BCSIR

Chemometric model for rapid detection of urea and hydrogen peroxide in milk

M. Basak[1], M. N. Uddin[2]*, R. A. Jahan[3], A. A. Rana[1] and M. M. Karim[1]

[1]*Department of Applied Chemistry and Chemical Engineering, University of Dhaka, Bangladesh*
[2]*Bangladesh Council of Scientific and industrial Research (BCSIR), Dhaka-1205, Bangladesh*
[3]*Centre for Advanced Research in Sciences, University of Dhaka, Dhaka-1000, Bangladesh*

Abstract

A simple and cost effective method has been developed for determination of adulteration in milk with urea and hydrogen peroxide by using chemometric modeling with Fourier Transform Infrared (FTIR) spectroscopic data. Milk samples were purchased from a dairy farm (South Banasree, Dhaka, Bangladesh) and spiked at different concentrations of urea and hydrogen peroxide. Spectral data of all samples were collected using ATR-FTIR spectrophotometer. After acquisition of spectral data, they were preprocessed with transformation techniques such as multiplicative scatter correction (MSC) and savitzky-golay derivative. The predictive performance of principal component regression (PCR) and partial least-square regression (PLSR) methods were assessed by relative prediction errors and recoveries (%) were compared . PLSR shows better prediction efficiencies over PCR with R^2 value 99% for urea and 95% for hydrogen peroxide.Six brands of commercial milk samples have been evaluated by this method and the samples contain 21.66-44.73 mg urea and 1.62-2.86 mg hydrogen peroxide in 100 ml milk. This method can be easily used in the quality assessment of milk.

Keywords: Fourier-Fransform Infrared spectroscopy (FTIR); Multiplicative Scatter Correction (MSC); Savitzky-Golay derivative; Partial Least-Square Regression (PLSR)

Introduction

Milk is a significant dietary sustenance and prominent dairy item consumed by people of all ages throughout the world. Liquid milk has been a significant nutrient source for thousand years (Zhang *et al*., 2014). Now-a-days food adulteration is a worryingly increasing problem in Bangladesh. So, it has immense importance to identify the adulterant in milk commercially available in our country.

To satisfy the developing need, milk and its items have been adulterated by addingcheap materials to increase the amount for economic valuethat eventually decrease the quality of milk. Urea, starch/blotting paper, water, glucose/sugar, caustic soda, melamine, refined vegetable oil (cheap cooking oil), hydrogen peroxide, white paint and detergent or

shampoo are the most common adulterants found in milk. These components reduce the nutritious value and impose risk to health (Kishor and Thakur, 2015; Das *et al.,* 2015). Among various adulterants used in adulteration of milk, urea and hydrogen peroxide have been chosen to identify in this study as there are very few techniques available for these ingredients.

Urea is generally added to increase the solid non-fat (SNF) value which is an important nutritional parameter of raw milk. Urea being a soluble agent is commonly found adulterant in milk and other dairy products. Presence of such adulterant eventually affects the quality of milk. The normal concentration of urea in milk is expected to be around

*Corresponding author e-mail: m2nashir@yahoo.com

(10-14) mg/100ml (Banupriya *et al*., 2014). Urea higher than 14mg/100ml is treated as added urea.

Hydrogen peroxide is added as a preservative in milk. It has been broadly utilized for conservation of raw milk because of its bactericidal properties (Haddadin *et al.,* 1996). High concentrations of hydrogen peroxide bring about changes in the milk protein. Excess amount of it can cause harmful impacts on dietary benefit of milk, for example, the

The above mentioned methods are slow and time consuming and most of them involve complex separation methods and need the use of different chemicals that are harmful to the environment. These drawbacks can be overcome by using the chemometric methods which is less time consuming. Besides other applications, chemometric method has been applied in assessing the quality of juice (Uddin *et al.,* 2017; Cozzolino *et al.*, 2011), meat (Kamruzzaman *et al.*, 2012), edible Vegetable

Table I. Composition of the calibration samples through Orthogonal Experimental Design

Sample ID	Urea (mg/100ml)	Hydrogen Peroxide (ml/100ml)	Sample ID	Urea (mg/100ml)	Hydrogen Peroxide (ml/100ml)
1	72	1.9	19	88.5	3.0
2	111	2.4	20	124.5	3.3
3	120	1.1	21	100.5	4.0
4	144	0	22	120	0.6
5	79.5	2.1	23	43.5	1.7
6	90	1.0	24	66	3.6
7	46.5	0	25	0	2.9
8	0	1.2	26	25.5	1.4
9	19.5	1.5	27	18	2.9
10	78	0	28	24	3.2
11	106.5	0	29	0	0
12	0	2.1	30	22.5	0.6
13	21	2.3	31	0	3.8
14	58.5	0.5	32	21	0
15	36	2.4	33	52.5	0.7
16	0	0.6	34	64.5	1.1
17	133.5	1.8	35	54	1.9
18	108	0.5			

degradation of folic acid, which is a basic nutrient to human body. Moreover, severe gastrointestinal problems can be caused by the ingestion of hydrogen peroxide at high levels.

Most of the investigative techniques proposed for distinguishing and evaluation of urea and hydrogen peroxide in milk are dependent on chromatographic and spectroscopic procedures (Ramesh *et al.,* 2015). Among those, quantitative estimations have been performed in several investigations (Banupriya *et al*., 2014). Another method based on UV-Vis spectroscopic technique has also been performed by measuring the intensity of light in a part of the spectrum, especially transmitted or emitted by a particular spectrum (Ramesh *et al.,* 2015).

oils (Karim *et al.,* 2015). But there has not been enough research work carried out to determine urea and hydrogen peroxide in cow milk by chemometric analysis with the help of ATR-FTIR technique. So it has become necessary to develop a method to determine urea combined with hydrogen peroxide adulterants in milk.

Materials and methods

Reagents, chemicals and samples

Raw cow milk was supplied by a dairy farm near South Banasree, Dhaka, Bangladesh. Six different local commercial samples were purchased from the local shops. Acetone (HPLC-Grade), urea and hydrogen peroxide with

highest purity were used for spiking.

Preparation of standard solutions

Total 35 preparations (calibration set) of standard solutions (Table I)containing mixture of milk, urea and hydrogen peroxide were prepared for the development of chemometric calibration model. Raw milk was mixed with urea and hydrogen peroxide in different concentration by following Orthogonal Experimental Design (OED). This experimental design was performed by using software SPSS (version 22.0).

FT-IR analysis

Transmittance measurement of the milk samples was carried out by using a ATR-FTIR spectrophotometer (Model- IR Prestige 21, Shimadzu, JAPAN). Spectral data of raw milk sample and standard samples were collected using Fourier transform infrared (FTIR) spectrophotometers in the wave number range of 4000-700 cm^{-1}. The instrument was standardized before each day's analysis. Weighing of the samples was carried out by using a standard calibrated balance (Model-H0001, A & D Company Ltd., JAPAN). A vortex mixture of Benchmark Scientific Inc. USA (Model-BV 1000) was also used to mix the sample solutions. Obtained data were processed by a licensed copy of CAMO the Unscrambler (Ver. 10.5).

Spectral acquisition

Infra-Red spectra were recorded by transferring each sample (approximately 0.5ml) to the ATR plate attached to the FTIR spectrophotometer. Multiple spectra (three times) were collected for each sample. All spectra were acquired in the range of 4000 to 700 cm^{-1} at a resolution of 1cm^{-1}. Acetone was used to clean the ATR stage 1 and dried after each experiment to ensure a clean surface of the crystal so as to obtain the best possible spectra.

Preprocessing of spectral data

The acquired IR data was preprocessed with some transformation to extract the meaningful information and to enhance the predictive capability for determining the analytical parameters out of the spectral data(Ahmad *et al.,* 2016).This preprocess is done for introducing changes in the values of the variables so as to make them better suited for an analysis. A wide range of transformations can be applied to data before they are analyzed. In this study, transformation techniques such as Multiplicative Scatter Correction (MSC), Savitzky-Golay derivative and their combinations were applied and their efficiency was assessed.

Multiplicative Scatter Correction (MSC) is a transformation technique that is used to compensate for additive and/or multiplicative effects in spectral data. MSC was initially intended to manage multiplicative scattering alone. However, a number of similar effects can be successfully treated with MSC, such as path length problems,interference, offset shifts, etc.The idea behind MSC is that the two effects, amplification (multiplicative) and offset (additive), should be removed from the data table to avoid that they dominate the information (signal) in the data table.

Another transformation technique Savitzky-Golay algorithm depends on performing a least squares linear regression fit of a polynomial around each point in the spectrum so as to obtain smooth data. It is showed by Savitzky and Golay that a set of integers could be derived and utilized as weighting coefficients to do the smoothing operation. In the study, Savitzky-Golay smoothing with 1st and 2nd derivatives and their combined treatments have been applied.

Chemometric method

Chemometrics emerged from chemistry has presented new techniques equipped for managing a lot of compound information by means of multivariate data analysis. In recent years, multivariate chemometric strategies appear to be the techniques showing the best execution in terms of complex compounds. In each spectrum there is enormous number of absorbance value for each wave point. Each wave point or data point is considered as spectroscopic variable. These variables are tremendous in number and are mutually correlated. In this situation Ordinary Least Square (OLS) method cannot be used as there is a problem of singularity. So we are going to use Principal Component Analysis (PCA), Principal Component Regression (PCR) and Partial Least Square Regression (PLSR) in our study and compared their prediction efficiency.

Partial Least-Square Regression (PLSR)

After pre-processing of the spectral data, PLSR has been used to predict different parameters of milk. A recently developed generalization technique of multiple linear regressions (MLR) is partial least squares (PLS) regression (Martens and Naes, 1996; Wold *et al.,* 2001; Brereton, 2000; Naes *et al.,* 2002). PLS regression is specifically noteworthy on the grounds that, unlike MLR, it can examine data with strongly collinear (correlated), noisy and redundant variables (X variables or wavelengths) and furthermore model several characteristics (Y values) at the same time (Lindgren *et al.,* 1993; Tenenhaus *et al.,* 2015; Fornell and Cha, 1994; Geladi, 1988).

For soft modeling in research and industrial applications, PLSR is used where features from principal component analysis and multiple regressions is generalized and combined by PLS regression. A set of dependent variables from a set of independent variables or predictors is analyzed and predicted by PLS regression (Uddin *et al.*, 2019).

The fitness of the model was evaluated by the coefficient of determination (equation 1) and the root means square error of the cross validation (equation 2). RMSECV is the square root of average squared difference between outputs and targets. Root mean square error can be interpreted as the percentage of variance in the dependent variable that can be explained by the predictors. R^2 is a statistical measure of how close the data are fitted to the regression line.

$$R^2 = 1 - \sum_{i=1}^{n}(m_i - p_i)^2 / \sum_{i=1}^{n}(m_i - \bar{m})^2 \qquad (1)$$

$$RMSECV = \sqrt{\sum_{i=1}^{n}\frac{(m_i - p_i)^2}{n}} \qquad (2)$$

Here m_i, p_i, m and n are measured, predicted, mean value and total number of spectra in data set respectively.

Results and discussions

Spectra of raw milk solution and together with a mixture of milk, urea and hydrogen peroxide are shown in the following Fig. 1. It can be seen from the figure that the spectra of raw milk and the mixture have overlapped significantly. The quantification of the adulterants in the mixture by traditional univariate calibration method is interrupted by the overlapping spectra. So, multivariate calibration technique was carried out.
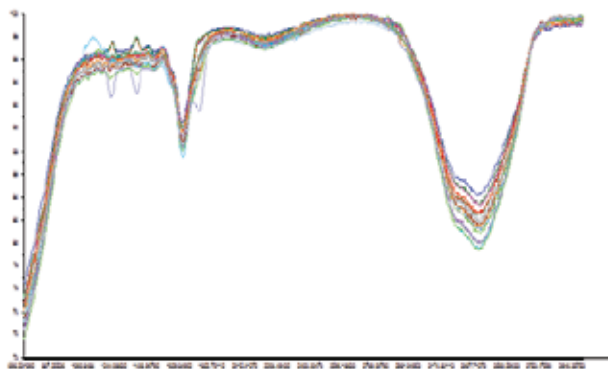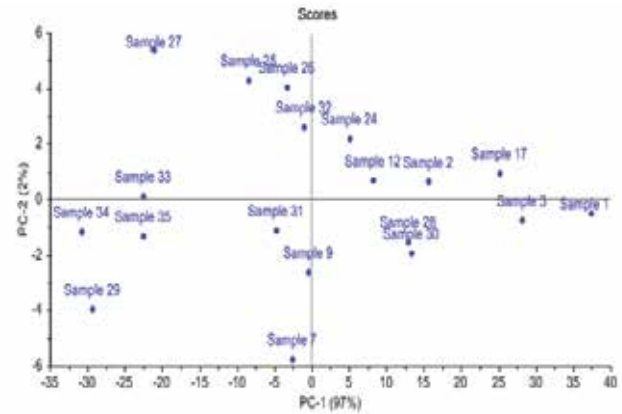

Fig. 2. Association between samples (score plot)

Fig. 2. shows the plot of the first principal component (PC1) against the second one (PC2) for the calibration mixtures. The plot shows the data distribution is called the score plot. From principal component analysis (PCA) it is shown that the first principal component (PC1) expresses 97% of the variation whereas PC2 express 2% of the total variation. So the sum of 99% is good enough for an excellent representation of the dimensional variable spaces in the two dimensional projection.

Fig. 3. represents the influence plot. During calibration, outliers related to the spectra arise in a PCA scores plot as points outside the normal range of variability. Outliers are cases that do not correspond to the model fitted to the bulk of the data. Outliers may be generated by typing errors, interface errors, file transfer, sensor malfunctions and fouling, poor sensor calibration, poor sample preparation or
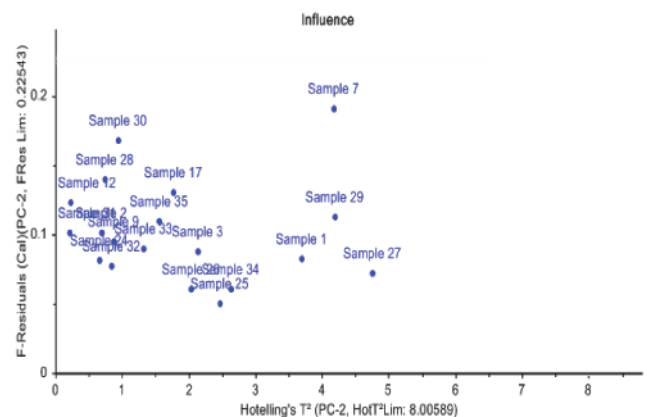

Fig. 1. Spectral presentation of transmittance against wavenumber for the raw and standard sample data

Fig. 3. Influence plot

poor sample presentation. PCA is very sensitive to outliers and can lead to misleading results when outliers are present.From the influence plot (Fig. 3.), we can see that no outlier in the samples considered in the study.

Explained variance plot predicts that highest explained variance is found with 7 PCs. First seven principal components express 99.88% of total variations for calibration data and 99.65% for validation data.

Spectral values and concentrations of urea and hydrogen peroxide in milk obtained from the experimental design are considered as data matrix for the study. 75% mixture solutions were used for the development of the model and remaining 25% mixture solutions were used for the validation of the developed model.

Validating a model means checking how well the model will perform on new data. A regression model is typically made to do predictions in the future. The validation of the model estimates the uncertainty of such future predictions. If the uncertainty is reasonably low, the model can be considered valid.

There are two model parameters considered in the sample. The slope of a line is a number which describes both thedirection and the steepness of the line. The predicted Y-value of the model is plotted against the measured Y-value. This is a good way to check the quality of the regression model. If the model gives a good fit, the plot will show points close to a straight line through the origin and with slope equal to 1. Another parameter offset isa short distance measured perpendicularly from a main survey line. In the regression model, the point where a regression line crosses the ordinate (Y-axis) is called offset.

The developed models are assessed by different parameters to see the effectiveness of the model for predictive purpose. Here two model efficiency parameters such as Root Mean Square Error (RMSE)and Coefficient of Multiple Determination (R^2) were considered. For urea the RMSE value of PLSR is 35.10 which is considerably high. Another parameter R^2 is 0.24 or 24% which is relatively poor. Therefore, the model is not a good one to predict urea in milk. For hydrogen peroxide the RMSE value is 1.02 which is low. Another parameter R^2 is 0.22 or 22% which is relatively poor. Therefore, the model is not a good one to predict hydrogen peroxide in milk to assess the quality. In order to improve the predictability of the models, some preprocessing techniques have been applied.
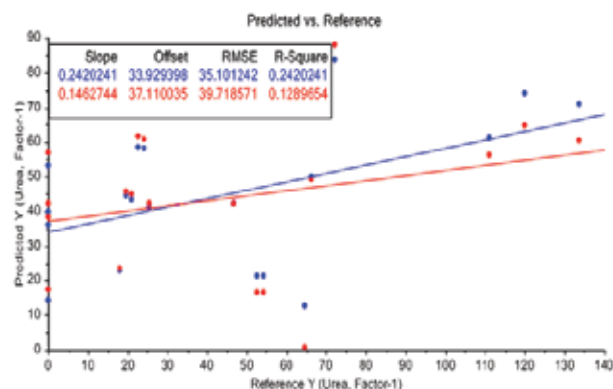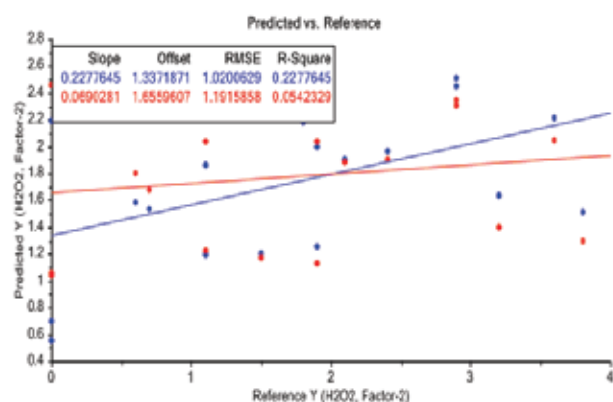

Fig. 4. Predicted vs. reference plot for Urea

Fig. 5. Predicted vs. reference plot for H$_2$O$_2$

As raw data are not good enough for predictive purpose, by doing Multiplicative Scatter Correction (MSC) treatment,very satisfactory result is found. This model can be used for industrial and commercial purpose.

In this model for urea, the RMSE value is 3.35 which is comparatively low. Another parameter R^2 is 0.99 or 99% which is relatively high. Therefore, the model is a good one to predict urea in milk to assess the quality. In the model for H$_2$O$_2$ the RMSE value is 1.04 which is comparatively low. Another parameter R^2 is 0.19 or 19% which is relatively poor. Therefore, the model is not a good one to predict hydrogen peroxide in milk to assess the quality. MSC treatment gives a satisfactory model for the prediction of urea. But this preprocessing technique does not provide a satisfactory model for the prediction of hydrogen peroxide. So another preprocessing technique in combination with MSC have been carried out.
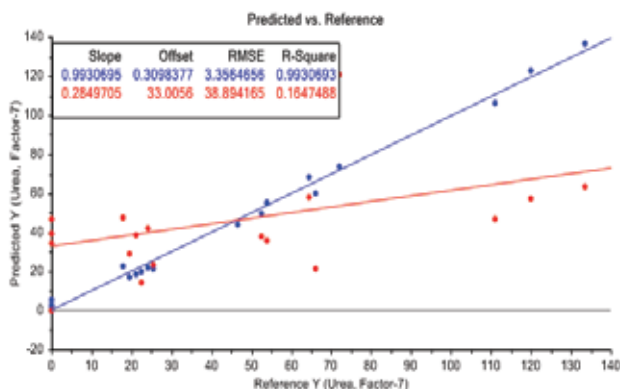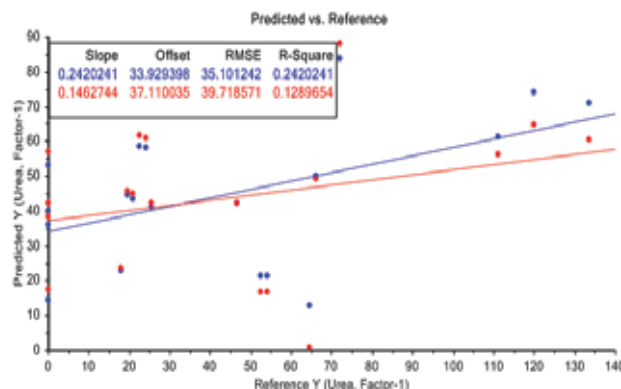
Fig. 6. Predicted vs. reference plot for urea

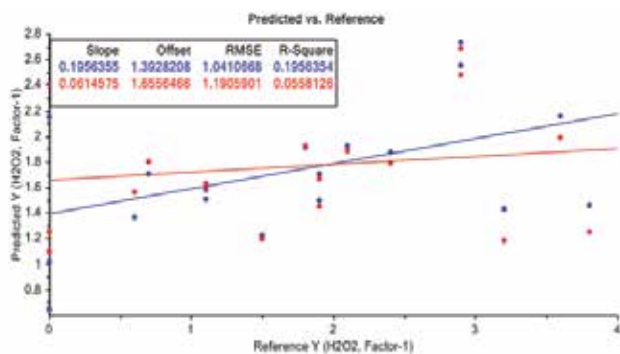

Fig. 8. Predicted vs. reference plot for urea
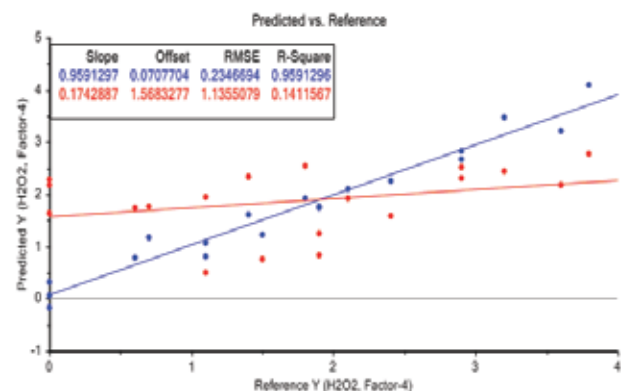

Fig. 7. Predicted vs. reference plot for H$_2$O$_2$

Fig. 9. Predicted vs. reference plot for H^{2}O^2

By doing combined treatment of MSC and Savitzky-Golay 2nd derivative, very satisfactory result is found.

In the model for urea, the RMSE value is 35.10 which is comparatively high. Another parameter R^2 is 0.24 or 24% which is relatively poor. So the model is not a good one to predict urea in milk. In the model for hydrogen peroxide, the RMSE value is 0.23 which is comparatively low. Another parameter R^2 is 0.95 or 95% which is relatively high. So the model is a good one to predict hydrogen peroxide in milk.

Comparison

A comparative study about the methods used is carried out to determine which method fits the best to determine the quantity of the components in the commercial samples. So a comparison between the raw and treated data for the values of RMSE and R^2 of the methods used in prediction is given in Table II.

About 99% R^2 for urea from the MSC preprocessing and 95% R^2 for hydrogen peroxide from the combined treatment of

MSC and Savitzky-Golay 2nd derivative preprocessing is obtained. So the prediction about the amount of components in the commercial samples will be sufficiently accurate.

Predicted results

The commercial samples are then run into the model to predict the amount of urea and hydrogen peroxide in them. The predicted data is given in Table III.

These are the amount of urea and hydrogen peroxide found in commercial samples. As we stated before, urea in cow milk varies from 10-14 mg/100ml. urea content above 14mg/100ml is treated as added Urea. From the above table, it is evident that these commercial samples contain added urea. It is also seen that these commercial samples contain little amount of hydrogen peroxide too.

Table II. Comparison between raw and treated data for calibration and validation data by PLSR

Data	Components	RMSE	R^2
Raw data	Urea	35.10	0.24
	H_2O_2	1.02	0.22
Treated data with mSC	Urea	3.35	0.99
	H_2O_2	1.04	0.19
Treated data with MSC+S-G^{2nd}	Urea	35.10	0.24
	H_2O_2	0.23	0.95

Table III. Predicted values of urea and hydrogen peroxide in commercial samples

Commercial samples	Amount of urea (mg/100ml)	Amount of hydrogen peroxide (mg/100ml)
Sample 1	34.48	2.25
Sample 2	24.33	2.0
Sample 3	21.66	1.62
Sample 4	39.60	2.35
Sample 5	44.73	2.86
Sample 6	35.43	2.32

Conclusion

From this study, we can conclude that FTIR spectroscopic data and PLS regression can be used to determine urea (R^2=99%) and hydrogen peroxide (R^2=95%) as adulterant in milk when the spectroscopic data are pretreated with MSC and MSC+S-G filtering.The proposed technique is rapid, non-dangerous, straightforward and easy to utilize. Compared to traditional methods, multivariate data analysis combined with FTIR instrumental techniques gives a new and a better insight into complex problems by measuring a great number of chemical compounds at once.The chemometric methods developed in the present study can be used to analyze a large number of samples within a very short time with great precision.

References

Ahmad MH, Nache M, Waffenschmidt S and Hitzmann B (2016), A fluorescence spectroscopic approach to predict analytical, rheological and baking parameters of wheat flours using chemometrics, *Journal of Food Engineering* **182**: 65-71. DOI: org/10.1016/j. jfoodeng. 2016.03.006

Banupriya PCRS, Supriya TV and Varshitha V (2014), Comparison of different methods used for detection of urea in milk by quantification of ammonia, *Int J Adv Res Elect, Electron InstrumEng* **3(3)**: 7858-63.

Brereton RG (2000), Introduction to multivariate calibration in analytical chemistry, *The Analyst* **125**: 2125-2154.

Cozzolino D, Cynkar WU, Shah N and Smith P (2011), Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality, *Food Research International* **44**(7): 1888-1896. DOI: org/10.1016/j.foodres.2011.01.041

Das S, Sivaramakrishna M, Biswas K and Goswami B (2015), A low cost instrumentation system to analyze different types of milk adulteration, *ISA transactions* **56**: 268-275. DOI: org/10.1016/j.isatra.2014.11.021

Fornell C and Cha J (1994), Partial least squares: Advanced methods of marketing research, *Advanced methods of marketing research* **407**: 52-78.

Geladi P (1988), Notes on the history and nature of partial least squares (PLS) modelling, *Journal of Chemometrics* **2(4)**: 231-246.

Haddadin MS, Ibrahim SA and Robinson RK (1996), Preservation of raw milk by activation of the natural lactoperoxidase systems, *Food Control* **7(3)**: 149-152. DOI: org/10.1016/0956-7135(96)00023-0

Kamruzzaman M, ElMasry G, Sun DW and Allen P (2012), Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis, *Analytica Chimica Acta* **714**: 57-67. DOI: org/10.1016/j.aca.2011.11.037

Karim MM, Karim SFE, Rana AA, Masum SM, Mondol A and Israt SS (2015), ATR-FTIR spectroscopy and chemometric techniques for the identification of edible vegetable oils, *Bangladesh J. Sci. Ind. Res.* **50**: 233-240. DOI: org/10.3329/bjsir.v50i4.25830

Kishor K and Thakur R (2015), Analysis of milk adulteration using mid-IR spectroscopy, *International Journal on Recent and Innovation Trends in Computing and Communication* **3**(10): 5890-5895.

Lindgren F, Geladi P and Wold S (1993), The kernel algorithm for PLS, *Journal of Chemometrics* **7**(1): 45-59.

Martens H and Naes T (1996), Multivariate calibration. John Wiley and Sons Ltd., Co., New York.

Naes T, Isaksson T, Fearn T and Davies T (2002), A user-friendly guide to multivariate calibration and classification.Chichester, NIR Publications, UK. p 420. DOI: 10.1255/978-1-906715-25-0

Ramesh KB, Gandhi KK, Valecha P, Shradda PK and Sushma M (2015), Quantification of Urea in Milk- A review of Existing Methods, *International Journal of Engineering Research and General Science* **3**(2): 809-814.

Tenenhaus M, Vinzi VE, Chatelin YM and Lauro C (2005), PLS path modeling, *Computational statistics & data analysis* **48**(1): 159-205.

Uddin MN, Mondol A, Karim MM, Jahan RA and Rana AA (2019), Chemometrics assisted spectrophotometric method for simultaneous determination of paracetamol and caffeine in pharmaceutical formulations, *Bangladesh J. Sci. Ind. Res.* **54(3)**: 215-222. DOI: org/10.3329/bjsir.v54i3.42673

Uddin MN, Majumder AK, Abdullah ATM and Kabir MA (2017), Chemometrics Assisted Method for Classification of Mango Juice as Adulterated or Safe with over Use of Artificial Colours by UV Spectroscopic Data, *Journal of Food and Nutrition Sciences* **5**(2): 51-56. DOI: 10.11648/j.jfns.20170502.15

Wold S, Sjöstrom M and Eriksson L (2001), PLS-regression: A basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems* **58**: 109-130.

Zhang, LG, Zhang X, Ni LJ, Xue ZB, Gu X and Huang SX (2014), Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy, *Food chemistry* **145**: 342-348. DOI: org/10.1016/j.foodchem.2013.08.064