Article

Variational Mode Decomposition Weighted Multiscale Support Vector Regression for Spectral Determination of Rapeseed Oil and Rhizoma Alpiniae Offcinarum Adulterants

Xihui Bian [1,2,3,*], **Deyun Wu** [1], **Kui Zhang** [1], **Peng Liu** [1], **Huibing Shi** [3], **Xiaoyao Tan** [1] **and Zhigang Wang** [1]

1 State Key Laboratory of Separation Membranes and Membrane Processes,
   School of Chemical Engineering and Technology, Tiangong University, Tianjin 300387, China;
   wudeyun1221@163.com (D.W.); zhangkui@dicp.ac.cn (K.Z.); liupeng012@tiangong.edu.cn (P.L.);
   tanxiaoyao@tiangong.edu.cn (X.T.); wangzhigang@tiangong.edu.cn (Z.W.)
2 State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China
3 Shandong Provincial Key Laboratory of Olefin Catalysis and Polymerization,
   Shandong Chambroad Holding Group Co., Ltd., Binzhou 256500, China; huibing.shi@chambroad.com
* Correspondence: bianxihui@mail.nankai.edu.cn

Abstract: The accurate prediction of the model is essential for food and herb analysis. In order to exploit the abundance of information embedded in the frequency and time domains, a weighted multiscale support vector regression (SVR) method based on variational mode decomposition (VMD), namely VMD-WMSVR, was proposed for the ultraviolet-visible (UV-Vis) spectral determination of rapeseed oil adulterants and near-infrared (NIR) spectral quantification of rhizoma alpiniae offcinarum adulterants. In this method, each spectrum is decomposed into K discrete mode components by VMD first. The mode matrix Uk is recombined from the decomposed components, and then, the SVR is used to build sub-models between each Uk and target value. The final prediction is obtained by integrating the predictions of the sub-models by weighted average. The performance of the proposed method was tested with two spectral datasets of adulterated vegetable oils and herbs. Compared with the results from partial least squares (PLS) and SVR, VMD-WMSVR shows potential in model accuracy.

Keywords: variational mode decomposition; support vector regression; adulteration; quality control; chemometrics

1. Introduction

Quality control is a critical analytical topic, especially regarding foods and herbs that play an essential role in everyday life. Adulteration is one of the major challenges in the quality control of foods and herbs. Some unscrupulous vendors dilute them with cheap alternative food sources or fraudulently label and sell low-quality products as premium ones to make more profit [1]. These not only damage the quality and nutritional value of foods but also are detrimental to consumer health. With a series of food adulteration incidents per year that lead to severe health impacts and economic costs, the problem of food fraud has become more sinister and devastating in the globalized food supply chain [2–4]. An important index for evaluating the nutritional value and quality of foods and herbs is to determine the contents of the main components in them. However, it is difficult to determine adulterants in the final product as some foods or herbs are similar in appearance but vary greatly in price [5,6]. In order to protect consumer rights and ensure food safety, there is an urgent need to develop a simple and reliable method to meet the accurate quantitative analysis of food and herb adulteration.

Various techniques have been applied to detect adulteration in foods and herbs, such as chromatography, mass spectrometry and capillary electrophoresis. These methods allow for a relatively accurate determination of the samples [7,8]. However, most of them

are time-consuming and expensive and require a high degree of technical expertise [9]. Spectroscopic techniques, especially ultraviolet-visible (UV-Vis) and near-infrared (NIR) spectroscopy, have been rapidly developed in scientific research and industrial production because of their non-contact, environmentally friendly and low-cost advantages [10–13]. Since the original spectra usually contain a large amount of signal overlap, background and noise information unrelated to the target, chemometric models need to be integrated to improve and expand the potential applications of the spectroscopic techniques.

Chemometrics, as an effective support means, has been developed extensively in analytical chemistry [14–17], especially in multivariate calibration methods for spectral data analysis, such as partial least squares (PLS) and support vector regression (SVR) [18,19]. PLS is a commonly used modeling method because of its practicality and versatility, but it may produce undesirable prediction results when dealing with strongly nonlinear issues [18]. SVR has the capability to solve both linear and nonlinear multivariate regression problems with a simple process [20]. These modeling approaches predict unknown samples by constructing one model, but the prediction performance of only a single model that is built between spectra and targets tends to be poor when the training set is small or the samples are outliers [21].

Ensemble modeling has gained increasing attention in the multivariate calibration for quantitative analysis [22,23]. Compared with the prediction of a single model, ensemble modeling achieves a greater accuracy and more robust results by combining the predictions of multiple sub-models to produce the final prediction [21]. One of its key points is the generation of training sub-sets that can be produced from samples, variables or both directions, such as bagging, cluster and boosting [24–26]. However, most spectra are essentially localized and have varying localization in time and frequency. These traditional ensemble strategies are all generated sub-models from the original data that do not use both time and frequency information of the signal simultaneously [27]. Due to the complexity of the spectra, if the original signal is decomposed by mathematical transformation before ensemble calibration, better results may be obtained. There is different information hidden in the data that can be revealed by converting signals from the original data space to other spaces through a certain mathematical transformation.

Three decomposition strategies are widely applied for the signal process, that is, the Fourier transform (FT) [28], wavelet transform (WT) [29] and empirical mode decomposition (EMD) [30]. FT portrays well the frequency domain information of signals, but it does not provide time domain information and can only deal with stationary and linear signals [31]. WT has displayed its modeling effectiveness owing to its capacity for time-frequency resolutions. Nevertheless, WT is not a self-adaptive decomposition that needs to choose wavelet filters and scales for a given application to obtain an optimal result [32]. EMD is a useful technique for processing non-stationary and nonlinear signals and decomposes the signal into a finite number of intrinsic mode functions (IMFs) [30]. Although this self-adaptive decomposition method is a potent tool for the multiscale analysis of data without the trouble of selecting the filters or scales, the existence of mode mixing and end effect in the EMD process will lead to the distortion of IMF components [33]. Therefore, it is necessary to develop a new mathematical transformation for the signal process, which can make up for the deficiency in the above methods.

Variational mode decomposition (VMD) is a new adaptive signal decomposition strategy that is particularly suitable for nonlinear and non-stationary signals [34]. It not only has a good separation effect on the noise in signals but also effectively suppresses the mode mixing and end effect [35]. Using VMD, a series of mode components can be decomposed from the complex spectra according to the inherent characteristics of the signals. Previously, various studies reported that VMD has been used successfully in multiple fields due to its efficiency superiority, such as the forecast of stock prices [36], wind speed forecasting [37] and fault diagnosis [38]. However, there are very few reports in the literature that use the VMD algorithm for ensemble modeling in the spectral determination

of food. Since VMD can fully utilize the information embedded over the frequency and time domains of spectral signals, it was introduced in the generation of sub-models.

Herein, a weighted multiscale SVR modeling method based on VMD for improving the prediction accuracy of food and herb adulterants is proposed and referred to as VMD-WMSVR. Firstly, each spectral signal is decomposed by VMD and then *K* mode components with different central frequencies are obtained. After recombining these mode components, SVR is used to establish sub-models for each mode. Finally, the predictions of each sub-model are weighted and averaged to obtain the ultimate prediction result. The spectral datasets of adulterated vegetable oils and herbs were investigated using this method. The performance of the method was evaluated based on the root mean square error of prediction (RMSEP) and correlation coefficient (R) and compared with results derived from single PLS and SVR models.

2. Materials and Methods

2.1. Sample Preparation

For adulterated vegetable oils, the sample consists of six different vegetable oils bought in different markets in the municipality of Tianjin. These are sesame oil, soybean oil, corn oil, peanut oil, rapeseed oil and sunflower oil. The six pure oils were blended in different mass proportions in order to form 51 adulterated vegetable oil samples. Each oil content is within the range of 0–100% (g/g) with an interval of ca. 2%. Before measurement, these samples were well shaken and sonicated in an ultrasonic instrument (SK6200HP, Kudos Ultrasonic Instrument Company, Shanghai, China) for 30 min to further mix and eliminate air bubbles. In this study, rapeseed oil was taken as the analysis target.

For adulterated herbs, the sample includes pure herbs. These are *Panax notoginseng* (PN), rhizoma alpiniae offcinarum (RAO), rhizoma curcumae (RC) and *Curcuma longa* (CL) purchased from various pharmacies in Tianjin. Since the herbs have a certain amount of moisture, they were dried at 60 °C to a constant weight. These herbs were ground into powder, passed through a 120-mesh stainless steel sieve and stored in sealed plastic bags measuring 60 mm × 100 mm. The four processed herbs were mixed at different mass percentages and ensured that the total mass fraction of the four herbs in each sample was 100%. There were 75 samples in the adulterated herbs dataset and studied with the content of RAO.

2.2. Spectral Collection

Two small spectral datasets were experimentally investigated. A UV-Vis spectrophotometer (Evolution 300, Thermo Fisher, Waltham, MA, USA) was used for the adulterated vegetable oils in order to obtain the spectra of 51 samples in the wavelength range from 200 to 800 nm with an interval of 1 nm. The average spectrum of three parallel measurements was used for each sample. There is a negative absorbance for the 200–380 nm wavelengths, which seems to have no useful information. When the absorbance is above four, there is an obvious noise phenomenon and the absorption peaks are mainly present at 380–800 nm. Thus, Figure 1a mainly shows the absorption peaks at wavelengths of 380–800 nm. The adulterated herbs were measured from 12,000 to 4000 cm^{-1} at 2 cm^{-1} intervals on a Vertex 70 NIR spectrometer (Bruker Optics Inc., Ettlingen, Germany). Figure 1b shows the NIR spectra of the samples.

Each spectrum of the adulterated samples in the same dataset is similar. Therefore, it is necessary to combine multivariate calibration with spectroscopy to achieve an accurate quantitative analysis. Before calculation, the two datasets were divided into the training and prediction set by the Kennard–Stone (KS) algorithm. KS is the most widely used grouping method in chemometrics, which usually yields good grouping results. For the vegetable oil dataset, 34 and 17 samples are used as the training and prediction sets, respectively. For the herb dataset, 50 and 25 samples are used as the training and prediction sets, respectively.
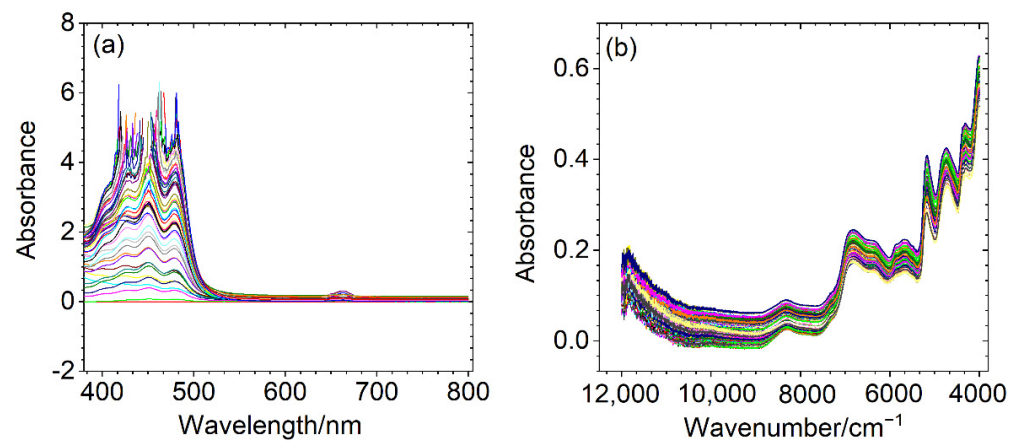
Figure 1. Measured spectra for adulterated vegetable oil (**a**) and herb (**b**) datasets.

2.3. Variational Mode Decomposition (VMD)

VMD is a powerful technique for signal analysis, which depends on the frequency information of the signal. The basic idea of the VMD algorithm is to construct and solve variational problems. For the construction of the variational problem, the purpose of VMD is to decompose the spectral signal **X** into a number of K discrete mode components u_k around the center frequency ω_k. At the same time, the sum of each mode is equal to the input signal **X**. The constrained variational model consists of the following target function.

$$\begin{cases} \min_{\{u_k, \ \omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum_k u_k = \mathbf{X} \end{cases} \tag{1}$$

where $\{u_k\} = \{u_1, \ldots, u_K\}$ is the mode ensemble obtained by decomposition, $\{\omega_k\} = \{\omega_1, \ldots, \omega_K\}$ represents the center frequency of each mode component, δ is the Dirac function, $\|\cdot\|_2$ is the L2 distance, $*$ is the convolution, j is the imaginary unit and **X** is a $[m \times n]$ matrix containing n spectral responses of m samples.

By introducing Lagrange multipliers and quadratic penalty terms, the above problem can be transformed into an unconstrained variational problem. An alternate direction method of multipliers (ADMM) is used to solve the saddle points of the multipliers' function. $\{u_k\}$, $\{\omega_k\}$ and the Lagrange multiplier are updated continuously in the frequency domain until the optimal solution of the variational problem is obtained. Finally, the results are derived by a FT. Please refer to Ref [34]. for the detailed algorithm.

2.4. Support Vector Regression (SVR)

SVR is a machine learning algorithm based on the principle of structural risk minimization and function approximation. It is specifically used to obtain predictive models via a number of identified support vectors and nonlinear kernel functions. The main process of SVR is to map the input data into a high-dimensional space by kernel functions. Then, the optimal hyper-plane is found in this feature space and a model is built to solve the linear regression problem. With strict statistical theory, SVR is able to be trained with few samples. Least square SVR (LSSVR) is one of the SVR algorithms. It can transform the quadratic programming problem into the problem of solving linear equations to reduce the complexity of computation. The Lagrangian function that is constructed to solve the linear system is as follows:

$$\begin{bmatrix} 0 & \mathbf{I}_n^{\mathrm{T}} \\ \mathbf{I}_n \mathbf{K} + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b_0 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{2}$$

where $\mathbf{I}_n$ is a $[n \times 1]$ vector, **K** is a $[n \times 1]$ kernel matrix, T is a transpose of a matrix or vector, γ is a weight vector, **b** is regression vector and b_0 is the model offset.

In this study, the Gaussian radial basis function (RBF) kernel function was used:

$$k_{i,j} = e^{\frac{-|x_i - x_j|^2}{2\sigma^2}} \tag{3}$$

where x_i and x_j denote the measured spectra of different samples and σ is the kernel width parameter. As we can see from Equations (2) and (3), the performance of the SVR model is mainly affected by two parameters, namely, γ and σ^2. More details are provided in Refs [22,39].

2.5. Variational Mode Decomposition Weighted Multiscale Support Vector Regression (VMD-WMSVR)

Motived by the advantages of VMD and SVR, a novel ensemble modeling method (VMD-WMSVR) is proposed for the spectral quantitative analysis of food and herb adulterants. This method includes the calibration and prediction stages. The schematic diagram of the proposed method is shown in Figure 2. Details of the process are described as follows.
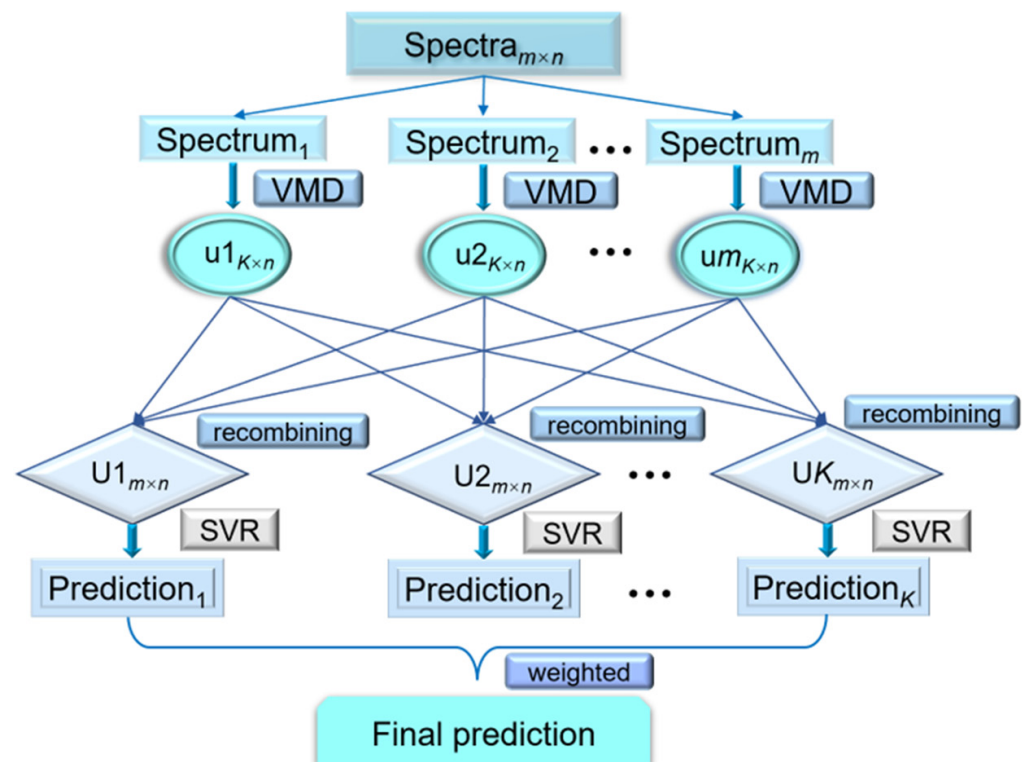

Figure 2. The schematic diagram of VMD-WMSVR.

(1) Each spectrum of the training set is decomposed by VMD into K discrete mode components uk ($k = 1, 2, \ldots, K$). Different K values of different samples have a large impact on the predictive stability of the proposed model. Moreover, too many mode components may destroy the linear relationship between the signal and the target value. The mode number K needs to be predetermined.

(2) Then, K mode components uk of the ith ($i = 1, 2, \ldots, m$) spectral signal are assigned sequentially to the ith row of each corresponding mode matrix Uk, i.e., the mode components uk are recombined to derive K modes Uk. In this way, each Uk contains the same number of samples and variables as the training set.

(3) SVR is used to build sub-models between each Uk and the target values. Overall, K multiscale regression sub-models are established for calibration.

(4) In the process of prediction, the spectral decomposition and recombination of the prediction set is the same as that of the training set. VMD-WMSVR is used for predicting

the samples in prediction set. Each sub-model gives a prediction and all the predictions are weighted and averaged to obtain the final prediction result. Sub-model weights are the inverse of the fourth power of the root mean square error of cross-validation (RMSECV).

3. Results and Discussion

3.1. The Mode Number K

In VMD-WMSVR model, the mode number K is a key parameter that needs to be set before the algorithm runs. Too few numbers of K may cause multiple components of the signal to be contained in one mode concurrently, resulting in insufficient decomposition. The model with too many numbers of K will create over decomposition problems and false modes [40]. In order to obtain the proper K value, the variation in the RMSEP with the mode number K for the two datasets is presented in Figure 3a,b, respectively.
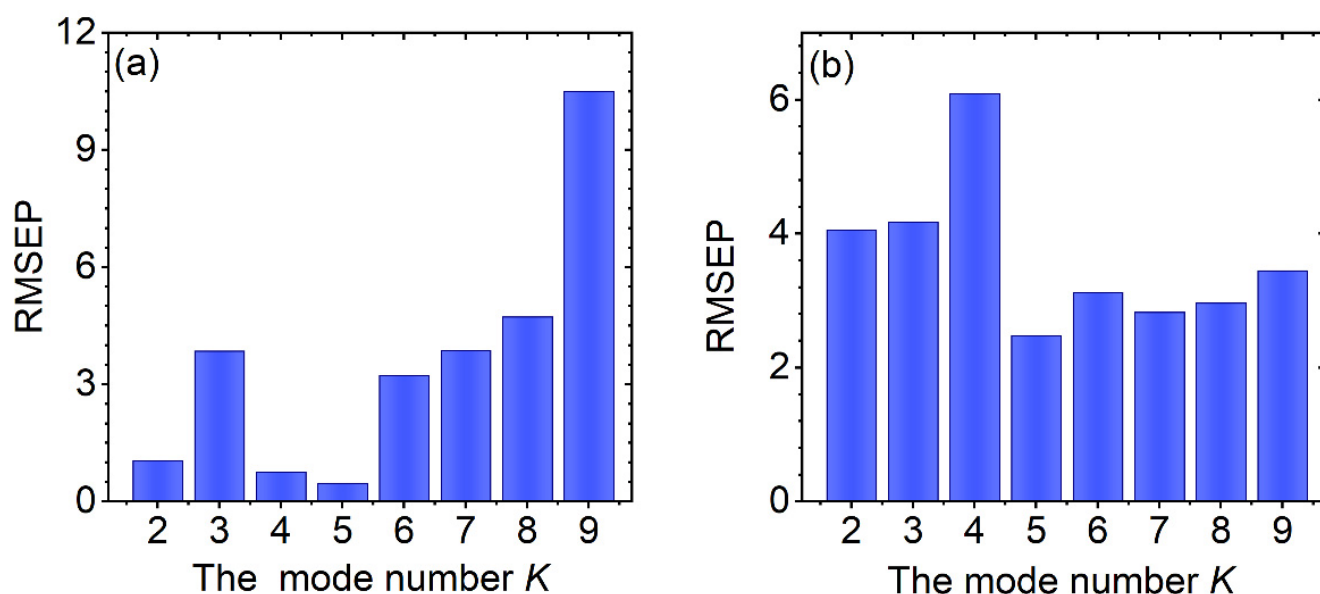

Figure 3. Variation in the RMSEP of VMD-WMSVR modeling with the mode number K for the adulterated vegetable oil (**a**) and herb (**b**) datasets.

VMD was applied multiple times with different K values for each spectrum and the result of the minimum RMSEP was considered the appropriate K value. Figure 3a shows that the RMSEP first presents a downward trend with the increase in K. When the K value reaches 5, the RMSEP obtains the minimum value and the prediction accuracy is the highest. After that, the RMSEP increases, especially when K is 9. This indicates that it may have mode mixing or pure noise modes that do not contribute much to the target value of interest. Figure 3b shows a similar trend to Figure 3a. As K is 5, the RMSEP reaches its minimum value. The smaller the RMSEP, the higher predictive accuracy of the model. Hence, the mode number K was set to 5 for both datasets.

3.2. The Spectral Decomposition of VMD

With the determination of K, each spectrum of the training set is decomposed by VMD and obtains K discrete mode components uk using this method. The UV-Vis spectra of 34 samples for the adulterated vegetable oils in the training set are decomposed. To illustrate the decomposition result, sample No. 2 is used. Figure 4a demonstrates that the original UV-Vis spectra are decomposed into five u components, which are graphically explained in the extracted order. This order represents the change in frequency from the lowest frequency to the highest. Different frequency blocks may contain different information and contribute varyingly to the model. The first three u components fluctuate slightly with a low number of peaks and the wavelength fluctuation in the range of 560–800 nm is gentle, which may contain some useful information. For u4, it almost

fluctuates symmetrically near zero over the entire wavelength range by detailed observation. In addition, big peaks alternate with small ones, making it difficult to determine whether they are noise or not. The variation in frequency for u5 is higher and the pronounced peak number increment is observed compared with the former u components, which behave more like noise.
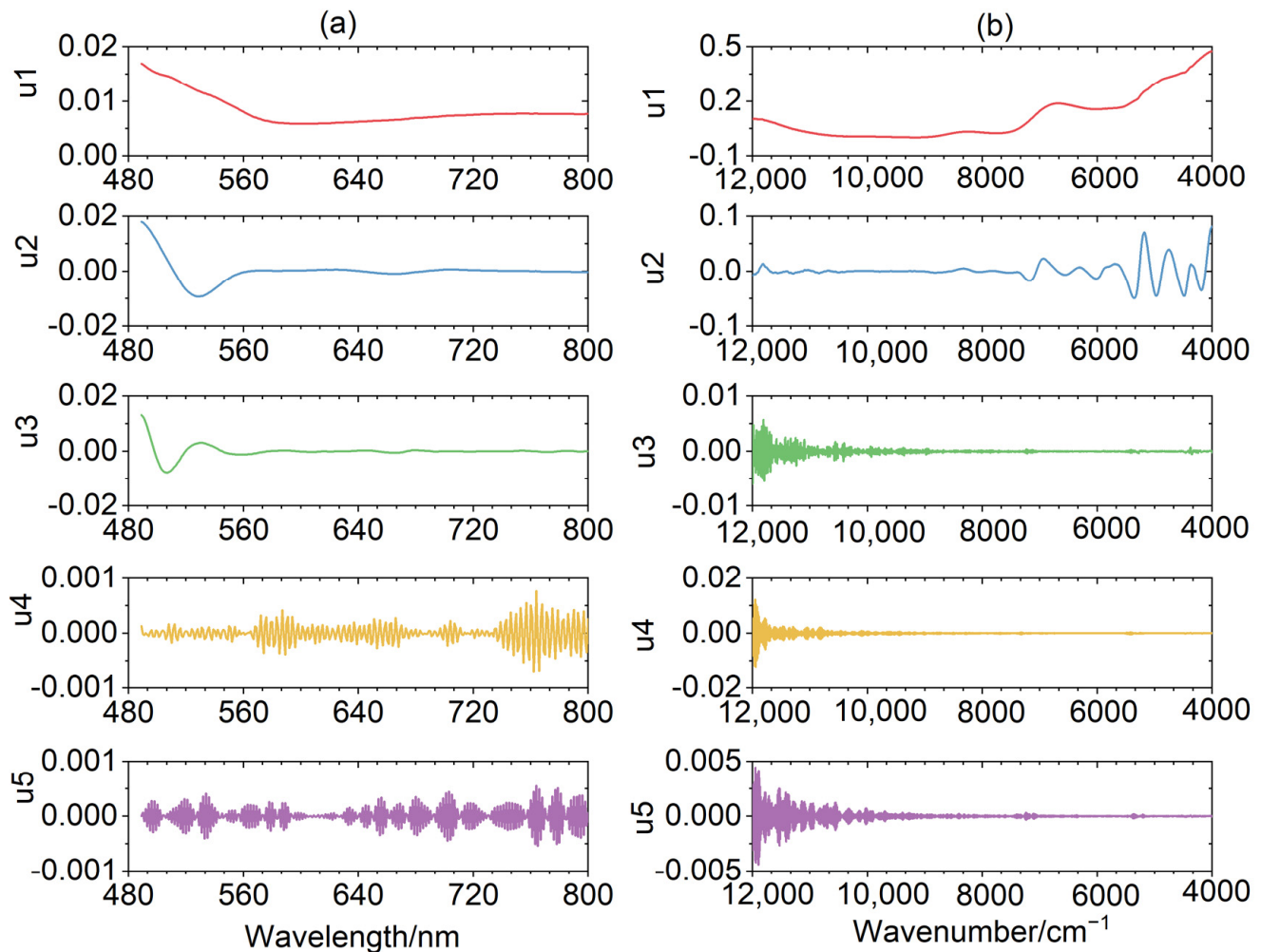

Figure 4. VMD diagram of the spectrum for sample No. 2 in the adulterated vegetable oil (**a**) and sample No. 17 in the adulterated herb (**b**) datasets.

For the adulterated herb dataset, each NIR spectrum of 50 samples in the training set is decomposed. Sample No. 17 is taken as an example. Figure 4b shows u1 oscillates slowly over the whole wavenumber with few peaks. Compared with u1, u2 changes more frequently and contains more peaks. Both of them have minor changes between 12,000 and 8000 cm^{-1} and have big fluctuations between 8000 and 4000 cm^{-1}, which may include a lot of helpful information. The last three u components have similar trends throughout the wavenumber range, fluctuating almost symmetrically around zero. The variation in frequency is prominent, changing rapidly from 12,000 to 10,000 cm^{-1}, which may have noise interference. In short, it can be seen from Figure 4 that the first u is a low-frequency component with a clear linear characteristic and a highly noticeable trend. As the order increases, the u component frequency becomes higher and higher, appearing with more irregularities and higher degrees of complexity. Since VMD is a mathematical decomposition, not all mode components have a well-defined chemical meaning for the spectral signal. The low-frequency and high-frequency mode components can be distinguished by observing their variation regularity combined with their variance values [41]. The low-frequency mode component changes gently with a large variance,

while the high-frequency mode component oscillates almost symmetrically at zero with a small variance.

3.3. Comparison of the Predicted Results

In order to evaluate the predictive ability of the proposed method, PLS and SVR are used for comparison. The parameters of PLS and SVR were optimized at first. For PLS, the Monte Carlo cross validation (MCCV) combined with the F-test was used to determine latent variables (LVs). The optimal LV for the adulterated vegetable oil and herb datasets is 5 and 4, respectively. For SVR, there are two parameters (γ, σ^2) that need to be predetermined. The particle swarm optimization (PSO) algorithm was adopted and the RMSEP was used as the evaluation standard for the parameters optimization. The optimal γ and σ^2 for the adulterated vegetable oil and herb datasets are 222.74, 227.37 and 247.46, 106.22, respectively. The relationship between the prepared and the predicted values for the prediction set by PLS (a), SVR (b) and VMD-WMSVR (c) for the two adulterated datasets is shown in Figure 5andFigure 6, respectively. The RMSEP and R of the prediction set are used as indicators to validate the performance of the models.
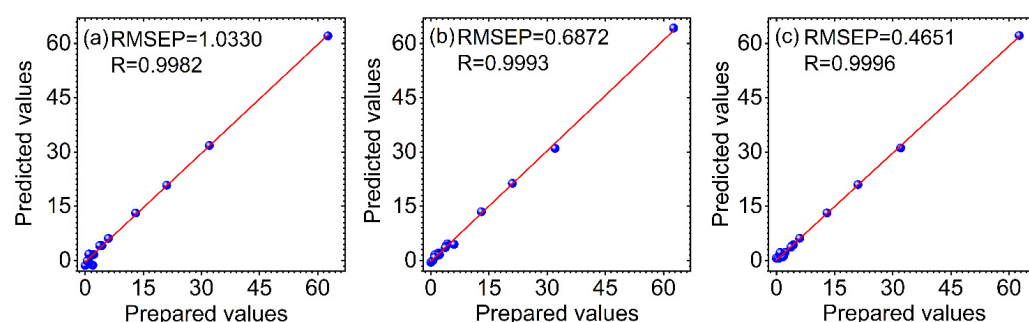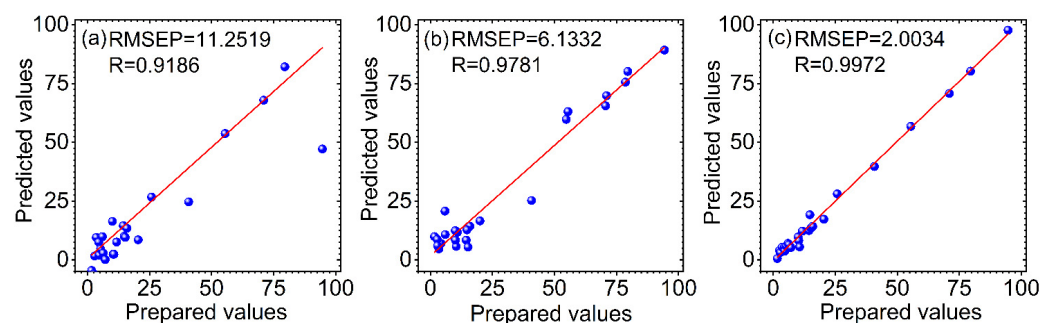

Figure 5. The relationship between the prepared and the predicted values for the prediction set by PLS (**a**), SVR (**b**) and VMD-WMSVR (**c**) for the adulterated vegetable oil dataset.

Figure 6. The relationship between the prepared and the predicted values for the prediction set by PLS (**a**), SVR (**b**) and VMD-WMSVR (**c**) for the adulterated herb dataset.

It can be found that the R values of the three methods are all above 0.9, indicating that these modeling methods combined with spectroscopy are effective for the quantitative analysis of rapeseed oil and RAO adulterants. However, the high benchmark of R leads to little room for improvement in all approaches, so their improvement is not significant from the perspective of R values alone. The variation in the RMSEP was used as the main criterion for comparison of the different methods. The RMSEP is a measure of the deviation between the predicted and prepared values. The smaller its value, the closer the predicted value is to the prepared value. For the vegetable oil dataset, it is observed from Figure 5 that SVR has a lower RMSEP and a higher R compared with PLS, demonstrating that SVR is superior to PLS. Among the three methods, VMD-WMSVR has the lowest RMSEP and the highest R. This indicates that the adaptive spectral decomposition can further improve the prediction ability of SVR and PLS. Thus, VMD-WMSVR has the best prediction, which

is attributed to the original spectra of VMD. Figure 6 shows that compared with PLS, the RMSEP for the herb dataset under SVR is reduced by 45%. This also indicates that SVR is better than PLS in modeling results. There is a good linearity between the prepared values and the predicted values in Figure 6c. Compared with PLS and SVR, the RMSEP of VMD-WMSVR is reduced by 82% and 67%, respectively. Therefore, the prediction results of both datasets suggest the potential of VMD-WMSVR in improving the predictive accuracy.

4. Conclusions

In summary, this work presented a new chemometric methodology named VMD-WMSVR and was applied for two spectral datasets to achieve the quantification of vegetable oil and herb adulterants. On the one hand, VMD is designed to make full use of the information by decomposing the original spectra adaptively into multiple mode components with different frequencies. The modeling technique can improve the accuracy of predictions compared with a single PLS and SVR. On the other hand, it is a non-destructive and efficient method for the determination of rapeseed oil and RAO adulterants without the use of reagents and the generation of harmful residues, which protects the environment. However, the performance of predicting actual new samples was not discussed in this paper and should be further studied in the future.

References

1. Nehal, N.; Choudhary, B.; Nagpure, A.; Gupta, R.K. DNA barcoding: A modern age tool for detection of adulteration in food. *Crit. Rev. Biotechnol.* **2021**, *41*, 767–791. [CrossRef]
2. He, Y.; Bai, X.L.; Xiao, Q.L.; Liu, F.; Zhou, L.; Zhang, C. Detection of adulteration in food based on nondestructive analysis techniques: A review. *Crit. Rev. Food Sci. Nutr.* **2020**, *61*, 2351–2371. [CrossRef] [PubMed]
3. von Wuthenau, K.; Muller, M.-S.; Lina, C.; Marie, O.; Markus, F. Food authentication of almonds (*Prunus dulcis* Mill.). Fast origin analysis with laser ablation inductively coupled plasma mass spectrometry and chemometrics. *J. Agric. Food Chem.* **2022**, *70*, 5237–5244. [CrossRef]
4. Li, D.; Zang, M.W.; Wang, S.W.; Zhang, K.H.; Zhang, Z.Q.; Li, X.M.; Li, J.C.; Guo, W.P. Food fraud of rejected imported foods in China in 2009–2019. *Food Control* **2022**, *133*, 108619. [CrossRef]
5. Lim, K.; Pan, K.; Yu, Z.; Xiao, R.H. Pattern recognition based on machine learning identifies oil adulteration and edible oil mixtures. *Nat. Commun.* **2020**, *11*, 5353. [CrossRef]
6. Negi, A.; Pare, A.; Meenatchi, R. Emerging techniques for adulterant authentication in spices and spice products. *Food Control* **2021**, *127*, 108113. [CrossRef]
7. Czerwenka, C.; Muller, L.; Lindner, W. Detection of the adulteration of water buffalo milk and mozzarella with cow's milk by liquid chromatography-mass spectrometry analysis of beta-lactoglobulin variants. *Food Chem.* **2010**, *122*, 901–908. [CrossRef]
8. Acosta, G.; Arce, S.; Martinez, L.D.; Llabot, J.; Gomez, M.R. Monitoring of phenolic compounds for the quality control of melissa officinalis products by capillary electrophoresis. *Phytochem. Anal.* **2012**, *23*, 177–183. [CrossRef] [PubMed]
9. Hong, E.; Lee, S.Y.; Jeong, J.Y.; Park, J.M.; Kim, B.H.; Kwon, K.; Chun, H.S. Modern analytical methods for the detection of food fraud and adulteration by food category. *J. Sci. Food Agric.* **2017**, *97*, 3877–3896. [CrossRef]
10. Bian, X.H.; Lu, Z.K.; van Kollenburg, G. Ultraviolet-visible diffuse reflectance spectroscopy combined with chemometrics for rapid discrimination of Angelicae Sinensis Radix from its four similar herbs. *Anal. Methods* **2020**, *12*, 3499–3507. [CrossRef]

11. Torrecilla, J.S.; Rojo, E.; Dominguez, J.C.; Rodriguez, F. A novel method to quantify the adulteration of extra virgin olive oil with low-grade olive oils by UV-Vis. *J. Agric. Food Chem.* **2010**, *58*, 1679–1684. [CrossRef] [PubMed]

12. Li, X.; Zhang, L.X.; Zhang, Y.; Wang, D.; Wang, X.F.; Yu, L.; Zhang, W.; Li, P.W. Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils. *Trends Food Sci. Technol.* **2020**, *101*, 172–181. [CrossRef]

13. Lohumi, S.; Lee, S.; Lee, H.; Cho, B.K. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends Food Sci. Technol.* **2015**, *46*, 85–98. [CrossRef]

14. Rivera-Perez, A.; Romero-Gonzalez, R.; Frenich, A.G. Feasibility of applying untargeted metabolomics with GC-Orbitrap-HRMS and chemometrics for authentication of black pepper (*Piper nigrum* L.) and identification of geographical and processing markers. *J. Agric. Food Chem.* **2021**, *69*, 5547–5558. [CrossRef] [PubMed]

15. Rios-Reina, R.; Garcia-Gonzalez, D.L.; Callejon, R.M.; Amigo, J.M. NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. *Food Control* **2018**, *89*, 108–116. [CrossRef]

16. Chen, Z.W.; Harrington, P.D. Self-optimizing support vector elastic net. *Anal. Chem.* **2020**, *92*, 15306–15316. [CrossRef]

17. Jia, W.; Dong, X.Y.; Shi, L.; Chu, X.G. Discrimination of milk from different animal species by a foodomics approach based on high-resolution mass spectrometry. *J. Agric. Food Chem.* **2020**, *68*, 6638–6645. [CrossRef]

18. Shao, X.G.; Bian, X.H.; Liu, J.J.; Zhang, M.; Cai, W.S. Multivariate calibration methods in near infrared spectroscopic analysis. *Anal. Methods* **2010**, *2*, 1662–1666. [CrossRef]

19. Olivieri, A.C. Analytical advantages of multivariate data processing. One, two, three, infinity? *Anal. Chem.* **2008**, *80*, 5713–5720. [CrossRef]

20. Ying, Y.W.; Jin, W.; Yu, H.X.; Yu, B.W.; Shan, J.; Lv, S.W.; Zhu, D.; Jin, Q.H.; Mu, Y. Development of particle swarm optimization-support vector regression (PSO-SVR) coupled with microwave plasma torch-atomic emission spectrometry for quality control of ginsengs. *J. Chemom.* **2017**, *31*, e2862. [CrossRef]

21. Bian, X.H.; Diwu, P.Y.; Liu, Y.R.; Liu, P.; Li, Q.; Tan, X.Y. Ensemble calibration for the spectral quantitative analysis of complex samples. *J. Chemom.* **2018**, *32*, e2940. [CrossRef]

22. Li, Y.K.; Shao, X.G.; Cai, W.S. A consensus least squares support vector regression (LS-SVR) for analysis of near-infrared spectra of plant samples. *Talanta* **2007**, *72*, 217–222. [CrossRef] [PubMed]

23. Xu, L.; Ye, Z.-H.; Yan, S.-M.; Shi, P.-T.; Cui, H.-F.; Fu, X.-S.; Yu, X.-P. Combining local wavelength information and ensemble learning to enhance the specificity of class modeling techniques: Identification of food geographical origins and adulteration. *Anal. Chim. Acta* **2012**, *754*, 31–38. [CrossRef]

24. Hu, Y.; Peng, S.L.; Peng, J.T.; Wei, J.P. An improved ensemble partial least squares for analysis of near-infrared spectra. *Talanta* **2012**, *94*, 301–307. [CrossRef] [PubMed]

25. Granato, D.; Santos, J.S.; Escher, G.B.; Ferreira, B.L.; Maggio, R.M. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends Food Sci. Tech.* **2018**, *72*, 83–90. [CrossRef]

26. Jiang, Y.Y.; Ge, H.Y.; Zhang, Y. Quantitative analysis of wheat maltose by combined terahertz spectroscopy and imaging based on Boosting ensemble learning. *Food Chem.* **2020**, *307*, 125533.

27. Bian, X.H.; Li, S.J.; Lin, L.G.; Tan, X.Y.; Fan, Q.J.; Li, M. High and low frequency unfolded partial least squares regression based on empirical mode decomposition for quantitative analysis of fuel oil samples. *Anal. Chim. Acta* **2016**, *925*, 16–22. [CrossRef]

28. Cadet, F.; Fontaine, N.; Vetrivel, I.; Chong, M.N.F.; Savriama, O.; Cadet, X.; Charton, P. Application of fourier transform and proteochemometrics principles to protein engineering. *BMC Bioinform.* **2018**, *19*, 382. [CrossRef]

29. Liu, Z.C.; Cai, W.S.; Shao, X.G. A weighted multiscale regression for multivariate calibration of near infrared spectra. *Analyst* **2009**, *134*, 261–266. [CrossRef]

30. Wu, X.Y.; Bian, X.H.; Lin, E.; Wang, H.T.; Guo, Y.G.; Tan, X.Y. Weighted multiscale support vector regression for fast quantification of vegetable oils in edible blend oil by ultraviolet-visible spectroscopy. *Food Chem.* **2021**, *342*, 128245. [CrossRef]

31. Perez-Canales, D.; Alvarez-Ramirez, J.; Jauregui-Correa, J.C.; Vela-Martinez, L.; Herrera-Ruiz, G. Identification of dynamic instabilities in machining process using the approximate entropy method. *Int. J. Mach. Tools Manu.* **2011**, *51*, 556–564. [CrossRef]

32. Liu, Y.; Cai, W.S.; Shao, X.G. Intelligent background correction using an adaptive lifting wavelet. *Chemom. Intell. Lab. Syst.* **2013**, *125*, 11–17. [CrossRef]

33. Wu, W.H.; Chen, C.C.; Jhou, J.W.; Lai, G. A rapidly convergent empirical mode decomposition method for analyzing the environmental temperature effects on stay cable force. *Comput. Aided. Civ. Infrastruct. Eng.* **2018**, *33*, 672–690. [CrossRef]

34. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE. Trans. Signal Proces.* **2014**, *62*, 531–544. [CrossRef]

35. Chen, Q.M.; Chen, J.H.; Lang, X.; Xie, L.; Rehman, N.U.; Su, H.Y. Self-tuning variational mode decomposition. *J. Franklin Inst.* **2021**, *358*, 7825–7862. [CrossRef]

36. Lahmiri, S. Intraday stock price forecasting based on variational mode decomposition. *J. Comput. Sci. Neth.* **2016**, *12*, 23–27. [CrossRef]

37. Hu, H.L.; Wang, L.; Tao, R. Wind speed forecasting based on variational mode decomposition and improved echo state network. *Renew. Energ.* **2021**, *164*, 729–751. [CrossRef]

38. Wang, Z.J.; Wang, J.J.; Du, W.H. Research on fault diagnosis of gearbox with improved variational mode decomposition. *Sensors* **2018**, *18*, 3510. [CrossRef]

39. Thissen, U.; Pepers, M.; Ustun, B.; Melssen, W.J.; Buydens, L.M.C. Comparing support vector machines to PLS for spectral regression applications. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 169–179. [CrossRef]
40. Nazari, M.; Sakhaei, S.M. Successive variational mode decomposition. *Signal Process.* **2020**, *174*, 107610. [CrossRef]
41. Tan, C.; Wang, J.Y.; Wu, T.; Qin, X.; Li, M.L. Determination of nicotine in tobacco samples by near-infrared spectroscopy and boosting partial least squares. *Vib. Spectrosc.* **2010**, *54*, 35–41. [CrossRef]