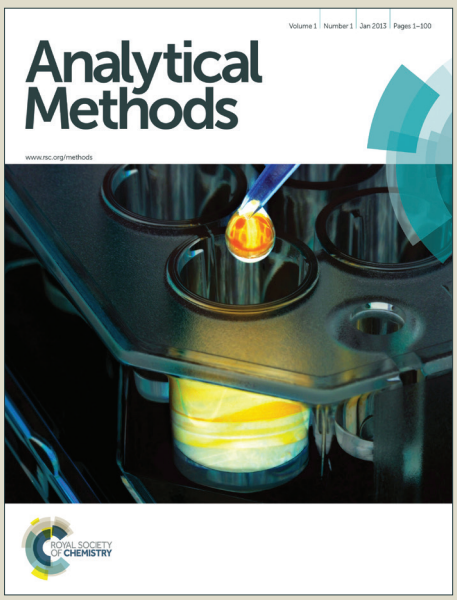
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Analytical Methods RSC Publishing

PAPER

Cite this: DOI: 10.1039/x0xx00000x

Adaptive Wavelet Packet Transform for Support Vector Machine Modeling as Globally Optimized by Particle Swarm Optimization Algorithm

Jin-Wen Zhang, Ji-Dong Wu, Li-Juan Tang*, Jian-Hui Jiang, Guo-Li Shen, Ru-Qin Yu*

DOI: 10.1039/x0xx00000x

www.rsc.org/methods

Multivariate calibration algorithms and statistical methods are typically required in quantitative near infrared (NIR) spectroscopic methods to relate NIR spectral response to chemical or physical properties of the samples. Based on two different version of particle swarm optimization (PSO) algorithms, an adaptively configured wavelet packet transform (WPT) based support vector machine (SVM) was developed and applied to quantitative near-infrared (NIR) spectra data analysis. WPT has been proved to be an effective method in extracting feature information from raw NIR spectra. It is also useful in noise suppression and data compression. SVM is famous for its ability of function approximation and remarkable generalization performance. For SVM modeling on the basis of WPT, discrete PSO and continuous PSO were used synchronously to optimize the structure of a WPT tree and the parameters of SVM synergistically according to the performance of the total model. It, thus, enables an adaptive and parameter-free model construction technology for NIR spectral data analysis. The performance of the proposed Ad-WPT-SVM is investigated using two real data sets. The results of different methods are compared, indicating the proposed method holds great potential for robust and reliable quantitative NIR spectral data analysis.

.

1. Introduction

Near-infrared (NIR) spectroscopy is an invaluable tool which provides cheap and rapid estimation of sample components in diverse fields like pharmaceutical industry[1-3], medical diagnostics (including blood sugar and oximetry)[4] food and agrochemical quality control[5] and so on. Commonly, regression modeling is necessary in quantitative NIR spectral analysis[6]. Nevertheless, it is not easy to reveal the concealed quantitative relationship between spectral information and sample properties because in most NIR spectra broad, weak, non-specific and overlapping bands may result in significant noises, serious non-linearity or poor relevant information. An effective way to solve this problem is variable selection or information extraction. Some methods including stepwise regression analysis (SRA)[7], uninformative variables elimination (UVE)[8-10], interval partial least squares (iPLS)[11,12] and moving window partial least squares regression (MWPLSR)[13-15] were developed to extract parts of relevant wavelengths or informative wavelength intervals for modeling. Variable-weighting was also suggested to magnify the utility of the relevant wavelengths and restrain the impact of irrelevant ones[16]. However, it is still a great challenge in quantitative analysis of NIR.

Discrete wavelet transform (DWT) is a great useful time-frequency analysis tool with many successful applications in NIR analysis[17], such as describing and clarifying local data structures, suppressing noise, correcting baseline, compressing database, et al. What's more, wavelet packet transform (WPT), the generalization of DWT, has been proved to be even more versatile[18].

State Key Laboratory of Chemo/Biosensing and Chemometrics, Laboratory of Tobacco Chemistry, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, P. R. China. Tel.: 86-731-88664085; Fax.: 86-731-88821916, E-mail addresses: rqyu@hnu.edu.cn (R. Q. Yu), tanglijuan@hnu.edu.cn(L. J. Tang).

Rather than DWT only decomposing approximation coefficients each time, WPT continuously decomposes detail coefficients as well by passing a signal through more filters than DWT[19], which makes WPT more flexible for the analysis of smooth signals typically like NIR spectra. Nevertheless, because continuously decomposing of approximation and detail coefficients is enabled in WPT, it will result in different wavelet packet trees. Different wavelet packet trees determine different combinations of coefficients for original signals and the combinations of coefficients impact the performance of the quantitative model straightforwardly, therefore, it is critical to find out the best decomposition tree in terms of a convenient criterion.

To solve this problem, herein, it is developed that adaptive wavelet packet transform followed by support vector machine (Ad-WPT-SVM) modeling as globally optimized by particle swarm optimization (PSO) algorithm for NIR spectral data analysis. In this method, the introduction of global stochastic optimization technique PSO allows flexible wavelet packet tree construction according to the performance of the total quantitative model, compared with a commonly used approach constructing a decomposition tree just via analyzing the signal itself based on Shannon entropy. The problem of wavelet packet decomposition tree construction can be processed as a discrete optimization issue using a discrete version of PSO. Besides using a discrete version of PSO to seek the optimal wavelet packet tree for a WPT-SVM model in NIR spectra analysis, PSO also can be used to search the other parameters of the model, such as penalty constant, kernel width in kernel transform[20,21]. The direct benefit is that it makes WPT-SVM to be an adaptive parameter-free method for NIR spectra analysis, without any parameters to be adjusted. To demonstrate the performance of the developed Ad-WPT-SVM, two real NIR spectra data sets were used as case studies. We also compared the performance of the developed method with that of WPT-SVM models whose decomposition trees were determined using a decision rule of Shannon entropy and other parameters were estimated by gird search.

2. Theory

2.1 Wavelet packet and wavelet packet tree

The idea of wavelet packet transform is the same as wavelet transform. In wavelets, via using a set of basis functions any arbitrary signal is decomposed into a series of coefficients (approximation coefficients and detail coefficients) in a new time-frequency space to exhibit some features in the new feature space that are inconspicuous in the initial domain. The difference of wavelet packet from wavelets is detail coefficients and approximation coefficients are decomposed continuously rather than only approximation coefficients further decomposed in wavelets. In other words, wavelet packet transform provides appropriate time resolution as well as frequency resolution simultaneously, which makes WPT more flexible to balance the time resolution and frequency resolution than DWT. Because of this difference, WPT accordingly describes the original signal better and more exactly when treating smooth signals. The decomposition procedure of WPT from a standpoint of signal processing is shown in Figure 1. An original signal, on the first decomposition level, is first decomposed into two parts, lower frequencies and higher frequencies. Though this, the information in lower frequency part is represented by approximation coefficients (A1), while that in higher frequencies is depicted by detail coefficients (D1). Then, the part of higher frequencies as well as the part of lower frequencies is decomposed respectively on the second decomposition level. This procedure can be carried out up to the maximum decomposition level.

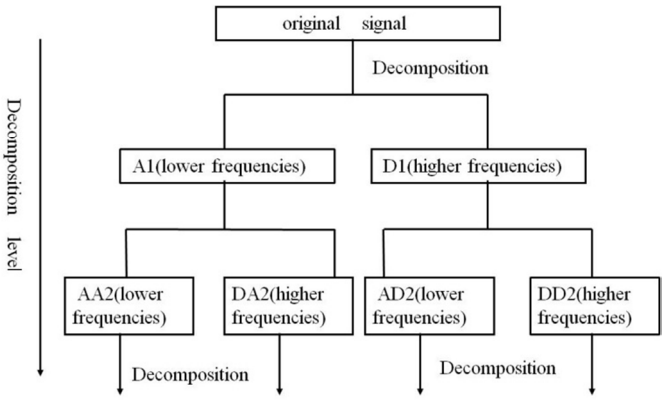


Figure 1. wavelet packet transform(WPT) decomposition.

In the view of information, the information contained in the original signal is equal to A1 + D1, A1 + AD2 + DD2, AA2 + DA2 + D1, or AA2 + DA2 + AD2 + DD2 in Figure 1. That is, an original signal could be decomposed into different ways through WPT. While the final tree adopted determines the coefficients which would be used for modeling and influence the performance of regression model directly, it is extremely critical to find out the optimal decomposition mode (wavelet packet tree) in terms of a convenient criterion. In general, Shannon entropy that describes information-related property of the signal is chosen as the criterion. The best tree is considered as the one with minimum entropy, or in other words, the one whose regularity or information approaches the maximum. This criterion works well in many applications. However, it is obvious a Shannon entropy criterion decides a tree independent of the final quantitative model. Thus, it is difficult for Shannon entropy criterion to provide desirable decomposition tree according to the performance of the total model. In the present study, a new criterion based on a modified PSO is proposed for flexible and convenient tree structural design in consideration of the performance of the total quantitative model built using SVM technique.

2.2 Support Vector Machine

Support vector machine (SVM)[22] is gaining popularity due to many attractive features and promising empirical performance. Consider the problem of approximating the set of data with a linear function,

$$\mathbf{y}=\mathbf{w}^{\mathrm{T}}\mathbf{X}+b \quad (1)$$

where **w** is the weight vector to be identified in the function, and b is the threshold. Based on structural risk minimization (SRM) principle, the optimal regression function is given by the minimum of the cost function Φ,

$$\Phi=\frac{1}{2}\mathbf{W}^{\mathrm{T}}\mathbf{W}+\mathrm{C}\frac{1}{I}\sum_{i=1}^{I}L_{\varepsilon}\ (y_i\ -\ y_{0i}) \quad (2)$$

where,

$$L_{\varepsilon}\ (y_i-y_{0i})=\begin{cases}|\ y_i\ -\ y_{0i}\ |-\ \varepsilon & |y_i\ -y_{0i}|\geq\varepsilon\\ 0 & \text{otherwise}\end{cases} \quad (3)$$

is the ε-insensitive loss function measuring the error between the given observations ($\mathbf{y}_o$) and the estimated ones (**y**) and ε is the tolerance zone. I is the number of the training compounds; $1/2\mathbf{w}^{\mathrm{T}}\mathbf{w}$ is used as a measure of the model complexity, defining the structure risk of a SVM model. A penalty constant C is introduced to determine the trade-off between the empirical error and the model complexity. Minimizing the cost function Φ under penalty constant C is to reduce the complexity as well as the empirical error of a model. As defined above, the w and b in Eq. (1) can be obtained by solving a quadratic programming problem. Extension of this linear technique to nonlinear regression can be performed in a straightforward manner by substituting a so-called kernel function K(xi, x) for the inner product <xi, x>. Although many functions can be used as the kernel function, Gaussian radial basis function transform is frequently utilized if the knowledge of a problem dealt with is lacking, K(xi, x)=exp(-||xi - x||2/(2σ2)). To determine the parameters in a SVM model, such as penalty constant C, tolerance zone ε and kernel width σ in Gaussian function transform, a continuous version of PSO is considered.

2.3 Particle Swarm Optimization Algorithms

PSO[23,24]is an evolutionary computation technique derived from simulating the behavior of birds searching food. In PSO, the potential solutions called particles fly through the problem space by following the current optimum particles. Each particle keeps track of its coordinate in the problem space which is associated with the best solution (fitness) it has achieved so far. This value is called personal best position (pBest) for particle i represented as **p**i = (pi1, pi2, …,piD). Another best value that is tracked by the particle swarm optimizer is the best value obtained so far by all particles in the solution space, called global best position (gBest) which is represented as **p**g = (pg1, pg2, …, pgD).Each particle updates its velocity **v**i = (vi1, vi2, …,viD) and position **x**i = (xi1, xi2, …, xiD) by tracking these two best values according to the following equations:

$$v_{id}(\text{new}) = w \times v_{id}(\text{old}) + c_1 \times r_1 \times (p_{id}- x_{id}) + c_2 \times r_2 \times (p_{gd}- x_{id}) \quad (4)$$

$$x_{id}(\text{new}) = x_{id}(\text{old}) + \mu \times v_{id}(\text{new}) \quad (5)$$

where w is an inertia weight which is brought into Eq. (4) to balance the global search and local search, r1 and r2 are random numbers between 0 and 1. Two positive constants, c1 and c2, called learning factors are introduced, and generally both take the integer value 2. In Eq. (5), μ is the time parameter determining the different flying time for each particle. The particle swarm optimization concept consists of, at each time step, changing the velocity of each particle toward its *p*Best and *g*Best locations. Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *p*Best and *g*Best locations.

Mostly, PSO is carried out in a continuous real-number space as described above, so-called a continuous version of PSO. For discrete optimization issues which can be expressed in an integer string varying from 0 to k, a discrete version of PSO is developed by Yu et al. Such PSO is still on the basis of the information-sharing mechanism of the continuous PSO and the pattern of updating particle by following two best positions. In a different way, a particle of discrete PSO represents changes of its site that should be an integer, and the velocity represents the probability of site x_{id} taking the integer value from 1 to k respectively. The velocity v_{id} of every individual is a random number in the range of (0,1). The resulting change in position is then defined by the following rule,

If $(0<v_{id}\leq a)$, then $x_{id}(\text{new}) = x_{id}(\text{old})$ (6)

If $(a<v_{id}\leq(1+a)/2)$, then $x_{id}(\text{new}) = p_{id}$ (7)

If $((1+a)/2<v_{id}\leq 1)$, then $x_{id}(\text{new}) = p_{gd}$ (8)

where a is a random value in the range of (0,1) named static probability. Static probability a starts with a value of 0.5and decreases to 0.33 when the iteration terminates. Though the velocity in the modified discrete PSO is different from that in continuous PSO, an information sharing mechanism and updating model of particle by following the two best positions is the same in the two versions of PSO.

2.4 WPT-SVM as Globally Optimized by PSO (Ad-WPT-SVM)

The WPT analysis provides us the feature coefficient spaces of NIR spectra, where we can relate the quantitative property with the NIR spectra using a modeling technique of SVM. To build a WPT-SVM model which can display good performances both in model training and prediction, we should provide optimal decomposition tree for WPT and optimal parameters for SVM first. Rather than using a strategy that estimates decomposition tree and SVM parameters separately, we optimize the WPT-SVM quantitative model globally via invoking PSO algorithms. The problem of constructing the wavelet packet decomposition tree actually can be solve as a discrete optimization issue by using a discrete version of PSO, and on the other hand the parameter determination of SVM can be treated as a continuous optimization issue by using a continuous version of PSO.

While building a WPT-SVM model in NIR spectra analysis, as two different versions of PSO are simultaneously employed, each particle in the solution space, accordingly, consists two parts: one is encoded as a string of integers representing the construction of the wavelet packet tree and the other is encoded as a string of continuous real numbers representing the parameters of SVM.

When using discrete integers to encode the wavelet packet decomposition tree, it is treated as a completely symmetrical tree for some conveniences, shown in Figure 2. For a full symmetric tree with L levels, the number of nodes of the tree can be easily measured as $\sum_{l=1}^{L}2^{l-1}$, and for any node with child nodes, the sequence number of its two child nodes would be twice and twice plus 1 as that of itself (named parent node).

Because of these attributes of a completely symmetrical tree, it will be extremely convenient to coding and decoding particles in a discrete version of PSO. To represent the tree structure, the discrete part of each particle is encoded as a binary bit with a length of $\sum_{l=1}^{L} 2^{l-1}$ for a tree with L levels, indicating whether the corresponding node exists or not. A bit of 0 in this part implies that the corresponding node does not exist at all. Of course, its descendent nodes are automatically excluded in the final decomposition tree. This also means that the wavelet packet coefficients associated with these nodes are useless, and vice versa, a bit of 1 indicates the existence of the associated node. According to the information provided by the corresponding wavelet packet decomposition tree, the wavelet packet coefficients for these nodes also will be reserved. By restoring the tree via decoding the optimal particle, the coefficients of all the bottom nodes of this optimal wavelet packet decomposition tree should be kept to build the optimal quantitative model. For enhancing the optimization efficiency, the first three bits in this part of a particle are always set to be 1 to guarantee the decomposition of the first level of a signal.

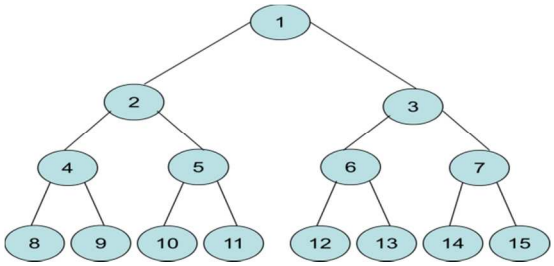


Figure 2. A full symmetric tree with 4 levels.

The second part coding as continuous real numbers of each particle is made up of three subparts representing the parameters of SVM, the penalty constant C, the tolerance zone ε and the kernel width σ in a Gaussian function transform, respectively. In addition, for further improving the ability of the PSO to overcome local optima, ten percent of particles are randomly selected and forced to fly at random, not following the two best positions.

Simultaneously employing the two versions of PSO to synchronously optimizing the particles makes the construction of decomposition tree and model parameter selection to be an entirety. Moreover, as PSO determines the optimal decomposition tree in accordance with the model performance, the noise in the NIR spectra (usually related to the detail coefficients of the highest frequency) will be automatically suppressed corresponding to the modeling requirements via giving the related node to be 1 or 0 in the tree optimizing process. Therefore, other noise suppression method is not required any more. In short, using the two versions of PSO for WPT-SVM enables the construction of an adaptive parameter-free quantitative model for NIR spectra analysis according to the performance of the total model.

To measure the particles of PSO in the solution space of WPT-SVM, an objective function also should be designed. Since the present method is oriented to improve the performance of a quantitative model, it is considered to both enhance the training of a calibration set and the prediction of an independent validation set by minimizing the following objective function:

$$\mathrm{Re}=\sqrt{\frac{\mathrm{RSSC}+\mathrm{RSSV}}{I+I_v}} \quad (9)$$

where RSSC is the sum of squared residual of the original calibration set and RSSV the sum of squared residual of the independent validation set, I and I_v are the size of calibration set and validation set, respectively.According to this objective function, the two versions of PSO optimize the model simultaneously, until the minimum error criterion is attained or the number of cycles reaches a user defined limit.

3. Data sets

3.1 Meat data

The data containing 240 samples of meat are recorded on a Tecator Infratec Food and feed analyzer working in the wavelength range 850–1050 nm by the near infrared transmission (NIT) principle[25]. Each sample contains finely chopped pure meat with different contents of moisture, fat and protein. For quantitative model building and evaluating, the total 240 samples are split into a calibration set with 120 samples, a validation set with 60 samples and a prediction set with 60 samples using a DUPLEX method[26].

3.2 Corn data

The corn data consists of 80 samples measured on NIR spectrometers with a wavelength range of 1100-2498nm recorded[27]. The protein and moisture contents for each sample are included. According to DUPLEX, the whole samples are divided into a calibration set, a validation set and a prediction set with 40, 20 and 20 samples, respectively.

4. Results and discussion

4.1 Meat data

The performance of Ad-WPT-SVM was firstly examined by modeling the protein, fat and moisture of meat data. As comparisons, partial least square (PLS), back propagation neural network (BPNN), and WPT-SVM using Shannon entropy criterion for tree structure determination were also employed to build the quantitative models. Table 1 summarizes the results of all the four modeling methods for protein, fat and moisture content prediction in meat.

Table 1. The RMSE value of each model for the content analysis of protein, fat and moisture in meat.

	Protein		Fat		Moisture	
	Cal set	Pre set	Cal set	Pre set	Cal set	Pre set
PLS	0.9754	0.9573	2.2376	2.0119	1.3372	1.4233
BPNN	1.6337	1.4486	4.1960	4.0982	2.6545	2.8042
WPT-SVM	0.8107	0.8673	1.4333	1.5226	1.1589	1.2245
Ad-WPT-SVM	0.6724	0.7566	0.3820	0.4918	0.7491	0.8128

note: RMSE represents for root mean square error; Cal set represents for calibration set and Pre set for prediction set.

For PLS models, the number of latent variables is determined using the prediction performance of calibration and validation sets. The root mean square error (RMSE) of PLS in protein content analysis was 0.9754 for the calibration set and was 0.9573 for the prediction set. The correlation between the calculated and observed values of protein content is shown in Figure 3a. For BPNN, 20 hidden nodes were used for each model, and validation set was used to estimate the proper iteration number for reducing the risk of over fitting or under fitting of the models. However, the results of BPNN were poorer than those obtained by PLS, as shown in Table 1 and Figure 3b. The RMSE was 1.6337 for the calibration set and 1.4486 for the prediction set, both much larger than those of PLS. WPT-SVM, of which the decomposition tree was determined using a decision rule of Shannon entropy and the parameters of SVM were estimated by gird search, gave RMSEs of 0.8107 and 0.8673 for the calibration set and prediction set, respectively. The correlation between the calculated and observed values of content was shown in Figure 3c. Such results were better than those of PLS and BPNN. This improvement may benefit from the feature extraction of WPT and the modeling technique of SVM.

To be an adaptive modeling method, the best decomposition tree and whole of the parameters of SVM including the variables weights, kernel width σ, penalty constant C and tolerance zone ε were optimized by discrete and continuous versions of PSO algorithms via minimizing the objective function defined in Eq. (7). Parameter optimization of multivariate calibration aims at finding the best available values of a function or a set of functions, and automatic parameter optimizing is able to provide an objective solution to a problem. The parameters optimized by PSO for both WPT and SVM were list in Table S1 in supporting information. With the optimal decomposition tree, RMSE for calibration set of 0.6724 was obtained by Ad-WPT-SVM and that for the prediction set was 0.7566. The correlation between the calculated and experimental values is shown in Figure 3d.

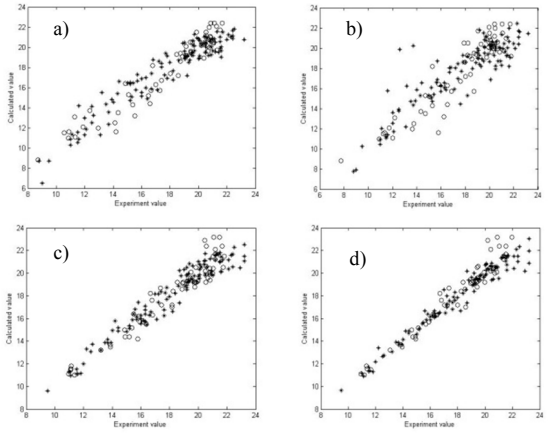


Figure 3. Correlation between the calculated values acquired from each model and the observed values of the protein in meat: (a) PLS, (b) BPNN, (c) WPT-SVM, (d) Ad-WPT-SVM.

Compared with WPT-SVM, Ad-WPT-SVM provided further enhanced performance for both the calibration set and the prediction set which displayed as much smaller RMSEs and better correlation between the calculated and experimental values, exhibiting that the proposed algorithm had better precision in modeling and superior generalization in prediction. Additional, Ad-WPT-SVM also exhibited the best performance in fat and moisture content prediction of meat data set as shown in Table 1. The *R*-square scores of all the models were also summarized in Table S2 in supporting information. As expected, the proposed Ad-WPT-SVM had the best *R*-square scores compared with the models built using other methods. It can be concluded that since the structure of the decomposition tree and the parameters were optimized synergistically using the PSO algorithm according to the performance of total model, it guaranteed Ad-WPT-SVM to be an automatic method with desirable performance in quantitative analysis of NIR spectral data.

4.2 Corn data

The same modelling strategies, PLS, BPNN, WPT-SVM with Shannon entropy criterion and Ad-WPT-SVM were used for the corn data to fit the contents of protein and moisture. The statistical results were included in Table 2.

Table 2. The RMSE value of each model for the content analysis of protein and moisture in corn.

	Protein		Moisture	
	Cal set	Pre set	Cal set	Pre set
PLS	0.0582	0.0678	0.1021	0.0900
BPNN	0.1520	0.1486	0.1468	0.1587
WPT-SVM	0.0152	0.0148	0.0740	0.0983
Ad-WPT-SVM	0.0066	0.0081	0.0067	0.0060

note: RMSE represents for root mean square error; Cal set represents for calibration set and Pre set for prediction set.

For the quantitative analysis of protein in corn, it could be found that PLS gave a RMSE of 0.0582 for the calibration set and a RMSE of 0.0678 for the prediction set. BPNN showed worse results than PLS, a RMSE of 0.1520 for the calibration set and that of 0.1496 for the prediction set. Such comparison is clearly exhibited in Figure 4a and 4b. Improved results were obtained using WPT-SVM, of which the model used Shannon entropy as the criterion for decomposition tree decision. A RMSE of 0.0152 for the calibration set and that of 0.0148 for the prediction set were obtained. Ad-WPT-SVM displayed even better performance both for modeling and prediction with the PSO optimized parameters listing in Table S1 in supporting information. Low RMSEs for the calibration set and prediction set were achieved with the values of 0.0066 and 0.0081, respectively. The correlation between the calculated and experimental values of protein contents is exhibited in Figure 4d, which is correspond to the results in Table 2. Compared with WPT-SVM with Shannon entropy as shown in Figure 4c, Ad-WPT-SVM improved the performance of the quantitative model further. Also, Ad-WPT-SVM exhibited the best performance in moisture content analysis with RMSE of 0.0067 for calibration set and 0.0060 for prediction set, both lower than those of PLS, BPNN and WPT-SVM. Table S3 in supporting information shows the *R*-square scores of all the models. These results further demonstrated that optimizing synergistically the structure of WPT and SVM parameters improved the performance of adaptive WPT-SVM, then exhibiting a great potential in NIR spectral quantitative analysis.

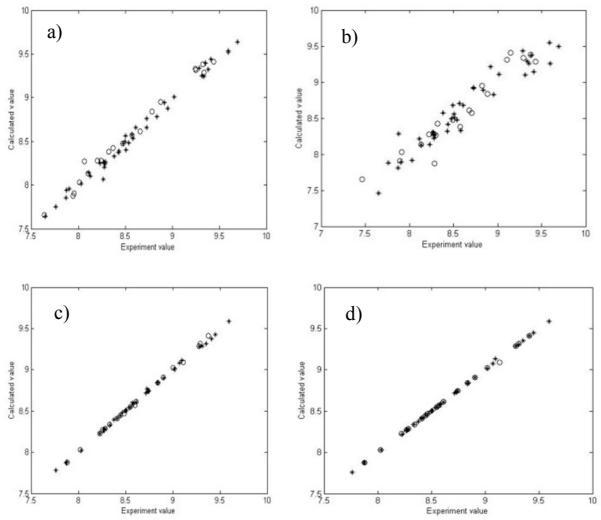


Figure 4. correlation between the calculated values acquired from each model based on NIR data and observed values of the protein in corn: (a) PLS, (b) BPNN, (c) WPT-SVM, (d) Ad-WPT-SVM.

5. Conclusion

In this paper, it was proposed adaptive wavelet packet transform for support vector machine modeling as globally optimized by particle swarm optimization for NIR spectral data analysis. Wavelet packet transform is very effective in information extraction especially for smooth singles like NIR spectra, and the use of PSO enabled synergistic optimization of the WPT structure and all parameters of SVM according to the performance of the total model. The performance of Ad-WPT-SVM was evaluated by using two NIR data sets. The results revealed that the proposed method offered enhanced performances both in model training and prediction due to the option of synergetic optimization of WPT-SVM according to the performance of the total model via synchronously using discrete and continuous versions of PSO. Therefore, it may provide a promising tool for flexible and robust NIR spectra analysis.

Acknowledgments

This work was supported by NSFC (21205034, 91317312, 21190041, 21405041, 21307029), Doctoral Fund of Ministry of Education of China (New Teachers, 20120161120032), Hunan Provincial NSFC (13JJ4031), Fundamental Research Funds for the Central Universities and Young Scholar Support Program of Hunan University.

Notes and references

1 J. Luypaert, D. L. Massart, and Y. V. Heyden, *Talanta.,* 2007, **72**, 865-883.
2 C. Gendrin, Y. Roggo, and C. Collet, *J. Pharm. Biomed. Anal*., 2008, **48**, 533-553.
3 S. Sasic, D. Blackwood, A. Liu, H. W. Ward, and H. Clarke, *J . Pharm. Biome . Anal*., 2015, **103**, 73-79.
4 A. C. Ehlis, S. Schneider, T. Dresler, and A. J. Fallgatter, *NeruoImage* ., 2014, **85**, 478-488.
5 D. Wu, J. Y. Chen, B. Y. Lu, L. Xiong, Y. He, and Y. Zhang, *Food. Chem*., 2012, **135**, 2147-2156.
6 A. Durand, O. Devos, C. Ruckebusch, and J. P. Huvenne, *Anal. Chim. Acta.*, 2007, **595**, 72-79.
7 S. E. Pratsinis, M. D. Zeldin, and E. C. Ellis, *Environ. Sci. Technol*., 1988, **22**, 212-216.
8 W. S. Cai, Y. K. Li, and X. G. Shao, *Chemom. Intell. Lab. Syst*., 2008, **90**, 188-194.
9 X. G. Shao, F. Wang, D. Chen, and Q. D. Su, *Anal. Bioanal. Chem.,* 2004, **378**, 1382-1387.
10 Q. J. Han, H. L. Wu, C. B. Cai, L. Xu, and R. Q. Yu, *Anal. Chim. Acta.,* 2008, **612**, 121-125.
11 L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck and S. B. Engelsen, *Appl. Spectrosc*., 2000, **54**, 413-419.
12 R. Leardi, and L. Norgaard, *Chemom*., 2004, **18**, 486-497.
13 J. H. Jiang, R. J. Berry, H. W. Siesler, and Y. Ozaki, *Anal. Chem*., 2002, **74**, 3555-3565.
14 S. J. Qin, *Comput. Chem. Eng*., 1998, **22**, 503-514.
15 Y. P. Du, Y. Z. Liang, J. H. Jiang, and R. J. Berry, *Anal. Chim. Acta.,* 2004, **501**, 183-191.
16 A. Rinnan, M. Andersson, C. Ridder, and S. B. Engelsen, *J. Chemom*., 2014, **28**, 439-447.
17 J. Trygg, and S. Wold, *Chemom. Intell. Lab. Syst*., 1998, **42**, 209-220.
18 O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, *Seizure*., 2015, **26**, 56-64.

19 B. Walczak, B. V. D. Bogaert, and D. L. Massart, *Anal. Chem*., 1996, **68**, 1742-1747.
20 Q. Shen, J. H. Jiang, and C. X. Jiao, *Eur. J. Pharm. Sci*., 2004, **22**, 145-152.
21 D. K. Agrafiotis, and W. Cedeno, *J. Med. Chem*., 2002, **45**, 1098-1107.
22 P. Rebentrost, M. Mohseni, and S. Lioyd, *Phys. Rev. Lett*., 2014, **113**,13-26.
23 L. Y. Chuang, H. W. Chang, C. J. Tu, and C. H. Yang, *Comput. Biol. Chem*., 2008, **32**, 29-38.
24 Y. Q. Luo, X. G. Yuan, and Y. J. Liu, *Comput. Chem. Eng*., 2007, **31**, 153-162.
25 P. Rios, and J. A. Stuart, *Environ. Sci. Technol*., 2003, **37**, 5463-5470.
26 J. J. Kelly, and C. H. Barlow, *Anal. Chem*., 1989, **61**, 313-320.
27 A. Boggess, and F. J. Narcowich, *Phys. Today*., 2002, **55**, 63-70.