

# Feasibility of terahertz spectroscopy for hybrid purity verification of rice seeds

Yaling Yang<sup>1</sup>, Shengling Zhou<sup>2\*</sup>, Jie Song<sup>2</sup>, Jie Huang<sup>2</sup>, Guanglin Li<sup>2</sup>, Shiping Zhu<sup>2</sup>

(1. College of Physical Science and Technology, Southwest University, Chongqing 400715, China;

2. College of Engineering and Technology, Southwest University, Chongqing 400715, China)

**Abstract:** The purity of hybrid rice seeds reflects the typical consistency of seed varieties in characteristics. The accuracy and reliability of seed purity detecting are of great significance to ensure the quality of seeds. In this study, the feasibility of identifying the purity of hybrid rice seeds, Xinong 1A/89, by terahertz (THz) time-domain spectroscopy system combined with chemometrics was explored. Three quantitative identification models for testing the purity of Xinong 1A/89 hybrid rice seed were developed and compared by THz absorption spectroscopy with extreme learning algorithm (ELM), Principal cComponent Regression (PCR) and Partial Least Squares Regression (PLSR). Experimental results showed that comparing with classical PLSR and PCR models, ELM presents a better feasibility and stability. For the testing set, the quantitative prediction result of ELM ( $E_{Loo}=2.005 \times 10^{-5}$ ,  $R^2=96.75\%$ ) is significantly better than those of PCR ( $E_{Loo}=7.346 \times 10^{-5}$ ,  $R^2=88.10\%$ ) and PLSR ( $E_{Loo}=8.007 \times 10^{-5}$ ,  $R^2=87.03\%$ ). The results highlight the feasibility of THz spectroscopy combined with ELM as an efficient and reliable method for verification of hybrid rice seeds.

**Keywords:** purity detection, hybrid rice seeds, terahertz spectroscopy, extreme learning algorithm

**DOI:** 10.25165/j.ijabe.20181105.3898

**Citation:** Yang Y L, Zhou S L, Song J, Huang J, Li G L, Zhu S P. Feasibility of terahertz spectroscopy for hybrid purity verification of rice seeds. *Int J Agric & Biol Eng*, 2018; 11(5): 65–69.

## 1 Introduction

The purity identification of hybrid rice seeds is the most important and core content in the industrial development of rice seeds<sup>[1]</sup>. So far, many scholars have devoted themselves to the research of seed purity identification technology and seed authenticity. The purity testing technologies have been further developed. However, the traditional detection techniques, field planting and DNA molecular marker identification, is confined by long cycle, high cost, strict requirement for technician and experimental conditions. Thus, it is necessary to explore a more efficient, convenient, reliable and accurate hybrid rice seeds purity identification technology. As efficient detection technologies, spectral technologies have drawn wide attention due to the on-line detection and efficient sample preparations. Hyperspectral image<sup>[2]</sup>, visible+near infrared<sup>[3,4]</sup>, and near infrared<sup>[5,6]</sup> techniques combined with chemometric methods have shown their success in the purity detection of hybrid seeds. Although many of the spectral technologies mentioned above have been used to quantitative analysis of seed purity, little attention has been paid to the use of terahertz (THz) spectroscopy for purity detection of hybrid seeds. More importantly, many theoretical studies<sup>[7,8]</sup> showed that most biological molecules, such as DNA components, protein and amino acids exhibit fingerprint spectra in the THz

region. All of these features predict its good application prospect in hybrid seed purity identification<sup>[9-12]</sup>. Thus, in this study we explored the feasibility of employing THz spectroscopy combined with different chemometrics for the quantitative identification of the purity of hybrid rice seeds. This work is expected to become a beneficial supplement to the existing researches.

## 2 Materials and methods

All THz spectra of samples were collected by T-Spec THz-TDS (Vilnius, Lithuania) with the transmission method. By scanning fast optical delay line in 10 Hz frequency, the waveform of electrical field of THz radiation was build. During the test, in order to avoiding or reducing the influence of the water vapor in the air, the spectrometer main unit is enclosed in a closed box filled with dry nitrogen, and the relative humidity was controlled below 5%.

The hybrid rice seed Xinong 1A/89, provided by the Rice Research Center of Southwest University, was chosen as the research object. By mixing with their female parent seeds, a set of 9 rice seed samples with the purity of 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% and 100% were prepared. Considering that the THz beam spot size with a diameter of 2 mm is very small compared to the seed size, the difference of detection site and individual character may lead to various results and subsequently reduce the generalization ability and stability of the prediction model. To eliminate the effects of these unfavorable factors<sup>[13-15]</sup>, the mixed rice seeds were dried, crushed, ground and then pressed into 13 mm diameter tablets in 1.0 mm thickness. Each tablet with the purity between 60% and 100% at intervals of 5% was prepared by compressing the powders in a pellet press, and applying a pressure of 10 t for 2 min. This series of pretreatment ensure uniform mixing of hybrid rice seeds and their parent seeds and distribution, which may guarantee accuracy reliability of analyzes results. When measuring, firstly we measured THz

**Received date:** 2017-10-08      **Accepted date:** 2018-02-17

**Biographies:** Yaling Yang, Master, research interest: application of biophysics, Email: 13883606889@126.com; Jie Song, Master, research interest: intelligent detection, Email: sj2008@swu.edu.cn; Jie Huang, PhD, research interest: biochemical sensing, Email: jiehuang@swu.edu.cn; Guanglin Li, PhD, research interest: intelligent detection, Email: liguanglin@swu.edu.cn; Shiping Zhu, PhD, research interest: agricultural intelligent detection, Email: zpswu@126.com.

\***Corresponding author:** Shengling Zhou, PhD, research interest: terahertz spectrum analysis, College of Engineering and Technology, Southwest University, Chongqing 400716, China. Tel: +86-13883738593, Email: swuzhou1@163.com.

reference signal  $e_r(t)$  without placing a sample, which provided the spectral characteristics of the noise level of the system. It allows determining the dynamic range of the system as the difference (in dB) between the maximum signal available and the noise level. For T-Spec system, the peak dynamic range (in power) is around 40 dB. And then under the same experimental conditions, each tablet sample was placed in a custom-designed sample holder on an  $x$ - $y$  scanning harness to ensure that the THz wave was normal to the tablet surfaces. By the controlling of step motor the tablet sample was moved in a  $3 \times 3$  grid to acquire 9 separate sample signals  $e_s(t)$  across the tablet. Transfer functions  $H(\omega)$  for all the investigated samples were calculated by taking the Fourier transform after de-convolving measured time-domain signals with the reference signal, according to Equation (1):

$$H(\omega) = \frac{FFT(e_s(t))}{FFT(e_r(t))} = |H(\omega)|e^{j\varphi(\omega)} \quad (1)$$

$$n_s(\omega) = n_0 - \frac{\omega d}{c} \varphi(\omega) \quad (2)$$

$$\alpha(\omega) = \frac{2}{d} \ln \left[ \frac{4n_0 n_s(\omega)}{|H(\omega)|(n_s(\omega) + n_0)^2} \right] \quad (3)$$

where,  $H(\omega)$  is the sample transmission function;  $|H(\omega)|$  is amplitude spectrum and  $\varphi(\omega)$  is phase spectrum;  $d=1.0$  mm is the sample thickness;  $n_0$  is the refractive index of nitrogen (The default is 1),  $n_s(\omega)$  and  $\alpha(\omega)$  are the refractive index and absorption coefficient of the sample separately. The refractive index and the absorption coefficient can be deduced according to Equations (2) and (3). THz absorption spectrum curve was calculated for hybrid rice Xinong 1A/89 with different purity, as shown in Figure 1.

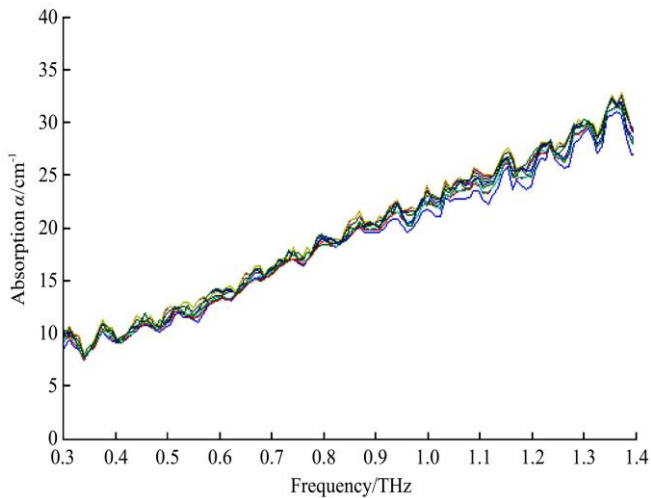


Figure 1 Mean THz absorption spectrum for different purity hybrid rice seeds of Xinong 1A/89

In the following analysis, all 81 THz spectra obtained from 9 tablet samples with various purities were divided into the training set and predicting set with a ratio of 2:1. For every tablet sample with different purity, the absorption spectra corresponding to 6 sample points were chose randomly to make up the training set, while the rest 3 absorption spectrums form the predicting set.

Figure 1 shows that the hybrid rice seeds have complex components, and the absorption spectra of each component overlap and interfere with others. Thus there was no obvious absorption peak observed in the absorption spectra. What's more, component of hybrid rice seeds is very close to their female parent seeds, and spectral absorptions of seed samples with various purities are very similar. It is impossible to discriminate one from

others directly by absorption spectrum. Hence, mathematical methods such as chemometric and pattern recognition are needed to process the absorption spectra and establish regression models to predict seed purity of hybrid rice. The challenge for any THz regression modeling is to find the appropriate modeling algorithm and explore the optimum operating parameters as these factors determine the prediction accuracy of the model and become the key problem to be solved in the modeling process. For the sake of tackling this problem, three different quantities modeling algorithms, ELM, PLSR and PCR, were used and compared for the regression analysis of THz spectra. In order to obtain a good generalization performance Leave-one-out (LOO) cross-validation procedure was used to select the optimum parameters for each model and the parameter that produced the least mean squared error of prediction (MSEP) was selected as optimum value when training the model.

### 3 Establishment and verification of regression model

#### 3.1 Extreme learning machine algorithm

ELM as a simple and fast learning algorithm for machine learning was originally proposed by Professor Huang of Singapore<sup>[16]</sup>. Its salient feature is that the input weights and bias of hidden nodes can be randomly generated instead of being exhaustively tuned, by which only the output weights need to be calculated using the Moore-Penrose generalized pseudoinverse.

Consider a regression task where we have a training set  $S$  with  $N$  samples,  $S = \{(\mathbf{x}_j, \mathbf{t}_j) | j=1, 2, \dots, N\}$ . Here  $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jd}]^T \in R^d$  is the input terahertz feature vector of the sample, and  $\mathbf{t}_j \in R$  is real purity value of the sample that  $\mathbf{x}_j$  belongs to. The ELM with  $L$  hidden nodes and activation function  $G(\mathbf{a}, b, \mathbf{x})$  can be expressed as:

$$\mathbf{y}_j = \sum_{i=1}^L \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}_j) = \mathbf{h}(\mathbf{x}_j) \boldsymbol{\beta} \quad (4)$$

where,  $G(\mathbf{a}, b, \mathbf{x}) = \frac{1}{1 + \exp(-(\mathbf{a} \cdot \mathbf{x} + b))}$ , and  $\mathbf{a}_i = [a_{i1}, a_{i2}, \dots, a_{id}]$  is

the input weights that connects the  $i^{\text{th}}$  hidden node with input nodes respectively.  $b_i$  is the bias of the  $i^{\text{th}}$  hidden node.  $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  is output weight of between the hidden layer and the output layer,  $\mathbf{y}_j$  is the actual output corresponding to input  $\mathbf{x}_j$ . For the training set  $S$ , Equation (4) can be overwritten as:

$$\mathbf{Y} = \mathbf{H} \boldsymbol{\beta} \quad (5)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \mathbf{h}(\mathbf{x}_2) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L} \quad (6)$$

where,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_L]^T$ ,  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^T$ .

The components of hybrid rice seeds are complex and diverse. The differences of the detection sites and individual characters will affect the stability of the identification results of the model. Thus, The ELM regression model for hybrid rice should meet the high steady-state precision while also ensuring generalization ability. According to Bartlett theory<sup>[17,18]</sup>, for the same training set, the smaller the output weight norm  $\|\boldsymbol{\beta}\|$  is, the better the generalization performance of the model. In order to obtain the best compromise between generalization performance and training accuracy, a penalty coefficient  $\lambda$  is introduced. Therefore, the optimization objective of the ELM algorithm aiming to reach the best training performance but also the smallest norm of output weights can be rewritten as:

$$\min_{\beta \in \mathbb{R}^{L \times m}} \frac{1}{2} \|\mathbf{H}\beta - \mathbf{T}\|_2^2 + \frac{\lambda}{2} \|\beta\|_2^2 \quad (7)$$

The solution of Equation (7) can be obtained as:

$$\begin{cases} \beta = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T \mathbf{T}, & \text{if } 0 < L \leq N \\ \beta = \mathbf{H}^T (\mathbf{H} \mathbf{H}^T + \lambda \mathbf{I})^{-1} \mathbf{T} & \text{else} \end{cases} \quad (8)$$

where,  $\mathbf{I}$  is the identity matrices.

### 3.2 Optimization and realization of the extreme learning machine model

Aim to minimum train error and improve the robustness and generalization performance of the model penalty factor  $\lambda$  is introduced to the ELM scheme. However, this brings a new problem how to seek the optimal parameters  $\lambda$  and  $L$  for a specific task.

To address this question, the Leave-One-Out (LOO) cross validation approach is adopted for parameter optimization.  $MSEP$  is used to assess the performance of regressions.

$$MSEP = \frac{1}{N} \sum_{j=1}^N (\mathbf{t}_j - \mathbf{y}_j)^2 \quad (9)$$

The smaller the  $MSEP$  is, the better the prediction accuracy is. To simplify the computation of the LOO algorithm, we adapt the linear regression theory during the model training process, which improve the efficiency in computing  $MSEP$  of LOO approach and free from repeated training. For the training set  $S$  with  $N$  samples, the calculation of  $MSEP$  can be rewritten as below:

$$E_{LOO} = \frac{1}{N} \sum_{j=1}^N \left( \frac{\mathbf{t}_j - \mathbf{y}_j}{1 - \mathbf{HAT}_{jj}} \right)^2 \quad (10)$$

where,  $\mathbf{HAT}_{jj}$  denotes the diagonal pixel of the  $j^{\text{th}}$  row and the  $j^{\text{th}}$  column in  $\mathbf{HAT}$  matrix, which defined as  $\mathbf{HAT} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ .

In order to achieve high steady-state precision and well generalization performance, the penalty factor  $\lambda$  and the number of hidden nodes  $L$  of ELM need to be selected appropriately. In this paper, a wide range of  $\lambda$  and  $L$  (81 different values of  $L$ ,  $L \in [20:1:100]$ , and 81 different values of  $\lambda$ ,  $\lambda \in [e^{-6.0}, e^{-5.9}, \dots, e^{1.9}, e^{2.0}]$ ) are tried for the purity regression model of hybrid rice seeds based on THz spectra.

Figure 2 shows the performance of ELM with sigmoid kernel corresponds to various combinations of  $\lambda$  and  $L$ . As can be seen from this figure, the performance of ELM is not sensitive to the number of hidden nodes  $L$ , and it can achieve high precision and well generalization performance as long as  $L$  is large enough. In other words, only one parameters  $\lambda$  need to be specified by users. Figure 3 depicts the recognition error and training time via different number of hidden nodes and optimal  $\lambda$  with using the LOO cross-validation approach.

From Figure 3, it can be clearly seen that the  $E_{LOO}$  of the ELM model about the training set and predicting set rapidly decreases with the increasing of  $L$  when  $L \in [20, 70]$ . But when  $L > 70$ , the  $E_{LOO}$  for the training set and predicting set tends to be stable. While, the training time of the model monotonically increases with  $L$ . Finally, considering the balance of training efficiency, recognition accuracy and generalization ability of ELM model,  $L_{opt} = 70$ ,  $\lambda_{opt} = e^{-5.7}$  are chosen for purity prediction of hybrid rice seed Xinong 1A/89, and in this condition, the  $E_{LOO}$  for the training data and the predicting data are  $E_{LOO}^{train} = 1.951 \times 10^{-5}$ ,  $E_{LOO}^{test} = 2.062 \times 10^{-5}$ , respectively, and the corresponding training time is 0.0042 s.

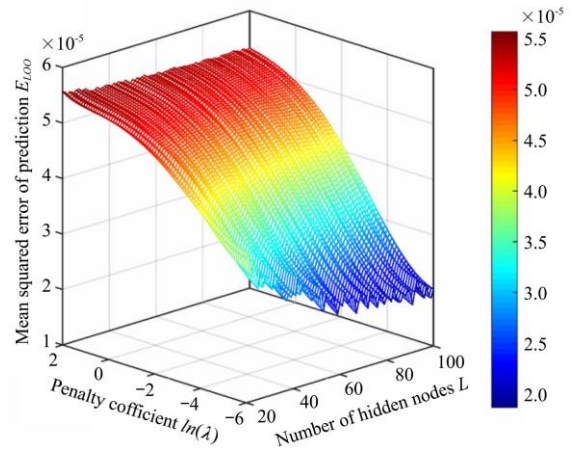


Figure 2 Performance of ELM respect to combinations of  $\lambda$  and  $L$

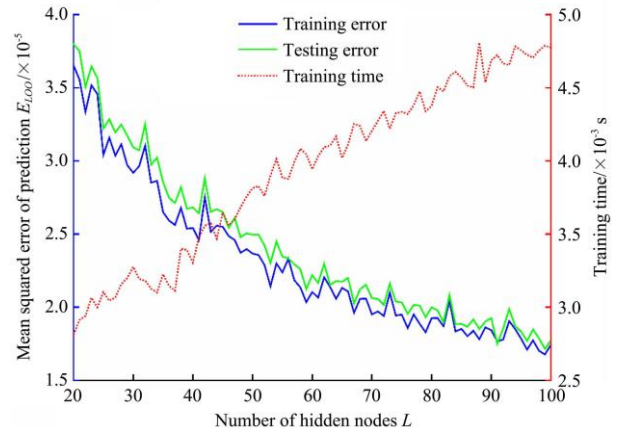


Figure 3 Recognition performance and training time under different  $L$  in ELM

### 3.3 Optimization and realization of PLSR and PCR model

To further explore the feasibility of terahertz spectroscopy for hybrid purity verification of race seed combined with different chemometric tools, two other powerful and widely used multivariate calibration method, PLSR and PCR, were also employed. Both PCR and PLSR methods have been applied diffusely and successfully in the quantitative analysis of THz spectroscopy data because of their ability to overcome problems common to this data such as collinearity, band overlaps and interactions and ease of their implementation. However, different from the ELM method, which build the model with whole spectral range, PCR and PLSR reduce the dimension of the spectroscopy data and compresses the information into a few new variables called principal components.

In principle, the number of principal component often affects the prediction accuracy also degrade accuracy of PCR and PLSR. If the number of principal components used in modeling is too small, the insufficient fitting may occur, that is, the selected principal components do not adequately reflect the special information which made the hybrid rice different from others and the model prediction accuracy will be reduced. Conversely if the number of components used in modeling is too large, some factors that represent the interference factors, non-modeled interferences, background variations and interactions may be introduced into the model, which may also degrade model performance.

Therefore, in order to build well-fitted model and avoid non-modeled interferences with the training set,  $MSEP$  combined with LOO cross-validation,  $E_{LOO}$  was also be used herein for choosing the optimal number of principle components in PLSR and

PCR. As shown in Figure 4, the prediction errors  $E_{Loo}$  of PLSR model and PCR model reach the minimum value when the 5 and 8 principal components were selected respectively.

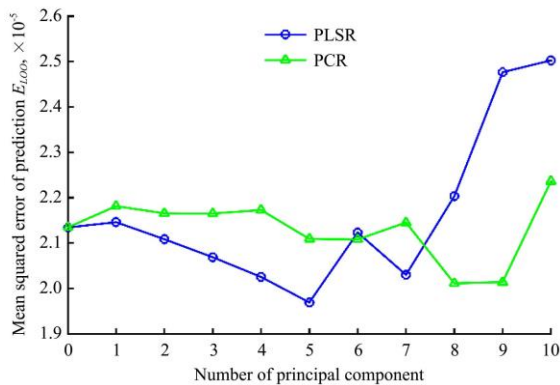


Figure 4 Estimated  $E_{Loo}$  using PLSR and PCR for the purity of training set under different number of principle components

## 4 Results and analysis

### 4.1 Experimental results

Figure 5 shows the predicted purity and the desired purity for the training set and testing set with PCR, PLSR and ELM model under the selected parameters. The results of PLSR and PCR are offset horizontally for clarity. The vertical ordinate represents the predicted results and the horizontal ordinate represents the real purity of the seed samples. The corresponding  $E_{Loo}$  and  $R^2$  obtained for the training and testing set are summarized in Table 1.

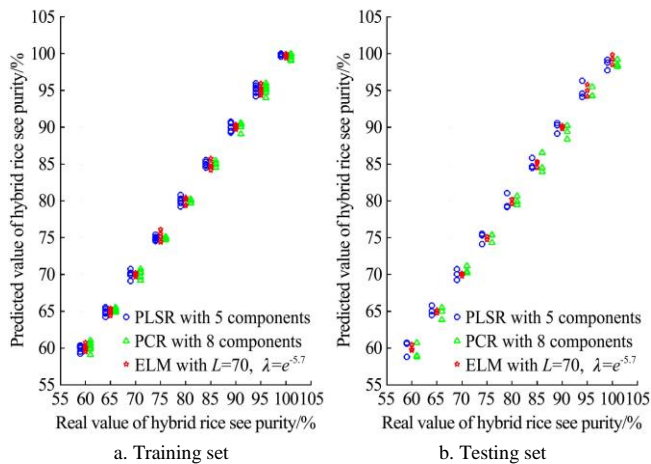


Figure 5 Purity prediction results of the PLSR, PCR and ELM model for the training set and testing set

Table 1 Statistical quantities,  $E_{Loo}$  and  $R^2$ , for the ELM, PCR and PLS-based training and testing set

Method	Training set		Testing set	
	$E_{Loo}^{train}$	$R^2$	$E_{Loo}^{test}$	$R^2$
ELM	$1.906 \times 10^{-5}$	0.9989	$2.005 \times 10^{-5}$	0.9675
PLSR	$2.097 \times 10^{-5}$	0.9987	$7.346 \times 10^{-5}$	0.8810
PCR	$1.978 \times 10^{-5}$	0.9988	$8.007 \times 10^{-5}$	0.8703

Here,  $R^2$  represents coefficient of determination, which provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. It is defined as:

$$R^2 = 1 - \frac{\sum_{j=1}^N (\mathbf{t}_j - \mathbf{y}_j)^2}{\sum_{j=1}^N (\mathbf{t}_j - \bar{\mathbf{t}})^2} \quad (11)$$

The larger  $R^2$  is, and the smaller the standard error is, the higher the precision of the model is assumed to be.

### 4.2 Discussion

Figure 5 and Table 1 show the regression results and the statistical quantities of models for the training set and testing set under the selected modeling parameter respectively. As it can be observed that, with the selected modeling parameters, all the models can achieve a similar good performance for the training set. All  $E_{Loo}^{train}$  according to LOO cross-validation are about  $2.0 \times 10^{-5}$ , and the  $R^2$  are about 99.9% for ELM, PLSR and PCR. However, when these calibrated models were used for the quantization of purity of hybrid rice seeds in testing set that did not have contribution in the model training steps, the performance of PLSR and PCR are obviously worse than that in the training set. The  $R^2$  and  $E_{Loo}^{test}$  obtained by PLSR from the analysis of selected 5 principle components are just 88.10% and  $7.346 \times 10^{-5}$  where those of PCR are only 87.03% and  $8.007 \times 10^{-5}$ , respectively. This deviation might be caused by many factors, such as limitation of the PLSR and PCR in complex and small sample calibration system, random noise, and so on, which made the conventional approaches more difficult to finish such task. While for ELM, the  $E_{Loo}^{test}$  and decision coefficient  $R^2$  of the testing set are  $2.005 \times 10^{-5}$  and 96.75%, which are no significant differences with the testing set, indicating that the ELM model had a certain feasibility and stability.

Besides, from the Figure 5, we should also note that when the purity is in 60%-90% the prediction error of the ELM model is less than 0.5%, and the  $E_{Loo}^{test}$  is only about  $6.3 \times 10^{-6}$ . However, when the purity of the seed reach above 90%, the  $E_{Loo}^{test}$  is about  $3.22 \times 10^{-5}$ , and the prediction error of the ELM model reach up to 2.3%, which is slightly below national standards.

## 5 Conclusions

Combined with three chemometric methods, ELM, PLSR and PCR, this paper explores the purity detection of hybrid rice seeds by using the new spectral detection method-terahertz time-domain spectroscopy. In order to build well-fitted model and avoid non-modeled interferences with the training set,  $MSEP$  combined with LOO cross-validation was be used for choosing the optimal parameters when training the models. Compared with classical PLSR and PCR models, ELM presents a better feasibility and stability. The quantitative prediction result of ELM ( $E_{Loo}=2.005 \times 10^{-5}$ ,  $R^2=96.75\%$ ) is significantly better than PCR ( $E_{Loo}=7.346 \times 10^{-5}$ ,  $R^2=88.10\%$ ) and PLSR ( $E_{Loo}=8.007 \times 10^{-5}$ ,  $R^2=87.03\%$ ). This result highlight the feasibility of THz spectroscopy combined with ELM as an efficient and reliable method for verification of hybrid rice seeds.

It is worth to point out that we use the mass ratio instead of quantity ratio to estimate the purity of the hybrid rice seeds, thus when the purity of seed reaches above 90%, the prediction accuracy of the model decreases slightly below national standards. However, as an important supplement of infrared spectroscopy, the THz spectrum technology makes up for the shortcomings of the traditional methods for detecting the purity of hybrid rice seeds and laid a foundation for the rapid detection of the purity of hybrid rice seeds. As a whole, THz spectroscopy is a very promising method which should be verified using various kinds of rice seeds before recommending its use in operational practice.

## Acknowledgements

This work is supported by Application Development Programs of Chongqing Science and Technology Commission (Grant No. cstc2014yykfA80006), Fundamental Research Funds for the Central Universities (Grant No. SWU117029) and National Natural Science Foundation of China (Grant No. 61401373 and 31771670).

## [References]

- [1] Lai Y H. Identification of the authenticity and purity of hybrid rice seed. *China Seed Industry*, 2015; 6: 19–21. (in Chinese)
- [2] Jia S Q, Liu Z, Li S M, Li L, Ma Q, Zhang X D, et al. Study on method of maize hybrid purity identification based on hyperspectral image technology. *Spectroscopy and Spectral Analysis*, 2013; 33(10): 2847–2852. (in Chinese)
- [3] Liang L, Yang M H, Liu Z X, Xu H W, Liu F H, He Q Z, et al. Purity measurement of hybrid rice seed Yixiang 725 with visible-near infrared reflectance spectra. *Spectroscopy and Spectral Analysis*, 2009; 29(11): 2962–2965. (in Chinese)
- [4] Farhadi M, Tigabu M, Stener L G, Odan P C. Feasibility of visible + near infrared spectroscopy for non-destructive verification of European × Japanese larch hybrid seeds. *New Forests*, 2016; 47(2): 271–285.
- [5] Attaviroj N, Kasemsumran S, Noomhorm A. Rapid variety identification of pure rough rice by fourier-transform near-infrared spectroscopy. *Cereal Chemistry*, 2011; 88(5): 490–496.
- [6] Li T X, Jia S Q, Liu X, Zhao S Y, Ran H, Yan Y L, et al. Maize hybrid seed purity identification based on near infrared reflectance (NIR) and transmittance (NIT) spectra. *Spectroscopy and Spectral Analysis*, 2015; 35(12): 3388–3392. (in Chinese)
- [7] Zhou S L, Zhu S P, Li G L, Huang J, Yang Y L. Study on the measurement and optimization of soybean oil optical spectrum in THz range. *Spectroscopy and Spectral Analysis*, 2016; 36(4): 924–928. (in Chinese)
- [8] Zhou S L, Valchev D G, Dinovitsier A, Chappell J M, Iqbal A, Ng B W-H, et al. Terahertz signal classification based on geometric algebra. *IEEE Transactions on Terahertz Science & Technology*, 2016; 6(6): 793–802.
- [9] Deshmukh Y, Khare P, Nadaf A B, Patra D. Discrimination between 2AP producing and non-producing rice rhizobacterial isolates using volatile profiling: a chemometric approach. *Journal of Chemometrics*, 2016; 29(12): 648–658.
- [10] Xu W D, Xie L J, Ye Z Z, Gao W L, Yao Y, Chen M, et al. Discrimination of transgenic rice containing the Cry1Ab protein using terahertz spectroscopy and chemometrics. *Sci Rep.*, 2015; 5: 11115.
- [11] Xu W D, Liu X, Xie L J, Ying Y. Comparison of Fourier transform near-infrared, visible near-infrared, mid-infrared, and Raman spectroscopy as non-invasive tools for transgenic rice discrimination. *Transactions of the ASABE*, 2014; 57(1): 141–150.
- [12] Xie L J, Yao Y, Ying Y B. The application of terahertz spectroscopy to protein detection: A review. *Applied Spectroscopy Reviews*, 2014; 49(6): 448–461.
- [13] Li T J, Liu J J, Shao G F, Fan L L. A novel THz spectroscopy recognition method for transgenic organisms based on APSO combined with SVM. *Optics & Spectroscopy*, 2016; 120(4): 660–665.
- [14] Shen X C, Li B, Li X, Long Y. Identification of transgenic and non-transgenic cotton seed based on terahertz range spectroscopy. *Transactions of the CSAE*, 2017; 33(S1): 288–292. (in Chinese)
- [15] Ge H Y, Jiang Y Y, Lian F Y, Zhang Y, Xia S H. Characterization of wheat varieties using terahertz time-domain spectroscopy. *Sensors*, 2015; 15(6): 12560–12572.
- [16] Gao H, Huang G B, Song S J, You K. Trends in extreme learning machines: A review. *Neural Networks*, 2015; 61(C): 32–48.
- [17] Baum E B, Haussler D. What size net gives valid generalization? *Neural Computation*, 2014; 1(1): 151–160.
- [18] Heeswijk M V, Miche Y. Binary/ternary extreme learning machines. *Neurocomputing*, 2015; 149: 187–197.