

The Application of Classical Least Square Algorithm in the Quantitative Analysis of Lime in Wheat Flour by ATR-MIR Spectroscopy

Dong Wang^{1,2,3}, Donghai Han¹, Zhihong Ma³, Ligang Pan³, Liu Zhao³, Ping Han³, Jihua Wang^{2,3*}

¹ College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, P. R. China

² Beijing Research Center for Information Technology in Agriculture, Beijing 100097, P. R. China

³ Beijing Research Center for Agri-food Testing and Farmland Monitoring, Beijing 100097, P. R. China

{wangd@nrcita.org.cn wangjh@nrcita.org.cn}

Abstract. In this thesis, classical least square (CLS) regression was applied in the attenuate total reflection mid-infrared (ATR-MIR) spectra data processing of lime in wheat flour to develop the corrected peak height model of the raw spectra, the corrected peak area model of the raw spectra, the corrected peak

Supported by Beijing Postdoctoral Research Foundation; China Postdoctoral Science Foundation (49th) (No. 20110490318); Postdoctoral Science Foundation of Beijing Academy of Agriculture and Forestry Science; Special Research of Public Welfare Industry, the Ministry of Agriculture of People's Republic of China (No. 201203046) and National High Technology Research and Development Program 863 (2010AA10Z201).

Author: Dong WANG, Ph.D., major: NIR spectroscopy analysis and NIR micro-imaging analysis, College of Food Science & Nutritional Engineering, China Agricultural University, Beijing 100083, P. R. China, Email: wangd@nrcita.org.cn;

* Corresponding Author: Jihua WANG, Researcher, Ph.D., major: Agricultural information technology, Beijing Research Center for Information Technology in Agriculture, Beijing 100097, P. R. China, Email: wangjh@nrcita.org.cn.

Address: Beijing 2449-26 mailbox, City of Beijing, P. R. China, Postcode: 100097 (Beijing Research Center for Information Technology in Agriculture).

height model of the 2nd derivative spectra and the corrected peak area model of the 2nd derivative spectra respectively. The result indicated that the correlation coefficients of the four models mentioned above are 0.9648, 0.9696, -0.9646 and -0.9599 respectively. *F-test* result indicated that a very remarkable correlation exists between the estimated and specified values of the calibration set and external validation set. The detection limits of the four models mentioned above are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which can fulfill the demand for the rapid quality safe screening of wheat flour in the market. This method, to some extent, can provide some references for not only the design and manufacturing of the special MIR instrument for the quality safe control of wheat flour in the market but also the quantitative determination of banned additives in wheat flour.

Key words: Classical Least Square, Attenuate Total Reflection, Mid-Infrared Spectroscopy, Wheat Flour, Lime

1 Introduction

Wheat flour is one of the important food sources for human beings, of which the quality safety is highly focused by all quarters of society. However, the quality safety incidents about wheat flour have been reported frequently in recent years, such as lime added in it, which results in not only negative influence but also being harmful to the customers' health. Therefore, the development of rapid, simple, practical and low-cost determination method for detecting lime in wheat flour becomes particularly urgent and very important.

Among the common instrument analytical methods, mid-infrared (MIR) spectroscopy analysis is widely applied in the fields of the identification and characterization of microbes[1], detection and identification of bacteria[2], quality analysis of edible oil[3, 4], quantitative determination of anthocyanins in cherry[5] with the advantages of distinguishable peaks, strong characteristics, high sensitivity, rapid analysis and non-invasive analysis.

However, the traditional MIR analysis involves the sample preparation such as tableting with KBr, which is difficult for the operators. Attenuated total reflection (ATR) technology can acquire the information of samples according to the reflective signals of the sample surface, which is free of sample preparation. ATR technology

has the advantages of high sensitivity, distinguishable peaks, user-friendly control and non-interference by humidity. Based upon the advantages mentioned above, ATR technology is now being widely applied in the fields of analysis of juice sweetness[6], biology structure studies[7], material study[8] and so forth. Therefore, it is convenient to collect the MIR spectra of powder samples, such as wheat flour, by ATR technology.

It is known by the experiment data that lime, such as calcium oxide, calcium hydroxide, has the characteristic absorption peaks in the MIR range, which is the theoretical foundation of the qualitative and quantitative analysis to it. The characteristic absorption peak in MIR range of wheat flour is far away from that of lime, i.e. the interference from wheat flour is not very obvious. Based on the MIR spectral characteristics of wheat flour and lime, CLS algorithm has a certain potential application in the quantitative analysis of lime in wheat flour by MIR spectra data.

Classical least square (CLS) regression is a data processing method applied in the data regression of analytical chemistry commonly[9, 10]. The basic principle of CLS is that to calculate the evaluated values of the regression coefficients between the independents and dependents according to a group of obvious values in order to decrease the errors between the specified values and the predicted values furthest. The obvious errors of CLS obey three hypotheses, i.e. the obvious values are independent each other with the same precision, the obvious errors are non-systematical, the obvious errors are as the law of normal distribution[11]. Compared to partial least square regression and principal component regression, CLS regression is easy to understand relatively. Meanwhile, it is easier to achieve CLS algorithm on the special MIR detection instruments, which is in favor of the promotion of this technology.

In this research, ATR-MIR spectra of the wheat samples mixed with lime were collected. The quantitative determination of lime in wheat flour by ATR-MIR combined with CLS algorithm was studied. The corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra were developed respectively. The external validation set was employed to evaluate the prediction performance of the models. This method is of great importance to the rapid quantitative screening of lime in wheat flour. This research, to some extent, will provide some references to not only the design and manufacturing the special MIR spectrometer for the quality safety screening of wheat flour, but also the quantitative determination of other banned additives in wheat flour.

2 Experiment

2.1 Instrument and Materials

Instrument: FT-MIR spectrometer, type: Spectrum 400, Perkin Elmer, U.S..

The ATR-MIR spectra are collected by FT-MIR spectrometer combined with ATR accessory, of which, the reference spectrum is collected by air, the wavenumber range is 4000 cm^{-1} - 540 cm^{-1} , the wavenumber resolution is 4 cm^{-1} , the time of accumulation is 240. The reference spectrum is collected once every 30 minutes.

Materials: wheat flour milled by the winter wheat from Changping experimental base of Beijing, without any additives. Calcium oxide and calcium hydroxide: A.R..

2.2 Method

The quantitative calibration models of lime are developed by the corrected peak height and corrected peak area of lime in ATR-MIR spectra combined with the algorithm of classical least square. The external validation set, of which the average and standard deviation are similar to the calibration set, is employed to evaluate the prediction performances of the models. The contents of lime in both the calibration set and external validation set are distributed in the range of 0.6 % - 16.0 % uniformly. The mass of each sample is about 50 grams. The mass data of wheat flour and lime in each sample are recorded according to analytical balance accurately. The information of the content of lime in the samples is showed in table 1.

Tab. 1. The information of the content of lime in the wheat flour samples (%)

Items	Calibration set	External validation set
Sample size	30	10
Minimum	0.6831	0.9643
Maximum	15.8145	14.1465
Average	5.2319	5.3556
Standard deviation	4.36	4.49

In order to decrease the error of peak height and peak area measurement, the

corrected peak height / area models of the 2nd derivative spectra are discussed despite the corrected peak height / area models of the raw spectra. The 2nd derivative [12] is one of the data preprocessing method that is applied in the data preprocessing in MIR data commonly, which can eliminate the baseline drifting and baseline translation.

The corrected height is the distance between the peak top of the target peak and the connecting line of the two base points around the target peak, while the corrected area is the integral value between the spectrum curve and the connecting line of the two base points around the target peak. The sketch maps of the measurement of corrected peak height and corrected peak area of the raw and the 2nd derivative spectra are showed as figure 1.

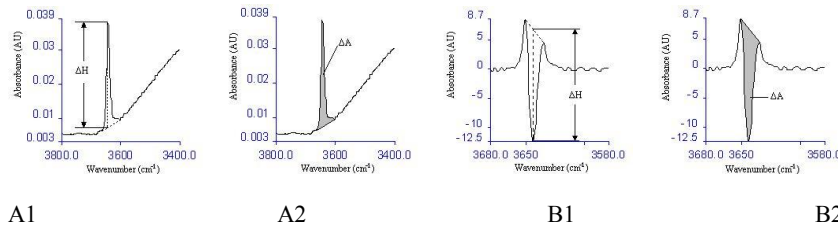


Fig. 1. the sketch maps of the measurement of the corrected peak height (A1) and the corrected peak area (A2) of the raw spectra and the corrected peak height (B1) and the corrected peak area (B2) of the 2nd derivative spectra

Correlation coefficient (r) and root mean square error of calibration ($RMSEC$) are regarded as the parameters to evaluate the models' performance; while standard error of prediction (SEP) and ratio performance deviation (RPD) are regarded as the parameters to evaluate the models' prediction performance; the statistic F -test value is regarded as the parameter to evaluate the regression significant of regression equation.

Looking up in the F -test value table according to the sample size of the calibration set and external validation set, $F_{(1,28)} = 7.64$, $F_{(1,8)} = 11.26$ when the significance level $\alpha = 0.01$.

3 Result and Discussion

3.1 Development of Calibration Models

Some of the ATR-MIR spectra of the samples from calibration set are showed as figure 2, from which, it can be seen clearly that the characteristic peak of lime exists at 3642 cm^{-1} .

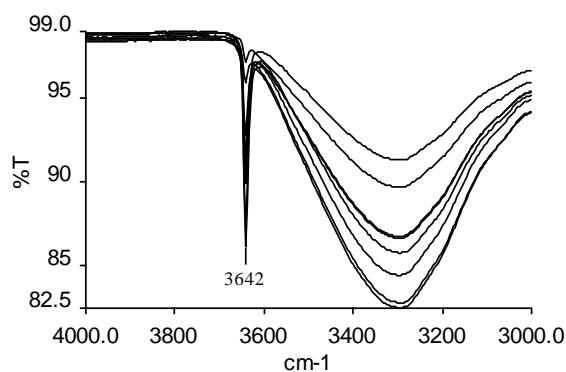


Fig. 2. Some of the ATR-MIR spectra of calibration set

The calibration models of lime in wheat flour are developed by the content of lime, i.e. the specified values, in wheat flour and the corrected peak height of the raw spectra data, the corrected peak area of the raw spectra data, the corrected peak height of the 2nd derivative spectra data and the corrected peak area of the 2nd derivative spectra data respectively. The sample size of the calibration set is 30. The correlations between the specified values and the corrected peak height of the raw spectra, the corrected peak area of the raw spectra, the corrected peak height of the 2nd derivative spectra and the corrected peak area of the 2nd derivative spectra are showed as figure 3, while the regression equations, correlation coefficients (r) and F -test values (F) of the four calibration models are showed in table 2.

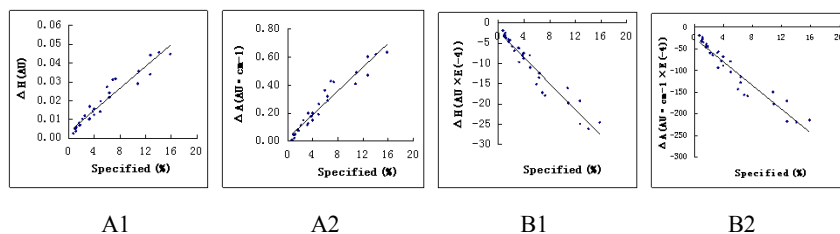


Fig. 3. The correlations between the specified values and the corrected peak height of the raw spectra (A1), the corrected peak area of the raw spectra (A2), the corrected peak height of the 2nd derivative spectra (B1) and the corrected peak area of the 2nd derivative spectra (B2)

Tab. 2. The regression equations, correlation coefficients (r) and F -test values (F) of the calibration models

Models	Regression equation	r	F
Corrected peak height of raw spectra	$y = 0.0029x + 0.0034$	0.9648	377
Corrected peak area of raw spectra	$y = 0.0426x + 0.0183$	0.9696	440
Corrected peak height of 2 nd derivative spectra	$y = -1.6070x - 2.1049$	-0.9646	375
Corrected peak area of 2 nd derivative spectra	$y = -13.6530x - 25.6610$	-0.9599	328

It can be seen from table 2 that the absolute values of the correlation coefficients of the models are no less than 0.95; the F -test values of the models are all much larger than that looked up in F -test value table ($F_{(1,28)} = 7.64$ when the significance level $\alpha = 0.01$), which indicated that a very remarkable correlation exists between the specified values and the corrected peak heights / areas of the raw spectra or the corrected peak heights / areas of the 2nd derivative spectra, which indicates that CLS regression has a certain potential application in the quantitative analysis of lime in wheat flour by ATR-MIR spectra data.

The correlations between the specified values and estimated values by self-prediction of the models are showed as figure 4.

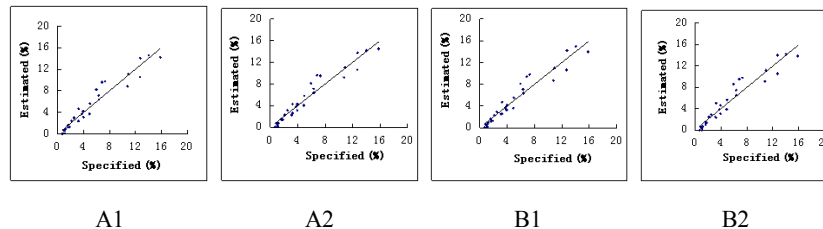


Fig. 4. The correlations between the estimated and specified values of the calibration set

A1: the corrected peak height model of the raw spectra, A2: the corrected peak area model of the raw spectra, B1: the corrected peak height model of the 2nd derivative spectra, B2: the corrected peak area model of the 2nd derivative spectra

It can be seen from figure 4 that, for each model, a good correlation exists between the specified values and estimated values by self-prediction of the model. The regression equations, correlation coefficients (r_c), root mean square errors of calibration ($RMSEC$), F -test values (F_c) of the four groups of data by self-prediction are showed as figure 3.

Tab. 3. The regression equations, correlation coefficients (r_c), $RMSEC$ and F -test values (F_c) of the calibration set by self-prediction

Models	Regression equation	r_c	$RMSEC$	F_c
Corrected peak height of raw spectra	$y = 1.0020x + 0.0061$	0.9649	1.17	378
Corrected peak area of raw spectra	$y = 0.9981x + 0.0198$	0.9699	1.07	444
Corrected peak height of 2 nd derivative spectra	$y = 0.9988x + 0.0106$	0.9647	1.17	376
Corrected peak area of 2 nd derivative spectra	$y = 0.9947x + 0.0497$	0.9606	1.23	334

It can be seen from figure 3 that for all of the regression equations by self-prediction, the slopes are all close to 1, the intercepts are close to 0, the correlation coefficients are no less than 0.96, the F -test values are much larger than that looked up in F -test value table ($F_{(1,28)} = 7.64$ when the significance level $\alpha = 0.01$), which indicates that, for each model, a very remarkable correlation exists between the specified values and estimated values by the self-prediction of the calibration model, i.e. CLS regression has a certain potential application in the quantitative analysis of lime in wheat flour by ATR-MIR spectra data.

3.2 External Validation

The external validation set is employed to evaluate the prediction performance of the

calibration models. The sample size of the external validation set is 10. The correlation between the specified values and estimated values by the corrected peak height model of the raw spectra, the corrected peak area model of the raw spectra, the corrected peak height model of the 2nd derivative spectra, the corrected peak area model of the 2nd derivative spectra are showed as figure 5.

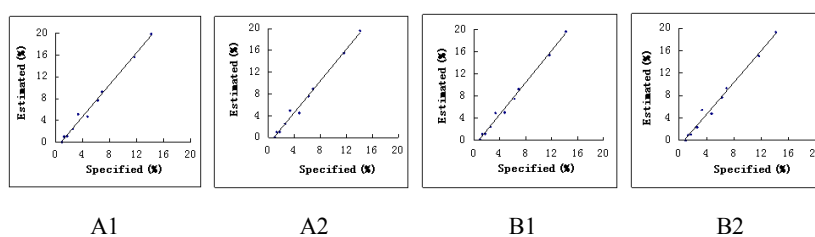


Fig. 5. The correlations between the estimated and specified values of the external validation set

A1: the corrected peak height model of the raw spectra, A2: the corrected peak area model of the raw spectra, B1: the corrected peak height model of the 2nd derivative spectra, B2: the corrected peak area model of the 2nd derivative spectra

It can be seen from figure 5 that a good correlation exists between the estimated and specified values of the external validation set for each model. The regression equations, correlation coefficients (r_p), standard errors of prediction (SEP), ratio performance deviations (RPD) and the F -test values (F_p) of the external validation set are showed in table 4.

Tab. 4. The regression equations, correlation coefficients (r_p), SEP , RPD and F -test values (F_p) of the external validation set

Models	Regression equation	r_p	SEP	RPD	F_p
Corrected peak height of raw spectra	$y = 1.4658x - 1.1709$	0.9954	2.59	1.73	862
Corrected peak area of raw spectra	$y = 1.4471x - 1.1580$	0.9954	2.47	1.81	862
Corrected peak height of 2 nd derivative spectra	$y = 1.4425x - 1.0632$	0.9964	2.48	1.81	1119
Corrected peak area of 2 nd derivative spectra	$y = 1.4224x - 1.0096$	0.9935	2.42	1.85	607

It can be seen from table 4 that for all the regression equations of external validation set, the correlation coefficients between the estimated and specified values are no less than 0.99, standard errors of prediction are no more than 2.6, ratio performance deviation are no less than 1.7, *F-test* values are much larger than that looked up in *F-test* value table ($F_{(1,8)} = 11.26$ when the significance level $\alpha = 0.01$), which indicates that a very remarkable correlation exists between the estimated and specified values of the external validation set for each model, i.e. the CLS regression has a certain potential application in the quantitative prediction of lime in wheat flour by ATR-MIR spectra data.

Comparing the four models, it can be seen that the differences among the four models are not significant in regression equations, correlation coefficients and the prediction results of the external validation set, the reason of which is that the characteristic peak of lime in ATR-MIR spectrum (3642 cm^{-1}) is far away from the absorption peak of wheat flour in ATR-MIR spectrum (3300 cm^{-1}), therefore, the interference from the spectral background becomes weak. According to that the detection limit is three times the *RMSEC*[11], the detection limits of the corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which are much less than that added in wheat flour for the purpose of weight increment. This method can fulfill the demand of rapid quality screening of wheat flour in the market.

4 Conclusion

Classical least square algorithm was applied in the development of the quantitative calibration models of lime in the wheat flour samples. The result indicated that the CLS regression has a certain potential application in the quantitative determination of lime in wheat flour by ATR-MIR spectra data. The detection limits of the corrected peak height / area models of the raw spectra and the corrected peak height / area models of the 2nd derivative spectra are 3.51 %, 3.21 %, 3.51 % and 3.69 % respectively, which are much less than that added in wheat flour for the purpose of weight increment. This method can fulfill the demand of rapid quality screening of

wheat flour in the market, which is of great and positive meaning to the monitoring of wheat flour in the market and the guarantee of the customers' health. This research, to some extent, will also provide some references to the quantitative determination of other banned additives and the design and manufacturing of the special MIR spectrometer for the quality screening of wheat flour in the market. The method to decrease the detection limit is still being researched.

Acknowledgment. This thesis is supported by Beijing Postdoctoral Research Foundation; China Postdoctoral Science Foundation (49th) (No. 20110490318); Postdoctoral Science Foundation of Beijing Academy of Agriculture and Forestry Science; Special Research of Public Welfare Industry, the Ministry of Agriculture of People's Republic of China (No. 201203046) and National High Technology Research and Development Program 863 (2010AA10Z201).

References

1. Cledir Santos, Marcelo E. Fraga, Zofia Kozakiewicz, Nelson Lima: Fourier transform infrared as a powerful technique for the identification and characterization of filamentous fungi and yeasts[J]. *Research in Microbiology*, 161, pp. 168 -- 175 (2010)
2. Hamzah M. Al-Qadiri, Mengshi Lin, Anna G. Cavinato, Barbara A. Rasco: Fourier transform infrared spectroscopy, detection and identification of *Escherichia coli* O157:H7 and *Alicyclobacillus* strains in apple juice[J]. *International Journal of Food Microbiology*, 111, pp. 73 -- 80 (2006)
3. Gozde Gurdeniz, Banu Ozen: Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data[J]. *Food Chemistry*, 116, pp. 519 -- 525 (2009)
4. Rubén M. Maggio, Teodoro S. Kaufman, Michele Del Carlo, Lorenzo Cerretani, Alessandra Bendini, Angelo Cichelli, Dario Compagnone: Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares[J]. *Food Chemistry*, 114, pp. 1549 -- 1554 (2009)
5. C. S. Pappas, C. Takidelli, E. Tsantili, P. A. Tarantilis, M. G. Polissiou: Quantitative determination of anthocyanins in three sweet cherry varieties using diffuse reflectance infrared Fourier transform spectroscopy[J]. *Journal of Food Composition and Analysis*, 24, pp. 17 -- 21 (2011)

6. S. N. Jha, S. Gunasekaran: Authentication of sweetness of mango juice using Fourier transform infrared-attenuated total reflection spectroscopy[J]. *Journal of Food Engineering*, 101, pp. 337 -- 342 (2010)
7. Vassiliki A. Iconomidou, Martha E. Georgaka, Georgios D. Chryssikos, Vassilis Gionis, Persefoni Megalofonou, Stavros J. Hamodrakas: Dogfish egg case structural studies by ATR FT-IR and FT-Raman spectroscopy[J]. *International Journal of Biological Macromolecules*, 41(1), pp. 102 -- 108 (2007)
8. Sune D. Ebbesen, Barbara L. Mojet, Leon Lefferts: In situ ATR-IR study of nitrite hydrogenation over Pd/Al₂O₃[J]. *Journal of Catalysis*, 256(1), pp. 15 -- 23 (2008)
9. Jing Li, Meiyun Zhang, Quanmin Li: Spectrophotometric determination of cefotaxime sodium with potassium ferricyanide. *Chinese Journal of Applied Chemistry*[J], 28(1), pp. 88 -- 93 (2011) (in Chinese)
10. Jianping Yu, Shanling Hu, Yan Lin, Xuzhong Luo, Tenggao Zhu: Determination of tungsten in rare earth-magnesium master alloy by spectrophotometry with thiocyanate and acid digestion[J]. *Chinese Journal of Spectroscopy Laboratory*, 27(6), pp. 2529 -- 2532 (2010) (in Chinese)
11. Wanzhen Lu, Hongfu Yuan, Guangtong Xu, Dongmei Qiang: *Modern Near-infrared Spectroscopy Analysis Technology*[M], China Petrochemical Press, pp. 124 -- 182 (2000) (in Chinese)
12. Cecilia Shiroma, Luis Rodriguez-Saona: Application of NIR and MIR spectroscopy in quality control of potato chips[J]. *Journal of Food Composition and Analysis*, 22, pp. 596 -- 605 (2009)