approaches infinity, and vice versa. The hyperparameter $\boldsymbol{\alpha}$ corresponding to the characteristic peaks can be learned by placing hyperpriors and maximizing their posterior probability.

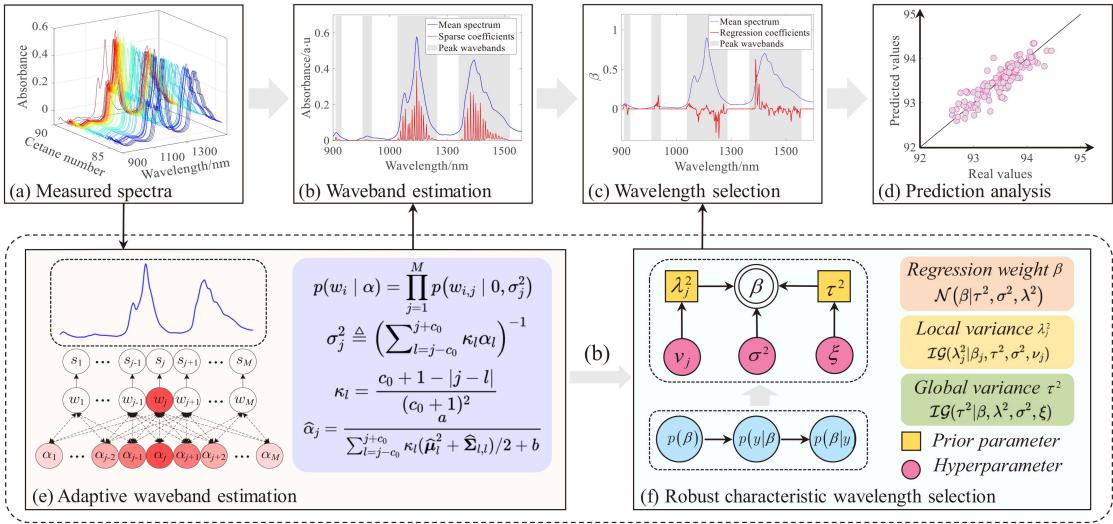


Fig. 1. Schematic diagram of the AWE-RCWS method.

B. BI of NIR Modeling

For $\mathbf{X}_s \in \mathbb{R}^{N \times K}$ determined by the SBL framework, where K denotes the number of candidate wavelengths, considering the linear regression model

$$\mathbf{y} = \mathbf{X}_s \boldsymbol{\beta} + \epsilon \quad (4)$$

where $\mathbf{y} = \{y_1, \dots, y_N\}^T$ is the response vector, $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_K\}^T$ is regression coefficients, ϵ is the error vector that follows a Gaussian distribution with zero mean and variance σ_ϵ^2 , and the superscript T denotes the transpose of a matrix or vector.

Using the ordinary least squares (OLS) method, the regression coefficients can be calculated as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \mathbf{y}. \quad (5)$$

The OLS method often has a low bias but large variance, and it is challenging to identify characteristic wavelengths with strong effects from high-dimensional input [24]. The BI has been applied for modeling and sparse variable selection to improve prediction accuracy, where the hyperparameters are inferred in a probability distribution form by introducing reasonable prior knowledge. For given dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, the likelihood distribution can be written as

$$p(\mathbf{y} | \mathbf{X}_s; \boldsymbol{\beta}) = \prod_{i=1}^N p(\mathbf{y}_i | \mathbf{x}_i^s; \boldsymbol{\beta}). \quad (6)$$

Based on the prior and likelihood beliefs of the coefficient distribution, the posterior distribution can be obtained as

$$p(\boldsymbol{\beta} | \mathbf{X}_s, \mathbf{y}) = \frac{1}{Z} p(\mathbf{y} | \mathbf{X}_s; \boldsymbol{\beta}) p(\boldsymbol{\beta}) \quad (7)$$

where Z is a normalization constant and independent of $\boldsymbol{\beta}$. Finally, given a new input $\hat{\mathbf{x}}$, the predictive distribution is computed as the expected value with the posterior distribution

$$p(\hat{\mathbf{y}} | \mathbf{X}_s, \mathbf{y}, \hat{\mathbf{x}}) = \int p(\hat{\mathbf{y}} | \hat{\mathbf{x}}, \boldsymbol{\beta}) p(\boldsymbol{\beta} | \mathbf{X}_s, \mathbf{y}) d\boldsymbol{\beta}. \quad (8)$$

III. PROPOSED AWE-RCWS METHOD

This section describes the proposed AWE-RCWS method in detail. As the flowchart shown in Fig. 1, the proposed AWE-RCWS method is a hybrid wavelength selection method consisting of the AWE method shown in Fig. 1(e) for peak waveband estimation and the RCWS method shown in Fig. 1(f) for characteristic wavelength selection. First, for the measured spectra shown in Fig. 1(a), the peak wavebands shown in Fig. 1(b) can be estimated by the proposed AWE method. Then, the characteristic wavelengths in Fig. 1(c) are determined by the proposed RCWS method. Finally, the prediction model shown in Fig. 1(d) is developed for the attributes of interest.

A. Adaptive Waveband Estimation

In the traditional SBL framework, as shown in Fig. 2(a), the sparse coefficients w_j and its hyperparameter α_j for each wavelength s_j are independent of each other, which is difficult to capture the inherent coupling properties of the measured spectra \mathbf{x} . For the spectral signal \mathbf{s} , due to the difficulty of calculation, the local coupling characteristics cannot be directly reflected in the sparse coefficient \mathbf{w} . Alternatively, as shown in Fig. 2(b), it can be represented by a coupled dependence of the variance hyperparameter α

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{j=1}^M p(w_j | \dots, \alpha_{j-1}, \alpha_j, \alpha_{j+1}, \dots) \quad (9)$$

and

$$\begin{aligned} & p(w_j | \dots, \alpha_{j-1}, \alpha_j, \alpha_{j+1}, \dots) \\ &= \mathcal{N}(w_j | 0, (\dots + \kappa_{j-1}\alpha_{j-1} + \kappa_j\alpha_j + \kappa_{j+1}\alpha_{j+1} + \dots)^{-1}) \end{aligned} \quad (10)$$

where $\{\kappa_{j-1}, \kappa_j, \kappa_{j+1}\}$ are the coupling correlation parameters representing the correlation between the coefficient w_j and its

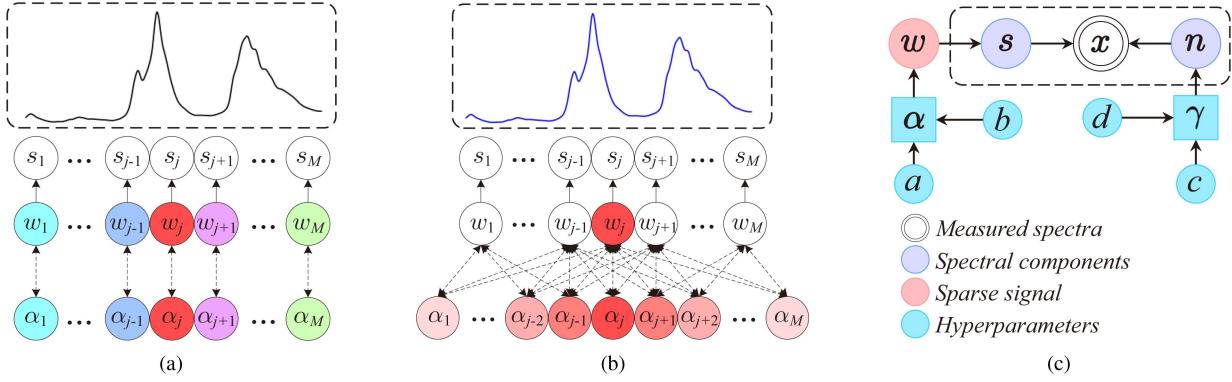


Fig. 2. Structures of (a) independent sparse model, (b) pattern-coupled sparse model, and (c) hierarchical model of spectral sparse representation, where \mathbf{x} is the measured spectra, s denotes the spectra signal, w represents the sparse coefficient, n is the noise, and $\{\alpha, \gamma, a, b, c, d\}$ are hyperparameters.

neighboring coefficients $\{w_{j-1}, w_{j+1}\}$, defined as follows:

$$\kappa_l = \frac{c_0 + 1 - |j - l|}{(c_0 + 1)^2} \quad (11)$$

where c_0 is the number of coupling hyperparameter.

For α and noise precision $\gamma = \sigma^{-2}$, they are generally modeled as follows:

$$p(\boldsymbol{\alpha}) = \prod_{j=1}^M \text{Gamma}(\alpha_j | a, b) \quad (12)$$

$$p(\gamma) = \text{Gamma}(\gamma | c, d) \quad (13)$$

where $\{a, b, c, d\}$ are hyperparameters.

Based on hierarchical model shown in Fig. 2(c) and BI workflow, the sparse coefficients w can been calculated as

$$\begin{aligned} p(w|x, \boldsymbol{\alpha}, \gamma) &= \frac{p(x|w, \gamma)p(w|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma)}{p(x|\boldsymbol{\alpha}, \gamma)} \\ &\propto p(x|w, \gamma)p(w|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma) \\ &= \mathcal{N}(w | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned} \quad (14)$$

and

$$\boldsymbol{\mu} = \gamma \boldsymbol{\Sigma} \mathbf{A}^T \mathbf{x} \quad (15)$$

$$\boldsymbol{\Sigma} = (\gamma \boldsymbol{\mathbf{A}}^T \boldsymbol{\mathbf{A}} + \boldsymbol{D})^{-1} \quad (16)$$

where $\boldsymbol{D} = \text{diag}\{\psi_1, \psi_2, \dots, \psi_M\}$ is a diagonal matrix with its i th diagonal element is calculated as $\psi_j = \sum_{l=j-c_0}^{j+c_0} \kappa_l \alpha_l$.

Given the estimated hyperparameters, the maximum a posterior (MAP) estimate of w is its posterior mean

$$\hat{w}_{\text{MAP}} = (\gamma^{-1} \boldsymbol{D} + \boldsymbol{\mathbf{A}}^T \boldsymbol{\mathbf{A}})^{-1} \boldsymbol{\mathbf{A}}^T \mathbf{x}. \quad (17)$$

The proposed AWE method can adaptively determine the peak and nonpeak wavebands according to the estimated sparse coefficients \hat{w} . In this section, the expectation–maximization (EM) is applied to maximize the posterior probability. The EM steps are briefly discussed as follows.

1) **E-Step:** Given the current estimates of the parameters $\{\alpha, \gamma\}$, we compute the expected value of the log-posterior.

For $\boldsymbol{\alpha}$ and γ , the posterior distribution can be expressed as

$$p(\boldsymbol{\alpha}, \gamma | \mathbf{x}, \mathbf{w}) \propto p(\mathbf{x}|\mathbf{w}, \gamma)p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma) \quad (18)$$

and it can obtained by EM as follows:

$$\begin{aligned} Q(\boldsymbol{\alpha}, \gamma | \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) &= E_{\mathbf{w}|\mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}} [\log p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})] \\ &\quad + E_{\mathbf{w}|\mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}} [\log p(\mathbf{x}|\mathbf{w}, \gamma)p(\gamma)]. \end{aligned} \quad (19)$$

For hyperparameters $\boldsymbol{\alpha}$

$$\begin{aligned} E_{\mathbf{w}|\mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}} [\log p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})] \\ = \int p(\mathbf{w} | \mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) \log[p(\mathbf{w}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})] d\mathbf{w} \\ = \log p(\boldsymbol{\alpha}) + \int p(\mathbf{w} | \mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) \log p(\mathbf{w}|\boldsymbol{\alpha}) d\mathbf{w}. \end{aligned} \quad (20)$$

Substituting (10) into (20) and ignoring the term independent of $\boldsymbol{\alpha}$, the Q -function can be reformulated as

$$\begin{aligned} Q(\boldsymbol{\alpha} | \boldsymbol{\alpha}^{(t)}) &= \log p(\boldsymbol{\alpha}) + \frac{1}{2} \sum_{j=1}^M \left(\log \psi_j \right. \\ &\quad \left. - \psi_j \int p(\mathbf{w} | \mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) w_j^2 d\mathbf{w} \right). \end{aligned} \quad (21)$$

As shown in (14), the posterior $p(\mathbf{w} | \mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)})$ obeys the multivariate Gaussian distribution with its mean and covariance matrix given by (15) and (16), respectively. Therefore, we have

$$\begin{aligned} \int p(\mathbf{w} | \mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) w_j^2 d\mathbf{w} &= E_{\mathbf{w}|\mathbf{x}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}} [w_j^2] \\ &= \hat{\mu}_j^2 + \hat{\Sigma}_{j,j} \end{aligned} \quad (22)$$

where $\hat{\mu}_j^2$ denotes the j th entry of $\hat{\mu}$, and $\hat{\Sigma}_{j,j}$ denotes the j th diagonal element of the covariance matrix. Based on the specified prior (12), the Q -function can rewritten as

$$\begin{aligned} Q(\boldsymbol{\alpha} | \boldsymbol{\alpha}^{(t)}) &= \sum_{j=1}^M \left(a \log \alpha_j - b \alpha_j + \frac{1}{2} \log \psi_j \right. \\ &\quad \left. - \frac{1}{2} \psi_j (\hat{\mu}_j^2 + \hat{\Sigma}_{j,j}) \right). \end{aligned} \quad (23)$$

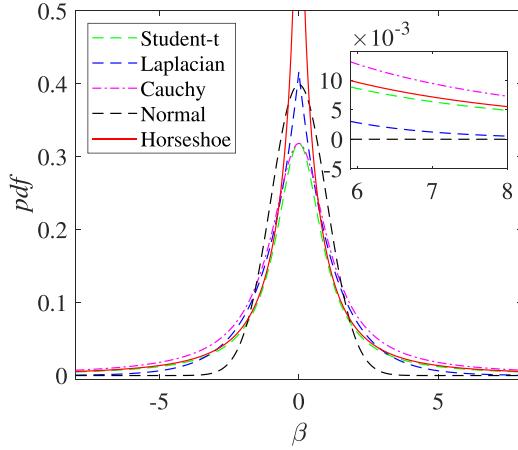


Fig. 3. Densities of different prior distributions.

For the parameter γ

$$\begin{aligned} E_{\mathbf{w}|\mathbf{x}, \alpha^{(t)}, \gamma^{(t)}} [\log p(\mathbf{x}|\mathbf{w}, \gamma)p(\gamma)] \\ = \frac{N}{2} \log \gamma - \frac{\gamma}{2} E_{\mathbf{w}|\mathbf{x}, \alpha^{(t)}, \gamma^{(t)}} [\|\mathbf{x} - \mathbf{A}\mathbf{w}\|_2^2] \\ - c \log \gamma - d\gamma. \end{aligned} \quad (24)$$

2) **M-Step:** As shown in (19), the estimation of α and γ are decoupled into the two independent terms. For hyperparameter α , due to the hyperparameters are coupled with each other, it is difficult to obtain an analytical solution with respect to α . Alternatively, a simple, analytical solution proposed in [14] is adopted for α_j optimization, and we have

$$\hat{\alpha}_j = \frac{a}{0.5\varphi_j + b}, \quad j = 1, \dots, M \quad (25)$$

where $\varphi_j \triangleq \sum_{l=j-c_0}^{j+c_0} \kappa_l (\hat{\mu}_l^2 + \hat{\Sigma}_{l,l})$.

For noise precision γ , computing the first derivative of (24) with respect to γ and setting it equal to zero, we get

$$\frac{1}{\gamma^{(t+1)}} = \frac{\|\mathbf{x} - \mathbf{A}\hat{\mu}\|_2^2 + (\gamma^{(t)})^{-1} \sum \rho_j + 2d}{N + 2c} \quad (26)$$

where $\rho_j \triangleq 1 - \psi_j \hat{\Sigma}_{j,j}$, $j = 1, \dots, M$.

For the high-dimensional spectra, the position information of the spectral peak can still be preserved by wavelength equidistant downsampling. Therefore, the peak waveband estimation can be achieved through the equidistant downsampling of spectra $\mathbf{x}[L]$ and the sparse coefficient upsampling $\hat{\mathbf{w}}[1/L]$ while avoiding high-dimensional matrix inversion, where $\mathbf{x}[L]$ denotes taking every L th wavelength and $\hat{\mathbf{w}}[1/L]$ denotes inserting $L-1$ variables between $\hat{\mathbf{w}}_j$ and $\hat{\mathbf{w}}_{j+1}$. The proposed AWE method is summarized in Algorithm 1.

Remark 1: For the proposed AWE method, the analytical method described Algorithm 1 is adopted for hyperparameter optimization, and related studies have verified its effectiveness in sparse learning [15], [16].

Algorithm 1: Proposed AWE Algorithm.

Input: NIR dataset \mathbf{x} , iteration times T , tolerance value ζ , sampling interval L ;

Output: Selected candidate wavebands \mathbf{x}^s ;

- 1: Spectra downsampling $\mathbf{x}[L]$, initialize $\alpha^{(0)}$, $\gamma^{(0)}$;
- 2: **while** $t \leq T$ and $e_w > \zeta$, **do**:
- 3: Calculate the posterior $p(\mathbf{w}^{(t)} | \mathbf{y}, \alpha^{(t)}, \gamma^{(t)})$;
- 4: $\hat{\mu}^{(t+1)} = \gamma^{(t)} \mathbf{\Sigma}^{(t)} \mathbf{A}^T \mathbf{x}$;
- 5: $\hat{\Sigma}^{(t+1)} = (\gamma^{(t)} \mathbf{A}^T \mathbf{A} + \mathbf{D})^{-1}$;
- 6: $\hat{\mathbf{w}}_{MAP}^{(t+1)} = (\mathbf{A}^T \mathbf{A} + (\gamma^{(t)})^{-1} \mathbf{D})^{-1} \mathbf{A}^T \mathbf{x}$;
- 7: Compute new estimation for α :
- 8: $\hat{\varphi}_j^{(t+1)} = \sum_{l=j-c_0}^{j+c_0} \kappa_l (\hat{\mu}_l^{(t+1)} + \hat{\Sigma}_{l,l}^{(t+1)})$;
- 9: $\hat{\alpha}_j^{(t+1)} = a / (0.5\hat{\varphi}_j^{(t+1)} + b)$;
- 10: Update the noise precision γ :
- 11: $\rho_j = 1 - \sum_{l=j-c_0}^{j+c_0} \kappa_l \hat{\alpha}_l^{(t+1)} \hat{\Sigma}_{l,l}^{(t+1)}$;
- 12: $\chi^{(t+1)} = \|\mathbf{y} - \mathbf{A}\hat{\mu}\|_2^2 + (\gamma^{(t)})^{-1} \sum \rho_j + 2d$;
- 13: $\gamma^{(t+1)} = (N + 2c) / \chi^{(t+1)}$;
- 14: $t = t + 1$ and $e_w = \|\hat{\mathbf{w}}^{(t+1)} - \hat{\mathbf{w}}^{(t)}\|_2$;
- 15: **end while**
- 16: Sparse coefficient upsampling $\hat{\mathbf{w}}[1/L]$;
- 17: Determine candidate wavebands $\{\mathbf{x}_j^s | \hat{\mathbf{w}}_j > 0\}_{j=1}^M$.

B. Robust Characteristic Wavelength Selection

For the dataset $\mathbf{X}_s \in \mathbb{R}^{N \times K}$ selected by the proposed AWE method, where K denotes the number of candidate wavelengths for all selected wavebands, the linear regression model is

$$p(\mathbf{y} | \mathbf{X}_s; \boldsymbol{\beta}, \sigma_\epsilon^2) = \mathcal{N}(\mathbf{X}_s \boldsymbol{\beta}, \sigma_\epsilon^2). \quad (27)$$

In the BI framework, it is generally assumed that the regression coefficients follow a sparse distribution. As shown in Fig. 3, however, a limitation of these distributions is that uniform shrinkage is performed to all the regression coefficients, which makes it difficult to promote a high level of sparsity while simultaneously avoiding overshrinkage of large coefficients. The horseshoe prior presented in [25] circumvents this problem by introducing $\mathcal{N}(0, \lambda_j^2 \tau^2 \sigma_\epsilon^2)$ as conditional prior for regression vector $\boldsymbol{\beta}$. Unlike traditional priors, the global-local shrinkage property of the horseshoe prior makes it particularly useful as a shrinkage prior for sparse regression. On the one hand, its flat, Cauchy-like tails allow the strong signals to remain large, i.e., unshrink [26]. On the other hand, its infinitely tall spike at the origin provides severe shrinkage for the zero elements of $\boldsymbol{\beta}$ [25]. In this work, the horseshoe distribution is chosen as the prior distribution for regression coefficients $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_K\}^T$

$$p(\beta_j | \lambda_j^2 \tau^2 \sigma_\epsilon^2) = \mathcal{N}(0, \lambda_j^2 \tau^2 \sigma_\epsilon^2) \quad (28)$$

where the variance σ_ϵ^2 follows Jeffreys' prior, i.e., $p(\sigma_\epsilon^2) = J(\sigma_\epsilon^2)$, $\tau^2 \sim \mathcal{C}^+(0, 1)$ is the global parameter that pulls all the weights globally toward zeros, while the local scales $\lambda_j^2 \sim \mathcal{C}^+(0, 1)$ allow some of the weights with big values to escape shrinkage. In order to obtain the analytical solution of regression coefficient $\boldsymbol{\beta}$, the auxiliary variables ν_j and ξ is introduced and

the revised hierarchy can be expressed as [27]

$$p(\lambda_j^2 | \nu_j) = \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\nu_j}\right) \quad (29)$$

$$p(\tau^2 | \xi) = \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\xi}\right) \quad (30)$$

where $p(\nu_1, \dots, \nu_K, \xi) = \mathcal{IG}(1/2, 1)$ and $\mathcal{IG}(\cdot)$ denotes the Inverse-Gamma distribution. The parameter posterior distribution can be obtained according to Bayes' theorem.

For the regression coefficient β , the conditional posterior distribution is

$$\begin{aligned} p(\beta | \mathbf{y}, \mathbf{X}_s; \sigma_\epsilon^2) &= \frac{p(\mathbf{y} | \mathbf{X}_s; \beta, \sigma_\epsilon^2) p(\beta | \lambda^2, \tau^2, \sigma_\epsilon^2)}{p(\mathbf{y} | \mathbf{X}_s; \sigma_\epsilon^2)} \\ &\propto p(\mathbf{y} | \mathbf{X}_s; \beta, \sigma_\epsilon^2) p(\beta | \lambda^2, \tau^2, \sigma_\epsilon^2) \\ &= \mathcal{N}(\mathbf{H}^{-1} \mathbf{X}_s^\top \mathbf{y}, \sigma_\epsilon^2 \mathbf{H}^{-1}) \end{aligned} \quad (31)$$

and

$$\mathbf{H} = \mathbf{X}_s^\top \mathbf{X}_s + \Lambda_*^{-1} \quad (32)$$

where $\Lambda_* = \tau^2 \Lambda$ and $\Lambda = \text{diag}\{\lambda_1^2, \lambda_2^2, \dots, \lambda_K^2\}$.

For the noise variance σ_ϵ^2

$$\begin{aligned} p(\sigma_\epsilon^2 | \mathbf{y}, \mathbf{X}_s; \beta) &= \frac{p(\mathbf{y} | \mathbf{X}_s; \beta, \sigma_\epsilon^2) p(\beta | \lambda^2, \tau^2, \sigma_\epsilon^2) p(\sigma_\epsilon^2)}{p(\mathbf{y} | \mathbf{X}_s; \beta)} \\ &\propto (\sigma_\epsilon^2)^{-(\frac{N+K}{2}+1)} \exp\left(-\frac{\|\mathbf{y} - \mathbf{X}_s \beta\|^2 + \beta^\top \Lambda_*^{-1} \beta}{2\sigma_\epsilon^2}\right) \\ &= \mathcal{IG}\left(\frac{N+K}{2}, \frac{\|\mathbf{y} - \mathbf{X}_s \beta\|^2 + \beta^\top \Lambda_*^{-1} \beta}{2}\right). \end{aligned} \quad (33)$$

Following the fast Bayesian approach in [28], the conditional posterior density of σ_ϵ^2 is

$$p(\sigma_\epsilon^2 | \mathbf{y}, \mathbf{X}_s; \beta) = \mathcal{IG}\left(\frac{N}{2}, \frac{\mathbf{y}^\top (\mathbf{I}_N - \mathbf{X}_s \Lambda_*^{-1} \mathbf{X}_s^\top) \mathbf{y}}{2}\right) \quad (34)$$

where $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix.

For the local hypervariance λ_j^2 , it can be calculated as

$$\begin{aligned} p(\lambda_j^2 | \beta_j) &= \frac{p(\beta_j | \lambda_j^2, \tau^2, \sigma_\epsilon^2) p(\lambda_j^2)}{p(\beta_j)} \\ &\propto (\lambda_j^2)^{-(1+1)} \exp\left(-\frac{1}{\lambda_j^2} \left(\frac{\beta_j^2}{2\tau^2 \sigma_\epsilon^2} + \frac{1}{\nu_j}\right)\right) \\ &= \mathcal{IG}\left(1, \frac{\beta_j^2}{2\tau^2 \sigma_\epsilon^2} + \frac{1}{\nu_j}\right), j = 1, 2, \dots, K. \end{aligned} \quad (35)$$

Similarly, the conditional posterior distribution for the global hypervariance τ^2 is

$$\begin{aligned} p(\tau^2 | \beta) &= \frac{p(\beta | \lambda^2, \tau^2, \sigma_\epsilon^2) p(\tau^2)}{p(\beta)} \\ &\propto (\tau^2)^{-(1+1)} \exp\left(-\frac{1}{\tau^2} \left(\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^K \frac{\beta_j^2}{\lambda_j^2} + \frac{1}{\xi}\right)\right) \end{aligned}$$

Algorithm 2: Proposed RCWS Algorithm.

Input: Dataset $\{\mathbf{X}_s, \mathbf{y}\}$, sampling times N_s ;

Output: Regression coefficient β ;

```

1: Initialize  $\sigma_\epsilon^2, \tau^2, \xi, \mathbf{v}$  and set  $i = 0$ ;
2: while  $i \leq N_s$ , do:
3:   Generate sample of  $\beta$ :  $\beta = \mathcal{N}(\mathbf{H}^{-1} \mathbf{X}_s^\top \mathbf{y}, \sigma_\epsilon^2 \mathbf{H}^{-1})$ 
4:   Generate sample of  $\sigma_\epsilon^2$ :
5:    $\sigma_\epsilon^2 = \mathcal{IG}(N/2, \mathbf{y}^\top (\mathbf{I}_N - \mathbf{X}_s \Lambda_*^{-1} \mathbf{X}_s^\top) \mathbf{y}/2)$ ;
6:   Update  $\beta, \sigma_\epsilon^2, \mathbf{v}$ ;
7:   Generate sample of  $\Lambda$ :
8:    $\lambda_j^2 = \mathcal{IG}(1, 1/\nu_j + \beta_j^2/2\tau^2 \sigma_\epsilon^2)$ ;
9:    $\Lambda = \text{diag}(\lambda_1^2, \lambda_2^2, \dots, \lambda_M^2)$ ;
10:  Generate sample of  $\mathbf{v}$ :  $v_j = \mathcal{IG}(1, 1 + 1/\lambda_j^2)$ ;
11:  Update  $\tau^2$  and generate sample:
12:   $\xi = \mathcal{IG}(1, 1 + 1/\tau^2)$ ;
13:  if  $i > b$  do:
14:    Store the learning estimate of  $\beta$ ;
15:  end if
16: end while

```

$$= \mathcal{IG}\left(1, \frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^K \frac{\beta_j^2}{\lambda_j^2} + \frac{1}{\xi}\right). \quad (36)$$

For the auxiliary variables ν_j and ξ , the conditional posterior distributions are

$$\begin{aligned} p(\nu_j | \lambda_j^2) &\propto p(\lambda_j^2 | \nu_j) p(\nu_j) \\ &\propto \nu_j^{-(1+1)} \exp\left(-\frac{1}{\nu_j} \left(1 + \frac{1}{\lambda_j^2}\right)\right) \\ &= \mathcal{IG}\left(1, 1 + \frac{1}{\lambda_j^2}\right), j = 1, 2, \dots, K \end{aligned} \quad (37)$$

and

$$\begin{aligned} p(\xi | \tau^2) &\propto p(\tau^2 | \xi) p(\xi) \\ &\propto \xi^{-(1+1)} \exp\left(-\frac{1}{\xi} \left(1 + \frac{1}{\tau^2}\right)\right) \\ &= \mathcal{IG}\left(1, 1 + \frac{1}{\tau^2}\right). \end{aligned} \quad (38)$$

The abovementioned conditional posterior distribution enables Gibbs sampling to obtain the target parameters. The proposed RCWS method is summarized in Algorithm 2.

Remark 2: Markov chain Monte Carlo method can be used for approximate statistical inference when complex probabilistic models make exact inference challenging. Gibbs sampling is often chosen because it can simplify the sampling process and avoid complex multidimensional integration [29].

IV. EXPERIMENTS SETTING

In the following section, we give details about datasets, including the measured NIR and benchmark datasets, comparative methods, evaluation metrics, and the parameter settings.

A. Experimental Datasets

1) *Shootout Dataset*: The shootout benchmark dataset consists of 655 pharmaceutical tablets and the corresponding active pharmaceutical ingredient [30]. The spectra were scanned from two spectrometers (Foss and Multitab Spectroscanners) and recorded from 600 to 1898 nm.¹ The original data consisted of two independent sets, and the first dataset that includes NIR between 600 and 1800 nm was used to predict the active pharmaceutical ingredient in this study.

2) *Diesel Fuel Dataset*: The diesel fuel dataset is consisted of 784 samples and six properties: bp50 (the boiling point at 50% recovery), CN (cetane number), d4052 (density), freeze (freezing temperature), total (total aromatics), and visc (viscosity) [31]. The NIR were collected from 900 to 1550 nm with 2 nm intervals, and only the CN with 381 samples is used in this article.²

3) *Tea Dataset*: The tea dataset containing 163 samples NIR and sucrose content was collected from a tea processing company in China. NIR containing overlapping peaks were scanned from 1000 to 2500 nm on an IRTtracer-100 spectrometer (SHIMADZU, Japan) and used to compare the wavelength selection capability of different methods in this work.

B. Comparative Methods

1) *PLS*: Partial least squares (PLS) is a classical linear regression method for NIR modeling, in which spectra and reference values are utilized to determine characteristic variables.

2) *Lasso*: Lasso is a classical WPS method, in which the L_1 penalty term is used to achieve variable selection [17].

3) *Elastic Net*: Elastic net is a classical WPS method, where the regularization parameter λ is used for regression coefficient penalty and strongly correlated wavelengths tend to enter or exit regression modeling process together [17].

4) *Backward interval-PLS (BiPLS)*: BiPLS is a classical WIS method for spectral analysis, in which the interval-PLS (iPLS) method is used for the spectra intervals division and then followed by backward elimination. The number of divided intervals (N_i) and the number of final selected intervals (N_{ki}) are main parameters of the BiPLS method [32].

5) *iPCPA*: iPCPA is a hybrid wavelength selection method in which iPLS is used as WIS for rough interval selection, and the permutation combination population analysis (PCPA) method is applied as WPS for fine wavelength selection [10]. The N_i and N_{ki} are the main parameters of the iPCPA method.

6) *iPLS-GA*: iPLS-GA is a hybrid method that combines iPLS with genetic algorithm (GA), which selects adjacent wavelengths during variable selection [14]. Similarly, the N_i and N_{ki} are the main parameters.

7) *AWE-PLS*: AWE-PLS combines the proposed AWE method with PLS, and its purpose is to verify the effectiveness of the proposed AWE method in peak waveband estimation by comparing the performance differences between two methods.

8) *SBL-RCWS*: The SBL-RCWS method is used to compare the difference in peak waveband estimation between the traditional SBL method and the proposed AWE method, the sparse representation using automatic relevance determination can refer to [18].

All experimental datasets are original NIR spectra, and baseline correction preprocessing is performed before spectral analysis. In addition, NIR datasets are randomly split into a training set (70%) and a prediction set (30%), and the sample index of different comparison methods are kept consistent within a single trial for fair comparison.

C. Model Performance Evaluation Metrics

The coefficient of determination (R^2) and root mean square error (RMSE), which, respectively, reflect the fitness and accuracy of the prediction model, are used to evaluate the model performance, defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (39)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (40)$$

where y_i is the real value, \hat{y}_i is the predicted value, \bar{y} is the mean of the real value, and n is the number of samples.

D. Parameter Setting

The penalty parameter λ of Lasso is determined by ten-fold cross-validation. For the AWE-RCWS method, the hyperparameters $\{a, b, c, d\}$ are initialized as 10^{-6} , the sampling interval L and coupling hyperparameter c_0 are set as 1, the b_i is set as $0.8N_s$ and only the number of Gibbs sampling (N_s) needs to be determined. For the SBL-RCWS method, the hyperparameters of sparse coefficient and noise variances are initialized as 10^{-4} and the parameters of RCWS are consistent with the AWE-RCWS method. Taking the real-world data as an example, the determination process of the main parameters of comparative methods is introduced in detail.

1) *Setting of Latent Variables LVs*: To determine the optimal LVs of the PLS method, the RMSE of cross-validation (RM-SECV) of different components performed on the tea dataset is shown in Fig. 4(a). It can be seen that as the LVs increases, the RMSECV decreases rapidly and tends to stable when the LVs is greater than 11. Thus, the LVs is set as 11 in the tea dataset.

2) *Setting of Regularization Parameter α* : To test the effect of α of Elastic net, we compare R^2 of the prediction set (R_P^2) under different α values. Fig. 4(b) shows the averaged R^2 over 50 independent trials. It can be seen that the R_P^2 increases when $\alpha \in [0.1, 0.6]$, and then begins to decrease as α continues to increase. Accordingly, the regularization parameter α is set to 0.6 for the tea dataset.

3) *Setting of Interval Number N_i* : For the BiPLS method, the R_P^2 of different N_i is shown in Fig. 4(c). It can be seen that for the tea dataset with overlapping peaks, R_P^2 keeps increasing when $N_i \in [30, 75]$ and R_P^2 decreases slightly when $N_i \geq 75$. Thus,

¹[Online]. Shootout: <http://www.idrcchambersburg.org/shootout2002.html>

²[Online]. Diesel fuel: <http://www.eigenvector.com/data/SWRI/index.html>

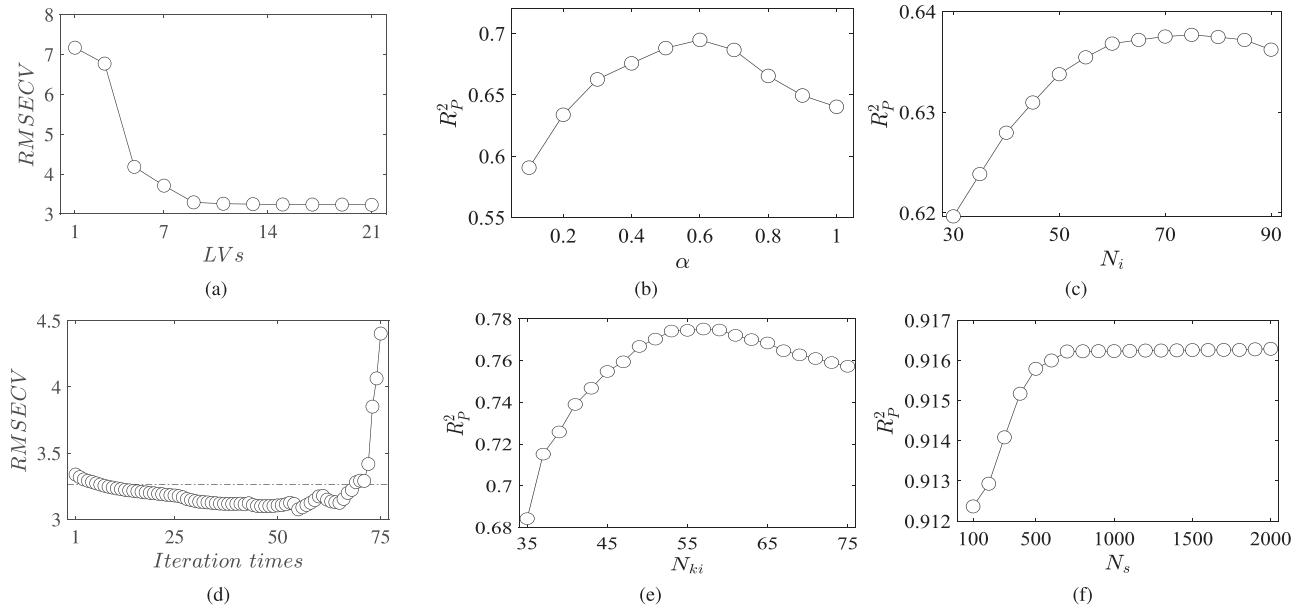


Fig. 4. Empirical parameter settings for the real-world tea dataset. (a) Latent variables LVs of the PLS method. (b) Regularization parameter α of the Elastic net method. (c) Number of divided interval N_i . (d) Cross-validation interval selection process of the BiPLS method (dashdotted line denotes mean RMSECV). (e) Number of kept intervals N_{ki} of the iPCPA method. (f) Sampling times N_s of the proposed AWE-RCWS method.

TABLE I
PARAMETERS SETTING FOR EXPERIMENTAL DATASETS

		Shootout	Diesel fuel	Tea
PLS	LVs	8	8	11
Lasso	λ	0.42	0.15	0.21
Elastic net	α	0.73	0.41	0.64
BiPLS	N_i, N_{ki}	30, 23	32, 27	75, 63
iPCPA	N_i, N_{ki}	30, 22	32, 24	75, 56
iPLS-GA	N_i, N_{ki}	30, 24	32, 25	75.59
AWE-PLS	LVs	8	7	13
AWE-RCWS	N_s	600	600	700

N_i is set as 75 in the tea dataset, and the cross-validation iteration process for interval selection is shown in Fig. 4(d). Finally, 63 intervals below mean RMSECV were used for modeling analysis.

4) *Setting of Kept Intervals N_{ki} :* Fig. 4(e) shows the effect of N_{ki} of the iPCPA method on the prediction performance. It can be seen that small N_{ki} easily produces an underfitting model, while an overly large N_{ki} will lead to the overfitting problem. Thus, the N_{ki} is set to 56 in this work.

5) *Setting of Sampling Times N_s :* The performance difference of different N_s is shown in Fig. 4(f). It can be seen that the model performance is insensitive to the setting of N_s when $N_s > 700$. To reduce computing cost, the N_s is set to 700.

Similarly, we have investigated main parameters of comparative methods on shootout and diesel fuel datasets and summarized in Table I.

V. EXPERIMENTAL RESULTS

In this section, experiment results are presented to illustrate the performance of the proposed method. All experiments were written in MATLAB (Version 2021b, MathWorks, Inc.) and conducted on a personal computer with 2.10 GHz Intel Core

TABLE II
COMPARISON RESULTS OF SHOOTOUT DATASET

	R_f^2	RMSET	R_p^2	RMSEP	Time(s)
PLS	0.9127	4.8921	0.9166	4.9863	0.05
Lasso	0.8657	6.4221	0.8645	6.5034	83.31
Elastic net	0.8837	5.7981	0.8782	6.0067	118.68
BiPLS	0.9231	4.5550	0.9187	4.8577	1140.03
iPCPA	0.9354	4.3093	0.9361	4.2246	661.58
iPLS-GA	0.9147	4.9654	0.9085	5.0817	282.39
AWE-PLS	0.9156	4.9349	0.9147	4.8957	4.91
SBL-RCWS	0.9312	4.3124	0.9305	4.3855	184.16
AWE-RCWS	0.9497	3.8073	0.9453	3.8453	192.92

i7-9700 CPU, 16 GB RAM, and a Microsoft Windows 10 operating system.

To illustrate the effectiveness of the proposed AWE method in peak waveband estimation, the original spectra, the mean spectrum, and the resulting sparse coefficients of three datasets are shown in Fig. 5. It can be clearly seen that the spectral peaks are sparse compared with hundreds of wavelengths. As shown in Fig. 5, for the mean spectrum represented by the red line, the sparse coefficients obtained by the proposed AWE method can accurately characterize the sparsity of the analyzed spectra, in which the sparse coefficients are clustered in the peak wavebands, while the coefficients of the nonpeak intervals are zeros.

A. Shootout Dataset

Comparative results of 100 randomized trials are summarized in Table II, where the bold values represent the optimal values. Compared with the PLS method with the shortest time, Lasso and Elastic net have poorer results on the shootout dataset. The BiPLS method has a similar performance to the PLS method, but it is the most time-consuming due to complex validation process.

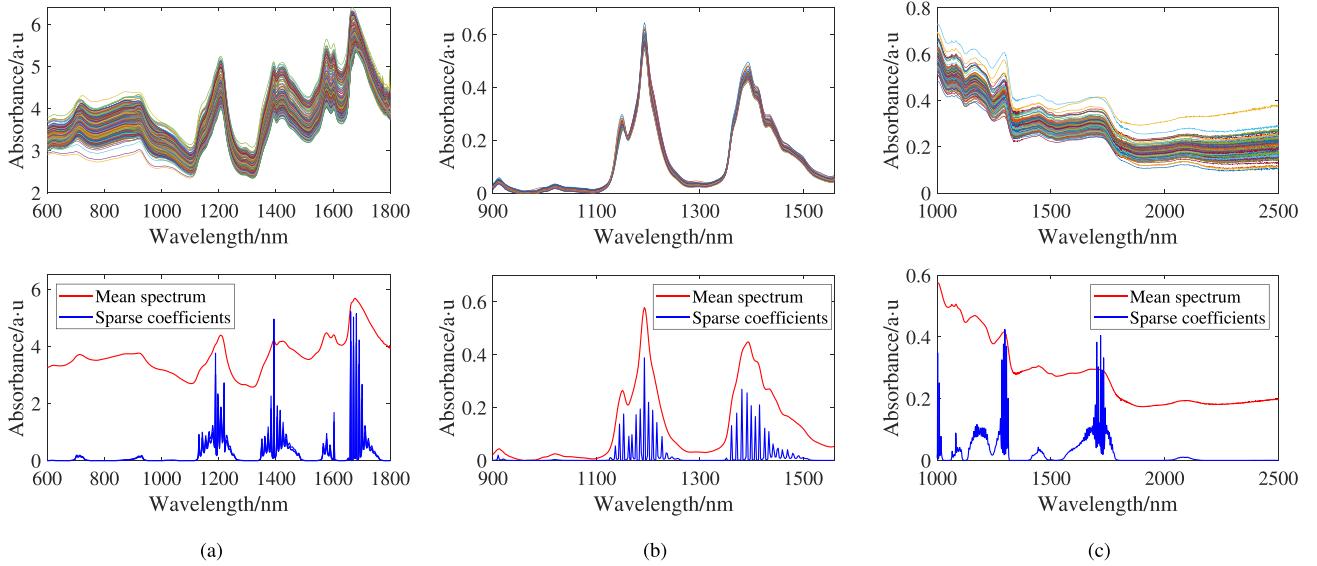


Fig. 5. Original spectra, mean spectrum (red line), and sparse coefficients (blue line) obtained by the AWE-RCWS method for: (a) shootout dataset, (b) diesel fuel dataset, and (c) tea dataset.

The performance improvement of iPCPA shows the efficiency of variable combination selection. The iPLS-GA method using adjacent variables selection performs worse than PLS. Although the AWE-PLS method has a similar prediction performance to the PLS method, the former only uses the estimated peak wavebands that consist of 345 wavelengths for modeling, while the latter is based on the entire spectrum that includes 600 wavelengths. The comparison results verify that the useful spectral information mainly lies in the peak wavebands and the effectiveness of the proposed AWE method in peak waveband estimation. The prediction accuracy comparison further verified the effectiveness of the proposed RCWS method in characteristic wavelength selection, where the proposed AWE-RCWS method obtained the best performance. Compared with the PLS method, the R_P^2 and RMSEP (R^2 and RMSE of prediction set) of the proposed AWE-RCWS method have improved and reduced 3.13% and 22.88%, respectively. Although the SBL-RCWS method performs well in the shootout dataset, but it is still inferior to the proposed method.

To analyze the rationality of the selected wavelengths, the regression coefficients of different methods are shown in Fig. 6, where the gray parts represent the intervals determined by different methods. It can be seen from Fig. 6(a) that, except for iPLS-GA, the regression coefficients estimated by the other six methods are all concentrated at 1200 nm, which is the first and second overtones of the C-H segment of methyl or methylene [33]. Nevertheless, many irrelevant variables are involved in BiPLS, while iPCPA and Lasso determine nine and six wavelengths, respectively. Although the wavebands centered at 1200 nm have been selected by the Elastic net and iPLS-GA methods, other intervals containing uninformative variables are also used for NIR modeling. Compared with AWE-RCWS, the SBL-RCWS method has selected more candidate wavebands, among which irrelevant variables may cause interference to subsequent analysis. Unlike other methods, only the peak

TABLE III
COMPARISON RESULTS OF DIESEL FUEL DATASET

	R_T^2	RMSET	R_P^2	RMSEP	Time(s)
PLS	0.9402	0.3453	0.9380	0.3587	0.06
Lasso	0.9656	0.2537	0.9512	0.3652	9.95
Elastic net	0.9797	0.2151	0.9705	0.2420	12.45
BiPLS	0.9313	0.3701	0.9307	0.3654	52.34
iPCPA	0.9742	0.2467	0.9730	0.2281	47.48
iPLS-GA	0.9122	0.4165	0.9038	0.4498	26.81
AWE-PLS	0.9391	0.3696	0.9368	0.3563	2.13
SBL-RCWS	0.9667	0.2631	0.9628	0.2765	32.11
AWE-RCWS	0.9796	0.2115	0.9772	0.2130	31.03

wavebands are selected in the proposed AWE-RCWS method, and the wavelengths at 1200 nm corresponding to the target attribute have large amplitudes, while other regression weights are zero or close to zero.

To more directly and deeply illustrate the effectiveness and robustness of the selected characteristic wavelengths, the mean and standard deviation of R_P^2 and RMSEP (under 100 random trials) of the model by gradually increasing characteristic wavelengths ranked according to their importance are shown in Fig. 7(a). It can be seen that the proposed AWE-RCWS method can converge quickly with a small standard deviation, which indicates that the proposed method can accurately determine the characteristic wavelengths containing the spectral characteristic information of the attribute of interest.

B. Diesel Fuel Dataset

As the comparison results summarized in Table III, it can be observed the proposed AWE-RCWS outperforms other methods. The PLS and AWE-PLS methods have similar prediction performance. Compared with AWE-PLS, the BiPLS method based on equal-length interval division performs worse and is time-consuming. In comparison, the iPLS-GA method performs worst because of the interference of uninformative variables in

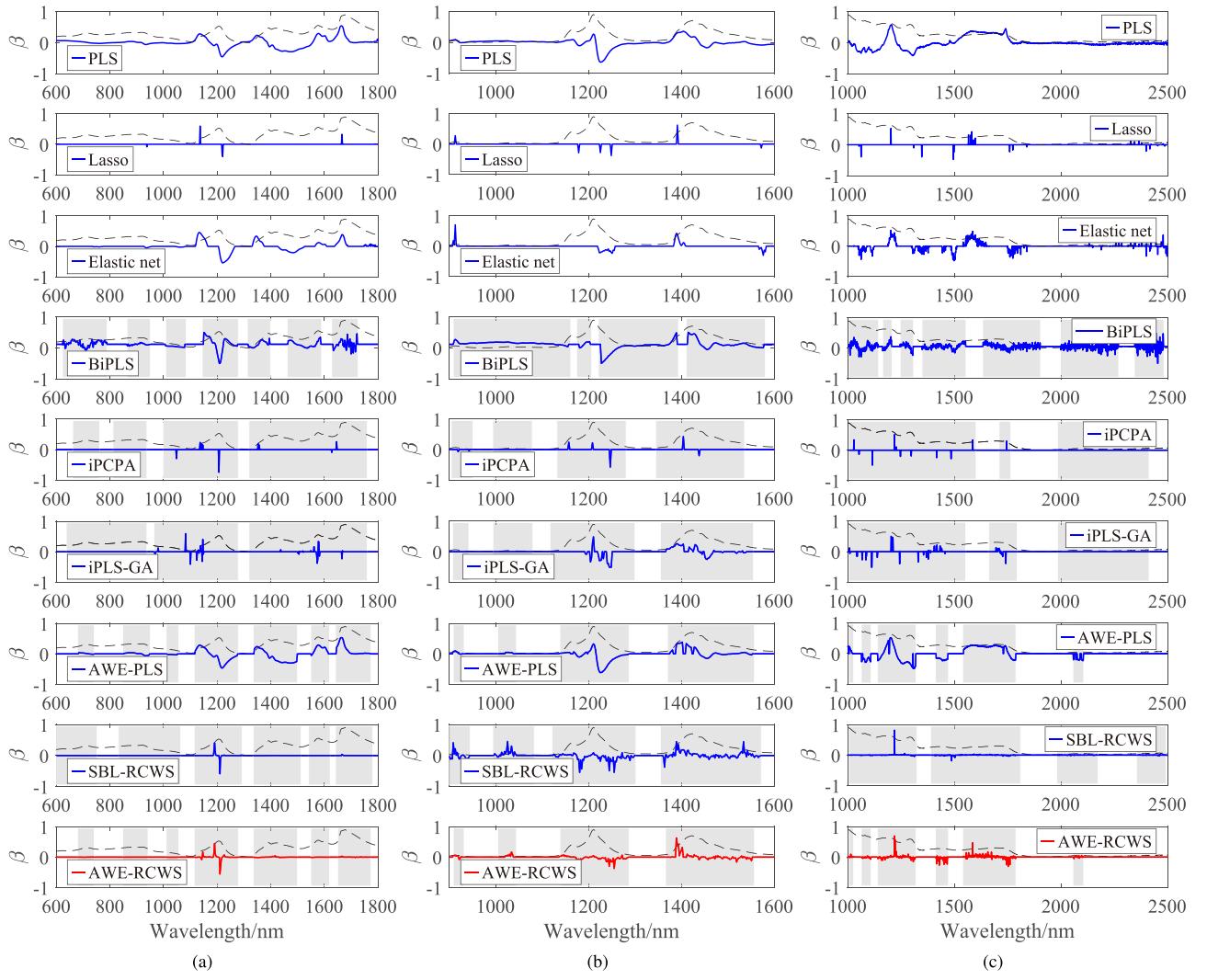


Fig. 6. Mean spectrum (dotted line), regression coefficients (solid line), and rough selected intervals (gray parts) of different methods for: (a) shootout dataset, (b) diesel fuel dataset, and (c) tea dataset.

the divided interval. Compared Lasso and Elastic net, the iPCPA and SBL-RCWS have better performance in cetane number prediction.

The estimated regression coefficients of comparative methods are shown in Fig. 6(b), where all regression coefficients are concentrated in regions around 910 nm, 1200 nm, and 1450 nm. Specifically, the iPLS-GA method fails to select the informative wavelengths between 900 and 1150 nm, which are the NIR characteristic bonds of aliphatic hydrocarbon, olefin, and alkyne [13]. In contrast, the uninformative wavelengths around 1580 nm are selected by Lasso, Elastic net, and BiPLS methods, which may result in the poor performance. Compared with SBL-RCWS, which contains more candidate wavelength intervals, while the characteristic wavelengths selected by the AWE-RCWS method lie in prominent characteristic bands for key hydrocarbon constituents, which makes it reasonable to use these variables to predict fuel properties well.

The mean and standard deviation of the R_P^2 and RMSEP with ranked wavelengths are shown in Fig. 7(b). Compared

with the SBL-RCWS method performs with a larger standard deviation, the AWE-RCWS method performs better with a relatively smaller standard deviation, demonstrating its effectiveness and robustness in waveband estimation and characteristic wavelength selection.

C. Tea Dataset

Table IV summarizes the results of different methods tested on the tea dataset. Among comparative methods, the BiPLS method performs worst. Lasso, Elastic net, iPCPA, and iPLS-GA methods have better results than PLS method, but the time consumption is larger than that of PLS. The SBL-RCWS method has better performance than other methods except the AWE-RCWS method. Compared with the PLS method, the R_P^2 and RMSEP of the AWE-RCWS method have increased and decreased by 63.03% and 56.16%, respectively. The significant improvement in prediction performance verifies the effectiveness of the

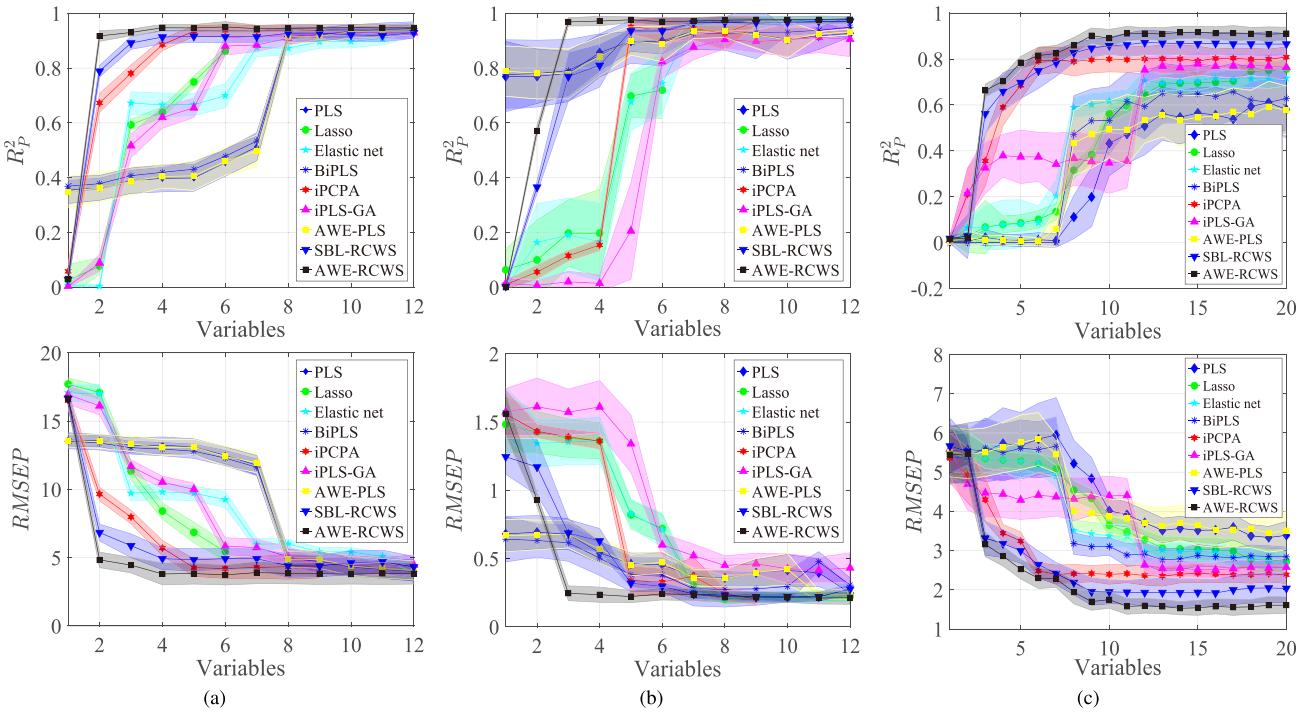


Fig. 7. R_P^2 and RMSEP under sorted characteristic wavelengths. (a) Shootout dataset. (b) Diesel fuel dataset. (c) Tea dataset.

TABLE IV
COMPARISON RESULTS OF TEA DATASET

	R_T^2	RMSET	R_P^2	RMSEP	Time(s)
PLS	0.5616	3.6475	0.5651	3.5182	0.55
Lasso	0.6902	3.0412	0.6487	3.2316	128.15
Elastic net	0.7025	2.9825	0.7055	2.9786	192.09
BiPLS	0.5426	3.6885	0.5405	3.6678	1945.98
iPCPA	0.7914	2.4232	0.7909	2.3076	388.56
iPLS-GA	0.7793	2.5142	0.7793	2.5143	1497.94
AWE-PLS	0.5576	3.6402	0.5607	3.5292	22.34
SBL-RCWS	0.8812	1.8491	0.8751	1.9022	175.32
AWE-RCWS	0.9133	1.5561	0.9172	1.5425	182.73

proposed AWE-RCWS method in NIR with complex overlapping peaks.

For the tea NIR, as shown in Fig. 5(c), there are multiple overlapping peaks in the spectra from 1000 to 1350 nm, and there is an independent spectral peak at 2200 nm. The existing iPLS-based methods, including BiPLS, iPCPA, and iPLS-GA methods, cannot accurately determine the spectral peak wavebands with different peak widths by dividing the tea data with overlapping peaks into equal-length intervals. As shown in the shaded part of Fig. 6(c), existing methods select the interval between 2000 and 2400 nm that contains many irrelevant wavelengths, which may cause interference for subsequent variable selection. In contrast, the proposed method can accurately estimate peak wavebands by adaptive learning, which fundamentally reduces the interference of the irrelevant variables in the nonpeak interval before subsequent analysis. In addition, for sucrose composed of C-H, O-H, C-C, and C-O, the proposed method can avoid excessive shrinkage of the relevant characteristic wavelengths while promoting

global irrelevant variable shrinkage through global-local shrinkage. As shown in Fig. 6(c), the proposed method can screen the spectral region of 1150–1400 nm that characterizes the second overtone and combination of first overtones of C-H, and the 1550–1600 nm wavelength that characterizes the key information of the C-H and C-C combination bands [34]. Due to the lack of effective wavelength screening mechanism, the PLS and BiPLS methods introduce much irrelevant information for sucrose prediction. For the Elastic net and iPLS-GA methods, group variables and adjacent variable selection mechanisms introduce many redundant variables. In contrast, the iPCPA method based on the wavelength combination effect will lose many relevant wavelengths. Although the SBL-RCWS methods has performed better than other methods, it introduces wavebands that contain nonspectral wavelength. Similarly, the differences in the means and standard deviations shown in Fig. 7(c) again demonstrate the accuracy and robustness of the proposed AWE-RCWS method in NIR modeling.

As the overlapping peaks become more severe, the performance difference between the compared methods increases significantly, which is attributed to the fact that the proposed method can adaptively estimate peak waveband estimation and perform robust wavelength selection. To further verify the robustness of the proposed AWE-RCWS method, the additive white Gaussian noise (AWGN) model is utilized to introduce white Gaussian noise to the original spectra, which is produced as AWGN(x , SNR) and SNR denotes the signal-to-noise ratio. Due to limited space, only the mean and standard deviation of R_P^2 and RMSEP obtained on the tea dataset are illustrated in Fig. 8. It can be seen that the proposed AWE-RCWS method outperforms existing methods and has relative small standard deviations under all

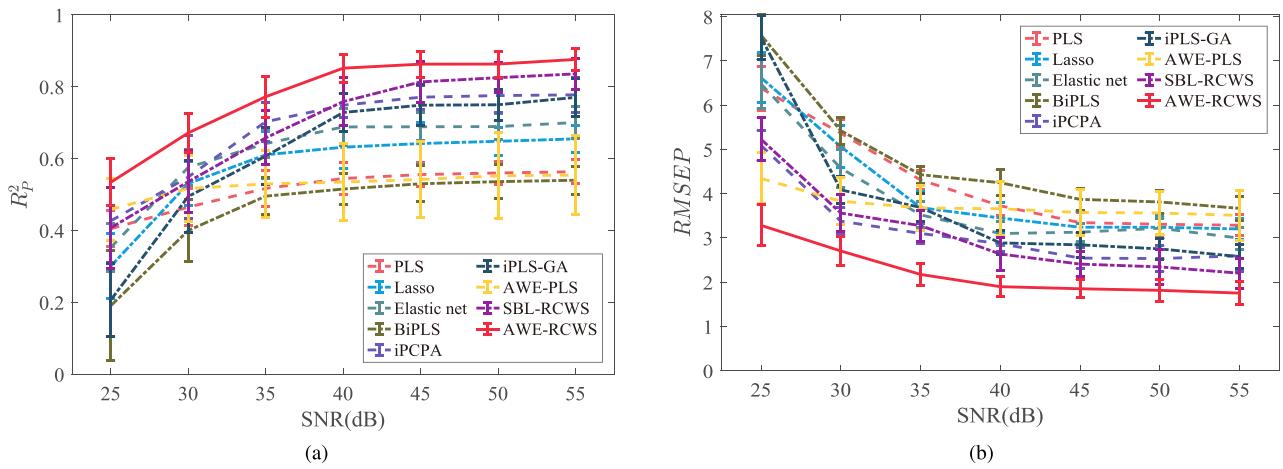


Fig. 8. Model comparison under different SNRs in tea dataset. (a) R^2_p , and (b) RMSEP.

SNRs, which indicates that the proposed method can robustly estimate spectral peak wavebands and determine characteristic wavelengths under measurement interference.

VI. CONCLUSION

NIR-based technologies have been widely used in various fields. This article proposes the AWE-RCWS method for peak waveband estimation and robust characteristic wavelength selection. Unlike existing approaches, the proposed method can accurately determine spectral peak wavebands through hierarchical pattern-coupled sparse learning. Furthermore, the BI framework with global-local shrinkage has been developed for simultaneously compressing irrelevant variables and relaxing shrinking wavelengths with significant explanatory power. Experimental results show that the proposed method can simultaneously improve the generalization and robustness of NIR models, which has excellent advantages in quality-related prediction and monitoring in practical applications.

This method is currently validated on small-scale datasets. We may explore large-scale industrial spectral dataset applications and online analysis in future work. We also plan to explore more suitable prior distributions by incorporating statistical analysis.

REFERENCES

- [1] K. Wang, W. Du, and J. Long, "Near-infrared wavelength-selection method based on joint mutual information and weighted bootstrap sampling," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5884–5894, Sep. 2020.
- [2] H. Yu, W. Du, Z. Lang, K. Wang, and J. Long, "A novel integrated approach to characterization of petroleum naphtha properties from NIR spectroscopy," *IEEE Trans. Instrum. Meas.*, vol. 70, May 2021, Art. no. 2510413.
- [3] C. Han, K. Müller, and H. Hang, "Enhanced performance of a brain switch by simultaneous use of EEG and NIRS data for asynchronous brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2102–2112, Aug. 2020.
- [4] D. Cassanelli, N. Lenzini, L. Ferrari, and L. Rovati, "Partial least squares estimation of crop moisture and density by near-infrared spectroscopy," *IEEE Trans. Instrum. Meas.*, vol. 70, Feb. 2021, Art. no. 1004510.
- [5] H. Wang et al., "Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues," *Trends Anal. Chem.*, vol. 153, pp. 116648–116663, Aug. 2022.
- [6] M. Yamada, W. Jitkrittum, S. Leonid, and M. Sugiyama, "High-dimensional feature selection by feature-wise kernelized Lasso," *Neural Comput.*, vol. 26, pp. 185–207, Feb. 2014.
- [7] Y. Yun, H. Li, B. Cheng, and D. Cao, "An overview of variable selection methods in multivariate analysis of near-infrared spectra," *Trends Anal. Chem.*, vol. 113, pp. 102–115, Apr. 2019.
- [8] J. Fu, H. D. Yu, Z. Chen, and Y. H. Yun, "A review on hybrid strategy-based wavelength selection methods in analysis of near-infrared spectral data," *Infrared Phys. Technol.*, vol. 125, pp. 104231–104243, Sep. 2022.
- [9] R. Leardi, "Application of genetic algorithm-PLS for feature selection in spectral data sets," *J. Chemometrics*, vol. 14, pp. 643–655, Sep. 2000.
- [10] J. Geng, C. Yang, Q. Luo, L. Lan, and Y. Li, "iPCPA: Interval permutation combination population analysis for spectral wavelength selection," *Anal. Chim. Acta*, vol. 1171, pp. 338635–338647, Aug. 2021.
- [11] H. Yu et al., "A bybrid variable selection strategy based on shrinkage of variable space in multivariate calibration," *Anal. Chim. Acta*, vol. 1058, pp. 58–69, Jun. 2019.
- [12] B. Deng, H. Lu, C. Tan, J. Deng, and Y. Yin, "Model population analysis in model evaluation," *Chemom. Intell. Lab. Syst.*, vol. 172, pp. 223–228, Jan. 2018.
- [13] H. Yu, X. Wang, F. Shen, J. Long, and W. Du, "Novel automatic model construction method for the rapid characterization of petroleum properties from NIR spectroscopy," *Fuel*, vol. 316, pp. 123101–123114, May 2022.
- [14] T. Pan, C. Wu, and Q. Chen, "Sparse reconstruction using block sparse Bayesian learning with fast marginalized likelihood maximization for NIR spectroscopy," *IEEE Trans. Instrum. Meas.*, vol. 71, Dec. 2022, Art. no. 2500410.
- [15] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 360–372, Jan. 2015.
- [16] L. Wang, L. Zhao, S. Rahardja, and G. Bi, "Alternative to extended block sparse Bayesian learning and its relation to pattern-coupled sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2759–2771, May 2018.
- [17] J. Lu, J. Ding, C. Liu, and T. Chai, "Hierarchical-Bayesian-based sparse stochastic configuration networks for construction of prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3560–3571, Aug. 2022.
- [18] E. Tipping, "Sparse Bayesian learning with the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.
- [19] L. Cheng, F. Yin, S. Theodoridis, S. Chatzis, and T. Chang, "Rethinking Bayesian learning for data analysis: The art of prior and inference in sparsity-aware modeling," *IEEE Signal Process Mag.*, vol. 139, no. 6, pp. 18–52, Nov. 2022.
- [20] T. Park and G. Casella, "The Bayesian Lasso," *J. Amer. Statist. Assoc.*, vol. 103, no. 482, pp. 681–686, Jan. 2012.

- [21] A. Yan, J. Guo, and D. Wang, "Robust stochastic configuration networks for industrial data modelling with Student's-t mixture distribution," *Inf. Sci.*, vol. 607, pp. 493–505, Aug. 2022.
- [22] G. Pillonetto and A. Yazdani, "Sparse estimation in linear dynamic networks using the stable spline horseshoe prior," *Automatica*, vol. 146, pp. 110666–110674, Oct. 2022.
- [23] H. Li, J. Dai, T. Pan, C. Chang, and H. So, "Sparse Bayesian learning approach for baseline correction," *Chemom. Intell. Lab. Syst.*, vol. 204, pp. 104088–104098, Jul. 2020.
- [24] F. Jerome, H. Trevor, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Statist. Softw.*, vol. 33, no. 1, pp. 1–22, Feb. 2010.
- [25] C. Carvalho, N. Polson, and J. Scott, "Handling sparsity via the horseshoe," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2009, vol. 5, pp. 73–80.
- [26] C. Carvalho, N. Polson, and J. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, pp. 465–480, Jun. 2010.
- [27] M. Enes and F. Schmidt, "A simple sampler for the horseshoe estimator," *IEEE Signal Process Lett.*, vol. 23, no. 1, pp. 179–182, Jan. 2016.
- [28] B. Rajaratnam, D. Sparks, K. Khare, and L. Zhang, "Uncertainty quantification for modern highdimensional regression via scalable Bayesian methods," *J. Comput. Graph. Stat.*, vol. 28, no. 1, pp. 174–184, Sep. 2018.
- [29] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection metropolis sampling within Gibbs sampling," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3123–3138, Jun. 2015.
- [30] D. Hopkins, "Shoot-out 2002: Transfer of calibration for content of active in a pharmaceutical tablet," *NIR News*, vol. 14, no. 5, pp. 10–13, Oct. 2003.
- [31] T. J. Rato and M. S. Reis, "SS-DAC: A systematic framework for selecting the best modeling approach and pre-processing for spectroscopic data," *Comput. Chem. Eng.*, vol. 128, pp. 437–449, Sep. 2019.
- [32] D. Ren, F. Qu, K. Lv, Z. Zhang, H. Xu, and X. Wang, "A gradient descent boosting spectrum modeling method based on back interval partial least squares," *Neurocomputing*, vol. 171, pp. 1038–1046, Jan. 2016.
- [33] Z. Wang et al., "Application of long-wave near infrared hyperspectral imaging for determination of moisture content of single maize seed," *Spectrochim. Acta A. Mol. Biomol. Spectrosc.*, vol. 254, pp. 119666–119678, Mar. 2021.
- [34] F. Simeone, C. Parrella, E. Schaffert, and B. Damasceno, "Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice," *Microchemical J.*, vol. 134, pp. 125–130, May 2017.



Yuqiang Li received the B.S. degree in automatic control and the M.S. degree in control theory and control engineering from Jiangsu University, Zhenjiang, China, in 2018 and 2021, respectively. He is currently working toward the Ph.D. degree in control science and engineering with the East China University of Science and Technology, Shanghai, China.

His current research interests include near-infrared spectroscopy, statistical learning with sparsity, and Bayesian optimization.



Wenli Du (Senior Member, IEEE) received the B.S. and M.S. degrees in chemical process control from the Dalian University of Technology, Dalian, China, in 1997 and 2000, respectively, and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2005.

She is currently a Professor and the Dean of the Graduate School and the Vice Dean of the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology. Her current research interests include control theory and applications, system modeling, advanced control, and process optimization.

Dr. Du was the recipient of the National Science Fund for Distinguished Young Scholars, five State Science and Technology Progress Awards, and 12 first prizes of provincial and ministerial-level Science and Technology Awards. She was the Executive Director for the Chinese Association of Automation and the Executive Director for the Chinese Association of Artificial Intelligence. She is an Associate Editor of seven international journals, including *Industrial and Engineering Chemistry Research*, *Computers and Chemical Engineering*, *Complex and Intelligent Systems*, and *Frontiers in Chemical Engineering*.



Xinjie Wang received the Ph.D. degree in control science and engineering from the College of Information Sciences and Technology, Donghua University, Shanghai, China, in 2020.

From 2021 to 2023, he held a postdoctoral position with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, China. He is currently with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China. His current research interests include computational modeling of plasticity and data-driven evolutionary optimization.



Huijing Yu received the B.S. degree in measurement control technology and instrument and the Ph.D. degree in control science and engineering from the East China University of Science and Technology, Shanghai, China, in 2017 and 2022, respectively.

She is currently with the China Aviation Lithium Battery Technology (Shenzhen) Company Ltd., Shenzhen, China, as a Senior Engineer in Big Data. Her current research interest is the application of artificial intelligence and smart manufacturing in new energy field.