

Principal Components Analysis the Biochemical Compounds Extracted from Dates Using Three Mixture Design and Identification by GC-MS

Taffaha. A. Arhouma¹, M. S. Hassan^{1,2}

¹Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan Malaysia

²Institute of Halal Research and Management (IHRAM), Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan Malaysia

jawadtaff.osama@yahoo.com

mohdsukri@usim.edu.my (Corresponding Author)

Abstract: *Principal Components Analysis (PCA) was performed on multivariate data of GC-MS results to study the relationship between compounds extracted with polarity of solvents mixture from the scores, loadings and plots. We used 13 mixture design of solvents on extraction of biochemical compounds in dates, We founded scores compounds are represented in four clusters (A, B, C and D) on the scores plot and loadings the PC1 and PC 2 accounted 70 % of the total variation with PC 1 having 42 % and PC 2 28 %.*

Keywords: Dates fruit, GC-MS and PCA principal components analysis

1. Introduction

Date fruits have phenolic compounds (mainly cinnamic acids) and flavonoids (flavones, flavonols and flavanones) that provide antioxidant activities as stated by Velioglu et al., 1998, Vayalil (2002) and Mansouri et al., 2005. The compositional and sensory characteristics of three native sun-dried date (*Phoenix dactylifera*) varieties cultivated in Oman, matching the antioxidant activity, anthocyanins, carotenoids, and phenolics for each diverse variety and as studied by Al-Farsi et al., in 2005 concluded that with the creation of the field of functional food and nutraceuticals, any evidences on the health-promoting components of dates will improve the knowledge and appreciation for the uses of dates in these health-promoting products.

GC-MS is one of the most widespread analytical techniques in many scientific fields owing to its high sensitivity low detection limit, rapid identification and having the ability of respectively analyzing number of ingredients analytes; therefore, it is the best appropriate to analyzing the volatile components (Chen et al., 2011).

PCA is probably the most popular multivariate statistical technique and it is widely used by almost all scientific disciplines. It is also likely to be the oldest multivariate technique, as its origin can be traced back to Pearson (1901), but its modern instantiation was formalized by Hotelling (1933) who also coined the term "principal component" (Abdi & Williams, 2010). Jolliffe (2002) also stated that the most generally accepted earliest descriptions of PCA were given by Pearson (1901) and Hotelling (1933). The aim of this study was to find the relationship between the compounds extracted from dates and polarity of solvents mixture using the scores and loadings plots.

2. Materials and Methods

2.1 Mixture Extraction

Methanol, hexane and chloroform were used for the extraction solvents of the mixture of the three different types of dates fruit. The usage of variety of solvents allows selecting the one with greatest ability to extract biochemical compounds in date fruits. The extraction was implemented by weighing 2 g of homogeneous sample and placed in 13 beakers and each was drenched into 20 mL of solvent for 2 h at room temperature. The solution was then filtered using Whatman filter paper No 1 before subjecting to rotary evaporator at 40 °C for concentration of the sample after which BSTFA was added for derivatization process before sent to GC-MS for compounds identification.

2.2 Data Analysis with PCA

PCA was performed with the aid of "THE UNSCRAMBLER[®]X" software (CAMO software version 10.1) on the multivariate data from GC-MS results.

3. Results and Discussion of GC-MS using Principal Component Analysis (PCA)

PCA is a powerful technique for pattern recognition that attempts to explain the variance of large set of inter-correlated variables and transforming into a smaller set of independent (uncorrelated) variables principal components (Svetlana et al., 2012). Explain the multivariate data of GC-MS results. Show the matrix of GC-MS results.

3.1 Scores

FIG. 1 shows the compounds are represented in four clusters (A, B, C and D) on the scores plot. The compounds with positive and high scores are denoted as (A) 1, 2, 3, 4, and 5, and are much close to one another. Compounds 6, 7 and 8

are also loaded and closely packed on PC 1. On the other hand, the compounds with high and negative impact on PC1 (B) include 27, 26, 25, 24, 23, 22, 21 and 20. These compounds are superimposed on one another. Also, on PC 1, compound 19 was negative and quite separate from the other samples. On PC 2, the compounds with the most influence are (D) 29, 30, 32 and 33. Also positively loaded on PC 2 are 9, 8, 5, 2, 3, 1 and 14.

The score plot is a two dimensional scatter plot (or map) of scores for two specified components (PCs) for PCA. The score plots show how well the data is distributed and gives

information in the samples (Wise et al., 2006). The scores plot (PC1, PC2) was used in this study because the two components which reveals more disparity in the data than any other pair of components. The nearer the samples are in the scores plot, the more alike they are with respect to the components concerned (i.e. they have close values for the corresponding variables). On the other hand, samples for which scores differ greatly are quite different from each other with respect to the variables. The score describes the major features of the sample, relative to the variables with high loadings on the PC (CAMO, 2011).

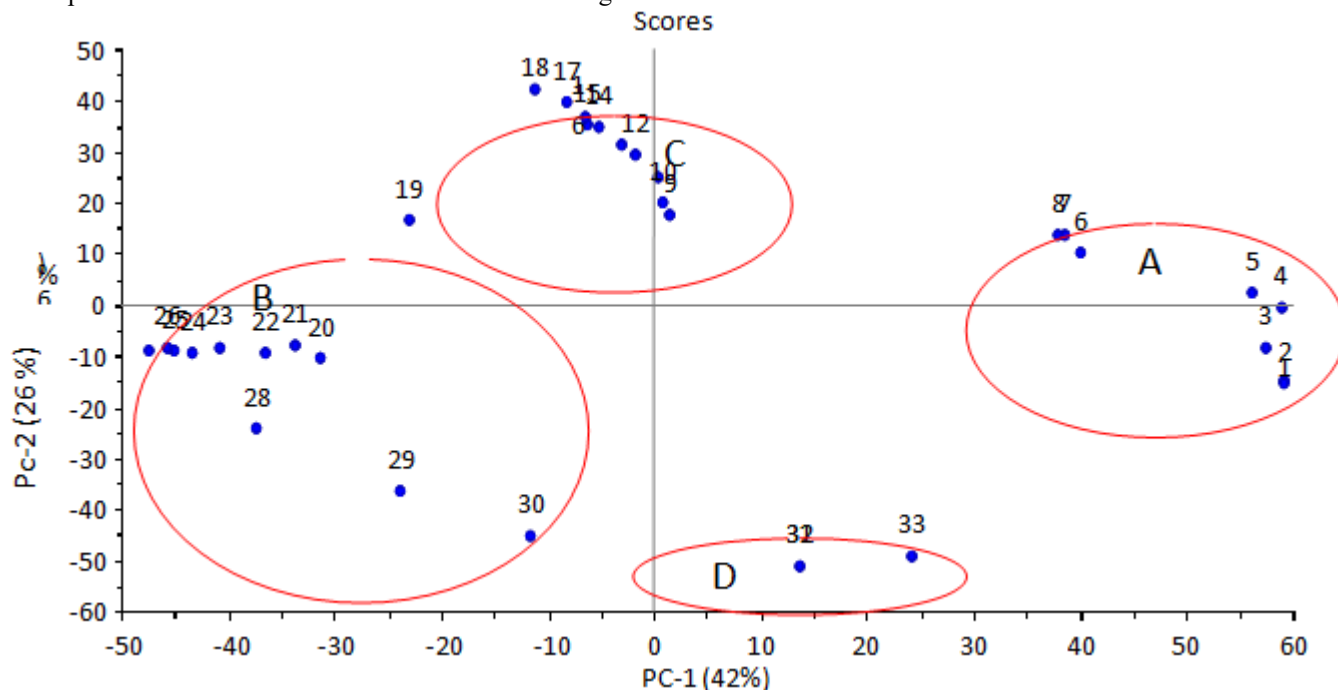


Figure 1: PCA Scores plot for the compounds extracted

3.2 Explained Loadings

FIG. 2 shows the PC 1 and PC 2 accounted 70 % of the total variation with PC 1 having 42 % and PC 2 28 %. On PC 1, (A) CH₁₀ + H₁₀, CH₁₄ + H₆ and (D) hexane were positively loaded. CH₁₀ + H₆ and CH₁₄ + H₁₀ were highly loaded and overlapped, indicating shared similarities. Also on PC 1, (B) M₁₀ + CH₁₀, M + CH + H, M₁₄ + H₆, M₁₀ + H₁₀ and methanol

were negatively loaded, with M₁₀ + H₁₀ and methanol superimposed on each other. Furthermore, the mixtures which are positive and those that are negative on PC 1 are both anti-correlated, indicating that the increase in one will lead to the decrease on the other. On PC 2, chloroform and M₆ + CH₁₄ were positively loaded and packed together while (C) M₁₄ + CH₆ and M₆ + H₁₄ were negatively loaded.

M = 100 % Methanol, CH = 100 % Chloroform, H = 100 % Hexane, M 14 + Ch 6 = 70 % Methanol + 30 % chloroform, M 6+ CH 14 = 30 % Methanol+70 % chloroform, M 10 + CH 10=50 % methanol+50 % chloroform, M 14 + H 6 = 70% Methanol +30 % Hexan, H 14+ M 6=70 % Hexan+30 % methanol , H10 + M 10 =50 % Hexan +50 % methanol, CH 14+ H 6= 70 % Chloroform +30 % Hexan, CH6 + H14 = 30 % Chloroform +70 % Hexan, CH10+ H10 = 50 % Chloroform + 50 % Hexane, M+CH+H = 33.3 % Methanol + 33.3 % Chloroform +33.3 % Hexan

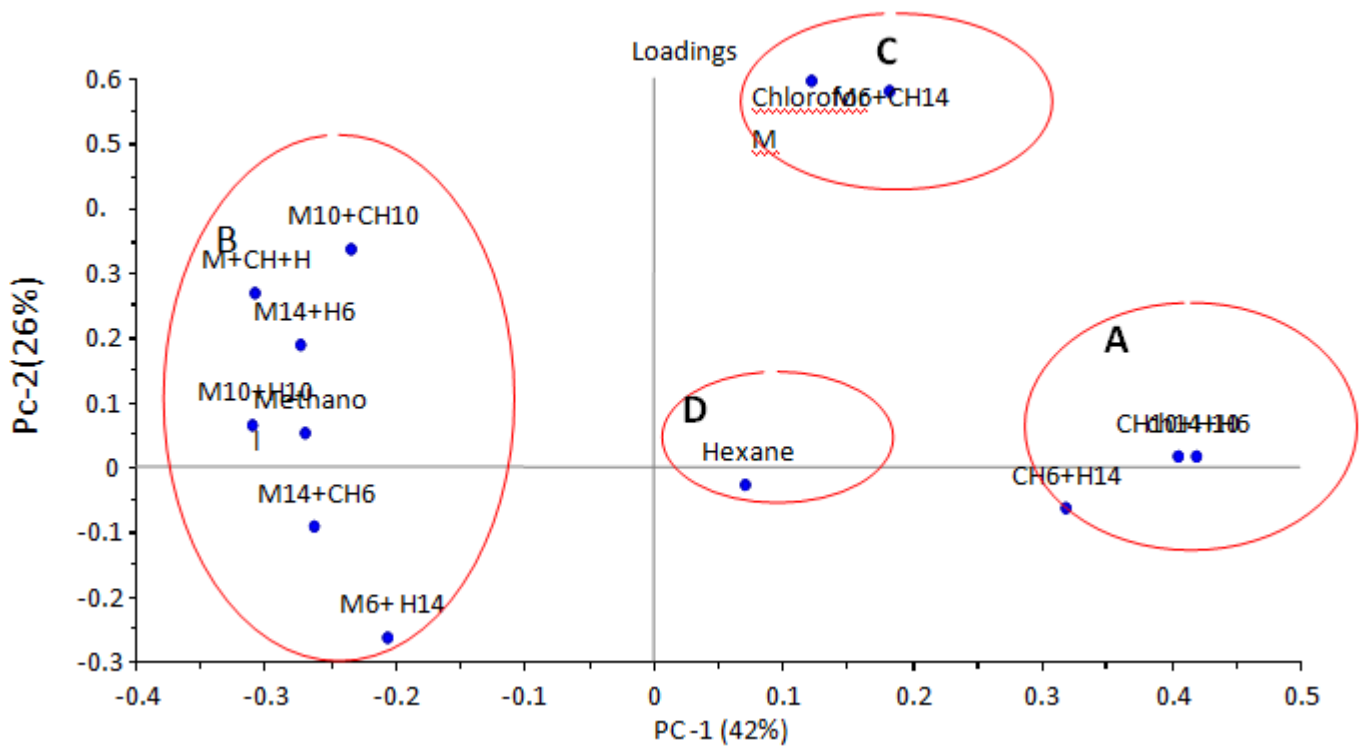


Figure 2: Loading plot of all variables on PC1 and PC2

3.3 Explained variance

FIG. 3 shows the blue curve indicates the calibrated variance while the red indicates the validated variance. From the

validated curve, it shows that PC 1 and PC 2 are enough to explain most variance as a sharp slope is encountered after the PC 2, after which a decline is visible.

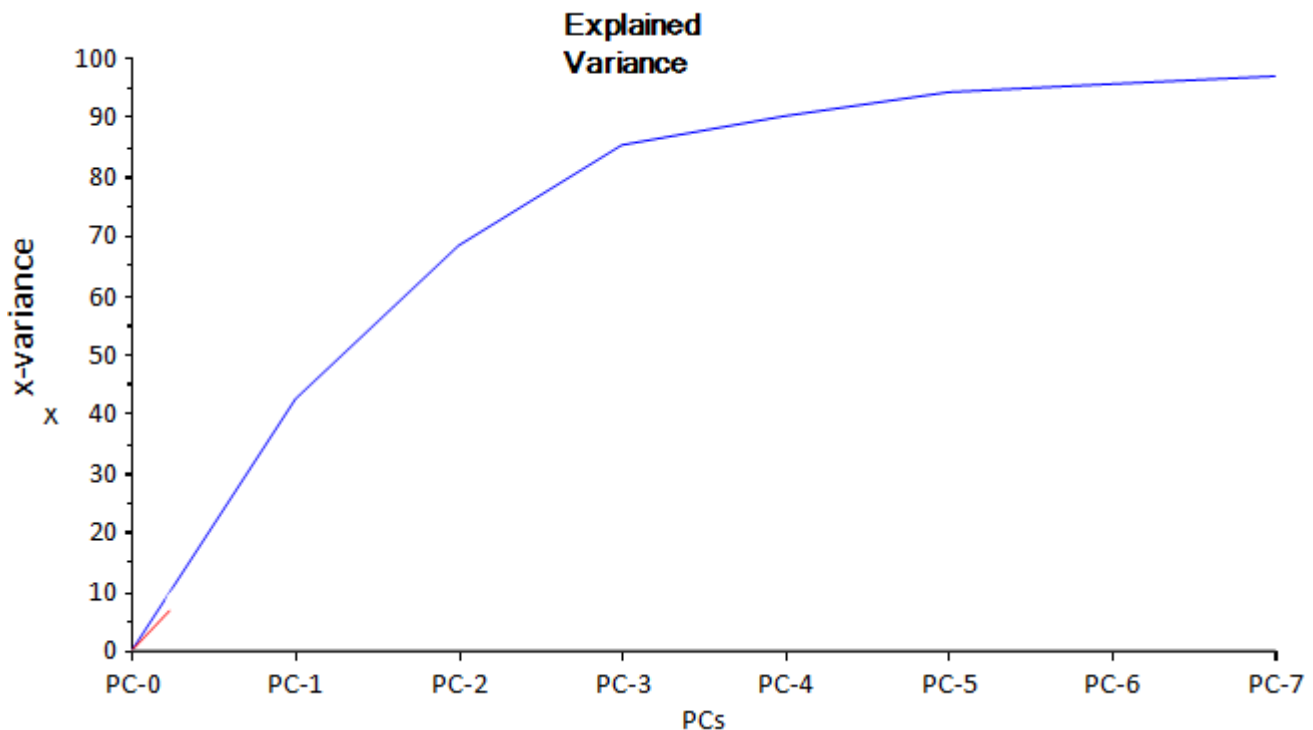


Figure 3: Cumulative variance.

3.4 Correlation Loadings

FIG. 4 shows the correlation loadings contain two ellipses, the inner and outer ellipses. The variables in the inner circle

indicate 50 % of explained variance while those in the outer ellipse indicate 100 % variance and have high contribution to PC1.

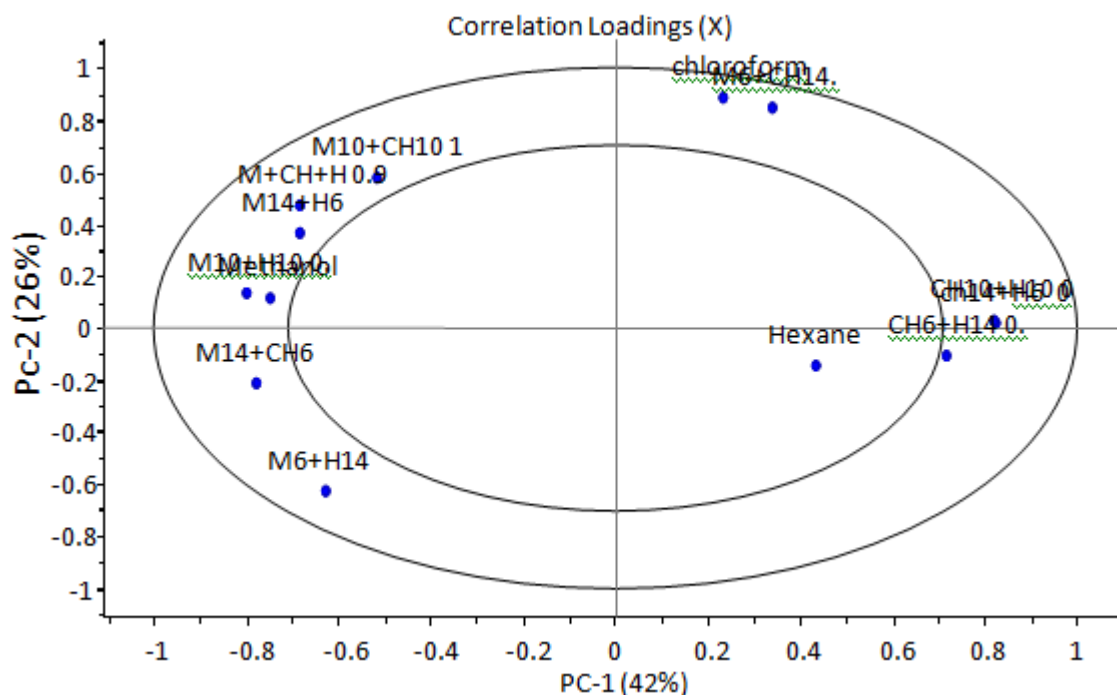


Figure 4: Correlation loadings plot of all variables along PC1 and PC2

3.5X-Loadings

FIG. 5 shows the two dimensional scatter plots of X-loadings describe the data structure in terms of variable contributions and correlations. Every variable analyzed has a loading on each PC and this reflects how much the individual variable contributes to that PC, and how well the PC takes into accounts the variation in the variable (Wise et al., 2006). The magnitude of the loadings indicates the

relative contribution of the individual variable to each PC based on the interrelationships among the variables, and the biological meaning is determined by eigenvectors (weight) and the PC scores. Variables on the component 1 vs. component 2 represent the largest variations in the data set. PC 1 is generally better correlated with the variables than PC 2; this is expected as PCs are extracted successively, each one accounting for as much of the remaining variance as possible (CAMO, 2011).

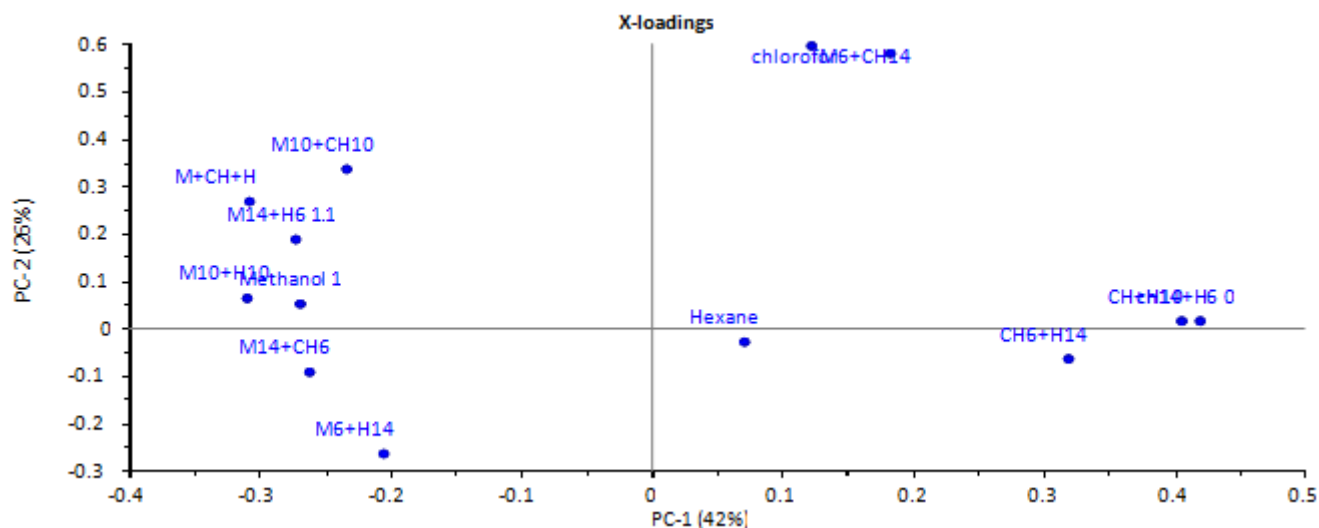


Figure 12: PCA X-loadings

4. Conclusion

Principal Components Analysis (PCA) was performed on multivariate data of GC-MS results to study the relationship between compounds extracted with polarity of solvents mixture from the scores, loadings and plots.

5. Acknowledgment

The authors wish to thank the Universiti Sains Islam Malaysia for supporting this research under NRGs sponsored by Ministry of Education, Malaysia.

References

- [1] Abdel-Zaher, A., R. Abdel & F. Elwasei. 2010. "Blockade of Nitric Oxide Overproduction and Oxidative Stress by *Nigella sativa* Oil Attenuates Morphine-Induced Tolerance and Dependence in Mice". *Neurochem Res.* Vol. 35: p. 1557–1565.
- [2] Abdi, H. 2003. "Multivariate analysis" in *Encyclopedia for Research Methods for the Social Sciences*. M. Lewis-Beck, A. Bryman, T. Futing (ed.). Thousand Oaks, CA: Sage Publications. p. 669-702.
- [3] Abdi, H. & L.J. Williams. 2010. "Principal component analysis". *Wiley Interdisciplinary Reviews: Computational Statistics*. Vol. 2. p. 433-459.
- [4] Abdullah, S. E., M. T. Said & H. A. Mohammed. 2011. "Phenolic Contents and Antioxidant Activity of Various Date Palm (*Phoenix dactylifera* L.) Fruits from Saudi Arabia". *Food and Nutrition Sciences*. Vol. 2: p. 8.
- [5] Adel, L. A. & N. B. Yassine. 2013. "Optimization of deglet-naur bate phenol extraction conditions". *World Academy of Science, Engineering and Technology*. Vol. 7: p. 691-694.
- [6] CAMO. 2011. The unscrambler® X 2009-2011. Version 10.1 (32-bit). CAMO Software AS.
- [7] CAMO. 2011. What is mmultivariate Analysis? USA: CAMO Software AS.
- [8] Jolliffe, I.T. 1986. *Principal Component Analysis*. Berlin: Springer-Verlag.
- [9] Jolliffe, I.T. 2002. *Principal Component Analysis*. New York: Springer.
- [10] Vayalil, P. 2002. "Antioxidant and antimutagenic properties of aqueous extract of date fruit (*Phoenix dactylifera* L. Arecaceae). *Journal of Agricultural and Food Chemistry*. Vol. 50: p. 610-617.
- [11] Velioglu, Y. S., G. Mazza, L. Gao & B. D. Oomah. 1998. "Antioxidant activity and total phenolics in selected fruits, vegetables, and grain products. *Journal of Agricultural and Food Chemistry*. Vol. 46 (10): p. 4113- 4117.
- [12] Wise, B. M., N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig & R. S. Kocih. 2006. *Chemometrics Tutorial for PLS-Toolbox and Solo*. USA: Eigenvector Research, Inc.