International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.10 (2016), pp.131-142
http://dx.doi.org/10.14257/ijmue.2016.11.10.12

# Research on Rapid Detection Method of Protein and Fat in Raw Milk Based on Mid-infrared Spectrum

Xiaoming Li[1], Guicheng Huo[2], Yan Wang[1], Hongmin Sun[1] and Qingming Kong[1*]

[1]*School of Electronic Engineering and Information Northeast Agricultural University, Harbin, 150030, China*
[2]*School of Food Science and Engineering Northeast Agricultural University Harbin, 150030, China*
*E-mail: lxm_hit@163.com*
**Corresponding Author*

***Abstract***

*In this paper raw milk was taken as the research object, using mid-infrared spectrum analysis method rapidly tests protein and fat content in raw milk through establishing quantitative analysis model. First of all, researchers collect 84 kinds of raw milk which is different areas and varieties in Heilongjiang province, and classify 84 samples into calibration sets and validation sets by standard chemical testing and spectral scanning on protein and fat, and calibration sets contain 64 samples, and validation sets contain 20 samples. Then, denoising by second derivative method which window size is 17 points, and selecting method by comparing SPA and GA characteristic wavelength respectively, and establishing PLS model and model demonstration. Eventually, researchers draw the conclusion that SPA wavelength selection method which value of m_max is 10 and GA wavelength selection method which value of Evaluat is 10d can effectively improve validation set model precision, but precision of the former is higher and characteristic wavelength is least, and its characteristic wavelength point reduces from 220 to 4, $R^2$ of protein model validation set is 0.8010929, RMSEP is 0.0207505, $R^2$ of fat model validation set is 0.91007989, RMSEP is 0.045329, relative standard deviation RSD<3%. These prove that the method for rapid testing of raw milk components is feasible.*

***Keywords****: raw milk, protein, fat, mid-infrared spectrum, successive projections algorithm*

## 1. Introduction

Dairy products quality safety incidents often happened in China in recent year. It doesn't only directly affect the physical and mental health of citizens, and seriously hurt consumer confidence in domestic dairy products quality and safety, greatly reduce the market share of domestic dairy products, and lead to huge economic losses to dairy products processing industry in China. Raw milk usually refers to the fresh milk, that is to say that it is fresh milk which is from cows breast extrusion without processing. It is base in the process of dairy production, therefore, raw milk components testing has vital significance to dairy products quality and safety control.

In the process of raw milk purchase and production management, manufacturers need rapidly and accurately test mainly content of raw milk, and protein and fat are the core indicators of measuring milk quality. Protein in raw milk mainly is casein protein, albumin, globulin, and lacto-protein *etc*. They are a kind of complete protein which digestibility is 98% above, and have very high biological titer, and contain all the essential amino acids which need the body growth. Fat in raw milk is natural fat which has unique biological properties and fat of high nutritional value. It contains short chain

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2016 SERSC

and medium-chain fatty acids of easy absorption and utilization, and its melting point is very low, and it is easy to digest and absorb because fat ball particles are minimal and highly dispersed milk melt shape, and contain a large number of fat-soluble vitamins.

At present, in aspect of raw milk components testing, traditional chemical analysis method is dominant, and they include Kjeldahl Nitrogen Determination method, Lowry method, and Rose-Gottlieb method, and Gerber's method, and Babcocok test method *etc.* [1-3]. As conventional testing method, they have mature technology and high accuracy, but they exist some problems, for example, testing process is complex, and testing time is longer, and testing cost is higher, and these methods can not realize online analysis and so on. Therefore, traditional chemical analysis method is very difficult to get a lot of samples' component in a short period of time.

Using infrared spectroscopy online testing component-content of raw milk is a hot issue. Comparing with traditional chemical analysis method, this method has the following advantages: testing speed is fast, no pollution, no sample pretreatment, and realizing online rapidly testing. So the method widely applies in raw milk component testing [4-8]. This paper tests protein and fat contents in raw milk by the Fourier transform infrared spectroscopy.

## 2. Materials and Methods

### 2.1. Experiment Sample Preparation

The paper collects 84 kinds of raw milk which is different areas and varieties in Heilongjiang province, and each variety is divided into equal two (350ml), and sample's label has not been used, and the total is 168 samples.

Chemical calibration of sample mainly aims at protein and fat in raw milk. Its determination method is as follows.

Principle: the protein in food is decomposed under the condition of catalytic heating, and ammonia and sulfuric acid which generate through decomposition combine and generate ammonium sulfate. Alkaline and distillation let ammonia dissociate, then it is titrated by hydrochloric acid standard solution after absorbing by boric acid, and the acid consumption multiplies by the coefficient factor is the protein content.

1) The determination method of protein-Kjeldahlmethod [9]

1) Reagents

① $CuSO_4$, $K_2SO_4$, $H_2SO_4$, NaOH

② Boric acid solution (20g/L) $H_3BO_3$: Taking 20 g boric acid, and adding water to dissolve and dilute to 1000 ml.

③ NaOH solution (400g/L), Taking 400 g NaOH and putting cold and diluting to 1000 ml after adding water dissolves.

④ Hydrochloric acid (HCI) standard solution (0.05mol/L)

⑤ Methyl red ethanol solution: Taking 0.1 g methyl red, and soluble in ethanol, and diluting to 100 ml with 95% ethanol.

⑥ Bromocresol green ethanol solution (1g/L) Taking 0.1 g bromocresol green, and soluble in ethanol, and diluting to 100 ml with 95% ethanol.

⑦ Methyl red - bromocresol green mixed indicator: mixed with 1:5 when using

2) Determination

① Sample processing (digestion): accurately taking hydrolyzed collagen 0.5 g (precision is 0.0001g), $CuSO_4$ 0.2g, and $K_2SO_4$ 6g into dry clean kjeldahl nitrogen pipe, and the pipe is put on digestive furnace, and adding $H_2SO_4$ solution 20ml and cover with the lid. Opening digestive furnace (set temperature 400℃), the instrument reaches setting temperature about 20 minutes, then starts timing, and cuts power and puts cold after sample digestion finishes (needs 2 to 3 hours, blue-green liquid and clear and transparent).

 Copyright ⓒ 2016 SERSC

② Distillation and absorption: The digested sample is transferred to 100ml volumetric flask, and adding water diluted to 100ml volumetric flask and shaken and spare, while doing blank experiment. Adding boric acid(20g/L) 50ml and 10D methyl red-bromocresol green mixed indicator to the flask, and opening the power switch of Kjeldahl Nitrogen Determination Instrument and the cooling water, and setting the alkali(40% NaOH solution) time and distillation time respectively is 7 seconds and 6 minutes. Taking the constant volume solution 10ml to digestive tube and placing to the corresponding position. Then closing safety door, and adding alkali and distillation after starting button and preheating 3 minutes. When absorption liquid is neutral, stopping absorption, and titrating receiving solution with 0.05mol/L hydrochloric acid standard solution, and the color from green to gray red is the end of the titration, and recording the volume of hydrochloric acid consumption.

3) Computation

$$X = \frac{N \times (V1 - V2) \times 0.014}{M \times 10/100} \times F \times 100$$

X: Percent of the protein in the sample, the unit is g;

N: Equivalent concentration of sulfuric acid or hydrochloric acid standard solution；

V1: Volume of sulfuric acid or hydrochloric acid standard solution for sample consumption, the unit is ml;

V2: Volume of sulfuric acid or hydrochloric acid standard solution for reagent blank consumption, the unit is ml;

0.014: 1N sulfuric acid or hydrochloric acid standard solution is equal to1ml the number of nitrogen;

M: Sample quality (volume), G (ML);

F: Conversion of nitrogen to protein. The content of the protein in nitrogen is generally 15 to 17.6 percent. Multiply 16% by a coefficient to compute the content of nitrogen. The coefficient is 6.25 for protein, 6.38 for dairy products , 5.70 for flour, 6.24 for cornand sorghum, 5.46 for peanut, 5.95 for meters, 5.71 for soybeans and their products, 6.25 for meat and meat products, 5.83 for barley, millet, oats, rye, and 5.30 for sesame seeds, sunflower.

(2) The determination method of fat-Soxhlet method [10]

Principle: the samples were extracted with anhydrous, petroleum or other solvents, and then were distilled. The rest matter is called fat or crude fat. The rest matter also contains pigment and volatile oil, wax, resin and other things. The fat from extraction method is free fat.

1) Reagents

① Anhydrous ether or petroleum ether

② Sea send: shown in GB 5009.3-85

③ Soxhlet extractor

2) Method

① Sample handling

Solid samples. Weigh 2-5g accurately, which are determined water. Mix with sea sand if necessary. Put all into cylinder filter.

Semi liquid or solid samples. Weigh 5.0-10.0g, put into evaporating dish, join the sand about 20g and dry in a boiling water bath. Then dry again at 95-105℃. Grind into fine, and move all into the filter tube. Wipe the evaporation dish and glass rod with the sample with ether cotton, and put the cotton into the filter paper tube.

② Extraction

Put filter tube into the fat extractor extraction tube. Connect receiving flask which is dried to constant weight. Add anhydrous ethyl ether and petroleum ether to the bottle volume 2 / 3 from the upper end of extracting the condenser tube. Heat on water bath in

Copyright ⓒ 2016 SERSC

order to make ethyl ether and petroleum ether continuous reflux extraction. The time is usually from 6 to 12h.

③ Weighing

Take the receiving flask down and recycle ethyl ether and petroleum ether. Dry in a water bath when the left ethyl ether is 1-2mL, and dry for 2 hours at 95-105℃. Weigh after cooling 0.5h.

3) Computation

$$X = \frac{m_1 - m_0}{m_2} \times 100$$

X-The content of fat in the sample (%);

m1-The quality of accept bottle and fat (g);

m2-The quality of accept bottle (g);

m-The quality of sample (Before determining water) (g).

### 2.2. Sample Classification

Ideal calibration sets should meet the following conditions. Samples in calibration set should include all the chemical compositions which may exist in measuring samples and their concentration (or properties) range should be over measuring samples. Physicochemical parameters of samples in calibration set should be evenly distributed. The sample composition is Gaussian distribution in the actual production process, especially in the large scale industrial processes. In addition, the number of samples in the calibration set should be enough to be able to statistically determine the quantitative functional relationship between the spectral variables and the physical parameters to be corrected.

Random Selection is difficult to get ideal sample set using the method of randomly selected samples. Only according to the nature of the distribution of calibration samples are also not getting satisfactory results, because there are possible existing great differences between two samples' spectral with a certain property. At present, Kennard-Stone selection method is the most commonly used method [11].

According to the normal form method, the raw milk is classified by the Kennard-Stone algorithm, and the most representative samples are selected as the calibration set. Sample distribution is as Table 2-1. The former is used as a calibration set for modeling, and the latter is used as a validation set for validation of the model. Value in table represents percentage of content.

Distribution uniformity of the calibration set and validation set is shown in Table 1. Validation set samples are distributed in the range of calibration set. Average of validation set is close to average of calibration set. Protein contents coner from 2.82 to 3.39, fat contents cover from 3.28 to 4.18.

**Table 1. Protein Content Statistics of Raw Milk**

| Sample Classification | Component | Samples No. | Range/(%) | Mean/(%) |
|---|---|---|---|---|
| Raw milk | Protein | 84 | 2.82-3.39 | 3.10 |
| Calibration sets | Protein | 64 | 2.82-3.39 | 3.11 |
| Validation sets | Protein | 20 | 2.94-3.16 | 3.07 |

 Copyright ⓒ 2016 SERSC

**Table 2. Fat Content Statistics of Raw Milk**

| Sample Classification | Component | Samples No. | Range/(%) | Mean/(%) |
|---|---|---|---|---|
| Raw milk | Fat | 84 | 3.28-4.18 | 3.71 |
| Calibration sets | Fat | 64 | 3.28-4.18 | 3.72 |
| Validation sets | Fat | 20 | 3.44-4.01 | 3.68 |

### 2.3. Spectral Acquisition

In the experiment using Thermo's Antaris II near-infrared spectrometers scan for raw milk, which range from 2564nm to 25000nm (3900cm$^{-1}$-400cm$^{-1}$) the resolution of 16cm$^{-1}$. Respectively, using sweep surface mode of liquid transmission, air is used as a comparison object sphere scanning. Before sample surface scanning, we set the number of background scan 64 times and experiment scan 64 times. Its sample scan results as shown in Figure 1, as sequence is raw milk spectrum.

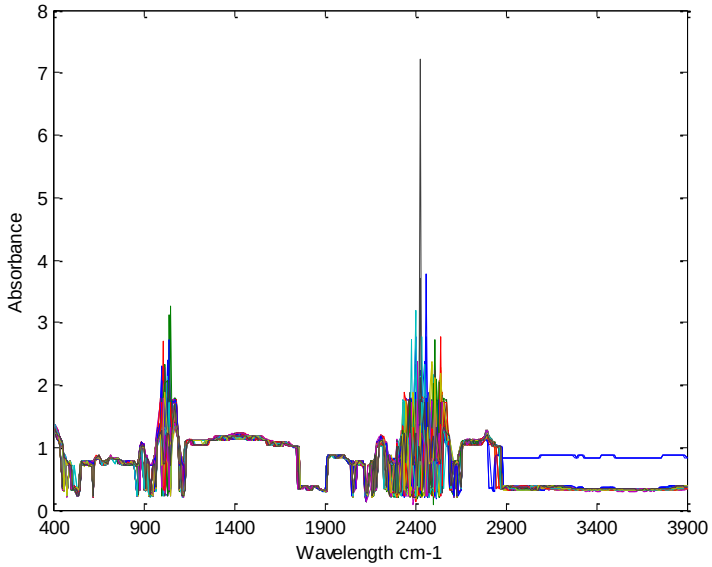



**Figure 1. Absorption Spectra of Raw Milk**

## 3. Results and Discussion

### 3.1. Spectral Denoising

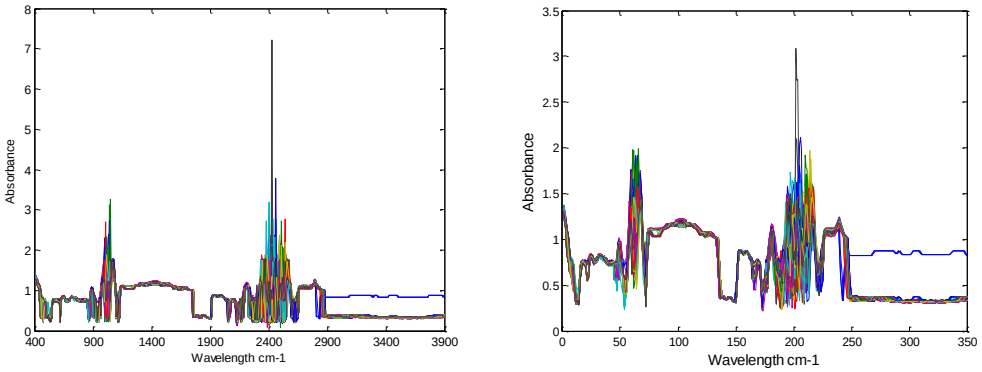
In addition to its own information, the near infrared spectrum of the sample also contains many other irrelevant information and noise, such as electrical signal noise, background noise and stray light. Therefore, how to eliminate the irrelevant information and noise of the spectral data is a very important pretreatment method in the process of establishing the model. In this paper, the SNV method and second derivative transformation method are compared.

(1) Standard normal variable transformation(SNV)

Standard normal variable transformation (SNV) is mainly used to eliminate the effect of solid particle size, surface scattering and optical path difference of spectra. Calculation formula of SNV is same as the standardization's. The difference between them is that the standard algorithm is to deal with a group of spectra and SNV is to deal with a whole spectrum. There are varieties of algorithms for process of normalization, including the

Copyright ⓒ 2016 SERSC

area of normalization, maximum normalization and average normalization method. The vector normalization is common used in the process of spectrum analysis. So SNV is used to remove the noise in this paper. The result after denoising is shown in Figure 2.



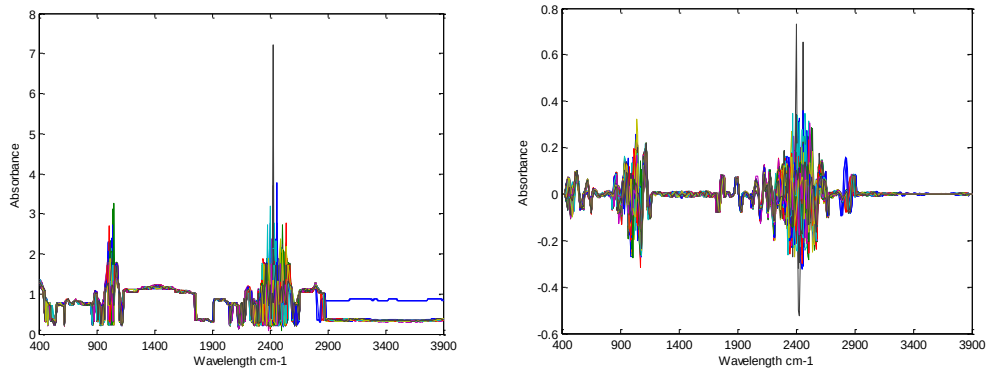

a. Original Spectral Image b. Spectral Image after SNV

**Figure 2. Original Spectral Image and Spectral Image after Smoothing**

(2) Derivative denoising

Derivative is also a very common denoising method, which can eliminate the baseline drift of spectrum and overcome the overlap of spectral bands. Derivative processing is divided into first-order derivative and second derivative, in which the former can remove drift independent from the wavelength, and the latter can remove drift dependent with wavelength. Direct finite difference method is common used in the derivation of the discrete spectrum. for the discrete spectrum XK, calculate first-order derivative and second derivative as follow formula when wavelength equals to k, differential width equals to g. Smooth before derivative, in order to eliminate the noise caused by spectral transform.

Savitzky-Golay convolution derivative method is used in this paper, result after derivative is shown in Figure 3.Smooth with 15 points after spectral derivative, in order to eliminate signal burr resulted by derivative. Verify derivative processing with 17 points after smoothing the spectrum. Results are shown in Figure 4 and Figure 5. Protein model interaction validation mean square error root (RMSECV) was 0.5669495, and the validation set mean square error root (RMSEP) was 0.0717114. Fat model interaction validation mean square error root (RMSECV) is 0.6796574, the validation set mean square error root (RMSEP) is 0.0940966.




a. Original Spectral Image b. Spectral Image after Derivative

**Figure 3. Original Spectral Image and Spectral Image after Derivative Transformation Smoothing**

 Copyright ⓒ 2016 SERSC

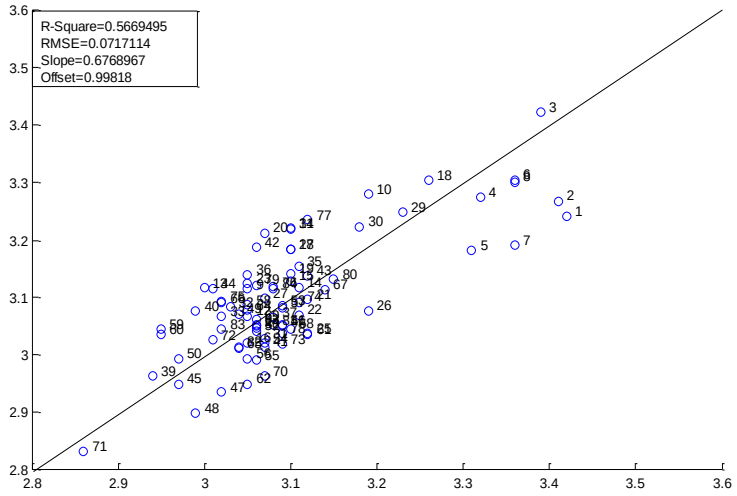



**Figure 4. Validation Results of Protein Calibration Model after Denoising**

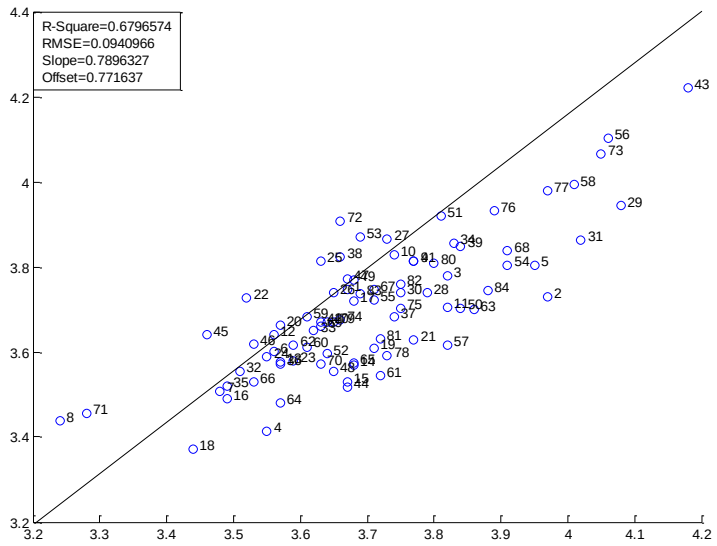



**Figure 5. Validation Results of Fat Calibration Model after Denoising**

### 3.2. Characteristic Wavelength Selection

(1) Successive projections algorithm (SPA)

Successive projections algorithm (SPA) is a forward selection method. it can find the minimum redundancy information variable set from the spectrum information to reduce the modeling process variables and colinearity between variables. The application of continuous projection algorithm can reduce extent model fitting process complexity, and speed up the fitting operation [12-13].

In this paper, according to the spectral information in different wavelength range, select the results of wavelength selection under the condition of m_max =10. Wavelength selection results are shown in Figure 6. The wavelength points reduce to 4 from 220. Fitting speed and efficiency of model is greatly accelerated. It could be seen from Table3

that the validation set coefficient $R^2$ increased to 0.785193, RMSEP reduces to 0.02093391, and validation set coefficient $R^2$ of fat model increased to 0.8456811, RMSEP reduces to 0.0649325.

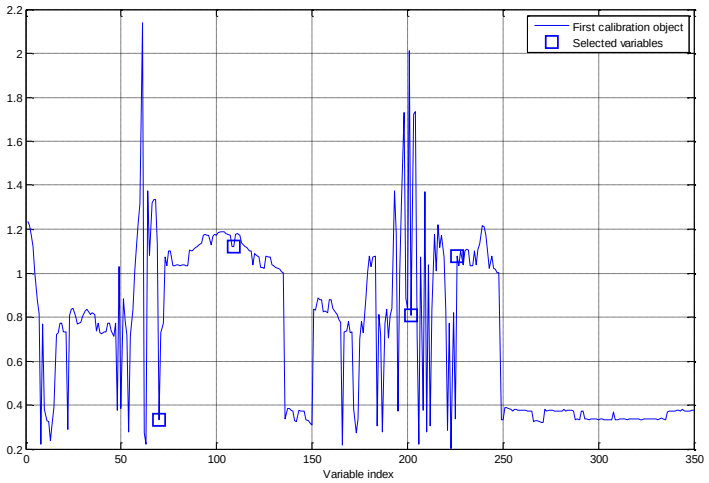



**Figure 6. Chosen Wavelength Points by SPA with M_Max=10**

**Table 3. Model Validation Results of SPA**

| Wavelength $cm^{-1}$ | Component | Factors | $R^2$ | RMSEP |
|---|---|---|---|---|
| 400-3900 | Protein | 7 | 0.785193 | 0.02093391 |
| 400-3900 | Fat | 11 | 0.8456811 | 0.0649325 |

(2) Genetic algorithm (GA)

Genetic algorithm (GA) is proposed by Holland in 1975 and based on natural selection and genetic regulation. GA keep good objective function value, eliminate worse variables, and ultimately achieve optimal results by using selection, crossover and mutation operation of genetic iterative.

Genetic algorithm (GA) will experience coding, initial population, selection, crossover and mutation and the fitness function in process spectral band selection. In this paper, we set maximum generations as 100, population size as 30, crossover rate as 50%, variation rate as 1%. Choose binary encoding and roulette method. according to individual fitness value from the current population size based on the mean square error (mse) to select, crossover restructuring mode selection single-point cross, cross before advanced individual matching, choose randomly matched way, mutation methods of uniform mutation, fitness function selection "F = RMSE"[14,15]. The wavelength selection result is shown in Figure 7.

 Copyright ⓒ 2016 SERSC

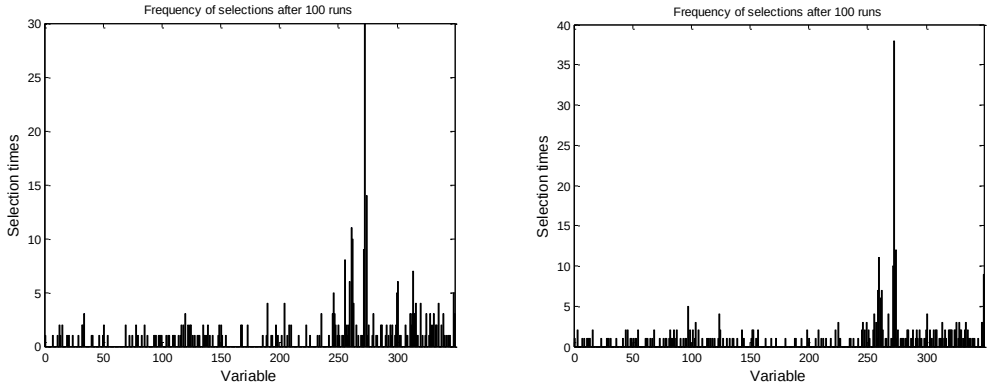



**Figure 7. Wavelength Selection Results by GA in the Condition Evaluate=10**

Validate the model with 10 evaluation times. Through the Table4 we found that when the estimated number is 10, protein model characteristic wavelengths reduced from 220 to 10, fat model characteristic wavelengths reduce from 220 to 11.Root mean square error (RMSEP) of protein model calibration set decreased to 0.0493237, and determination coefficient $R^2$ increased to 0.7193055, root mean square error (RMSEP) of fat model calibration set decreased to 0.0859908, and determination coefficient $R^2$ increased to 0.7310928.

**Table 4. Model Validation Results of GA**

| Wavelength ($cm^{-1}$) | Evaluate | Component | Wavelength points | $R^2$ | RMSEP |
|---|---|---|---|---|---|
| 400-3900 | 10 | Protein | 273,274, 260, 272, 349, 259, 262, 261, 97, 124 | 0.7193055 | 0.0493237 |
| 400-3900 | 10 | Fat | 273, 274, 261, 272, 260, 262, 246, 304, 349, 251, 287 | 0.7310928 | 0.0859908 |

### 3.3. Quantitative Analysis Model

Partial least squares regression (PLSR) is a new kind of multivariate statistics analysis methods. It mainly studies the more dependent variable regression modeling of multiple variables. Especially when the internal height linearly dependent variables, partial least squares regression method is more effective. In addition, the partial least-square regression better solve the problem such as sample number less than the number of variables. In this paper, quantitative analysis model is established by using PLS, validated by the validation set, the result is shown in figure. Choose 64 samples as calibration set, 20 samples as a validation set. From the Table4, we can see that determination coefficient $R^2$ of protein validation is 0.8010929, the root mean square error of validation set (RMSEV) is 0.0459126, determination coefficient $R^2$ of protein validation is 0.9100798, the root mean square error of validation set (RMSEV) is 0.0459126.

**Table 4. Model Validation Results of PLS**

| Wavelength cm-1 | Component | $R^2C$ | RMSEC | $R^2V$ | RMSEV |
|---|---|---|---|---|---|
| 400-3900 | Protein | 0.7717302 | 0.0459126 | 0.8010929 | 0.0207505 |
| 400-3900 | Fat | 0.8163694 | 0.0716456 | 0.9100798 | 0.0453295 |

Copyright ⓒ 2016 SERSC

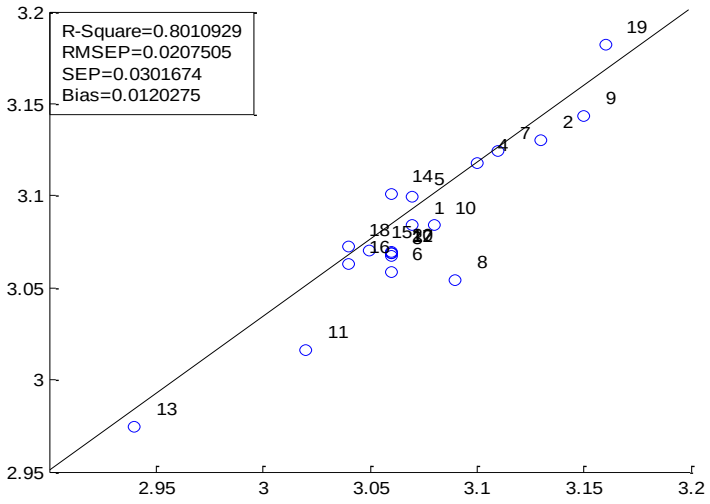



**Figure 8. Validation Results of Protein Model**

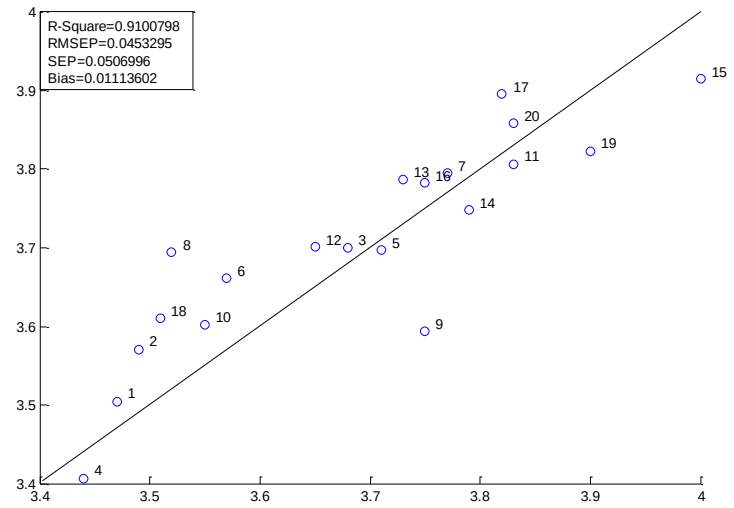



**Figure 9. Validation Results of Fat Model**

## 4. Conclusion

This paper has researched on raw milk component by mid-infrared spectrum analysis method (400-3900cm$^{-1}$), and respectively established quantitative model based on protein and fat in raw milk. Meanwhile, 84 samples were classified into calibration set and validation set, and they respectively contained 64 samples and 20 samples. Finally, validation set $R^2$ of protein validation model is 0.8010929 and validation set RMSEP is 0.0207505, and validation set $R^2$ of fat validation model is 0.91007989 and validation set RMSEP is 0.045329.


Copyright ⓒ 2016 SERSC

## Acknowledgements

This research was supported by Heilongjiang Postdoctoral Financial Assistance of China. (Grant No.LBH-Z12040) and Key Technologies R & D Program of Heilongjiang province of China (Grant No. GC12B305), and "Young Talents" Project of Northeast Agricultural University (Grant No. 14QC13).

## References

[1]. X. Jiaxi, "Protein detection method and determination of protein content in dairy products", University Chemistry, vol. 24, no. l, **(2009)**, pp. 66-69.
[2]. "State Technical Supervision Bureau of the people's Republic of China. General test method for infant formula powder (GB/T5413.1-5413.32)", Standards Press of China, **(1997)**, pp. 123-126.
[3]. Y. Shuhua, L. Jing and N. Wenji, "Method for determination of fat in milk and dairy products", Occupation and Health, vol. 21, no. 5, **(2005)**, pp. 698.
[4]. S. Shahabi, H. G. Bagheri and K. Ramazani, "Tensile bond strength of sealants following Er: YAG laser etching compared to acid etching in permanent teeth", Lasers Med Sci., vol. 27, no. 2, **(2012)**, pp. 371-375.
[5]. N. Six, J. J. Lasfargues and M. Goldberg, "In vivo study of the pulp reaction to Fuji IX, a glass ionomer cement", J Dent., vol. 28, no. 6, **(2000)**, pp. 413-422.
[6]. D. Lefier, R. G. Pin and S. Pochet, "Determination of fat, Protein, and laetose in raw milk by Fourier transform infrared spectroscopy and by analysis with a conventional filter-based milk analyzer", J. AOAC. Int., vol. 79, no. 3, **(1996)**, pp. 711-717.
[7]. R. G. Pina and D. Lefier, "Protein definition for milk Payment to farmers in milk Protein Definition and Standardization", International Dairy Federation: Brussels, Belgium, **(1995)**, pp. 16-23.
[8]. Y. Etzion, R. Linker and U. Cogan, "Determinations of Protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy", Journal Dairy Sci, vol. 87, no. 9, **(2004)**, pp. 2779-2788.
[9]. A. K. Hewavitharana and B. V. Brakel, "Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk", Analyst, vol. 122, no. 4, **(1997)**, pp. 701-704.
[10]. R. H. Wilson and H. S. TAPP, "Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods", Trends in Analytical Chemistry, vol. 18, no. 2, **(1999)**, pp. 85-93.
[11]. F. A. Inon, S. Garrigues and D. L. Guardia, "Nutritional parameters of commercially anailable milk samples by FTIR and chemometric techniques", Analytica Chimica Acta, vol. 513, no. 2, **(2004)**, pp. 401-412.
[12]. D. W. Templeton, A. D. Sluiter, T. K. Hayward, B. R. Hames and S. R. Thomas, "Assessing corn stover composition and sources of variability via NIRS"m Springer Science + Business Media B.V. vol. 16, no. 4, **(2009)**, pp. 621-639.
[13]. F. Xu, J. Yu, T. Tesso, F. Dowell and D. Wang, "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review", Applied Energy, vol. 104, **(2013)**, pp. 801-809.
[14]. H. Jiang, G. Liu, X. Xiao, C. Mei, Y. Ding and S. Yu, "Monitoring of solid-state fermentation of wheat straw in a pilot scale using FT-NIR spectroscopy and support vector data description", Microchemical Journal, vol. 102, **(2012)**, pp. 68-74.
[15]. G. W. Mathison, H. Hsu, R. Soofi-Siawash, G. Recinos-Diaz, E. K. Okine, J. Helm and P. Juskiw, "Prediction of composition and ruminal degradability characteristics of barley straw by near infrared reflectance spectroscopy", Canadian Journal of Animal Science, vol. 79, no. 4, **(1999)**, pp. 519-523.

## Authors

**Xiaoming Li**, (1978), male, Ph.D. associate professor, mainly engaged in the research and application of food quality and safety traceability technology.

Copyright ⓒ 2016 SERSC

 Copyright ⓒ 2016 SERSC