# NON-DESTRUCTIVE DETECTION OF MELAMINE IN MILK POWDER BY TERAHERTZ SPECTROSCOPY AND CORRELATION ANALYSIS ALGORITHM

**X. Sun*, K. Zhu, J. Hu, X. Jiang, Y. Liu**

*School of Mechatronics & Vehicle Engineering, East China Jiaotong University, Nanchang, 330013, China; e-mail: 874916937@qq.com*

*Investigations were initiated for developing a rapid and non-destructive detection method to measure the illegal additive of melamine into milk powder by using terahertz (THz) spectroscopy and the correlation analysis algorithm. The absorption coefficients exhibited a maximum absorption peak at 2.04 THz, which would normally increase along with the concentration of melamine additive. In the current study, correlation analysis was carried out to select a pair-variable at 2.04 and 2.34 THz for improving the predictive ability of the multiple linear regressions (MLR) model. Compared with the partial least square (PLS) model in full spectrum, the MLR model for powder samples could be considered successful in terms of quality control of milk powder with correlation coefficient ($R^2$) of 0.97 and root mean square error of prediction (RMSEP) of 1.38%. At the same time, the MLR model was simple and easier to interpret than the PLS one. The results of the research suggested that THz spectroscopy in combination with the correlation analysis algorithm has a significant potential in the quantitative analysis of the illegal additive of melamine in milk powder.*

***Keywords:*** *terahertz spectroscopy, melamine, correlation analysis, food safety.*

# НЕРАЗРУШАЮЩИЙ МЕТОД ОБНАРУЖЕНИЯ ДОБАВКИ МЕЛАМИНА В СУХОМ МОЛОКЕ С ИСПОЛЬЗОВАНИЕМ ТЕРАГЕРЦОВОЙ СПЕКТРОСКОПИИ И АЛГОРИТМА КОРРЕЛЯЦИОННОГО АНАЛИЗА

**X. Sun*, K. Zhu, J. Hu, X. Jiang, Y. Liu**

*УДК 543.42:547.491.8*

*Школа мехатроники и автомобилестроения, Восточно-китайский университет Цзяотун, Наньчан, 330013, Китай; e-mail: 874916937@qq.com*

*(Поступила 19 июля 2018)*

*Разработан быстрый и неразрушающий метод обнаружения добавки меламина в сухом молоке с использованием терагерцовой (ТГц) спектроскопии и алгоритма корреляционного анализа. Коэффициенты поглощения демонстрируют максимум при 2.04 ТГц, который обычно увеличивается вместе с концентрацией меламиновой добавки. Проведен корреляционный анализ для выбора пар переменных на 2.04 и 2.34 ТГц для улучшения прогнозирующей способности модели множественных линейных регрессий (MLR). По сравнению с моделью частичного наименьшего квадрата (PLS) в полном спектре модель MLR для образцов порошка показывает лучшие результаты с точки зрения контроля качества сухого молока с коэффициентом корреляции $R^2$ = 0.97 и среднеквадратичной ошибкой прогноза RMSEP = 1.38 %. В то же время модель MLR проще для интерпретации, чем PLS. Результаты исследования позволили предположить, что ТГц спектроскопия в сочетании с алгоритмом корреляционного анализа обладает значительным потенциалом в количественном анализе нелегальной добавки меламина в сухое молоко.*

***Ключевые слова:*** *терагерцовая спектроскопия, меламин, корреляционный анализ, безопасность пищевых продуктов.*

**Introduction.** Melamine (1,3,5-triazine-2,4,6-triamine), as a nitrogen-rich chemical substance with 66.6% nitrogen by weight, is sometimes used to fraudulently increase the perceived protein content in a variety of food products, such as milk, infant formula, pet food, biscuits, candy, and coffee drinks [1, 2]. Digestion of melamine by pets, cattle, and infants tends to cause serious health problems that may damage the reproductive system and kidneys, and in some cases even lead to death. Since compounds rich in nitrogen can mimic a high protein concentration, it is difficult to use standard methods (Kjeldahl, Dumas, and modified Lasssaigne) to distinguish between nitrogen from protein and non-protein source, leading to the widespread use of this chemical to adulterate milk-based products [3]. Therefore, a reliable and high-throughput screening method to detect this contaminant in food is required, especially when milk-based products are concerned.

Terahertz (THz) spectroscopy can be applied to detect target chemicals as a high-throughput screening and quantitative analysis method, with the advantages of fast analysis and simplicity [4]. Many biological molecules have unique spectral fingerprints in the frequency range of 0.1–3.0 THz, meaning that THz spectroscopy can be used to identify and to analyze them [5, 6]. Ever since THz spectra uncovered the linear relationship with overlapping and complexity of signals, quantitative analyses of target components have often been carried out using some chemometric methods [7]. Partial least squares (PLS) regression has also been used in the quantitative analysis of multi components based on the absorption coefficient in THz band, such as amino acid, tetracycline hydrochloride, and aflatoxin B1 [8–11]. Many biological molecules have unique fingerprint absorption peaks in THz band, indicating that multiple linear regression (MLR) can be used for the identification and the analysis of target chemicals with as few variables as possible in the process of handling THz spectra. In comparison with the full-spectra modelling method of PLS, MLR is more concise and easier to interpret [12]. Generally, a robust MLR model may be developed with a pair of spectral variables. However, there has been no publication on the application of a pair of THz spectral variables to quantitatively detect melamine in milk power.

Against this background, we were attempted to find the best frequency pairs by correlation analysis and to develop a robust MLR model. Through the research, the effect was evaluated for the detection of melamine in milk powder.

**Experimental.** *Samples preparation.* Melamine powder was purchased from Sigma Aldrich Corporation and used without further purification; milk powder was purchased from a local supermarket and was validated to be rid of melamine powder through high performance liquid chromatography (HPLC) analysis; the melamine power samples were crushed into small particles that were sufficiently smaller than the THz wavelength with the purpose of reducing the baseline offsets at higher frequencies. These particles were then mixed carefully with milk powder at several different concentrations (from 0.50% to 19.99%, g/100g), and three replicates were prepared for each concentration. Then the mixture was compressed into pellets with a diameter of 13 mm under a pressure of 10 MPa using a tablet press. The mechanically determined pellet thickness ranged from 1 to 2 mm, aiming at providing a sufficient path length in order to eliminate the effect of multiple reflections that occurred between the two surfaces of the pellet sample in the spectra.

*THz measurement.* The absorption spectra were recorded with the TAS7500SU THZ-time domain system (THz-TDS) provided by Advantest Corporation that worked in transmission mode. Details of this system are published in the literature [13]. The system includes two ultra-short pulse fiber lasers, which are ensured to be under synchronized control. The central wavelength and the maximum output power of these pulses are 1550 nm and 50 mW, respectively. These pulses provide extremely short pulse width less than 50 fs and low jitter below 50 fs. The system could achieve a sampling rate of 8 ms per scan and an ultra-wide frequency band extending to 7 THz. The experiment was carried out at room temperature under in a dry-air purged container with a relative humidity of 0%. Three measurements were recorded for each sample in order to reduce the possibility of random error. The reference waveform was collected when the THz pulses passed through a sample holder without any sample mounted in it.

*Parameters extraction.* A fast Fourier transform (FFT) is adopted to acquire the spectral distribution of the THz pulse in the frequency and described as

$$\tilde{E}(\omega) \equiv A(\omega)e^{-\phi(\omega)} = \int E(t)e^{-i\omega t}dt, \quad (1)$$

where $A(\omega)$ is the amplitude of the electric field, $\phi(\omega)$ is the phase of the electric field, and $E(t)$ is the time domain waveform.

The absorption coefficient ($\alpha$) of the sample could be calculated as

$$\alpha = (1/d)\ln(A_R/A_S), \quad (2)$$

where $A_R$ and $A_S$ are the amplitude of the reference and sample signal, respectively, and $d$ is the thickness of the sample.

*Correlation analysis.* A simple correlation analysis method using two-wavelength ratios and differences is used in order to find the best pair of wavelengths for the quantitative analysis of melamine [14]. A total of 260 spectral variables in the range of 0.75–2.73 THz is adopted for the following analysis. The correlation value $R^2$ of the first pair wavelengths is calculated as

$$R^2 = \frac{r_1^2 + r_2^2 - 2r_1 \times r_2 \times r_x}{1 - r_x^2}, \quad (3)$$

where γ is the melamine concentration, $x_1$ and $x_2$ are the absorption coefficients at frequencies 1 and 2 respectively, $r_1$ is the simple correlation between γ and $x_1$, $r_2$ is the simple correlation between γ and $x_2$, and $r_x$ is the simple correlation between $x_1$ and $x_2$. The result $R^2$ is the squared correlation for the two-variable equation predicting γ from $x_1$ and $x_2$. Correlation analysis and MLR were carried out via Matlab software.

**Results and discussion.** *Statistics of measured melamine concentration.* The melamine concentration values of 160 samples varied from 0.50% to 19.99%, the distribution of which was approximately normal around the averaged value of 10.25%. All samples were sorted according to the melamine concentration values with the purpose of avoiding bias in the subset divisions. One in every four samples was then divided into the prediction set in accordance with the rule that the melamine concentration values range of the calibration set should cover that prediction set. In the final stage, the samples were divided into calibration and prediction sets with the ratio around 3:1 for modelling the applicable model. There were 121 samples in the calibration sets. The remainder is then partitioned into the prediction set. Specific statistics of the calibration and prediction sets are listed in Table 1.

TABLE 1. Statistics of Calibration and Prediction Sets of Melamine Illegal Additive in Milk Powder

| Data set | $N$ | Range, % | Mean, % | SD, % | CV, % |
|---|---|---|---|---|---|
| Total | 160 | 0.50–19.99 | 10.25 | 5.79 | 56.49 |
| Calibration | 121 | 0.50–19.99 | 10.33 | 5.84 | 56.53 |
| Prediction | 39 | 0.50–19.99 | 10.00 | 5.70 | 57.00 |

N o t e. $N$ number of samples, SD standard deviation, CV coefficient of variation.

*Analysis of spectral properties.* The absorption coefficients of the mixture, milk powder and melamine samples in the 0.75–2.73 THz frequency region are shown in Fig. 1. Areas below 0.75 THz or beyond 2.73 THz were considered to be ineffective data, judged by a relatively low signal-to-noise ratio (SNR). Three absorption bands were observed in melamine and the mixture of melamine and milk powder in the frequency range of 2–2.75 THz. The absorption coefficient peaks at 2.04 and 2.28 THz for melamine and mixture were similar to those in the literature [15, 16]. The peak at 2.04 THz was attributed to the translational lattice vibration caused by the six hydrogen bonds stretching between intra-layer molecules [17]. By contrast, the peak at 2.28 THz was much more complicated, which may be clue to the rotation and torsion lattice vibration, resulting from the combination of hydrogen bonds, including inter-layer ones, and stretching and π-π stacking vibration. The peak at 2.68 THz was observed in our THz-TDS measurement and could be attributed to the translational lattice vibration caused by the π–π stacking vibration between the center molecule and a molecule from the lower layer. However, the peak at 1.43 THz may be caused by the milk powder. The absorption coefficient generally rose along with increase in melamine content. Therefore, the peaks of 2.04, 2.28, and 2.65 THz could be used to investigate the relationship between these significant peaks and melamine concentrations by adopting MLR method. The limit of detection (LOD) values were 15.00, 19.47, and 32.25% at 2.04, 2.28, and 2.65 THz. These results showed that a single peak was not qualified in predicting melamine in milk powder, because the intensity of a single peak might change slightly with the status of the sample. Hence, the combinations of peaks should be considered with the purpose of improving the predictive ability of the models.

MLR models were assessed by correlation coefficient ($R^2$), root mean square error of prediction (RMSEP), and LOD. A better MLR model would obtain higher $R^2$ and lower RMSEP and LOD values. According to this principle, further investigation should be executed to mine THz spectra for accuracy improvement. The limit of detection (LOD) with 99.86% confidence interval can be calculated from the MLR calibration curve based on the significant peaks in THz absorption coefficient [18]

$$\mathrm{LOD} = 3\sigma/m, \tag{4}$$

where $\sigma$ is the standard error of the predicted concentration and $m$ is the slope of the calibration curve in the MLR model, $\sigma$ equals to RMSEP.

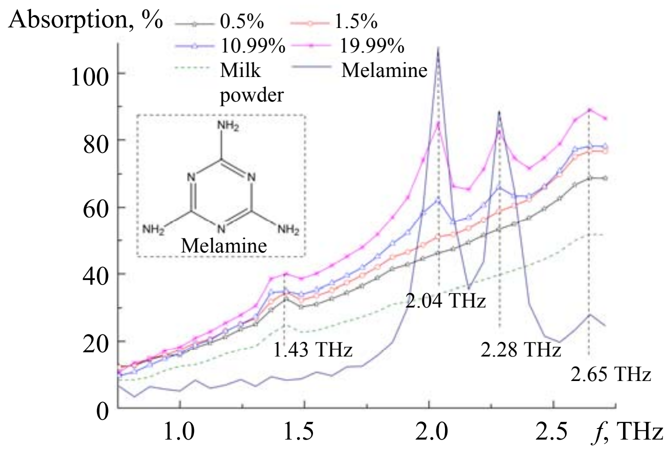


Fig. 1. Absorption spectra of melamine, milk powder and mixture in the region of 0.75–2.73 THz.

*MLR model with combinations of significant peaks.* The performance of the MLR model built with a single variable was easily affected by baseline shift, which arose from the scattering effect in the spectrum measurement. This interference could be reduced by increasing the number of variables. For improving the accuracy of the MLR model, several combinations of significant peaks were adopted to develop MLR models (Table 2). The MLR model with $R^2$ of 0.96, RMSEP of 1.62%, and LOD of 5.45% achieved better results than the others, which was built with the significant peaks of 2.04 and 2.65 THz. Compared with the PLS method in full spectrum, the MLR method was more suitable when the fingerprints spectrum was concerned because it was simple and easier to interpret. In this case, the accuracy of the MLR model was close to the PLS one. The $R^2$, RMSEP, and LOD of PLS model were 0.98, 1.21, and 3.78%, respectively. An alternative method was proposed to select a pair of significant variables automatically by correlation analysis, which was reasonable and could avoid artificial errors.

TABLE 2. Results of MLR Models Developed with Combinations of Significant Peaks in Calibration and Prediction Sets

| Frequency, THz | Model | Calibration | | Prediction | | LOD, % |
|---|---|---|---|---|---|---|
| | | $R^2$ | RMSEC, % | $R^2$ | RMSEP, % | |
| 2.04, 2.28 | $y = 0.019x_{2.04} - 0.016x_{2.28} + 0.03$ | 0.95 | 1.86 | 0.96 | 1.76 | 6.47 |
| 2.04, 2.65 | $y = 0.009x_{2.04} - 0.006x_{2.65} + 0.05$ | 0.96 | 1.59 | 0.96 | 1.62 | 5.45 |
| 2.28, 2.65 | $y = 0.013x_{2.28} - 0.011x_{2.65} + 0.06$ | 0.96 | 1.65 | 0.93 | 2.04 | 6.60 |

*Correlation analysis.* A simple correlation analysis was carried out in order to find a pair of absorption coefficients where the difference gave the best correlation for determining the melamine concentration of mixture samples, for the purpose of which all possible combinations of pair variables were tested in the extended regions of 0.75–2.73 THz. The pair variables that had the highest $R^2$ value were then selected as the best combination. Contour plots of the correlation coefficients for the two-variable ratios and difference are shown in Fig. 2a. For the ratios, four regions with relatively high correlation coefficient were observed. The highest correlation coefficient of 0.94 occurred at 2.04 and 2.34 THz. The absorption coefficient responses indicated the maximum absorption peak to be located at 2.04 THz, which presented the best linear behavior with a melamine concentration in the range of 0.75–2.73 THz. The reason for the 2.34 THz peak could be explained by the coefficient of variation (CV) plots in Fig. 2b. CV, defined as the ratio of standard deviation to mean, is a standardized measure of dispersion of a probability distribution or frequency distribution. The peak of 2.34 THz was adjacent to the location of the averaged CV at 2.38 THz, and its function was to modify the MLR model.

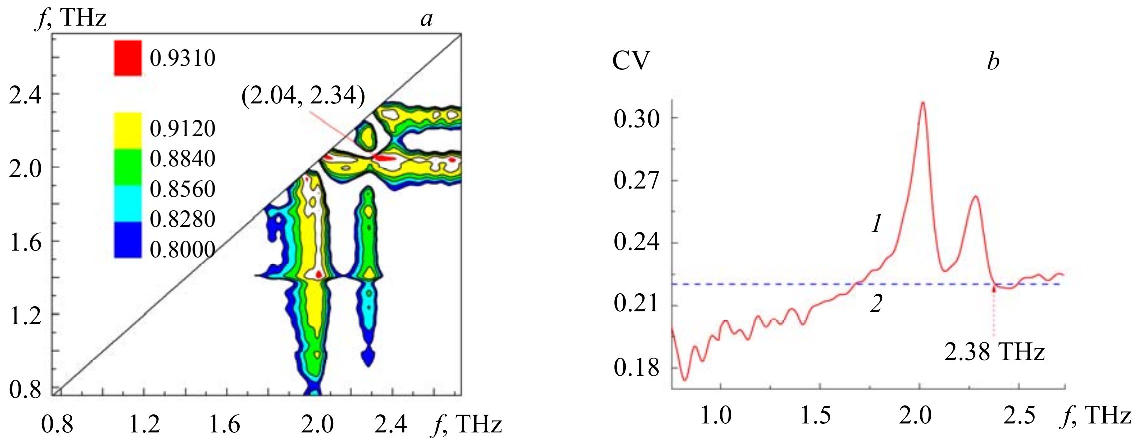


Fig. 2. Contour plots of correlation coefficient arising from correlation between a spectral variable pair combination (a) and plots of coefficient of variation (CV) (1) and averaged CV (2) in range of 0.75–2.73 THz (b).

*Comparison of MLR and PLS models.* The best MLR model was achieved with a pair of variables located at 2.04 and 2.34 THz. By comparison, the PLS model was also developed in full-spectrum region of 0.75–2.73 THz. The practical predictive ability of the MLR and PLS models was subsequently evaluated with 39 unknown samples in the prediction set which had not been used in the calibration. Figure 3 demonstrates the scatter plot between reference and THz measurement in the prediction set. The performance of these models can be evaluated via two statistical parameters, RMSEP and $R^2$. The lower the RMSEP is, the more accurate the predictions are. The mean error is to be preferred for the whole population rather than for a single sample. The $R^2$ between actual and predicted values is another typical parameter for assessing the performance of models. A value of $R^2$ close to 1 indicates a good linear dependence between the actual and the predicted values, namely the good predictability of the model. Compared with PLS, the MLR model could predict the accuracy with an RMSEP of 1.38% and $R^2$ of 0.97. In addition, the MLR model is simple and easier to interpret compared with the PLS one. To sum up the comparison analysis above, the RMSEP of 1.38% indicated that correlation and THz spectroscopy could qualify for predicting melamine in milk powder.

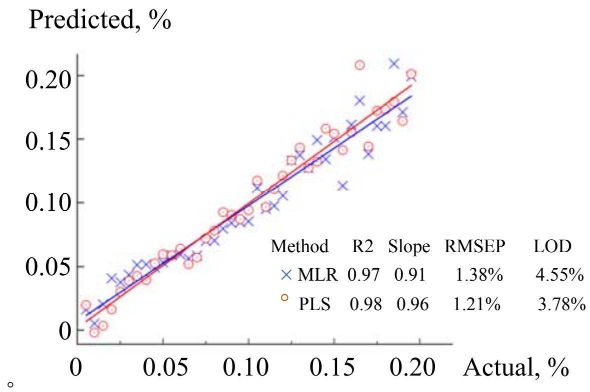


Fig. 3. Predicted results of PLS and MLR models.

**Conclusion.** The applicability of THz spectroscopy in combination with the correlation analysis for the detection of melamine has been presented. Results of the research showed that melamine exhibited three absorption peaks in the range of 0.75–2.73 THz and a maximum absorption peak located at 2.04 THz. A pair of absorption coefficients at 2.04 and 2.34 THz was adopted with the purpose of developing a robust MLR model. The performance of the MLR model was close to the PLS one with an RMSEP of 1.38%. This research indicated that THz spectroscopy with correlation analysis method could be used as a screening tool to rapidly and quantitatively detect melamine without the need for expensive and laborious chemical analysis.

**Acknowledgment.** The authors gratefully acknowledge the financial support of the Outstanding Youth Talent Program of Jiangxi Province (20171BCB23060) and the Education Department Project of Jiangxi Province (GJJ160478).

**Conflict of interest.** X. Sun, K. Zhu, J. Hu, X. Jiang, and Y. Liu declare that he has no conflict of interest.

**Ethical approval.** This article does not contain any studies with human or animal subjects.

## REFERENCES


1. S. K. Mathanker, P. K. Weckler, N. Wang, *Trans. ASABE*, **56**, 1213–1226 (2013).
2. S. Dhakal, K. Chao, J. Qin, M. Kim, D. Chan, *J. Food Meas. Char*., **10**, 374–386 (2016).
3. E. Domingo, A. A. Tirelli, C. A. Nunes, M. C. Guerreiro, S. M. Pinto, *Food Res. Int*., **60**, 131–139 (2014).
4. J. Qin, Y. Ying, L. Xie, *Appl. Spectrosc. Rev*., **48**, 439–457 (2013).
5. A. Redo-Sanchez, G. Salvatella, R. Galceran, E. Roldós, J. García-Reguero, M. Castellari, J. Tejada, *Analyst*, **136**, 1733–1738 (2011).
6. W. Liu, C. Liu, J. Yu, Y. Zhang, J. Li, Y. Chen, L. Zheng, *Food Chem*., **251**, 86–92 (2018).
7. S. H. Lu, B. Q. Li, H. L. Zhai, X. Zhang, Z. Y. Zhang, *Food Chem*., **246**, 220–227 (2018).
8. Z. Li, A. Guan, H. Ge, F. Lian, *Microchem. J*., **132**, 185–189 (2017).
9. H. Ge, Y. Jiang, F. Lian, Y. Zhang, S. Xia, *Food Chem*., **209**, 286–292 (2016).
10. J. Qin, L. Xie, Y. Ying, *Food Chem*., **170**, 415–422 (2015).
11. Y. Ma, Q. Wang, L. Li, *J. Quant. Spectrosc. Radiat. Transfer,* **117**, 7–14 (2013).
12. K. Liu, X. Chen, M. Li, H. Chen, X. Ruan, W. Liu, *Anal. Chim. Acta,* **858**, 16–23 (2015).
13. L. Jiang, M. Li, C. Li, H. Sun, L. Xu, B. Jin, Y. Liu, *J. Infrared Millim. Terahertz Waves,* **35**, 871–880 (2014).
14. K. Lee, S. Kang, S. R. Delwiche, M. S. Kim, S. Noh, *Sens. Instrum. Food Qual*., **2**, 90–96 (2008).
15. S. H. Baek, H. B. Lim, H. S. Chun, *J. Agric. Food Chem*., **62**, 5403–5407 (2014).
16. Y. H. Hwang, Y. H. Noh, D. Seo, H. M. Yeo, S. Kim, J. Park, H. S. Chun, K. Kwak, *Bull. Korean Chem. Soc*., **36**, 891–895 (2015).
17. M. He, M. Li, Z. Tian, W. Cao, J. Han, *Nucl. Sci. Technol*., **23**, 209–214 (2012).
18. B. Liu, P. Zhou, X. Liu, X. Sun, H. Li, M. Lin, *Food Bioprocess Technol.*, **6**, 710–718 (2013).