

## Research Article

# Nonlinear Multivariate Calibration of Shelf Life of Preserved Eggs (Pidan) by Near Infrared Spectroscopy: Stacked Least Squares Support Vector Machine with Ensemble Preprocessing

Lu Xu,<sup>1</sup> Si-Min Yan,<sup>1</sup> Chen-Bo Cai,<sup>2</sup> Xiao-Ping Yu,<sup>1</sup> Jian-Hui Jiang,<sup>3</sup>  
Hai-Long Wu,<sup>3</sup> and Ru-Qin Yu<sup>3</sup>

<sup>1</sup> Zhejiang Provincial Key Laboratory of Biometry and Inspection & Quarantine, College of Life Sciences, China Jiliang University, Hangzhou 310018, China

<sup>2</sup> Department of Chemistry and Life Science, Chuxiong Normal University, Chuxiong 675000, China

<sup>3</sup> State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China

Correspondence should be addressed to Xiao-Ping Yu; [yxp@cjlu.edu.cn](mailto:yxp@cjlu.edu.cn) and Ru-Qin Yu; [rquy@hnu.edu.cn](mailto:rquy@hnu.edu.cn)

Received 15 May 2013; Accepted 16 August 2013

Academic Editor: Pedro D. Vaz

Copyright © 2013 Lu Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper aims at developing a rapid and nondestructive method for analyzing the shelf life of preserved eggs (pidan) by near infrared (NIR) spectroscopy and nonlinear multivariate calibration. A major concern with a nonlinear model is that the noncomposition-correlated spectral variations among pidan objects of different batches and production dates would unnecessarily increase model complexity and cause overfitting and degradation of prediction. To reduce the negative influence of unwanted spectral variations, stacked least squares support vector machine (LS-SVM) with an ensemble of 62 commonly used preprocessing methods is proposed to automatically optimize data preprocessing and develop the nonlinear model. The analysis results indicate that stacked LS-SVM can obtain stable calibration model, and the prediction accuracy is improved compared with models with single-preprocessing methods. Since LS-SVM is much faster than its ordinary counterparts, stacked LS-SVM with ensemble preprocessing can be performed within an acceptable computational time. When the objects and spectral variations are very complex, the proposed method can provide a useful tool for data preprocessing and nonlinear multivariate calibration.

## 1. Introduction

Preserved egg (pidan) is a popular traditional food in South and East Asian countries, including China, South Korea, Thailand, and Japan. Although its nutritional value is slightly decreased compared with fresh eggs, it is favored and consumed in at least 30 different countries globally for its pleasant and particular taste, flavor, texture, and especially its extended shelf life compared with unprocessed eggs [1]. It is commonly recognized that the storage of pidan should not exceed six months. For pidan out of the original pickle, besides the degradation of food quality, the risk of contamination by microbes increases with storage, which has become a major safety concern with pidan products [2]. Therefore, rapid and reliable methods for analysis of pidan

shelf life are highly required to perform routine analysis of pidan products at the market.

As an important quality parameter, the shelf life of pidan has a very complex relationship with its internal chemical components, and it is still difficult to directly estimate the shelf life using the results of chemical or biological analyses [3]. Near infrared (NIR) spectroscopy combined with multivariate calibration methods has been widely used in rapid analysis of food quality parameters [4]. By developing the relationship between the measured NIR spectra and the reference values of sample parameters, future objects can be analyzed and predicted. For pidan, NIR reflectance techniques combined with chemometrics can achieve rapid and nondestructive analysis, which is very suitable because there are usually many batches of samples to be analyzed.

TABLE I: The analyzed pidan objects.

No.	Batch size <sup>a</sup>	Shelf life (days)	Brand	No.	Batch size	Shelf life (days)	Brand
1	36	28	A	16	29	131	B
2	34	44	A	17	29	148	B
3	34	60	A	18	26	166	B
4	34	87	A	19	24	188	B
5	34	104	A	20	16	208	B
6	33	120	A	21	32	27	C
7	34	137	A	22	30	42	C
8	32	155	A	23	30	58	C
9	26	177	A	24	30	85	C
10	23	197	A	25	30	102	C
11	32	38	B	26	30	118	C
12	30	55	B	27	30	135	C
13	30	71	B	28	27	153	C
14	29	98	B	29	25	175	C
15	29	115	B	30	26	195	C

<sup>a</sup>Number of preserved eggs.

To model the complex relationship between the measured NIR spectra and pidan shelf life, nonlinear multivariate calibration methods like artificial neural net (ANN) [5], kernel partial least squares (KPLS) [6, 7], and support vector machines (SVMs) [8, 9] are commonly used. Compared with linear models like principal components regression (PCR) [10] and partial least squares (PLS) [11], though nonlinear models have better model flexibility and accuracy when the data nonlinearity is significant, they are generally more complex and have a higher risk of overfitting. To develop a practical calibration model for analysis of pidan shelf life, sufficient and representative objects are collected to train the model. Therefore, proper data preprocessing is necessary to remove the noncomposition-correlated spectral variations among objects of different batches and with a wide range of shelf life, which would lead to overfitting of a nonlinear model.

Considering the lack of sufficient prior information and an incomplete knowledge of the raw NIR data, data preprocessing in multivariate calibration still depends on simple trial and error. Selecting the suitable preprocessing methods might be data dependent and require special expertise and experience of the chemometricians. An ensemble preprocessing method [12, 13] has been recently proposed by combining different single-preprocessing methods, such as smoothing and taking derivatives and standard normal variate (SNV) [14] with stacked regression [15–17]. This method can automatically select and weight the most suitable preprocessing methods and has been demonstrated to be effective in linear PLS calibration of NIR data [13].

In this paper, stacked least squares support vector machine (LS-SVM) [18, 19] with ensemble preprocessing is proposed to simultaneously optimize data preprocessing and develop a nonlinear calibration model. The proposed chemometric method and the results of data analysis will be presented in detail in the following subsections.

## 2. Theory and Method

**2.1. Pidant Samples and NIR Analysis.** Three brands of pidan objects are analyzed, and for each brand the NIR spectra are measured at certain time intervals. All the pidan objects are made of fresh duck eggs with unleaded processing methods. The detailed information concerning samples is presented in Table 1. Each pidan object is washed vigorously with deionized water and is left to dry completely before NIR analysis. During storage, the samples are stored in a cool, dark, and dry place with integral packaging. The NIR spectra are measured in the diffuse reflectance mode on a Bruker-TENSOR37 FTIR spectrometer (Bruker Optics, Ettlingen, Germany). A PbS detector is used with an internal gold background as the reference. The equatorial region of an egg directly contacts the NIR probe. For each spectrum, 64 scans are performed, and more scans cannot improve the signal significantly. Considering the possible differences in the internal composition of an egg, the average of three spectra measured around the equatorial region of each egg is used as the raw spectrum. The wavelength range is from 12,000 to 4,000  $\text{cm}^{-1}$  with a resolution of 8  $\text{cm}^{-1}$ .

**2.2. LS-SVM.** Data analysis is performed on MATLAB 7.10.0 (MathWorks, Sherborn, MA). The toolbox LS-SVMlab v1.8 [19] is used for basic LS-SVM algorithms. Here a brief introduction of LS-SVM is given, and, for a detailed description of LS-SVM algorithms, one can refer to [18, 19]. Consider a training set of  $n$  objects  $\{\mathbf{x}_k, \mathbf{y}_k\}_{k=1}^n$  with input features  $\mathbf{x}_k \in R^p$  and response values  $y_k \in R^1$ , where  $R^p$  and  $R^1$  are the  $p$ -dimensional and one-dimensional vector spaces, respectively. Nonlinear regression by LS-SVM can be written as

$$y = \mathbf{w}^T \varphi(\mathbf{x}) + b + e, \quad (1)$$

where  $\mathbf{w}$  is the regression coefficient vector,  $e$  is the model error,  $b$  is the constant term, and the function  $\varphi(\mathbf{x})$  maps the

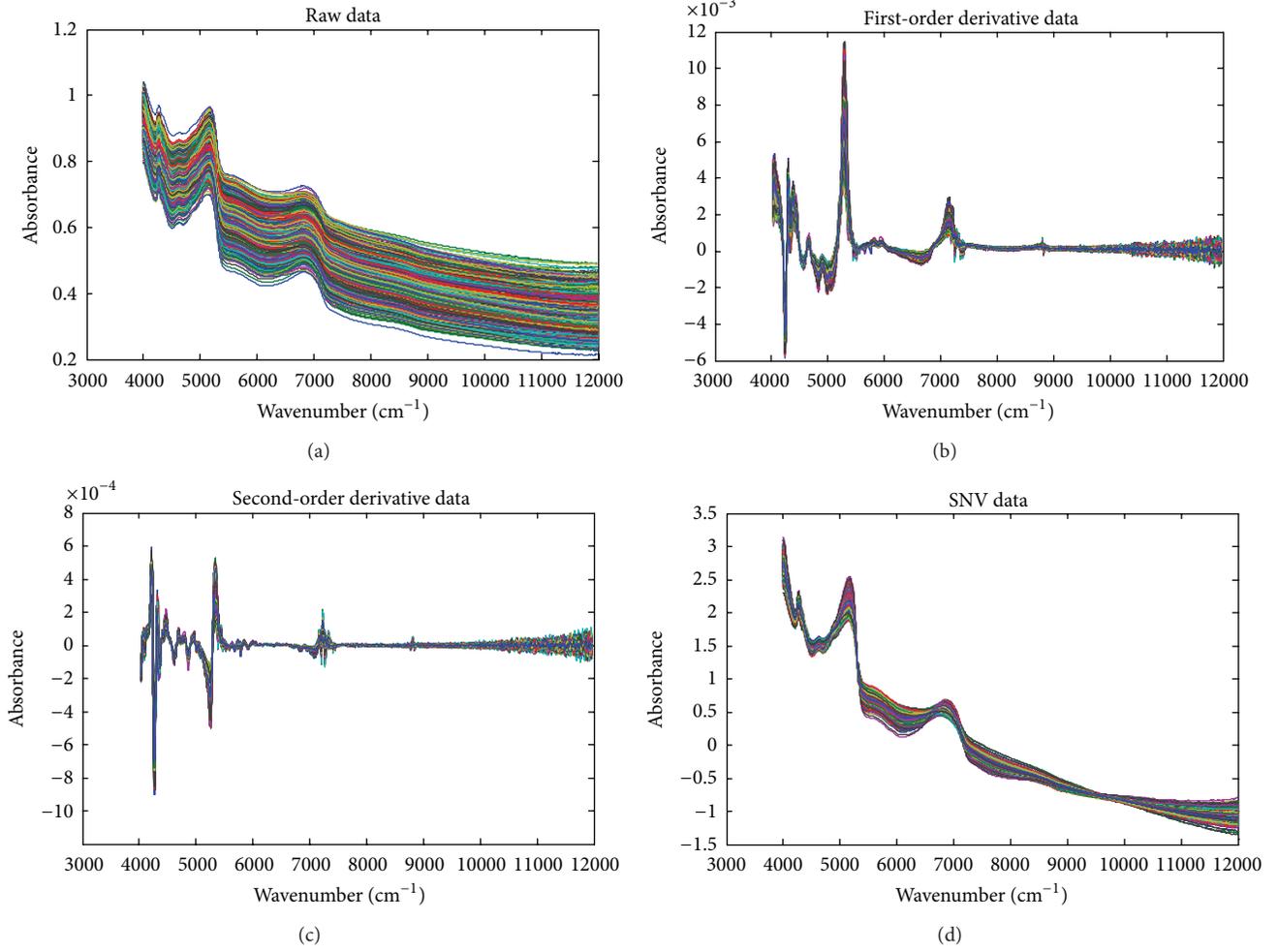


FIGURE 1: Some of the raw, first-order derivative, second-order derivative and SNV spectra of 884 pidan objects.

original data into a nonlinear space. LS-SVM solution can be obtained by solving the following optimization problem:

$$\begin{aligned} \text{Min } f(\mathbf{w}, b) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \gamma \frac{1}{2} \sum_{i=1}^n e_i^2 \\ \text{s.t. } y_k &= \mathbf{w}^T \varphi(\mathbf{x}_k) + b + e_k, \quad k = 1, \dots, n, \end{aligned} \quad (2)$$

where  $\gamma$  is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the learning errors.

In this paper, the frequently used Gaussian radial basis function (RBF) is used as nonlinear transformation:

$$\varphi(\mathbf{x}_i, \mathbf{x}) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}\|^2}{2\sigma^2}\right), \quad (3)$$

where  $\sigma$  is the kernel width parameter that controls the nonlinear nature of the regression. For LS-SVM, the two model parameters,  $\gamma$  and  $\sigma$ , should be optimized simultaneously.

**2.3. Stacked LS-SVM with Ensemble Preprocessing.** Automatic selection and weighting of different preprocessing methods

can be achieved by combining submodels with differently preprocessed NIR spectra. The ensemble model is the linear combination of submodels:

$$\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_l] \mathbf{q}, \quad (4)$$

where  $\mathbf{y}$  is the final prediction of response values by ensemble model,  $\mathbf{y}_i$  ( $i = 1, 2, \dots, l$ ) is the prediction of a submodel, and  $\mathbf{q}$  is the weighting vector. In this paper, 62 different preprocessing methods as listed in Table 2 (no. 1–62), including smoothing and taking first-order derivative (D1) and second-order derivative (D2); standard normal variate (SNV) and their combinations are used to preprocess the raw data, and 62 submodels are developed. Smoothing and taking D1 and D2 spectra are performed using the method proposed by Savitzky and Golay [20]. These methods are preferred because they are simple to compute and frequently used.

Monte Carlo cross-validation (MCCV) [21] stacked regression is used to derive the weighting vector  $\mathbf{q}$  by nonnegative least squares regression (NNLSR):

$$\mathbf{y}_{\text{mccv}} = [\mathbf{y}_{\text{mccv}1}, \mathbf{y}_{\text{mccv}2}, \mathbf{y}_{\text{mccv}3}, \dots, \mathbf{y}_{\text{mccv}l}] \mathbf{q}, \quad (5)$$

TABLE 2: Different preprocessing methods used for data fusion and the results of different models.

No.	Preprocessing <sup>a</sup>	RMSEC (days)	RMSEMCCV (days)	RMSEP (days)
1	S (5, 2) <sup>b</sup>	3.3	8.2	7.7
2	S (7, 2)	3.5	8.4	8.0
3	S (9, 2)	3.2	8.5	7.9
4	S (9, 3)	3.4	8.4	7.9
5	S (11, 3)	3.1	8.8	7.8
6	S (13, 3)	3.3	8.7	7.6
7	S (15, 3)	4.0	8.5	7.3
8	S (17, 4)	3.9	8.5	7.8
9	S (19, 4)	3.9	8.5	7.6
10	S (21, 5)	4.4	8.6	7.4
11	D1 (5, 2)	0.1	9.6	8.0
12	D1 (7, 2)	0.1	9.8	8.5
13	D1 (9, 2)	0.2	53.2	51.2
14	D1 (9, 3)	0.1	9.9	8.1
15	D1 (11, 3)	0.1	10.0	8.5
16	D1 (13, 3)	0.1	9.8	8.3
17	D1 (15, 3)	0.1	9.8	8.7
18	D1 (17, 4)	0.1	9.8	8.6
19	D1 (19, 4)	0.7	9.7	8.2
20	D1 (21, 5)	0.1	9.9	8.6
21	D2 (5, 2)	0.1	11.9	10.1
22	D2 (7, 2)	0.1	12.5	10.2
23	D2 (9, 2)	0.1	13.0	10.4
24	D2 (9, 3)	0.8	13.1	10.4
25	D2 (11, 3)	0.1	12.9	10.0
26	D2 (13, 3)	0.1	12.3	9.3
27	D2 (15, 3)	0.1	11.8	8.4
28	D2 (17, 4)	0.1	13.5	10.5
29	D2 (19, 4)	0.1	13.4	9.9
30	D2 (21, 5)	0.1	12.7	8.8
31	1 + SNV	3.2	9.1	8.3
32	2 + SNV	3.7	9.4	8.6
33	3 + SNV	3.7	9.6	8.6
34	4 + SNV	3.6	9.6	8.6
35	5 + SNV	3.5	9.7	8.5
36	6 + SNV	3.3	9.8	8.4
37	7 + SNV	3.4	9.7	8.1
38	8 + SNV	3.5	9.8	8.6
39	9 + SNV	3.5	9.8	8.4
40	10 + SNV	3.2	9.9	8.2
41	11 + SNV	1.1	9.7	8.5
42	12 + SNV	2.1	9.9	8.7
43	13 + SNV	1.4	9.8	8.9
44	14 + SNV	2.1	9.8	8.2
45	15 + SNV	1.7	9.9	8.7
46	16 + SNV	1.3	9.8	8.8
47	17 + SNV	1.3	9.9	9.2
48	18 + SNV	1.8	10.0	9.2
49	19 + SNV	1.9	9.9	8.8
50	20 + SNV	1.4	10.0	9.4

TABLE 2: Continued.

No.	Preprocessing <sup>a</sup>	RMSEC (days)	RMSEMCCV (days)	RMSEP (days)
51	21 + SNV	0.1	12.7	10.4
52	22 + SNV	0.1	13.1	10.4
53	23 + SNV	0.1	12.7	10.4
54	24 + SNV	0.1	12.8	10.6
55	25 + SNV	0.1	11.9	9.8
56	26 + SNV	0.3	11.0	9.3
57	27 + SNV	0.7	10.5	8.6
58	28 + SNV	0.1	12.5	10.7
59	29 + SNV	0.1	12.1	10.0
60	30 + SNV	0.1	11.8	9.1
61	Raw data	3.2	8.1	7.6
62	SNV	3.0	8.9	8.1
63	Ensemble model	2.3	7.2	6.2

<sup>a</sup>S: smoothing, D1: taking first-order derivative, and D2: taking second-order derivative.

<sup>b</sup>The numbers in the brackets indicate the number of data points and the order of fitting polynomial used in the algorithm by Savitzky and Golay.

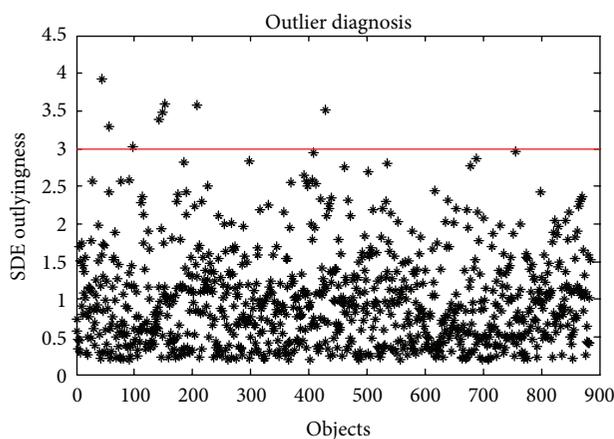


FIGURE 2: Stahel-Donoho estimates (SDE) of outlyingness values for 884 measured pidan objects based on raw NIR spectra.

where  $\mathbf{y}_{\text{mccv}}$  is the stacking of all the reference response values of the leave-out objects in MCCV and  $\mathbf{y}_{\text{mccvi}}$  ( $i = 1, 2, \dots, l$ ) contains the corresponding predicted values by submodels. This method can obtain the lowest root mean squared error of MCCV (RMSEMCCV) among all possible linear combinations under nonnegative constraints. Moreover, by multiple predictions of MCCV, the length of  $\mathbf{y}_{\text{mccv}}$  is very large, so  $\mathbf{q}$  tends to be sparse or most of its elements will be zeros, which means that just a few proper preprocessing methods are included and weighted in the final ensemble model.

The above-stacked LS-SVM with ensemble preprocessing can be developed in two steps. Firstly, a set of submodels are developed based on different preprocessing methods as in Table 2 (no. 1–62). In this paper, simplex method is performed to search for  $\gamma$  and  $\sigma$  that minimize the error of cross-validation for each submodel. Secondly, MCCV is performed for each submodel with the selected parameters, and the submodel weights are deduced as in (5).

### 3. Results and Discussion

The raw spectra, first-order derivative (D1), second-order derivative (D2), and SNV spectra of the pidan objects are plotted in Figure 1. As seen from Figure 1, the raw NIR spectra of pidan samples with different shelf life values are very similar in the wavelength range of  $4,000\text{--}12,000\text{ cm}^{-1}$ . Some of the peaks can be assigned as  $6,000\text{--}7,000\text{ cm}^{-1}$  (overlapping of the first overtone of O–H stretching and N–H stretching),  $5,160\text{ cm}^{-1}$  (the combination of the baseband of O–H stretching and the baseband of H–O–H deformation),  $4,870\text{ cm}^{-1}$  (combination of N–H stretching and deformation of peptide groups), and  $4,600\text{ cm}^{-1}$  (combination of C=O stretching and deformation of peptide groups). Although it is difficult to attribute each peak to certain chemical components, NIR can characterize the changes of some components like proteins and water involved in storage. Taking derivative spectra can significantly improve the peak resolution and highlight some detailed frequencies in the signals. SNV can remove some of the spectral variations, but the preprocessed spectra still have some backgrounds retained. Therefore, smoothing and taking derivatives coupled with SNV are also included as options of data preprocessing.

The Stahel-Donoho estimate (SDE) of outlyingness [22, 23] is performed for outlier diagnosis by projecting each high-dimensional object onto random directions for many times and estimating the robust location and scatter of projections. In this paper, the number of random projections is set to be 500. Figure 2 demonstrates the SDE plot of 884 measured pidan objects based on the raw NIR spectra. According to the  $3\sigma$  rule, an object with SDE value over 3 can be detected as an outlier. Eight outliers are detected and excluded from calibration and prediction.

The DUEPLX algorithm [24] is then performed on the raw NIR spectra of the remaining 876 objects to form representative data sets. 657 objects are used for training and 219 objects for prediction. To compare the performance of different data preprocessing methods on calibration, all

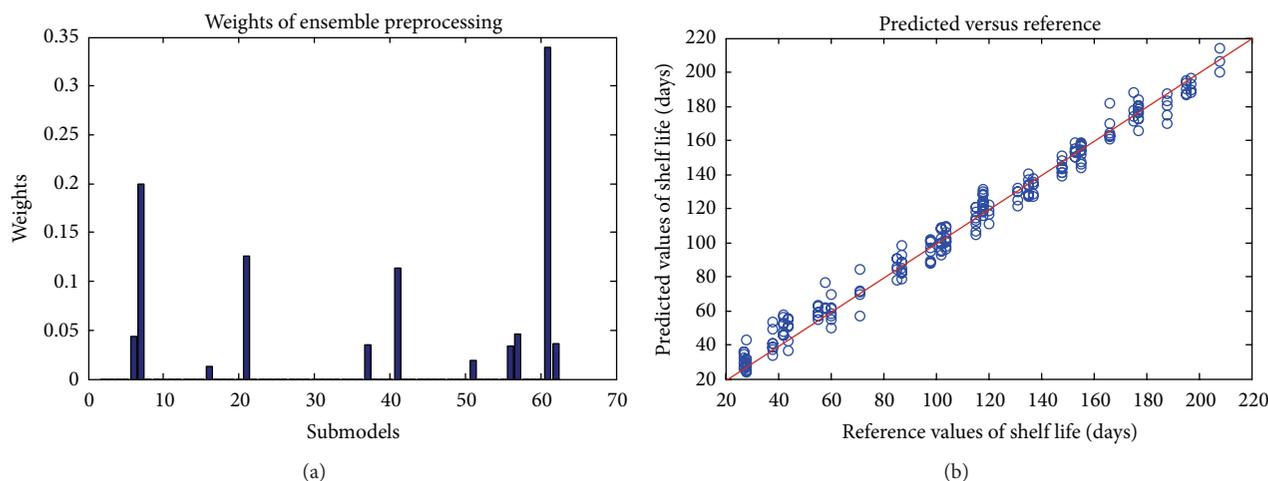


FIGURE 3: Submodel weights and predicted results obtained by stacking LS-SVM with ensemble preprocessing.

the models are trained and evaluated with the same training and test sets.

The parameters of LS-SVM submodels are optimized with simplex method to minimize the mean squared errors (MSE) of 20-fold cross-validation. The weights of submodels are estimated by MCCV stacked regression as in (5). Considering that the size of training data set is large ( $n = 657$ ), for MCCV, each time, a comparatively high percent (40%) of the training samples are left out for prediction, and the number of resampling is 50. With different preprocessing methods, LS-SVM and stacked models are developed, and the results are demonstrated in Table 2. The computation time for stacked LS-SVM with ensemble preprocessing is about 2 hours on our computer (Intel Core 2 Duo CPU E7500 @ 2.93 GHz, 1.96 GB memory). Considering the size of the original data (876 by 2074), the computational time is acceptable.

For submodels, submodel 13 in Table 2 has an abnormally large root mean squared errors of MCCV (RMSEMCCV) and root mean squared errors of prediction (RMSEP). This might be attributed to the underfitting caused by a large number of data points (9) and a low polynomial order (2). It can be seen that, with single preprocessing, the best predictions with RMSEP of 7.3 days are obtained by smoothing with 15 points and third-order polynomial. All the single-preprocessing methods do not significantly improve the prediction accuracy compared with the raw data (RMSEP = 7.6 days).

The ensemble preprocessing method obtains a reduced RMSEP of 6.2 days. The results of ensemble model are demonstrated in Figure 3. As seen from Figure 3, the distribution of submodel weights indicates that the ensemble model includes all types of single-preprocessing methods, such as smoothing, D1, D2, smoothing + SNV, D1 + SNV, and D2 + SNV. The results demonstrate that ensemble model can optimize data preprocessing by taking advantage of different preprocessing methods. Though, for all the models, RMSEMCCV is higher than RMSEP, RMSEMCCV and RMSEP show a similar tendency. Therefore, by minimizing RMSEMCCV, the ensemble model can obtain stable predictions of test objects.

## 4. Conclusions

Stacked LS-SVM regression with ensemble preprocessing is proposed as a data preprocessing and nonlinear calibration method. For NIR analysis of pidan shelf life, this method can automatically combine and weight a set of 62 single-preprocessing methods. The prediction accuracy was improved compared with the results using raw spectra and other single-preprocessing methods, including smoothing and taking derivatives, SNV, and their combinations. Since LS-SVM is much faster than its counterparts, the proposed stacked LS-SVM algorithm can be performed within an acceptable computational time. When the objects are very complex and it is necessary to reduce unwanted variations and the risk of overfitting, this method can provide a useful tool for data preprocessing and nonlinear multivariate calibration.

## Acknowledgments

This work is financially supported by the National Public Welfare Industry Projects of China (nos. 201210010, 201210092, and 2012104019), the National Natural Science Foundation of China (no. 31000357), Hangzhou Programs for Agricultural Science and Technology Development (no. 20101032B28), and the Key Scientific and Technological Innovation Team Program of Zhejiang Province (no. 2010R50028).

## References

- [1] J. Wang and D. Y. C. Fung, "Alkaline-fermented foods: a review with emphasis on pidan fermentation," *Critical Reviews in Microbiology*, vol. 22, no. 2, pp. 101–138, 1996.
- [2] J. Zhou, X. X. Xiang, and X. J. Zhang, "Etiology analysis of a case of food poisoning caused by pidan," *Chinese Journal of Health Laboratory Technology*, vol. 20, p. 422, 2010.
- [3] A. M. K. Pedro and M. M. C. Ferreira, "Multivariate accelerated shelf-life testing: a novel approach for determining the shelf-life of foods," *Journal of Chemometrics*, vol. 20, no. 1-2, pp. 76–83, 2006.

- [4] L. Xu, P. T. Shi, X. S. Fu et al., "Protected geographical indication identification of a chinese green tea (anji-white) by near-infrared spectroscopy and chemometric class modeling techniques," *Journal of Spectroscopy*, vol. 2013, Article ID 501924, 8 pages, 2013.
- [5] S. Curteanu and H. Cartwright, "Neural networks applied in chemistry. I. Determination of the optimal topology of multilayer perceptron neural networks," *Journal of Chemometrics*, vol. 25, no. 10, pp. 527–549, 2011.
- [6] B. Walczak and D. L. Massart, "The radial basis functions—partial least squares approach as a flexible non-linear regression technique," *Analytica Chimica Acta*, vol. 331, no. 3, pp. 177–185, 1996.
- [7] T. Czekaj, W. Wu, and B. Walczak, "About kernel latent variable approaches and SVM," *Journal of Chemometrics*, vol. 19, no. 5–7, pp. 341–354, 2005.
- [8] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1999.
- [9] H. Li, Y. Liang, and Q. Xu, "Support vector machines and its applications in chemistry," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 2, pp. 188–198, 2009.
- [10] R. de Maesschalck, F. Estienne, J. Verdú-Andrés et al., "The development of calibration models for spectroscopic data using principal component regression," *International Journal of Chemistry*, vol. 2, no. 19, 1999.
- [11] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [12] L. Xu, J.-H. Jiang, Y.-P. Zhou, H.-L. Wu, G.-L. Shen, and R.-Q. Yu, "MCCV stacked regression for model combination and fast spectral interval selection in multivariate calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 87, no. 2, pp. 226–230, 2007.
- [13] L. Xu, Y.-P. Zhou, L.-J. Tang et al., "Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration," *Analytica Chimica Acta*, vol. 616, no. 2, pp. 138–143, 2008.
- [14] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.
- [15] D. Wolpert, "A mathematical theory of generalization: part I," *Complex System*, vol. 4, no. 2, pp. 151–200, 1990.
- [16] D. Wolpert, "A mathematical theory of generalization: part II," *Complex System*, vol. 4, no. 2, pp. 200–249, 1990.
- [17] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, no. 1, pp. 49–64, 1996.
- [18] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [19] J. A. K. Suykens, T. van Gestel, J. de Brabanter, B. de Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.
- [20] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [21] Q.-S. Xu and Y.-Z. Liang, "Monte Carlo cross validation," *Chemometrics and Intelligent Laboratory Systems*, vol. 56, no. 1, pp. 1–11, 2001.
- [22] W. A. Stahel, *Robuste schätzungen: infinitesimale optimalität und schätzungen von kovarianzmatrizen [Ph.D. thesis]*, ETH Zürich, 1981.
- [23] D. L. Donoho, *Breakdown properties of multivariate location estimators [Ph.D. thesis]*, Harvard University, Boston, Mass, USA, 1982.
- [24] R. D. Snedecor, "Validation of regression models: methods and examples," *Technometrics*, vol. 19, no. 4, pp. 415–428, 1977.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

