COMPUTATIONAL ESTIMATION OF SOYBEAN OIL ADULTERATION IN NEPALESE MUSTARD SEED OIL BASED ON FATTY ACID COMPOSITION

K. SHRESTHA & B. DE MEULENAER
Research Group Food Chemistry and Human Nutrition,
Department of Food Safety and Food Quality, Ghent University,
Coupure links 653, 9000 Ghent, Belgium

INTRODUCTION

Adulteration of edible oil with cheaper and lower quality oil has been one of the interests of research for long time. The impossibility of physical separation of the individual oil from the blended oil creates complication on the estimation and detection capabilities (Vliet *et al.,* 2007). Various approaches have been made for the estimation of adulteration of different edible oils, and most of the methods are based on fatty acid composition using different chemometrics techniques and they are generally more complex (Vliet *et al.,* 2007; Marini *et al.,* 2004; Voncina *et al.,* 2005).

Mustard oil is a very common and popular oil in Nepal. Soybean oil, being comparatively cheaper one, could be most possible intentional adulterant on mustard oil. The present study is carried out in order to develop a method based on simple and easily available software to estimate soybean oil adulteration in the mustard oil using only fatty acid composition. This study also aims to study the variability in the fatty acid composition of mustard samples and its impact on the adulterants detection capability.

MATERIALS AND METHODS

Fatty acid composition (reference data set)

Two hundred and three fatty acid composition dataset of mustard seed oil, two of soybean oil, one of corn oil, one of sunflower oil and two of argemone oil were collected from different literatures. Analysis of fatty acid composition of six different mustard seed samples collected from Nepal, six different mustard seed samples collected from Tierenteyn, Belgium and one soybean oil sample collected from GB supermarket, Belgium were carried out in our laboratory. The compilation of all these fatty acid compositions was taken as the reference data set and was used to develop a method to predict the soybean oil adulteration in a simulated blend.

Fatty acids analysis

The boron trifluoride method was used for the preparation of fatty acid methyl esters (FAMEs) and fatty acids analysis was carried out by gas chromatography (AOCS, 1990).

Adulteration prediction

Minimisation of weighted sum of squared error method using Microsoft Excel solver was used for adulteration prediction. The weight was given for the reciprocal of variance of each fatty acid.

RESULTS AND DISCUSSION

Method development

Validation of theoretical background on real soybean mustard oil blend

Soybean and mustard oil blends containing 0, 5, 10, 20, 40, 60, 80 and 100 % of soybean oil were prepared in duplicate. The fatty acid composition of these blends were analysed and expressed as percentage of total fatty acid. The fatty acid composition dataset of pure soybean and mustard oil was taken as reference set and minimisation of SSE by Microsoft Excel solver was carried out to predict the percentage of soybean oil. The obtained result of predicted soybean oil and actual soybean oil along with 95% prediction interval (dashed line) has been shown in Figure 1 (A). The linear regression between actual and predicted soybean oil % showed the $R^2 > 0.99$. The slope was highly significant ($p < 0.001$) and was not significantly different from 1 ($p > 0.05$) and the intercept value was not significantly different from 0. These statistics assure the equivalency of actual and predicted soybean oil % and give support to theoretical background of the method.

PCA (principal component analysis) of fatty acid composition of reference dataset

PCA was carried out on the reference fatty acid composition data. Furthermore, K-mean clustering was carried out on that dataset and six clusters were found to be optimum based on SSE. The different clusters have been shown with different symbols in PCA plot (Figure 1 (B)). There were two clusters of high erucic mustard (cluster 1 and cluster 2), two clusters of low erucic mustard (cluster 3 and cluster 4), one cluster of all adulterant (soybean, sunflower, corn and argemone oil) (cluster 5) and one heterogeneous cluster of mustard (cluster 6). Mean SSE of cluster 1 to 5 were in the range of 25 to 82.15, while that of cluster 6 was 276.52. Since cluster 6 was not a homogenous cluster and constituted only around 3% of total dataset, this cluster was not used for further calculations. The two high erucic mustard clusters 1 and 2 were named H_1 and H_2, two low erucic mustard clusters 3 and 4 were named L_1 and L_2 respectively.

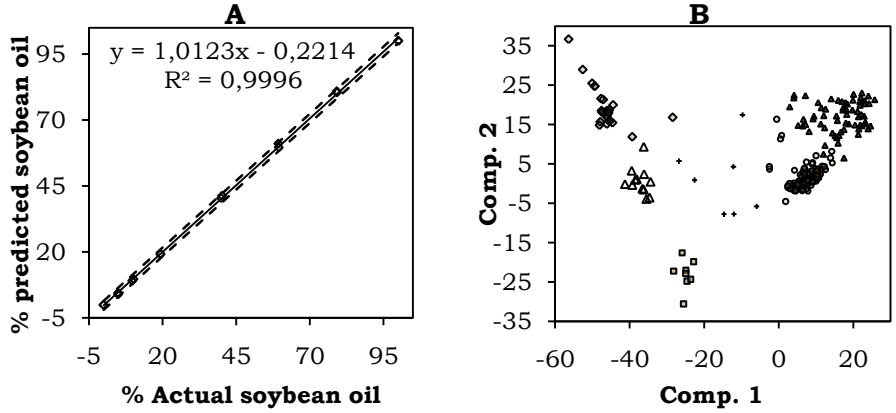

Fig. 1. (A) actual vs predicted percentage of soybean oil in the real blend by solver method (dotted lines represents 95% prediction interval) and (B) PCA

of all reference data set (○ cluster 1, ▲ cluster 2, ◇ cluster 3, △ cluster 4, □ cluster 5, + cluster 6).

Method on simulated soybean mustard blend

Division of data set into training and validation set

The division of whole reference data set of fatty acid composition into two sets was done for each cluster independently. Each cluster was divided randomly using random number generator of Microsoft Excel with discrete distribution of 0 and 1, giving equal probabilities (0.5) for both.

Development of solver method from the training data set

The training set contained H_1, H_2, L_1, L_2 and S clusters. The mean fatty acid composition of each cluster was taken as representative for that cluster. For each fatty acid, variance was calculated within a cluster and then pooled variance was calculated among clusters. The reciprocal of pooled variance was taken as weight and weighted SSE minimisation between observed and predicted fatty acid was performed. Finally, the sum of predicted percentages of H_1 and H_2 clusters was expressed as high erucic mustard oil percentage and that of L_1 and L_2 clusters was expressed as low erucic mustard oil percentage.

Solver options were: maximum time 500 sec, iterations 10000, precision 0.000001, convergence 0.0001, assume non-negative, estimate tangent, derivative forward and search conjugate method. Solver was always started beginning from equal percentage allocation from each cluster to allow equal probability of selection by solver in the blend. A macro was developed to automate the process of solver in Microsoft excel with above options.

Evaluation and validation of method by using mathematical simulated blend generated from training set and validation set

Blending Simulation

The random number generator of Microsoft Excel was used to randomly select as well as to assign the quantity for the preparation of simulated blends of mustard and soybean oil. We prepared 4000 simulated blend each from training set and validation set data.

Prediction based on method using solver

The solver method was used for the prediction of soybean, high erucic and low erucic mustard oil in the training set and validation set blends. The graphical representation of actual versus predicted percentage of soybean oil by using solver method along with 95 % prediction interval is shown in Figure 2. In all the cases, the value of R^2 was more than 0.97, and the slope was close to 1. The LODs based on 3 sigma limit on the validation set were found to be 13.64, 11.14 and 16.41 percent for soybean, high erucic and low erucic mustard oil respectively. The high amount of LOD is attributed to the high variability in the fatty acid composition of different samples. The 95 % prediction intervals for soybean, high erucic and low erucic mustard oil were 8.92, 7.28 and 10.72 percent respectively.

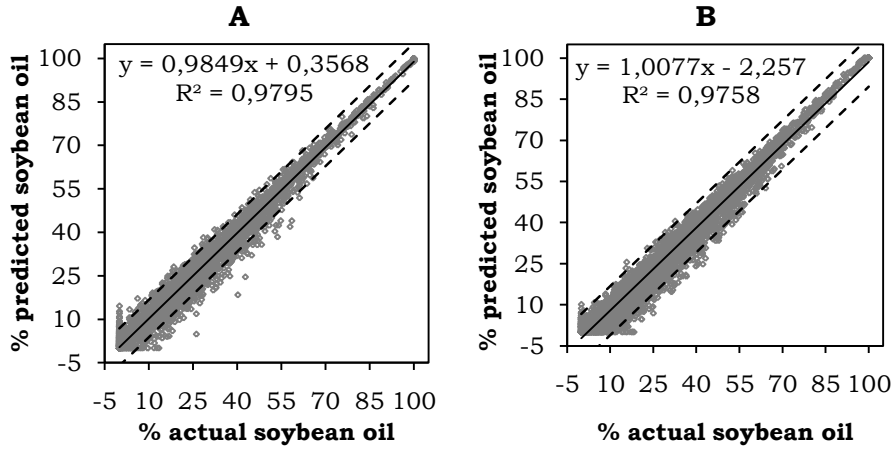

Fig. 2. Plot of actual vs. predicted % of soybean oil by solver method on the blend from training set (A) and validation set (B) (dotted lines represents 95% prediction interval).

CONCLUSIONS

One mathematical method was developed to estimate the soybean oil adulteration in mustard oil based on fatty acid composition. This method worked very well in the real laboratory blending study. The method was further studied on the extreme condition of variability observed in the literature. This method also worked on the simulated blends of different mustard and soybean oil. As expected, the detection capability decreased when simulated in extreme variability conditions observed in the literature data. The blending was simulated with samples from all around the world along with blending of up to 16 samples. These are generally extreme conditions. Normally, there are limited types of samples available in the Nepalese market, and only considering the variability within that will help to increase detection capability further. This method estimate adulterant as soybean oil. Further confirmation of type of adulterant could be performed by other appropriate analysis (e.g. tocopherol, sterol and other minor component analysis). Finally, one limitation of our method needs to be considered before using it. This method was developed based on consideration of 4 mustard clusters (two of high erucic H_1 and H_2 and two of low erucic cluster L_1 and L_2) and hence does not necessarily work for samples not falling within those clusters.

REFERENCES

VAN VLIET, M.H, VAN KEMPEN, M.P., REINDERS, M.J.T., DE RIDDER, D. (2005). Computational estimation of the composition of fat/oil mixtures containing interesterifications from gas and liquid chromatography data. J. Am. Oil Chem. Soc. **82** (10):707-716.

MARINI, F., BALESTRIERI, F., BUCCI, R., MAGRI, A.D., MAGRI, A.L., MARINI, D. (2004). Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. Chemom. Intell. Lab. Syst. **73** (1):85-93.

VONCINA, D.B., KODBA, Z.C., NOVIC, M. (2005). Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. Chemom. Intell. Lab. Syst. **75** (1): 31-43.

AOCS (1990). Fatty acid composition by GLC. AOCS Official Method Ce 1b-89.