

Establishment of Mathematical Relationships Between Smoke Point and Minor Compounds in Vegetable Oils Using Principal Component Regression Analysis

Yiting Wu^a, Bing Fang^{a*}, Hongyan Wu^b, MingZhang^c, Yuemin Shen^d

^aBeijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science & Nutritional Engineering, China Agricultural University, Beijing, 100083, China

^bMilitary General Hospital of Beijing PLA., Beijing, 100700, China

^cSchool of Food Science and Chemical Engineering, Beijing Technology and Business University, Beijing 100048, China

^dKey Laboratory of Functional Dairy, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, 100083, China.
49631263@qq.com

To satisfy consumers' requirement of less smoke, vegetable oils should possess certain thermal stability, which was evaluated by smoke point and dependent on the content of the minor compounds. As the proposal of the moderate refining technology, contents of the minor compounds such as free acid, peroxide, phosphorus, solvent, water and other volatile matters under the condition would change interactively, leading the establish of the vegetable oil standards difficult. The purpose of this work was to investigate the effects of the minor compounds on the smoke point. Soybean oil and rapeseed oil samples with different contents of free fatty acid, peroxide, phosphorus, solvent, moisture and volatile matter were collected and their mathematical relationships with smoke point was analyzed by principal component regression. Based on these results, a regression equation was obtained: $\text{Ln smoke point} = 4.790 - 0.021 \times \text{Ln (acid value)} + 0.014 \times \text{Ln (peroxide value)} - 0.051 \times \text{Ln (content of phospholipids)} - 0.105 \times \text{Ln (moisture and volatile matter)}$ ($R^2 = 0.848$). Furthermore, according to the correlation coefficient of the calculated and measured smoke point (0.972), the suitability of the regression model in terms of its ability to predict the smoke point of different oil samples was good.

1. Introduction

Vegetable oils played an important role in our diet. Although there were large difference between the diet culture of eastern and western countries, vegetable oils were mainly used by heating or frying. The high temperature during cooking would lead to the production of smoke containing the degradation products of oils, which were uncomfortable and harmful to health. To satisfy consumers' requirement, the refining technology of vegetable oils need to ensure the thermal stability of vegetable oils.

Smoke point was used to evaluate the thermal stability of vegetable oils, which corresponded to the lowest temperature when the thermal decomposer or impurities volatilize continuously. There are several non-triglyceride components in crude oils with a boiling point lower than that of triglycerides, such as free fatty acid, peroxide, phosphorus, solvent, water and other volatile matters (Yen et al., 1997). Refining process of vegetable oils could remove these non-triglyceride components, increasing the smoke point and the stability significantly (Akoh and Min, 2008).

The existing technology aimed to maximum remove the minor compounds to extend the shelf-life and thermal stability of vegetable oils, however, it also led to the loss of the micronutrients such as tocopherols and sterols, formation of the harmful trans fatty acids, polymeric triacylglycerols, and oxidized triacylglycerols (Alpaslan et al., 2001; De et al., 1999; Farhoosh et al., 2009). As the widely acceptance of moderate refining, the criteria for vegetable oils was discussed to relax, such as the acid value and the temperature and time during deodorization. Due to the refining process contained several consecutive steps, change of one technology

may influence several compounds, making the establishing of new vegetable oil standards difficult, especially the consistency of each quality indexes.

To maintain the thermo stability of vegetable oils under the moderate refining technology, it was necessary to clarify the effects of the changes of minor compound contents on smoke point. Although there were studies about the influences of certain minor compound on smoke point, such as free fatty acid and phosphorus, the synthetic effect of these non-triglyceride compounds and their mathematical relationship were unsolved. Therefore, in this study, we collected 69 soybean oil (SO) and 80 rapeseed oil (RO) samples with different contents of free fatty acid, peroxide, phosphorus, solvent, moisture and volatile matter from the main oil plants in China. Based on the results of the contents of these minor compounds and smoke point, a mathematical model was established using the principal component regression analysis.

2. Materials and Methods

2.1 Materials and Equipments

SO and RO samples were collected from the main oil plants in China, including Wilmar, Huifu and COFCO. All the other chemicals and solvents were from China. Ultraviolet-Visible spectrophotometer (UV2450, Shimadzu, Kyoto, Japan), Smoke point instrument (HLY-III, Qianjiang, Hangzhou, China).

2.2 Chemical analysis

Acid value, peroxide value, content of phospholipids, residual solvent, moisture and volatile matter of the 149 oil samples were measured according to the AOCS method Cd 3d-63, Cd 8-53, Ca 19-86, Ba 14-87 and Ac 2-41, respectively. Smoke point was measured according to AOCS method Cc 9a-48. All experiments were conducted with repeated treatments.

2.3 Principal components regression

Principal components regression was analyzed using the SPSS 16.0 software. Data was first converted into the logarithmic form and then the multicollinearity for each independent variable was diagnosed. Based on the standardized principal component results, the 'best' regression equation was established.

2.4 Validation of the regression equation

Besides the 149 samples, we also collected another 14 samples to evaluate the validation of the regression equation. The validation of the regression equation was evaluated by the bivariate correlation analysis of the calculated and measured value using the SPSS 16.0 software.

3. Results and Discussion

3.1 Measurement of the component and smoke point

The acid value (AV, X1, mg KOH/g oil), peroxide value (PV, X2, meq/kg), content of phospholipids (CP, X3, mg/100 g oil), content of residual solvent (RS, X4, mg/kg), content of moisture and volatile matter (MCM, X5, %) and smoke point (SP, Y, °C) of the total 149 samples were list in Table 1. It can be seen from Table 1 that the samples have a range of acid value, peroxide value, content of phospholipids, content of residual solvent, content of moisture and volatile matter and smoke point between 0.080-2.760 mg KOH/g oil, 1.400-7.800meq/kg, 0.001-155.000mg/kg, 0.064-2.302 mg/100g oil, 0.010-0.075% and 151.0-220.0 °C, respectively. sample 1-69 was soybean oil, sample 70-149 was rapeseed oil.

3.2 Principal components regression results

After running the SPSS linear regression procedure, obtain the results of correlation coefficient and collinearity diagnostics. The partial regression coefficients beta value of independent variables (X1, X4, and X5) are highly significant ($p < 0.05$). Beta value of X2 is equal to 0.046, indicating there is a positive correlation between peroxide value and smoke point (Y). The result is contrary to the common sense, which was due to the multicollinearities among the independent variables. Meanwhile, tolerance of X1, X2, X4, and X5 (0.042, 0.060, 0.051, and 0.120) was small (close to 0), the variance inflation factors (VIF) of these variables were large (24.011, 16.587, 19.751, and 8.299) (Liu et al, 2003). The eigenvalue of X5 is close to 0 (0.001), indicating the variables are highly inter-correlated and the matrix is said to be ill-conditioned. (Hao and Li, 2011). Furthermore, the condition index of X5 is more than 15 (66.117), indicating the variables were highly inter-correlated.

To solve the multicollinearity of variables, we use principal component regression to establish the regression model. The principal component regression is the method of selecting some principal components and then constructs a regression model regarding them as new explanatory variables (Liu et al., 2003; Wibowo and Yamamoto, 2012). The principal component analysis can gather highly correlated independent variables into a principal component, and all principal components are independent of each other (Kawano et al., 2015; Sainani et al., 2014).

Variables were first transformed to the standardized variables first using the SPSS descriptive statistics procedure and the factor of phospholipids was excluded. It was reported that smoke point decreased as the increasing of phospholipids, however, it needs to get to a certain value such as 30 mg/kg (Wang et al., 1999), even 100 mg/kg (Han et al., 2015), which was unpractical in refined vegetable oils. The standardized variables were then subjected to the factor analysis. The variance proportion of the first three principal components was 68.944%, 27.178%, and 3.463%, respectively, and the cumulative variance proportion was 99.584%. Therefore, we choose Component 1 (C1), Component 2 (C2) and Component 3 (C3) as the independent factor when subjected to the regression analysis. In Table 4-2, obtain coefficients related the four standardized independent variables to three principal components to create expressions of three principal components: $C1 = 0.916 \times Z(X1) + 0.800 \times Z(X2) + 0.719 \times Z(X4) + 0.872 \times Z(X5)$ (Eq (1)); $C2 = 0.353 \times Z(X1) - 0.569 \times Z(X2) + 0.684 \times Z(X4) - 0.413 \times Z(X5)$ (Eq (2)); $C3 = -0.171 \times Z(X1) - 0.177 \times Z(X2) + 0.098 \times Z(X4) + 0.261 \times Z(X5)$ (Eq (3)).

After running the SPSS linear regression procedure using the principal components C1, C2, and C3 as the independent factor, a equation was obtained $Z(Y) = -0.318 \times C1 - 0.222 \times C2 - 1.135 \times C3$ (Eq (4)) ($R^2 = 0.847$). The tolerances and VIFs of principal components C1, C2, and C3 are equal to 1, and their eigenvalues and condition indices are close to 1. These suggested that C1, C2, and C3 are independent of each other. Applied Equation 1-3 to Equation 4 and then get the standardized linear regression equation: $Z(Y) = -0.176 \times Z(X1) + 0.072 \times Z(X2) - 0.492 \times Z(X4) - 0.482 \times Z(X5)$ (Eq (5)).

Table 2: Correlations of variables

		Y	X1	X2	X4	X5
Y	Pearson Correlation	1.000	-0.864 **	-0.534 **	-0.807 **	-0.707 **
	Sig. (2-tailed)		0.000	0.000	0.000	0.000
	Sum of Squares and Cross-products	2.264	-16.759	-5.981	-17.693	-7.355
	Covariance	0.015	-0.113	-0.040	-0.120	-0.050
X1	Pearson Correlation	-0.864 **	1.000	0.558 **	0.878 **	0.611 **
	Sig. (2-tailed)	0.000		0.000	0.000	0.000
	Sum of Squares and Cross-products	-16.759	166.078	53.524	164.862	54.468
	Covariance	-0.113	1.122	0.362	1.114	0.368
X2	Pearson Correlation	-0.534 **	0.558 **	1.000	0.174 *	0.884 **
	Sig. (2-tailed)	0.000	0.000		0.034	0.000
	Sum of Squares and Cross-products	-5.981	53.524	55.425	18.828	45.530
	Covariance	-0.040	0.362	0.374	0.127	0.308
X4	Pearson Correlation	-0.807 **	0.878 **	0.174 *	1.000	0.367 **
	Sig. (2-tailed)	0.000	0.000	0.034		0.000
	Sum of Squares and Cross-products	-17.693	164.862	18.828	212.128	37.019
	Covariance	-0.120	1.114	0.127	1.433	0.250
X5	Pearson Correlation	-0.707 **	0.611 **	0.884 **	0.367 **	1.000
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	
	Sum of Squares and Cross-products	-7.355	54.468	45.530	37.019	47.835
	Covariance	-0.050	0.368	0.308	0.250	0.323

Note: ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed)

3.3 Establish of the linear regression equation

To obtain the general linear regression equation, standardized variables should be transformed to the normal variables. After running the bivariate correlations procedure, we can get the value of Sum of Squares and Cross-products of variables: $L_{YY}=2.264$, $L_{X_1X_1}=166.078$, $L_{X_2X_2}=55.425$, $L_{X_4X_4}=212.128$ and $L_{X_5X_5}=47.835$ (Table 2). Calculate the general partial regression coefficients b_i with $b_1' = -0.176$, $b_2' = 0.072$, $b_3' = -0.492$, and $b_4' = -0.482$ in the light of Equation 5: $b_1 = b_1' (L_{YY} / L_{X_1X_1})^{1/2} = -0.176 \times (2.264/166.078)^{1/2} = -0.021$, $b_2 = b_2' (L_{YY} / L_{X_2X_2})^{1/2} = 0.072 \times (2.264/55.425)^{1/2} = 0.015$, $b_4 = b_4' (L_{YY} / L_{X_4X_4})^{1/2} = -0.492 \times (2.264/212.128)^{1/2} = 0.051$, and $b_5 = b_5' (L_{YY} / L_{X_5X_5})^{1/2} = -0.482 \times (2.264/47.835)^{1/2} = -0.105$. The constant b_0 in accordance with Equation 5 and Table 4, $b_0 = Y_{\text{mean}} - \sum b_i X_{i\text{mean}} = 5.238 - [(-0.021) \times (-0.1237) + 0.015 \times 1.134 + 0.051 \times (-1.1747) + (-0.105) \times (-3.5224)] = 4.790$, and finally obtain the general linear regression equation: $Y = 4.790 - 0.021 \times X_1 + 0.014 \times X_2 - 0.051 \times X_4 - 0.105 \times X_5$ ($R^2=0.848$), that is Ln smoke point= $4.790 - 0.021 \times \text{Ln}(\text{acid value}) + 0.014 \times \text{Ln}(\text{peroxide value}) - 0.051 \times \text{Ln}(\text{content of phospholipids}) - 0.105 \times \text{Ln}(\text{moisture and volatile matter})$ ($R^2=0.848$) (Eq (6)).

3.4 Evaluation the validation of the regression equation

14 samples were left to evaluate the validation of the regression equation. The values of acid value, peroxide value, phospholipids content, residual solvent content, moisture and volatile matter content were applied to Equation 6 to obtain the calculated smoke point value (Table 3). The quality of the fit between the model-predicted and experimental values, and the suitability of the regression model in terms of its ability to describe the data were measured by the bivariate correlation of the calculated and measured value. The correlation coefficient of the calculated and measured smoke point was equal to 0.972 with high significance ($p < 0.001$), indicating a good predictive ability of the smoke point regression.

Table 3: Smoke points and chemical analysis results of soybean oil and rapeseed oil samples

	smoke point (°C)		acid value (mg KOH/g oil)	peroxide value (meq/kg)	residual solvent (mg/kg)	phospholipids (mg/100g oil)	moisture and volatile matter (%)
	calculated	measured					
1	228.9	219.5	0.081	1.462	2.545	0.119	0.011
2	211.4	215.0	0.181	4.418	0.001	0.066	0.030
3	198.9	203.5	0.152	4.950	106.576	0.111	0.044
4	194.8	201.0	0.140	5.200	155.000	0.132	0.050
5	205.5	200.5	0.169	4.629	44.296	0.083	0.036
6	199.5	191.5	0.240	5.800	0.001	0.196	0.030
7	217.7	213.5	0.127	1.029	0.019	0.255	0.010
8	211.4	207.0	0.162	1.183	0.121	0.300	0.012
9	186.9	185.0	0.405	2.245	0.822	0.609	0.026
10	163.8	175.0	0.974	4.738	2.470	1.334	0.057
11	162.6	159.5	1.275	4.022	0.001	2.074	0.046
12	161.8	156.5	2.090	6.100	0.001	1.741	0.050
13	158.8	155.0	2.080	7.200	1.850	1.528	0.065
14	156.0	153.5	1.627	6.800	0.617	1.760	0.075

4. Conclusions

In conclusion, this study provides a predicted equation to evaluate the smoke point according to the contents of minor compounds, which laid a theoretical basis for the determination of the quality indexes of vegetable oils under moderate refining technology. The general linear regression equation was obtained using the principal component regression analysis: $Y = 4.790 - 0.021 \times X_1 + 0.014 \times X_2 - 0.051 \times X_4 - 0.105 \times X_5$ ($R^2=0.848$), that is Ln smoke point= $4.790 - 0.021 \times \text{Ln}(\text{acid value}) + 0.014 \times \text{Ln}(\text{peroxide value}) - 0.051 \times \text{Ln}(\text{content of phospholipids}) - 0.105 \times \text{Ln}(\text{moisture and volatile matter})$ ($R^2=0.848$). Based on the correlation

coefficient of the calculated and measured smoke point ($R^2=0.972$), the predicted equation indicated a good predictive ability.

Reference

- Akoh C.C., 2008, Min D.B. Food lipids, chemistry, nutrition, and biotechnology [M]. New York, 3rd Edition, CRC Press. ISBN-13: 978-1420046632.
- Alpaslan M., Tepe S., Simsek, 2001, O. Effect of refining processes on the total and individual tocopherol content in sunflower oil[J]. *IntJFood SciTech*, 36(7): 737~739. DOI: 10.1111/j.1365-2621.2001.00523.x.
- De Greyt W.F., Kellens M.J., 1999, Huyghebaert A.D. Effect of physical refining on selected minor components in vegetable oils[J]. *Fett/Lipid*, 101(11): S428~S432. DOI: 10.1002/(SICI)1521-4133(199911)101:11<428:AID-LIPI428>3.0.CO;2-T.
- Farhoosh R., Einafshar S., 2009, Sharayei P. The effect of commercial refining steps on the rancidity measures of soybean and canola oils[J]. *Food Chem*, 115: 933~938. DOI: 10.1016/j.foodchem.2009.01.035.
- Han S.S., Meng Z., Li J.W., et al., 2014, Effects of phospholipids and free fatty acids on smoke point and oxidative stability of oils[J]. *China Oils and Fats*, DOI: 10.3321/j.issn:1003-7969(2014) 04-0023 -04. (Chinese)
- Hao W., Li, F. 2010, A new method for ill-conditioned diagnosis based on spatial analysis[J]. *JInformation and Computing Science*, DOI: 10.1016/j.sna.2012.11.047.
- Kawano S., Fujisawa H., Takada T., et al. 2015, Sparse principal component regression with adaptive loading[J]. *ComputStatData An*, Doi: 10.1016/j.csda.2015.03.016.
- Sainani K.L., 2014, Introduction to principal components analysis[J]. *PM&R*, DOI: 10.1016/j.pmrj.2014.02.001.
- Wang X.G., Zhou S.L., Xu X., et al., 1999, Effect of non-triglyceride composition on smoke point of oils[J]. *China Oils and Fats*, DOI: 10.3321/j.issn:1003-7969.1999.01.017. (Chinese)
- Wibowo A., Yamamoto Y., 2012, A note on kernel principal component regression[J]. *Math ComputModel*, DOI: 10.1007/s10598-012-9143-0.
- Yen G.C., Shao C.H., Chen C.J., et al., 1997, Effects of antioxidant and cholesterol on smoke point of oils[J]. *LWT-Food SciTech*, DOI: 10.1006/fstl.1996.0236.