Supplement for

LIBS Spectrum Baseline Correction Method based on Non-Parametric Prior Penalized Least Squares Algorithm

Shengjie Ma[a,b,c], *Shilong Xu*[*a,b,c], *Youlong Chen*[a,b,c], *Zhenglei Dou*[a,b,c], *Yuhao Xia*[a,b,c], *Wanying Ding* [a,b,c], *Jiajie Dong* [a,b,c] *and Yihua Hu*[*a,b,c]

a. State Key Laboratory of Pulsed Power Laser Technology, National University of Defense Technology, Hefei 230037, People's Republic of China

b. Key Laboratory of Electronic Restriction of Anhui Province, National University of Defense Technology, Hefei 230037, People's Republic of China

c. Advanced Laser Technology Laboratory of Anhui Province, National University of Defense Technology, Hefei 230037, People's Republic of China

Corresponding Author

*Email: skl_hyh@163.com; mail: xushi1988@nudt.edu.cn

1. Supplement for the improvement of the weighting method in Section 2.2

For the AsLS algorithm, the value of ω_i can be expressed as

$$\omega_i = \begin{cases} p & , y_i > z_i \\ 1-p, & y_i \le z_i \end{cases}, \tag{S.1}$$

where p is a fixed asymmetric parameter, it is recommended to set between 0.001 and 0.1. For the AsLS algorithm, it is essential to choose the values of p and λ reasonably to obtain a better correction result. However, the value of p is fixed, and it cannot be adjusted automatically.

For the airPLS algorithm, the weight coefficient can be described as

$$\omega_i^t = \begin{cases} 0 & , y_i^t \ge z_i^{t-1} \\ \exp\left(\dfrac{t\left|y_i^t - z_i^{t-1}\right|}{\left|d^-\right|}\right), & y_i^t < z_i^{t-1} \end{cases}, \tag{S.2}$$

where d^- is the negative part of d, and $d = y - z$ is the difference between the spectrum signal and the fitted baseline. t is the iteration number. In the airPLS algorithm, a different method is used to assign ω_i, and its value can be adaptively obtained.

For the IairPLS algorithm, the weight coefficient can be described as

$$\omega_i^t = \begin{cases} 0 & , y_i^t \ge z_i^{t-1} \\ \dfrac{t\exp\left(\dfrac{\left|y_i^t - z_i^{t-1}\right|}{\left|d^-\right|}\right)}{1+\exp\left(\dfrac{\left|y_i^t - z_i^{t-1}\right|}{\left|d^-\right|}\right)}, & y_i^t < z_i^{t-1} \end{cases}, \tag{S.3}$$

In order to better adapt to the noisy environment, the arPLS algorithm was proposed, and the weight ω_i is

$$\omega_i^t = \begin{cases} logistic\left(y_i^t - z_i^{t-1}, m_{d^-}, \sigma_{d^-}\right), & y_i^t \ge z_i^{t-1} \\ 1 & , y_i^t < z_i^{t-1} \end{cases}, \tag{S.4}$$

where m_{d^-} and σ_{d^-} are the mean and standard deviation of d^-, and the generalized logical function is defined as follows

$$logistic(d, m, \sigma) = \frac{1}{1+\exp\left(\dfrac{2\left(d-\left(-m+2\sigma\right)\right)}{\sigma}\right)}, \tag{S.5}$$

The weighting method in the arPLS algorithm considers the noise distribution along the baseline. It assumes that the noise is evenly distributed above and below the baseline in the region without the spectrum peaks.

For the IarPLS algorithm, the weight coefficient can be described as

$$\omega_i^t = \frac{1}{2}\left[1 - \frac{e^t\left[d-\left(-m+2\sigma\right)\right]/\sigma}{\sqrt{1+\left[e^t\left[d-\left(-m+2\sigma\right)\right]/\sigma\right]^2}}\right], \tag{S.6}$$

For the drPLS algorithm, the weight coefficient can be described as

$$\omega_i^t = \begin{cases} \dfrac{1}{2}\left[1 - \dfrac{e^t\left[d-\left(-m+2\sigma\right)\right]/\sigma}{1+\left|e^t\left[d-\left(-m+2\sigma\right)\right]/\sigma\right|}\right], & y_i^t \ge z_i^{t-1} \\ 1 & , y_i^t < z_i^{t-1} \end{cases}, \tag{S.7}$$

For the LSRPLS algorithm, the weight coefficient can be described as

$$\omega_i^t = \begin{cases} \dfrac{1}{2}\left[1 - \dfrac{10^t\left[d-\left(-m+2\sigma\right)\right]/\sigma}{1+\left|10^t\left[d-\left(-m+2\sigma\right)\right]/\sigma\right|}\right], y_i^t \geq z_i^{t-1} \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , y_i^t < z_i^{t-1} \end{cases}, \tag{S.8}$$

2. Supplement for the three multivariate analysis models in Section 4.3

2.1 The Partial Least Squares Regression (PLSR) model

In the PLSR model, the selection of principal components is very critical. If the principal components are too small, the effective information in the original independent variable matrix may be ignored, which will affect the fitting ability and prediction accuracy. Too many principal components may improve the R^2 of the prediction results, but may also introduce irrelevant noise and overfitting phenomenon. According to Table 5, the number of characteristic spectral lines selected for different elements varies from 4 to 12. In order to determine the number of principal components, the Cross Validation method is usually used. However, considering the small number of independent variables in the paper, we use an iterative approach to choose the number of principal components. Table S1 gives the R^2 of the seven elements under different principal components, and the best principal component can be selected by comparing the values of R^2.

Based on the results in Table S1. The principal components of Ca, Si, Mg, Al, Mn, P, Ti are 9, 5, 4, 4, 7, 3, 5, respectively.

Table S1 The R2 results of the seven element under different values of principal components.

Element	Number of inputs	Principal components	R^2
Ca	12	1	0.7402
		2	0.8968
		3	0.9109
		4	0.9006
		5	0.9300
		6	0.9251
		7	0.9376
		8	0.9422
		9	**0.9464**
		10	0.8748
		11	0.8712
		12	0.8710
Si	6	1	0.4840
		2	0.5525
		3	0.6318
		4	0.7372
		5	**0.7948**
		6	0.6650
Mg	4	1	0.8477
		2	0.9526
		3	0.9530
		4	**0.9761**
Al	4	1	0.7260
		2	0.7984
		3	0.8669
		4	**0.9140**
Mn	9	1	0.5528
		2	0.6710

		3	0.6784
		4	0.6880
		5	0.6902
		6	0.7054
		7	**0.7225**
		8	0.7132
		9	0.7133
P	4	1	0.6206
		2	0.7183
		3	**0.7254**
		4	0.7334
Ti	6	1	0.7299
		2	0.8654
		3	0.8720
		4	0.8926
		5	**0.9338**
		6	0.9050

2.2 The Support Vector Regression (SVR) model

The essence of the SVR model is to find a regression plane to which all the data has the closest distance. In the SVR model, the choice of the penalty factor and the radial basis kernel function is very important to construct regression function. The penalty factor represents the importance of SVR model to outliers and influences the prediction accuracy. The larger the penalty factor, the smaller the error of the training set, and it is more likely to overfit. The smaller the penalty factor is, the larger the error of the training set is, and it is easy to underfit. If the penalty factor is too large or too small, the generalization ability of the model will be weakened. The radial basis kernel function represents the distribution of data mapped to the high-dimensional feature space, which influences the training speed. The larger the radial basis kernel function is, the fewer support vectors are, and the faster the speed is. The smaller the radial basis kernel function, the more support vectors, the slower the speed.

For the selection of these two parameters, the Cross Validation and Grid Search method can be used, but it takes quite a long time to find the best parameters, especially in a large range. Therefore, the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) algorithm can be used to select suitable parameters, which are more efficient than the Grid Search algorithm. And in this paper, the PSO algorithm is adopted to optimize the two performance-influencing parameters in the SVR model. Table S2 gives the optimized results for the seven elements.

Table S2 The optimized parameters of SVR for seven elements.

Element	The penalty factor	The radial basis kernel function
Ca	724.0773	0.1767
Si	38.027	0.4659
Mg	100	0.1574
Al	256	0.0055
Mn	100	0.0010
P	16	0.0156
Ti	3.1454	0.2171

2.3 The Extreme Learning Machine (ELM) model

The Extreme Learning Machine (ELM) is a kind of single hidden layer neural network, whose main feature is that the weight and bias coefficient of the hidden layer node can be randomly generated, and will not be adjusted after generation. The hyperparameters of ELM model include the number of neurons in hidden layer, neuron activation function, etc. Table S3 gives the hyperparameters of ELM model.

Table S2 The hyperparameters of ELM model for seven elements.

Element	Number of neurons in hidden layer	Activation function
Ca	20	Sigmoid
Si	20	Sigmoid
Mg	20	Sigmoid
Al	20	Sigmoid
Mn	20	Sigmoid
P	20	Sigmoid
Ti	20	Sigmoid

2.4 Model performance evaluation

The Root Mean Square Error (RMSE), and the Correlation Coefficient (R^2) are often used to evaluate the prediction accuracy of the multivariate analysis model. R^2 represents the linear relationship between the actual value and the predicted value of the model. RMSE is used to measure the deviation between the actual value and the predicted value, which is to evaluate the performance of the model from the perspective of collation.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y_i}\right)^2}, \tag{S.9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N}}, \tag{S.10}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\overline{y_i}$ is the mean of y_i. N is the number of samples.

Finally, we introduce the Relative Standard Deviation (RSD). The RSD is used to evaluate the repeat ability of the LIBS measurement. The smaller the RSD is, the more precision the LIBS measurement is.

$$RSD(\%) = \frac{S}{\overline{y}} \times 100\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n-1}}}{\overline{y}} \times 100\% \tag{S.11}$$

where S is the standard deviation, y_i is the LIBS data, and $\overline{y}$ is the mean of the average value of n measurements. And n is 200 in this paper.