QUALITATIVE AND QUANTITATIVE ANALYSIS OF POTASSIUM SORBATE IN MILK POWDER USING TERAHERTZ SPECTRA **

F. Y. Lian [1,2*], M. X. Fu [1,2], D. G. Xu [1,2]

[1] Key Laboratory of Grain Information Processing and Control,
Henan University of Technology, Ministry of Education, Zhengzhou Henan, 450001, China
[2] Key Laboratory of Henan Province for Grain Photoelectric Detection and Control,
Henan University of Technology, Zhengzhou, 45001, China; e-mail: lfywork@163.com

Using the terahertz time-domain spectroscopy system (THz-TDS), the potassium sorbate content in milk powder has been investigated. Firstly, the THz time-domain spectra of the samples were transformed into frequency spectra by Fourier transform, and then, through a set of simplified formulas of the optical parameters, the THz absorption spectra and refractive index spectra were obtained. Based on the characteristic absorption peak of potassium sorbate, a quantitative analysis model of potassium sorbate in milk powder was constructed using Ordinary Least Squares (OLS) linear regression. In the case of low content (<1%) of potassium sorbate, a special THz absorption band (0.887–1.000 THz) was selected to construct a Partial Least Squares (PLS) model that is called Sub-PLS. In addition, for qualitative analysis, we used partial least squares discrimination analysis (PLS-DA). Experimental results show that the proposed Sub-PLS model is superior to the PLS mode based on full spectrum in accuracy, and the proposed PLS-DA is superior to other nonlinear regression methods such as support vector machine and neural network.

Keywords: *THz time-domain spectroscopy, potassium sorbate, ordinary least squares, partial least squares, partial least squares discrimination analysis.*

КАЧЕСТВЕННЫЙ И КОЛИЧЕСТВЕННЫЙ АНАЛИЗ СОРБАТА КАЛИЯ В СУХОМ МОЛОКЕ С ИСПОЛЬЗОВАНИЕМ ТЕРАГЕРЦОВОГО СПЕКТРА

F. Y. Lian [1,2*], M. X. Fu [1,2], D. G. Xu [1,2]

УДК 543.42

[1] Главная лаборатория по обработке и контролю информации Министерства образования, Чжэнчжоу, Хэнань, 450001, Китай
[2] Хэнаньский технологический университет, Чжэнчжоу, 45001, Китай; e-mail: lfywork@163.com

(Поступила 24 июня 2019)

Содержание сорбата калия в сухом молоке исследовано с помощью терагерцовой (ТГц) спектроскопии во временной области. ТГц спектры образцов преобразованы в частотные спектры с помощью преобразования Фурье. С применением набора упрощенных формул, описывающих оптические параметры, получены ТГц спектры поглощения и спектральные зависимости показателя преломления. На основе данных по интенсивности характерного максимума поглощения сорбата калия построена модель количественного анализа сорбата калия в сухом молоке. Метод линейных наименьших квадратов использован для оценки неизвестных параметров в модели линейной регрессии. В условиях низкого (<1%) содержания сорбата калия выбрана специальная полоса поглощения (0.887–1.000 ТГц) для построения модели частичной регрессии наименьших квадратов. Для качественного анализа использован дискриминантный анализ на основе частичных наименьших квадратов. Предложенная модель частичной регрессии наименьших квадратов превосходит по точности

** Full text is published in JAS V. 87, No. 4 (http://springer.com/journal/10812) and in electronic version of ZhPS V. 87, No. 4 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

модель линейных наименьших квадратов, основанную на полном спектре, а дискриминантный анализ на основе частичных наименьших квадратов превосходит другие методы линейной регрессии, такие как метод опорных векторов и нейронная сеть.

Ключевые слова: *терагерцовая спектроскопия во временной области, сорбат калия, метод наименьших квадратов, частичные наименьшие квадраты, дискриминантный анализ на основе частичных наименьших квадратов.*

Introduction. Nowadays potassium sorbate is commonly added to food as a preservative. However, according to China's national standards [1], potassium sorbate is banned in milk powder; therefore, a batch of milk powder will be declared illegal once potassium sorbate is detected. Nowadays, many methods have been used to detect banned additives in food. These methods mainly include gas chromatography [2], high performance liquid chromatography [3], chromatography-mass spectrometry [4], thin-layer chromatography, ion chromatography [5], and so on. These methods are divided into two categories: bioassay technique and physicochemical analysis. The bioassay technique mainly is used to analyze material composition using biological immunity or biosensor, while physicochemical analysis is mainly used to make qualitative and quantitative analysis using various analytical instruments. Each of these methods has its own characteristics and scope of application but, in reality, also has its disadvantages and limitations. Take the bioassay technique, for example: sample preparation and detection process are complicated and time-consuming. Physical detection methods can shorten test time and simplify detection procedure but cannot meet requirements because of poor accuracy; so, a detection method with speed, convenience, and high precision will be required for wide-scale milk-quality screening.

Terahertz spectra detection is a new technique in the field of spectral detection [6, 7]. Qin et al. reported the use of terahertz time-domain spectroscopy (THz-TDS) to detect tetracycline hydrochloride (TCsH) in infant milk powder for the first time [8]. A four kinds of TCsH exhibit unique spectral features in the region of 0.3–1.8 THz. The main spectral features of these TCsH were still detectable when mixed with infant milk powder with concentrations of 1–50%, even in ternary mixtures. Cui et al. studied the measurement of mixtures of melamine and milk powder using a THz ray [9] and found that there were two absorption peaks at 1.99 and 2.29 THz in all the spectra, which provided a method of detecting melamine in milk powder. Naito et al. inspected milk components by terahertz attenuated total reflectance (THz-ATR) spectrometer equipped with a temperature controller [10] and confirmed that THz spectroscopy has potential for quantitative analysis of milk fat, total solid, lactose, milk protein, casein, and somatic cells. Compared to traditional detection methods, THz detection has higher speed and can even be finished within minutes at room temperature. In particular, since it is a nondestructive detection method, samples do not have to be destroyed and consumed in THz detection process. Moreover, THz waves can easily penetrate common packing materials, but because of lower energy, it cannot harm the human body and is proven to be very safe [11]. These advantages of THz spectral detection provide a new quick and effective method for detecting potassium sorbate or sodium benzoate in milk when needed to check milk quality widely. We applied the THZ spectroscopy to the analysis of additives in milk powder, not only to provide a new rapid physical examination method but also, by using the chemometrics method, to give a better quantitative detection method. Literature research shows that there are few studies on checking milk powder using terahertz detection.

Experiment platform and data acquisition. *Experiment platform.* THz-TDS, named Z3, made by the Zomega Corp., USA, was used as experiment platform and is shown in Fig. 1. The mode-locked femtosecond laser beam is split into a pump and a probe beam by a polarizing beam splitter. A half-wave plate is used to change the intensity ratio between the pump and probe beam. The probe beam interacts with a linear slow-scan time delay stage to set the relative timing between the pump and probe impulses prior to activating the THz emitter. The THz emitter consists of a photoconductive dipole antenna fabricated on a LT-GaAs wafer. A high-voltage AC bias is applied across the electrodes of the antenna to accelerate photocarriers generated by the pump beam, creating a transient photocurrent that generates a free-space THz impulse in the forward direction. In the far-field, the radiated electric field is proportional to the first time, derivative of the current. A 25 mm FL TPX lens (both transparent to THz and visible light) is used to collimate the THz beam out of the Z3 THz emitter module. This system is a kind of time-domain spectrograph that produces and detects wide-band THz impulse and can do a spectral measurement on samples in the two modes of transmission and reflection. Its main optical parameters and components are as follows: frequency range 0.1–3.5 THz, spectral resolution ≥5 GHz (after fast Fourier transform), and dynamic range 85 dB. This device is not mobile and needs to be used in an ultraclean environment.

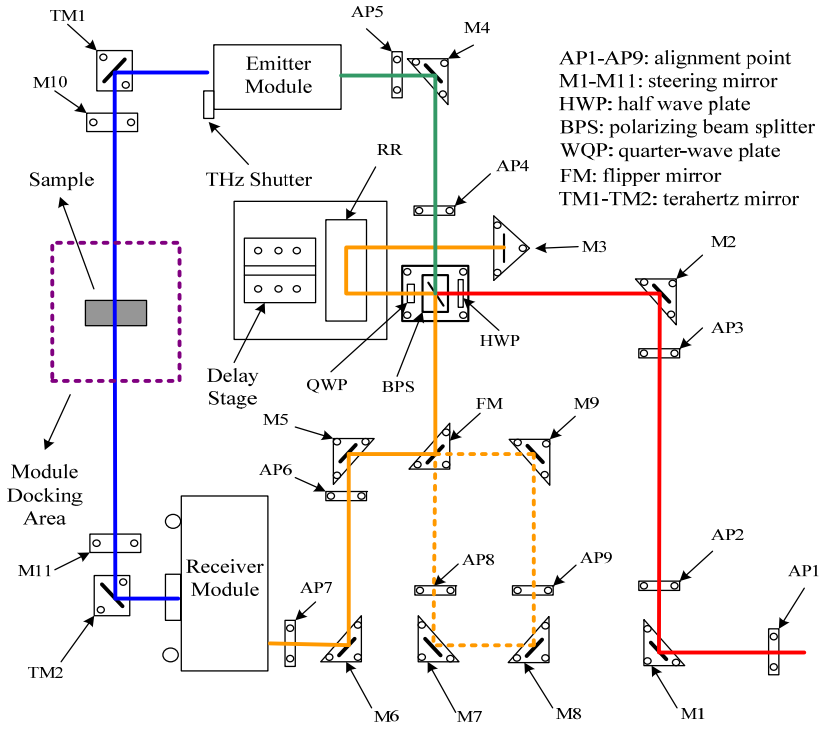

Fig. 1. Overhead view of the internal optics of the Z3-XL THz spectrometer.

Data acquisition. The optical parameters that describe the optical properties of macroscopic matter mainly include refractive index, extinction coefficient, absorption coefficient, dielectric constant, and so on. Various optical parameters can all be summarized as the complex refractive index $\tilde{n}(\omega)$, which can be expressed as follow:

$$\tilde{n}(\omega) = n(\omega) - jk(\omega), \tag{1}$$

where $n(\omega)$ and $k(\omega)$ are real refractive index and extinction coefficient, which describe the dispersion and absorption characteristics respectively, and are represented by the following formulas:

$$n(\omega) = \frac{\varphi(\omega)}{\omega d} c + 1, \tag{2}$$

$$k(\omega) = \ln\left[\frac{4n(\omega)}{\rho(\omega)(n(\omega)+1)^2}\right]\frac{c}{\omega d}, \tag{3}$$

where c is the speed of light, d is the thickness of the sample, ω is the signal frequency, $\varphi(\omega)$ is the phase difference of the THz electric field between reference signal and sample signal, and $\rho(\omega)$ is their amplitude ratio, which can be expressed as follows:

$$\rho(\omega) = 4n(\omega)e^{-k(\omega)\omega d/c} / (1+n(\omega))^2, \tag{4}$$

$$\varphi(\omega) = (n(\omega)-1)\omega d/c. \tag{5}$$

The absorption coefficient $a(\omega)$ and the extinction coefficient $k(\omega)$ are related as follows:

$$a(\omega) = 2\omega k(\omega)/c, \tag{6}$$

where the extinction coefficient can be calculated as follows:

$$a(\omega) = \frac{2k(\omega)\omega}{c} = \frac{2}{d}\ln\left[\frac{4n(\omega)}{\rho(\omega)(n(\omega)+1)^2}\right]. \tag{7}$$

Sample preparation. The raw materials used in the experiment are potassium sorbate and pure milk powder. Pure milk, of the Mengniu brand and produced by the Mengniu Dairy Group Co. Ltd, China, was purchased from the supermarket. Its food production license number is QS150105010003 and product standard code is Q/NOMR0006S. According to the market inspection report of the State Food and Drug Admini-

stration, the milk powder should not contain potassium sorbate. When preparing samples, potassium sorbate was added to the milk powder deliberately and mixed evenly according to the proportion of 0% (pure milk powder), 1, 4, 7, 9, 10, 15, 20, 30, 40, 50, and 100% (pure potassium sorbate). We weighed 160 mg from each proportional mixture to make slices. When making slices, 6 MPa pressure was maintained for 3 min on the tablet machine to make each slice's thickness around 1.1 mm. Finally, we used an electric vernier caliper to measure the thickness of each slice, recorded the data, and put the slices into sealing bag.

Experimental results and analysis. *The THz spectra of the samples.* We put the prepared samples into the THz time-domain spectroscopy system to acquire data under the conditions of 22°C and 4% RH. When testing, the time window was set at the point when the first echo wave was received, and each sample was scanned three times. Note that the reference signal should be acquired first before beginning a new scan. The system can calculate an average value of three times the THz spectra.

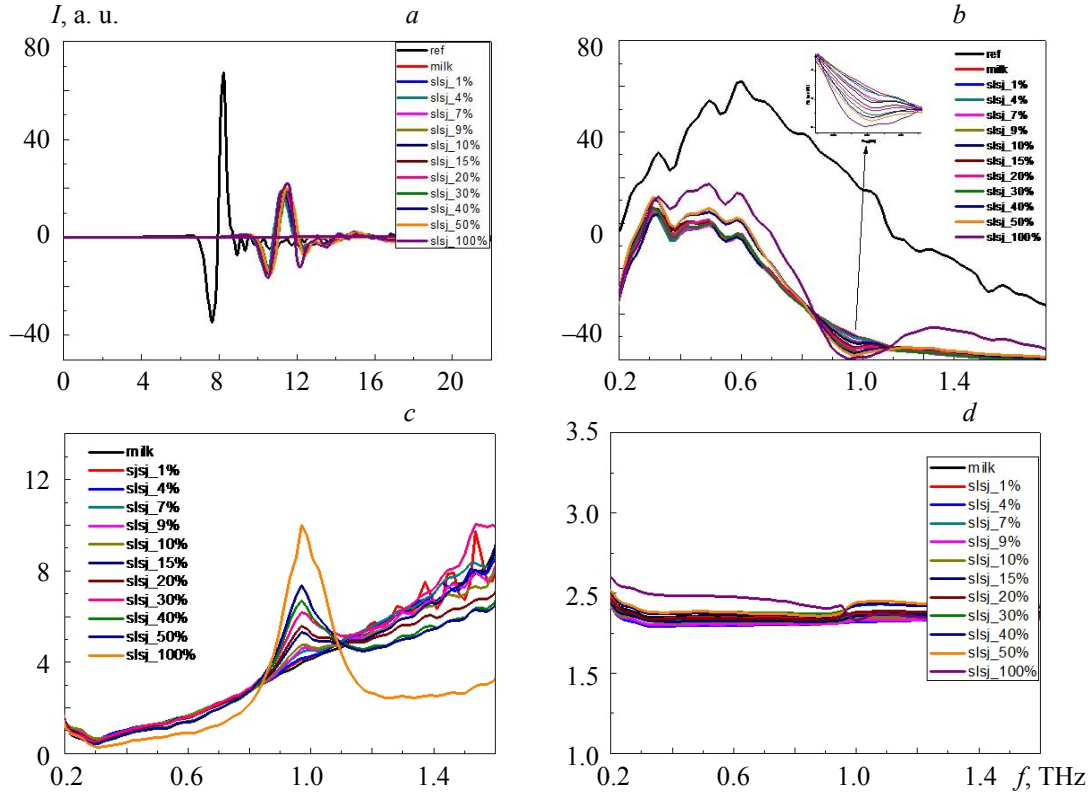

Fig. 2. The THz spectra of the mixture of milk powder and potassium sorbate: a) time-domain spectra; b) frequency spectra; c) absorption spectra; d) refractive index spectra.

The typical THz spectra of the samples are shown in Fig. 2a. It can be seen from Fig. 2a that, compared to the reference signal, all signals of the samples have a certain time delay, an apparent attenuation of amplitude, and a change in optical distance due to the absorption of samples to THz wave in their time-domain spectra. Figure 2 also shows that the time delay of pure potassium sorbate is longer, and its amplitude attenuation is smaller than that of other samples, which might be because the slice of pure potassium sorbate is thicker than other samples of the same quality and under the same pressure. The waveforms of 13 signals, including the reference signal, are very similar in shape and only show some changes in amplitude or phase, which demonstrated the stability of the system used. The stability of the system can reduce the error caused by the system itself and improve the testing precision.

Frequency spectra can be obtained via FFT of the time-domain signals of the samples, as shown in Fig. 2b. Compared to the reference signal, the amplitudes of the sample signals attenuate apparently due to the hard absorption of the THz wave. In the range of 0.2–1.6 THz, the amplitudes first increase and then shrink with frequency, and after 1.4 THz, the amplitudes shrink to zero gradually. At 0.972 THz, there are troughs for all samples except pure milk powder, and the troughs become smaller and smaller with rise in the content of potassium sorbate in the samples due to the absorption of potassium sorbate to THz wave.

Figure 2c shows the absorption coefficient spectra of the mixture of milk powder and potassium sorbate. It can be seen from the figure that, with increase in frequency, the absorption coefficients of the samples increase gradually. At 0.972 THz, some wave crests emerge in the spectra. These wave crests are the characteristic absorption peaks of potassium sorbate. It can also be seen clearly that, with increase in the concentration of potassium sorbate, the values of characteristic absorption peaks also increase gradually. Consequently, we regarded the characteristic absorption peak values at 0.972THz as a reference for quantitative analysis of potassium sorbate concentration in the mixture, but it is still necessary to do a further analysis of the linear relationship between the characteristic absorption peaks and the concentration of potassium sorbate. What needs to be pointed out is that the characteristic absorption peaks are not obvious and cannot predict the concentration of potassium sorbate when its concentration is under 1%. Therefore, chemometric methods will be necessary.

Figure 2d shows the refractive index spectra of the samples. It can be seen from Fig. 3 that, with increase in frequency in the effective THz wave band, the refractive indexes of the samples decline slowly, and at the characteristic absorption peaks, some broken lines emerge because of the absorption of the samples to THz wave, which are in accordance with the above analysis.

Quantitative analysis. *Least squares analysis.* Ordinary Least Squares (OLS) is a common method in linear regression analysis and can reveal the relationship between two variables through several groups of numerical value. OLS is often used in curve fitting [12], which was also adopted in this paper for predicting the concentration of potassium sorbate in milk powder.

To evaluate the performance of the model, some parameters should be introduced. In this paper, we adopted three parameters: R, which is called correlation coefficient; RMSE, which is the root-mean-square error; and MRE, which is the maximal relation error, to evaluate the performance of the model. The definitions of the three parameters are as follows [13]:

$$R = \frac{\sum_{i=1}^{n}(y_i^{\text{ref}} - \bar{y})(y_i^{\text{pre}} - \overline{y_i^{\text{pre}}})}{\sqrt{\sum_{i=1}^{n}(y_i^{\text{ref}} - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(y_i^{\text{pre}} - \overline{y_i^{\text{pre}}})^2}}, \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{\text{pre}} - y_i^{\text{ref}})^2}, \tag{9}$$

$$\text{MRE} = \max\left|\frac{y_i^{\text{pre}} - y_i^{\text{ref}}}{y_i^{\text{ref}}} \times 100\%\right|, \tag{10}$$

where y_i^{ref} is the reference value of the ith sample, y_i^{pre} is the predicted value of the ith sample, $\bar{y}$ is the reference value of all the samples, and $\overline{y_i^{\text{pre}}}$ is the predicted value of all the samples. The R values and the RMSE values of the calibration set (RMSEC) are calculated using the reference values and the prediction values, and the RMSE values of the calibration set that were obtained through cross-validation methodology (called RMSECV) are calculated using cross-validation methodology based on the reference values and the predicted values [14].

Generally speaking, the larger the model's R value, the closer the real value to the predicted value. The performance of the model may be evaluated using three parameters: correlation coefficient (R), mean square error (MSE), and maximum relative error (MRE). The larger the correlation coefficient, the smaller the MSE and MRE, and the better the performance of the established model. Similarly, the calibration precision of the calibration set is also evaluated using these three values. The larger the R value, the smaller the RMSEC value, which means a higher calibration precision for the established model. The real prediction accuracy of the model may be evaluated using the parameter MSE, and a smaller MSE value means a higher prediction accuracy [15].

Quantitative analysis using Least Squares. We tested 10 groups of milk powder samples with various concentrations of potassium sorbate. For each concentration of potassium sorbate, we prepared three samples, and each sample was tested three times, so we obtained a total of 30 sets of data. To perform the quantitative analysis for the 30 sets of data, we selected 10 sets of data as the prediction set according to the ratios that were mentioned above and selected other data sets as the calibration set.

As shown in Fig. 2, there is an obvious absorption peak in 0.972 THz in the absorption spectrum of potassium sorbate, and the amplitude of the absorption peak changes with the concentration of potassium sorbate. In this paper, we selected the value of the absorption coefficient in 0.972 THz as the input of the estab-

lished model. We tested each sample three times and obtained the three values of the absorption peak of each sample and calculated the average value for each concentration. Lastly, we established a relationship model between the concentrations of potassium sorbate and their absorption peaks at 0.972 THz using the least squares method.

We used the LS tool package of MATLAB to execute an OLS algorithm, which enabled us to analyze the average absorption coefficient and refractive index for the 10 groups of samples. The relationships between the concentration of potassium sorbate and the two optical parameters are shown in Fig. 3.

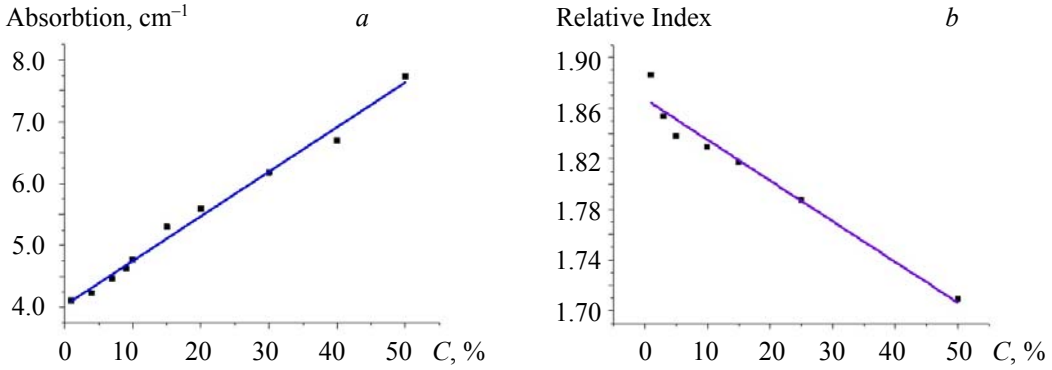

Fig. 3. The relationship between the concentration of potassium sorbate and the two optical parameters: (a) absorption coefficient, (b) refractive index of the samples (in 0.972 THz).

After matching, the linear regression function for the absorption coefficient is $y = 0.07219x + 4.02809$, and the linear regression function for the refraction index is $y = –0.00322x + 1.86707$, where the x coordinate represents the concentration values of potassium sorbate, and the y ordinate represents the average absorption coefficients and the refraction indexes of the samples in 0.972 THz.

As Fig. 3 shows, there is a linear relationship between the concentration values of potassium sorbate and the average absorption coefficients of the samples.

Based on the above discussion, we put the average absorption coefficients of the samples in the calibration set into the model to obtain the prediction values of concentration. Then, using the cross validation, method based on reserved concentration values in the calibration set, we obtained the cross-prediction concentration values and reference concentration values in the calibration set. Using Eqs. (8) and (9), we obtained the R values and RMSEC values of the calibration set, as well as those using the cross validations method. The results are shown in Table 1.

TABLE 1. Evaluation Parameters for the Two Regression Model

Parameter	Calibration set		Cross validation	
	R	RMSEC	R	RMSECV
Absorption coefficient	0.9904	0.1235	0.9803	0.5611
Refraction index	0.9623	0.40204	0.9547	0.55217

To evaluate the two proposed models further, we input the average absorption coefficients and the refraction indexes of the samples into the prediction set to obtain the prediction concentration values, and calculated the prediction error using Eq. (9). The prediction error values of the two models are about 5.7 and 7.1% respectively, which indicate that the models are effective for predicting concentrations using the least square method, and the regression model using absorption coefficients is more effective than that using refraction indexes. However, for the case where the concentration of potassium sorbate is under 1%, which is a very common situation, the characteristic absorption peak in 0.972 THz is not very obvious and even unidentifiable; thus, using the above models to predict concentration will produce relatively large errors. In order to solve this problem, it is necessary to use the chemometrics method to improve prediction performance; so, we propose the following model, called subinterval Partial Least Squares (PLS) based on the optimal THz wave band.

Subinterval PLS analysis. To improve the prediction accuracy under the condition of low content of potassium sorbate in milk powder, we divided the effective THz absorption spectrum (0.2 T–1.6 THz) into 16 equal subintervals. In each subinterval, an PLS algorithm was developed. The selected PLS model based on optimal subinterval was called Sub-PLS. The partition map of the full spectrum is shown in Fig. 5, where each histogram means the RMSECV value of each PLS model based on each subinterval, and the red line means the average absorption coefficient in the full spectrum.

In Fig. 4, the widths of all bar graphs are the same, and the height of each bar graph expresses the RMSECV value calculated using the Sub-PLS model. The eighth bar graph with the frequency range of 0.887–1.000 THz, which has the smallest RMSECV value, is selected as the optimal area for developing the Sub-PLS model. Experiments show that for predicting the concentration of potassium sorbate in milk powder, the prediction accuracy of the Sub-PLS is superior to the preceding OLS model.

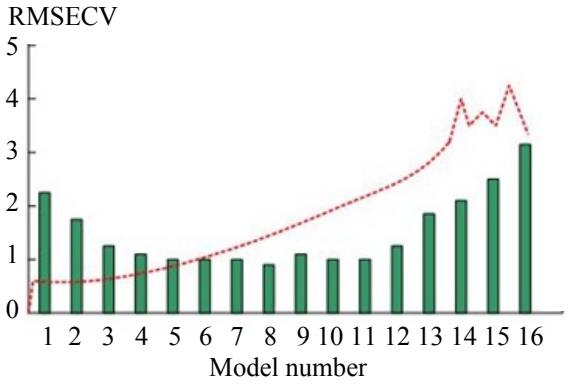

Fig. 4. RMSECV value of each PLS model and average absorption coefficient in the full spectrum.

We used the most effective part of the full spectrum (0.2–1.6 THz) to construct the Sub-PLS model to filter out noise or disturbance, which improved the prediction accuracy of the model. In the experiments we assigned a different number to each subinterval of the full spectrum to find the optimal subinterval. For example, we divided the full spectrum into 32, 16, and 8 intervals, and the optimal frequency bands of all divisions are shown in Table 2 by comparing RMSEC values and RMSECV values. As shown in Table 2, the optimal Sub-PLS model has the best RMSECV value (0.967) and RMSEC value (0.573).

TABLE 2. The Prediction Accuracy of the Optimal Sub-PLS Model

Interval number	Optimal frequency band, THz	RMSEC	RMSECV
32	0.731–0.956	0.768	1.260
16	0.887–1.000	0.573	0.967
8	0.675–1	0.837	1.237

Comparison with other algorithms. To verify the prediction performance of the proposed Sub-PLS model, we compared the Sub-PLS model to other common models that are widely applied in spectrum analysis. The comparative results are shown in Table 3. The Sub-PLS model has better performance in predicting low concentrations of potassium sorbate, and its accuracy reaches 92.8% due to maximum avoidance of noise or disturbance in constructing the model. The prediction results of other methods in Table 3 are all poor and below 82%. For different Sub-PLS models, using different frequency band, the prediction results are also very different. For example, in our experiments, the frequency band with the best prediction performance is in the range of 0.887–1.000 THz. In other words, comparison of the results, show that the prediction function of Sub-PLS model is superior to other models.

Incidentally, the Sub-PLS model proposed in this paper can also be applied to predict the presence of other food additives, such as sodium benzoate, sodium nitrite, and so on. In addition, for the application of other THz spectrum data, such as FIR spectrum [16–18], BWO spectrum [19, 20], spectrum based on CW parameters [21], and spectrum based on photomixing, the method used in this paper has also great potential and referential significance.

TABLE 3. Prediction Accuracy of Various Models

Model	Total prediction accuracy (%)	Positive prediction accuracy (%)	Negative prediction accuracy (%)
PCR	75.00	73.33	76.67
PLS	78.33	80.00	76.67
BP	81.67	76.67	86.67
Sub-PLS	92.08	91.67	93.33

Qualitative analysis. In most of cases, people care more about whether the milk powder contains potassium sorbate than the quantity of potassium sorbate in it. For this reason, we define samples with potassium sorbate content exceeding 1% as strong positive, samples with potassium sorbate content under 1% as weak positive, and samples with no potassium sorbate as negative, and identified these three classifications using partial least squares discrimination analysis (PLS-DA). PLS-DA is a multivariable statistical analysis method for discriminant analysis, the principle of which is forming a discrimination model by training different samples [22].

The judging steps using PLS-DA are as follows: construct the category variables using the calibration set; express category variables using the coding method; construct the model between the coding sequences and the spectral data; calculate the predicted value Y_p and the bias value D. The judgment standard is as follow: if $Y_p > 0.5$ and $D < 0.5$, then the sample belongs to this category; if $Y_p < 0.5$ and $D < 0.5$, then the sample does not belong to this category; if $D \geq 0.5$, then the model is unstable. Before constructing the PLS-DA model, we first coded the three categories (strong positive, weak positive, negative) as [100], [010], [001] respectively, and then constructed the PLS-DA model between the spectra and the category variables using the PLS regression method.

Table 4 shows that the model constructed using PLS regression has good correlation between the spectra and the category variables. All the correlation coefficients for the predicted category variables Y (Y_p) and reference category variables Y (Y_r) are above 0.94, which indicate that the model has a good degree of fitting.

TABLE 4. Calibration and Validation Results of PLS-DA Judgment Model

Sample set	Model parameters	Strong positive	Weak positive	Negative
Calibration set	R^2	0.963	0.914	0.967
	Correlation	0.976	0.943	0.981
	Bias	2.73×10^{-6}	-1.88×10^{-6}	2.11×10^{-6}
Validation set	R^2	0.957	0.908	0.959
	Correlation	0.967	0.925	0.960
	Bias	3.28×10^{-6}	4.33×10^{-6}	-2.79×10^{-6}

To verify the identification precision of the sample types, 100 samples in our test set were used to compare the performance of four algorithms titled BP, SVM, RBF, and PLS-DA, respectively. Out of 100 samples, there were 70 negative samples and 30 positive samples, including 10 strong positive samples. Four parameters, named true positive (TP), false positive (FP), true negative (TN), and false negative (FN), which were widely applied in the binary classification situation, were used in the experiments to evaluate the performance of the model. For the sake of simplicity, we divided all classifications into two kinds of binary classifications: positive and negative, and strong positive and weak positive. The four values of each classification algorithm for the two kinds of binary classifications are shown in Tables 5 and 6.

Three indicators were used to evaluate the classification performance of the four methods. They are sensitivity, specificity, and accuracy. Sensitivity represents the identification ability of the classifier to positive samples and can be calculated using the formula TP/P, where P is the total number of positive samples and is equal to the indicator recall, which is another common index. Specificity represents the identification ability of the classifier to negative samples and can be calculated using the formula TN/N, where N is the total number of negative samples. Accuracy is the most important evaluation indicator, which represents the performance of the classifier, and can be calculated using the formula (TP+TN)/(P+N). The values of the three indexes of each method for the two binary classifications are shown in Table 6.

TABLE 5. The values of TP, FP, TN, and FN of Each Classification Algorithm for Classification of Positive and Negative (100 samples, N = 70, P = 30)

Predicted value	Algorithm	True value	
		Positive	Negative
Positive	BP	23	8
	RBF	26	6
	SVM	25	4
	PLS-DA	28	1
Negative	BP	6	62
	RBF	3	65
	SVM	3	66
	PLS-DA	1	68

TABLE 6. The Values of Sensitivity, Specificity, and Accuracy for Each Classification Algorithm

Algorithm	Sensitivity (TP/P)	Specificity (TN/N)	Accuracy (TP+TN)/(P+N)
BP	0.77	0.87	0.85
RBF	0.87	0.93	0.91
SVM	0.83	0.94	0.91
PLS-DA	0.93	0.97	0.96

Table 6 shows that the PLS-DA method has the best classification accuracy. The SVM and RBF methods are also good, but the BP method is the worst. The over-fit phenomena of BP neural network limit its generalization ability. The reason that the classification accuracy of PLS-DA is better than SVM or RBF is that this method removes some redundant information or disturbance using PCA optimizing to make the data used in the classification to reflect the characteristic of the samples more effective. In particular, the experimental results show that when the potassium sorbate content in samples is lower than 0.05%, the probability of generating incorrect results is greatly increased, so we defined the detection limit as 0.05% in this research work.

Conclusions. Based on the THz absorption spectrum, we constructed the OLS regression model and the Sub-PLS model for the qualitative and quantitative analysis of potassium sorbate concentration in milk powder. It can be seen from the above discussion that the performance of the Sub-PLS model mainly is determined by the selected frequency band, and the appropriate THz band will help to improve the prediction accuracy. If the width of the selected band is too wide, the noise or irrelevant information will increase, which inevitably reduces the prediction accuracy. However, if the width of the selected band is too narrow, some characteristic features will be lost, which reduces the prediction accuracy as well. The prediction accuracy of the established models probably is affected by THz absorption spectrum itself. The retrieved THz spectrum may be affected by the experimental environment, experimental procedure, background noise, and other factors. Similarly, the samples themselves have also an impact on their THz spectra, such as the manufacturing conditions of the samples or the components of the samples. To construct better regression models, it is necessary to consider these factors.

Acknowledgments. This work was supported by the Key scientific research projects of institutions of higher learning (18A51003), Henan, China, Ministry of Education, Zhengzhou, China, and the National Natural Science Foundation under Grant 61805072, and the National Key Research and Development Program of China under Grant 2017YFD0401003. We thank Lijian Zhang, postgraduate, from Henan University of Technology, for helping us in some of the experiments. We also thank Hongyi Ge and Yuying Jiang, who worked for Henan University of Technology, for suggestions in the preparation of the manuscript. The authors declare no conflict of interest.

REFERENCES

1. National Standards for Food Safety/National Standard for Food Additives (GB 2760-2014), *The State Standard of the People's Republic of China*, **2**, 186–187 (2015).
2. R. Ascensión, S. Isabel, O. Manuel, G. Rafael, *J. Chem*., **38**, 1–8 (2014).

3. C. X. Yuan, Y. Y. Xie, R. S. Jin, *Food Anal. Methods*, **11**, 1–7 (2017).
4. M. G. Qian, H. Zhang, L. Q. Wu, *Food Chem.*, **166**, 23–28 (2015).
5. J. F. Song, C. Han, *Cereals Oils*, **5**, 83–87 (2007).
6. F. S. Vieira, C. Pasquini, *Anal. Chem.*, **86**, 3780–3786 (2014).
7. B. Ferguson, X. C. Zhang, *Physics*, **1**, 26–33 (2002).
8. J. Qin, L. Xie, Y. Ying, *Anal. Chem.*, **86**, N 23, 11750–11757 (2014).
9. Y. Cui, K. Mu, X. Wang, Y. Zhang, *Proc. SPIE*, **7385** (2009); doi: 10.1117/12.835293.
10. H. Naito, Y. Ogawa, K. Shiraga, N. Kondo, *IEEE/SICE Int. Symp. System Integration* (2011); doi: 10.1109/SII.2011.6147444.
11. K. Q. Wang, D. W. Sun, H. B. Pu, *Trends Food Sci. Technol*., **67**, 93–105(2017).
12. K. Li, J. X. Peng, E. W. Bai, *IEEE Trans. Autom. Control*, **50**, N 80, 1211–1216 (2005).
13. L. H. Xue, *China J. Guiyang Univ.*, **10**, N 4, 9–10 (2015).
14. J. R. Wang, Z. Y. Zhang, Z. W. Zhang, *China Spectrosc. Spectr. Anal.*, **36**, N 2, 316–321 (2016).
15. L. Z. Tan, S. Zhang, W. S. Zhong, *China Ship Electron. Eng.*, **1**, 20–22 (2016).
16. D. K. Pradhan, P. Misra, V. Sreenvas, *J. Appl. Phys*., **115**, N 24, 243904-1–243904-9(2014).
17. B. P. Gorshunov, V. I. Torgashev, E. S. Zhukova, *Nat. Commun*., **7**, 12842 (2016).
18. E. S. Zhukova, V. I. Torgashev, B. P. Gorshunov, *J. Chem. Phys*., **140**, N 22, 224317 (2014).
19. G. A.Komandin, S. V. Chuchupal, S. P. Lebedev, *IEEE Trans. Ter. Sci. Tech*., **3**, N 4, 440–444 (2013).
20. A. S. Prokhorov, E. S. Zhukova, A. A. Boris, *Quantum Electron*., **56**, N 8-9, 620–627 (2014).
21. S. Hayashi, K. Nawata, T. Taira, *Sci. Rep*., **4** (2014).
22. L. L. Chuen, L. Choong-Yeun, J. A. Aziz, *Analyst*, **143**, N 15, 3526–3539 (2018).