ISSN 2349-7823

International Journal of Recent Research in Life Sciences (IJRRLS)
Vol. 6, Issue 2, pp: (1-7), Month: April - June 2019, Available at: www.paperpublications.org

Authentication of honey samples from Eastern Mau forest Kenya by principal component analysis and cluster analysis of quantitative melissopalynological variables

Philip O. Onyango

Department of Botany, School of Physical and Biological Sciences, Maseno University, Kisumu City, Kenya

Abstract: **The aim of this work was to test the suitability of principal component analysis and cluster analysis for distinguishing 27 unifloral and multifloral honeys from Eastern Mau using melissopalynological parameters: Pollen density, pollen types, season and site of collection, honey type and Shannon Weaver diversity index of the pollen types. The extraction of six variables were well represented in the common factor space. Six principal components have their communalities explained. Principal component 1 explains upto 44.07% of the variance and has the highest Eigen value 2.64. Component 1 and 2 had Eigenvalues above 1.00 and both cumulatively accounted for 67.67% of the total variance. Pollen density, Season, pollen types, honey types are more correlated to the principal component 1, while sites and pollen Shannon Weaver diversity index were more correlated to principal component 2. Four of the six clusters created in cluster analysis had mixture of honey samples from various regions and Botanical origin. All samples collected from Mariashoni in April formed an isolated single cluster, all the samples (MA-S1-AP, MA-S2-AP,MA-S3-AP) being unifloral honey. Honey samples KA-S2-DE, KA-S3-DE, NE-S1-DE, NE-S2-DE , all collected in December formed a cluster of honeys from adjacent mesoregions of Nessuit and Kapkembu.**

Keywords: **Cluster analysis, Eastern Mau, Honey, Principal component analysis, Honey, Pollen.**

I. INTRODUCTION

Honey is the most improtant primary product of beekeeping. Honey adulteration is a reprehensible practise that involve incorporation of sugar syrups into the genuine product. Through apiculture practice, honey may be adulterated by use of sugars in feeding of bees. Sale of honey fraudulently labelled is the third way of honey adulteration. Differentiation between floral honey and honeydew honey is a response to consumer demands. Extensive analysis to prove the honey authenticity is key. Quality control protocols in combination with multivariate analyis , is capable of classifying honey according to their geographical origin , botanical origin , adulteration and chemical constituents (Kristina et al., 2008).

To understand and improve quality of honey, there is need for new tools capable of measuring a range of properties as well as large and complex data analysis techniques. Multivariate analysis involves the use of mathematical and statistical methods to look at the honey sample in its entirety and not just in its single component to untangle all the complicated interactions among individual constituents and understand their combined effects on the whole object. Currently, principal component analysis (PCA) or cluster analysis (CA), provides opportunities for analysis and classification of a food sample wholly (Corbella and Cozzolino, 2006). Therefore, it is not always necessary to determine all constituents of the sample to know whether it falls within a defined range or group. Instead, trends or correlation among individual quality characteristics can be used (Anklam, 1998). PCA serves for data reduction while losing only a small amount of information and presentation of data in just two dimensions corresponding to individual variable and type of honey (Kristina et al., 2008). Cluster analysis studies similarity of multidimensional objects that have natural tendency to group and consequently groups the same objects into clusters. This approach has been used in separating honey samples into a cluster of honeys like clusters of dandelion honeys (Beebe, et al., 1998)

ISSN 2349-7823

International Journal of Recent Research in Life Sciences (IJRRLS)
Vol. 6, Issue 2, pp: (1-7), Month: April - June 2019, Available at: www.paperpublications.org

Occassionally authors disgree on the characterisation of plants as nectar or pollen producers , the use of cluster analysis and discriminat analysis have minimised possiblity of subjective interpretations (Boudilio et al., 2002).The aim of this work was to test the suitability of PCA and CA for distinguishing unifloral and multifloral honeys of diffrent botanical and regional origin using melissopalynological parameters: Pollen density, pollen types, season and site of collection, honey type and Shannon Weaver diversity index of the pollen types.

II. METHODS

A. Study site

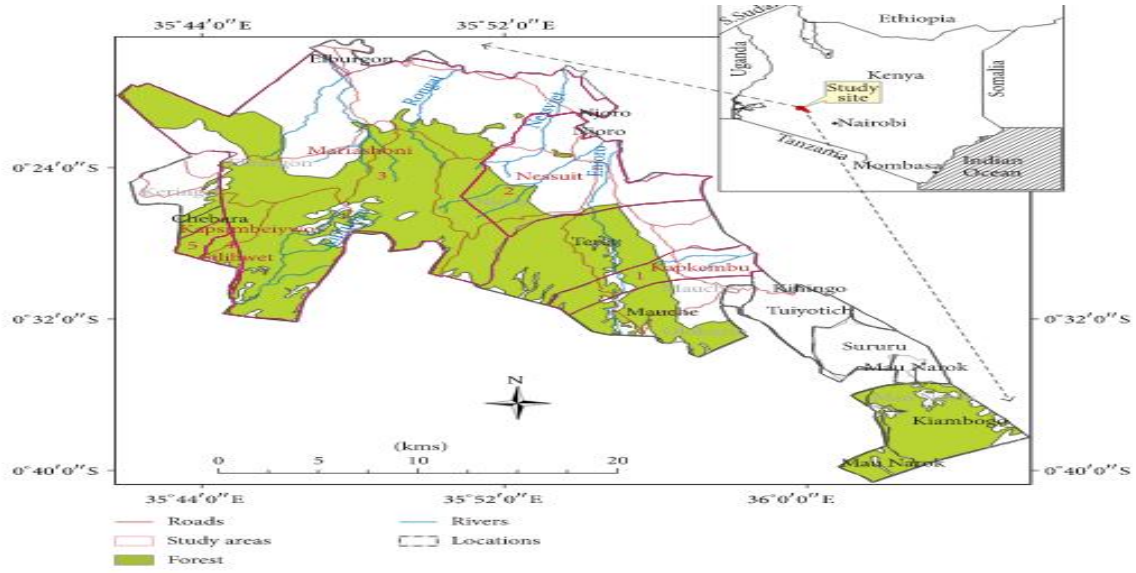

Figure 1. Map of study area in Eastern Mau forest Kenya.

Eastern Mau is one of the largest blocks in the Mau forest complex measuring about 65,921 ha. It lies on the rain shadow and receives bimodal rainfall varying from 1,000 mm-1500mm annually. The eastern Mau forest block is made up of ridges, summits, hilltops, steep slopes. The area is made up of class V vegetation of about 50-75% plant density. There is up to 40% dependence on honey production in Eastern Mau.

B. Sample collection and preparation

Three honey samples were collected from each of strata (Mariashoni, Kapkembu, and Nessuit) at the end of April, 2016; August 2016; December, 2016) from the hives of Bee keeping Ogieks of the Eastern Mau forest region. Only the honey strained by fine sieves or cheese-cloth were collected from the beekeepers, placed in sealed food grade screw cup bottles, and transported to the Maseno University Botany laboratory in cooler boxes. Samples from 3 beekeepers (three replicates) per population substratum were collected. Samples for further analysis were refrigerated at 3 ± 2^0C and stored in dark with screw cup bottles. *Laboratory sample* consisted of 100-200 g of honey. The laboratory sample was transformed into the test sample by thorough stirring. Granulated hard samples were softened by slight warming. Dirty samples were liquefied at 40°C and strained through cheese-cloth. Slides were prepared from 10.0 g of honey weighed and dissolved in 20 ml of hot distilled water at 39°C. The solution was then centrifuged for 10 min at 2500 r/min and decanted. The honey sugars was completely removed by dispersing again with 10 ml of distilled water. The solution was then poured into a centrifuge tube, and centrifuged for 5 min. The entire sediment was put on a slide using Pasteur pipettes and spread out over an area about 20 X 20 mm, using a thin glass. After drying by slight heating at 40°C, the sediment was mounted with glycerine gelatine, liquefied by heating in a water-bath at 40°C. The sediment constituents remaining in the tube were stirred again with a drop of distilled water, pipetted again, and the used pipette rejected to eliminate the contamination of pollens from other honeys. If the honey sample is poor in pollen 20 g was used. For samples rich in sediment, the residuum was spread under two cover glasses. The microscopical identification was based on the identification and counting of pollen grains and other particles in honey. Identification was done by reference to the literature and to comparative refrence slide preparations. A complete analysis involving the identification of all pollen grains and other microscopic constituents in

ISSN 2349-7823

International Journal of Recent Research in Life Sciences (IJRRLS)
Vol. 6, Issue 2, pp: (1-7), Month: April - June 2019, Available at: www.paperpublications.org

the sediment was carried out. Three degrees of accuracy including estimates, determination of frequency classes and counts expressed in percentage was used on two slides (prepared as above) independently, from the same honey. Any pollen of wind-pollinated or nectar-lacking plants were noted separately. Abortive and misshapen pollen grains were counted as far as they could be identified. Spores and honey dew elements were noted separately.

C. Botanical origin and pollen density through microscopical examination

The extent to which a given honey sample is derived from different plant sources was deduced from the frequencies of the pollens and honeydew elements in it. Honey was considered to have been produced mainly from one plant (unifloral honey) if the pollen of that plant is predominant. Honey was regarded honey dew only if ratio of HDE/P was equal to or greater than 3. Pollen reference slides were prepared.

500 pollen grains were counted for the determination of relative frequencies .Magnification of 400 to 1000X was used for identifying the various elements in the sediment. The Identification and counting of pollen grains is done in groups of 100, following 5 parallel equidistant lines uniformly distributed from one edge of the cover slip (22X22mm) to the other, until 500 grains are counted. Abortive, irregular or broken pollen grains are counted if they can be identified. Non-identifiable, non-identified grains, honeydew elements (HDE), i.e. fungal spores, hyphae and microscopic algae were noted separately. Pollen grain frequencies were estimated according to the following terms: "Very frequent" for grains constituting more than 45% of the total; "Frequent" for grains constituting 16-45% of the total; "Rare" for grains constituting 3-15% of the total and "Sporadic", for grains constituting less than 3%. The frequency classes were described as follows: "Predominant pollen" (more than 45% of the pollen grains counted); "Secondary pollen" (16-45%); "Important minor pollen" (3-15%) and “Minor pollen", (less than 3%). Honey with predominat pollen type was classified as Monofloral honey. For pollen grains that were not identified as far as the genus or species, a note was added after the scientific name, to indicate that the term was used in a wider meaning. The proportion of the HDE to the total frequency of pollen grains from nectar plants were described as follows: Practically none (0.00-0.09); Few (0.10-1.49), Medium quantity (1.50-2.99), numerous (3.00-4.49), Very numerous (>4.50). Estimates of the frequency of pollen grains of anemophilous and other nectar less plants were expressed as follows: "sporadic" (less than 3% of the total); "rare" (3-15%) ;"frequent" (16-45%); "Very frequent" more than 45%. The identification of pollen types was based on shape, morphological characteristics and size of the pollen grains . Pollen types identified by using reference pollen slides. Acetolysed anther material according to Erdtman (1960) , from Eastern Mau Apiflora observed in intial studies were used to develop reference slides. Fresh material from Musaceae and Lauraceae were only warmed with 2-5% KOH solution for 2 minutes instead of acetolysis and their slides sealed with paraffin wax.

D. Data analysis

Results of melissopalynological analysis of honeys were processed by PCA and CA using SPSS Version 18. For cluster analysis a data matrix of 27x6 was prepared from 27 objects (honey samples) and six variables (Sites, seasons, pollen types, pollen density, honey types and Shannon weaver diversity index). The pollen density was reduced in to Log.

III. RESULTS

Table 1. Melissopalynological characteristics of the 27 honey samples collected from Eastern Mau, Kenya

Sample	Site	Season	Pollen Types	Log(Pollen Density)	Honey Type	Shannon Weaver diversty index
KA-S1-AP	Kapkembu	April	13	4.961	M	3.036
KA-S2-AP	Kapkembu	April	10	5.004	M	2.708
KA-S3-AP	Kapkembu	April	12	5.009	M	2.936
MA-S1-AP	Marioshoni	April	15	5.192	U	2.832
MA-S2-AP	Marioshoni	April	15	5.205	U	1.925
MA-S3-AP	Marioshoni	April	15	5.194	U	2.276
NE-S1-AP	Nessuit	April	11	5.032	U	2.524
NE-S2-AP	Nessuit	April	12	5.008	M	2.370
NE-S3-AP	Nessuit	April	12	5.063	M	2.191

KA-S1-AU	Kapkembu	August	12	4.892	U	2.191
KA-S2-AU	Kapkembu	August	13	4.890	U	2.205
KA-S3-AU	Kapkembu	August	12	4.921	M	2.146
MA-S1-AU	Marioshoni	August	12	5.121	U	2.736
MA-S2-AU	Marioshoni	August	11	5.119	M	1.215
MA-S3-AU	Marioshoni	August	15	5.130	M	1.839
NE-S1-AU	Nessuit	August	14	4.953	M	1.936
NE-S2-AU	Nessuit	August	12	4.919	M	2.832
NE-S3-AU	Nessuit	August	13	4.990	M	2.020
KA-S1-DE	Kapkembu	December	11	4.764	U	3.332
KA-S2-DE	Kapkembu	December	10	4.680	M	2.168
KA-S3-DE	Kapkembu	December	10	4.692	M	2.168
MA-S1-DE	Marioshoni	December	9	4.898	M	2.639
MA-S2-DE	Marioshoni	December	10	4.888	M	1.677
MA-S3-DE	Marioshoni	December	8	4.893	M	2.431
NE-S1-DE	Nessuit	December	10	4.746	M	2.164
NE-S2-DE	Nessuit	December	8	4.820	M	1.475
NE-S3-DE	Nessuit	December	11	4.886	M	2.736

KEY: M=Multifloral, U=Unifloral

27 honey samples were collected. Seven honey samples were unifloral while 19 were multifloral. The number of pollen types in honey ranged from 8 to 15. Highest and lowest pollen density were observed respectively from Mariashoni and Kapkembu mesoregion.

Table 2. Communalities of principal component analysis

	Initial	**Extraction**
Sites	1.00	.61
Season	1.00	.76
Pollen types	1.00	.74
Pollen density(Log)	1.00	.86
Honey type	1.00	.53
Shannon Weaver	1.00	.56

Communalities as proportions of each variables variance that could be explained by the principal components (Latent continua) between 53% (Honey type) and 86% (Pollen density) could be explained by the principal components. The extraction of six variable (Sites, seasons, pollen types, pollen density, honey types and Shannon weaver diversity index were well represented in the common factor space.

Table 3. Total variance explained

	Initial Eigenvalues			**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.64	44.07	44.07	2.64	44.07	44.07	2.62	43.58	43.58
2	1.42	23.60	67.67	1.42	23.60	67.67	1.45	24.10	67.67
3	.76	12.74	80.41						
4	.66	10.92	91.33						
5	.34	5.70	97.04						
6	.18	2.97	100						

The number of principal components , 6 are as many as the number of variables whose communalities have been explained. Principal component 1explains upto 44.07% of the variance and has the highest Eigen value 2.64. Principal Component 2 accounts for 23.6% of the total variance and an Eigenvalue of 1.416. Component 1 and 2 had Eigenvalues above 1 and both cumulatively accounted for 67.67% of the total variance. Successive principal components accounted for less and less variances.

ISSN 2349-7823

International Journal of Recent Research in Life Sciences (IJRRLS)
Vol. 6, Issue 2, pp: (1-7), Month: April - June 2019, Available at: www.paperpublications.org

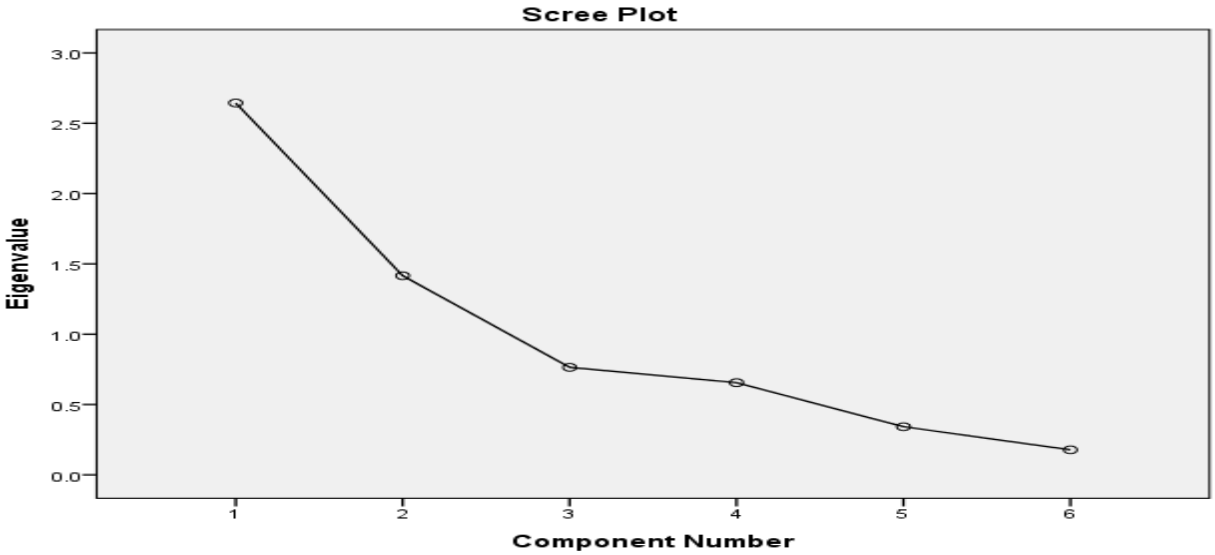

Figure 1. Scree plot of components against Eigen values.

The scree plot has graphed the eigenvalue against the component number. The first two components have the highest eigenvalues with each successive component explaining smaller and smaller amounts of total variance. The principal components analysis has redistributed the values of correlation matrix using eigenvalue decomposition to redistribute the variances to first component extracted.

Table 4. Component matrix of principal component analysis

	Component 1	Component 2
Pollen density(Log)	.88	-.31
Season	-.87	
Pollen types	.86	
Honey type	.59	.42
Sites		-.78
Shannon Weaver		.73

Table above shows the component loadings, the correlations between the six variables the principal components. Pollendensity, Season, pollen types, honey types are more correlated to the principal component 1, while sites and Shannon Weaver diversity index were more correlated to principal component 2. Two components have been extracted.

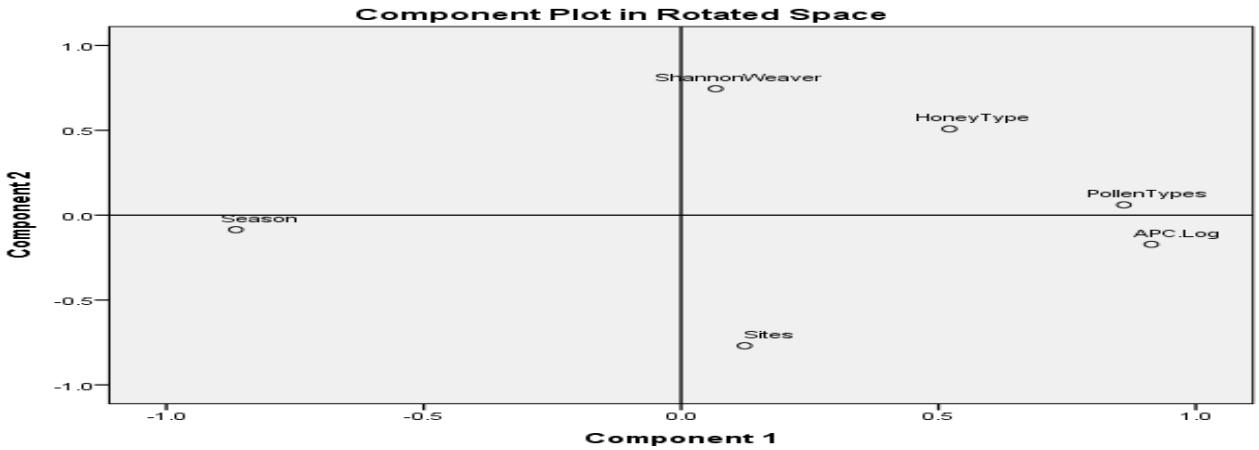

Figure 2. Saturation of variables in axes of first two principal components.

PCA disclosed two principle components

ISSN 2349-7823

International Journal of Recent Research in Life Sciences (IJRRLS)
Vol. 6, Issue 2, pp: (1-7), Month: April - June 2019, Available at: www.paperpublications.org

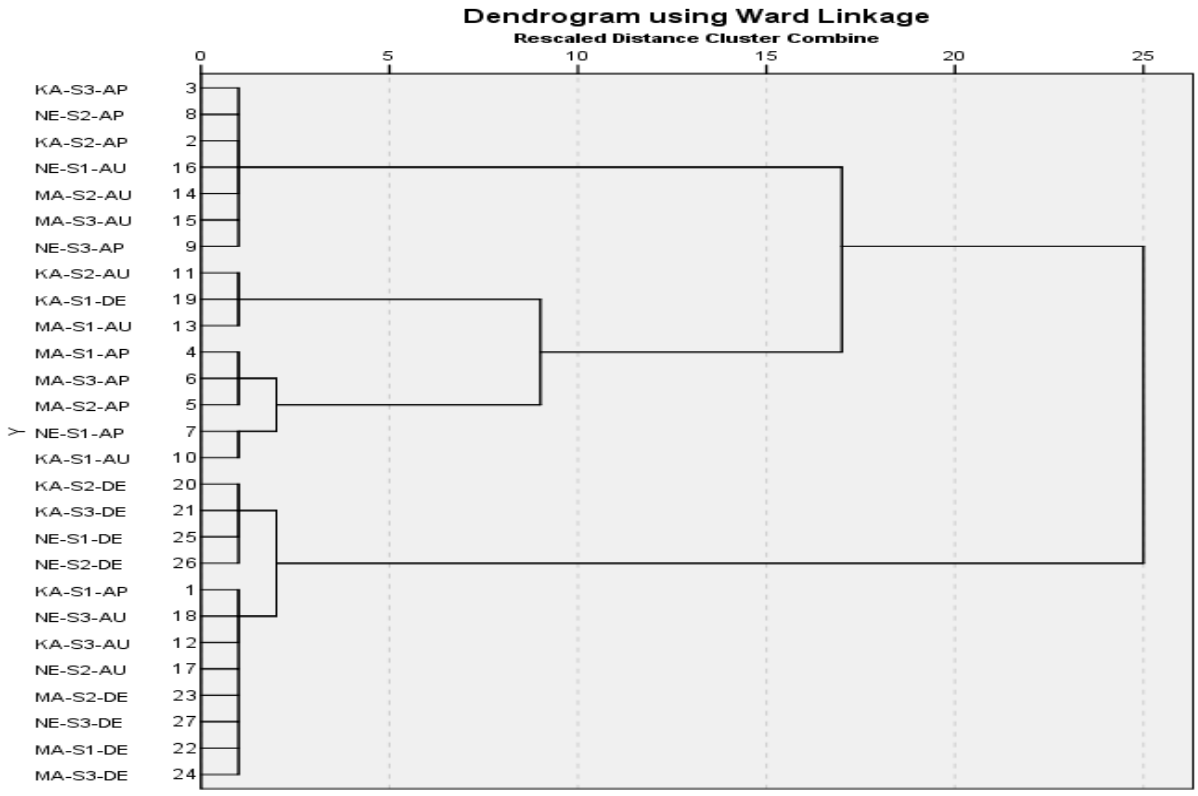

Figure 3. Dendroigram from cluster analysis of 27 honey samples from Eastern Mau forest.

6 clusters were created from the cluster analysis algorithm. The clusters had mixture of honey samples from various regions except for all samples collected from Mariashoni in April that formed an isolated single cluster, with all samples (MA-S1-AP, MA-S2-AP,MA-S3-AP) unifloral honey. Honey samples KA-S2-DE, KA-S3-DE, NE-S1-DE, NE-S2-DE , all collected in December formed a cluster. The cluusters did not exclusively bring together honey samples of a given type, of given plant origin, or season of collection.

IV. DISCUSSION

The selection of a small number of key variables in pricipal component analysis increases the reliability of mathematical classification , eliminates features with minor information and allows a visual examination of the data set by two dimensional plot of the key features. Principal component analysis was applied to autoscaled data in studies by Kuchla et al., (2015), data analysis showed 73.51% of the total variance was explained by the first two components. The first principal component (PC1 with 43.29% of total variance) was strongly influenced by 3 variables, while for the second principal component (PC2 with 30.22% of the total variance) 2 variables were more important. Similar results were observed in this study where 44.07% of total variance was explained by the first principal component, with 67.67% being explained by the first 2 principal components.

Cluster analysis have given a distinct discrimination of honeys with a wider geographical and botanical origin (Baudilio et al., 2002). Although proven to be very usefull, honeys coming from restricted area makes discrimination on the basis of geographical and botanical origin limited (Boudilio et al., 2002). The application of cluster analysis classification is further justified especially when presence of one pollen type has been used as a discriminator amongst honey samples and to relate a certain honey type to certain geomorphological zone. This, however doesnt take note of accidental contamination. As such reliance of rare occurence can not provide basis for sound classification (Boudilio et al., 2002).

It would be expected that honeys from one season group in to specific clusters . Four clusters did not conform to the seasonality of collection. The two clusters that conformed were clusters with samples from April and December. *Acacia* and *Eucalyptus* unifloral honey samples from Marioshoni clustered together. Other clusters combined both Unifloral and multifloral honey samples. Similar results have been observed in previous studies (Samir et al., 2007, Ana and Francisco,

2014). Three clusters did not conform to the seasonality of collection (Samir et al., 2007). This could be attributed to the long flowering period of part of the taxa and the varied frequency of occurence of pollen in honey from different areas during the same season (Samir et al., 2007). Classes obtained by cluster analysis have been reported to be composed of both unifloral honeys, multifloral honeys, and mixture of both monofloral and multifloral honeys (Boudilio et al., 2002), and with varying similarity magitudes (Ana and Francisco, 2014). Similar trends were observed in this study.

Fewer clusters than the sites from which the samples were collected were observed in this study. Similar trends were reported by Samir et al. (2007), attributed to border effect. Honey Samples collected closer to each other geographically (samples 20, 21, 25, and 26) from adjacent Nessuit and Kapkembu mesoregions, were classified within the same cluster due to similar pollen composition. A cluster analysis of the taxa using presence/absence data, also grouped honey samples of five geomorphological zones into four classes based on their pollen composition (Poderoso et al., 2012). In studies by Samir et al. (2007) in West india, a K means clustering of honey samples showed that though samples are collected in 3 different seasons , from different areas, most of the Summer and Autumn honeys have similar floristic composition (Boudilio et al., 2002). A similar trend was observed in this study in which April honey samples and December honey samples were clustered together separately. Cluster analysis by Boudilio et al., (2002) generated a dendogram of seven classes. Infrequent pollen contributed to the determination of the class in the honey samples. Presence of pollen types associated with endemic species show the use of native vegetation by honey bees corroborating the importance of these pollen types as geographic markers for honey samples from a region (Ana and Francisco, 2014).

V. CONCLUSION

The use of cluster analyis have proven useful in distinguishing honeys unifloral from multifloral honey from different seasons and sites of collection. The principal component analysis have demonstrated that pollen type, pollen density, honey type and season during which the honey is collected contributes 44.07% of the total variation and thus important in distinguishing the honeyfrom Eastern Mau forest. All Mariashoni unifloral honey samples collected are distinnguishable by cluster analysis from the rest of honey collected from other regions of Eastern Mau.

REFERENCES

[1] Ana, P. C. S. and Francisco de Assis Ribeiro dos Santos (2014). Pollen diversity in honey from Sergipe, Brazil. 53:2, 159-170. DOI:10.1080/00173134.2014.896941.

[2] Anklam, E.: A review of the analytical methods to determine the geographical and botanical origin of honey. Food Chemistry, 63, 1998, pp. 549–562.

[3] Boudilio, H., Rosa, M.V., Roberto ,S.M., Valentin, P. (2002). Characterisation of honeys by melissopalynology and statistical analysis. Canadian Journal of plant science. 82:pp75-82.DOI:10.4141/P00-187.

[4] Beebe, K. R. – Pell, R. J. – Sholtz, M. B. S.Chemometrics: a practical guide. New York : John Wiley & Son, Inc., 1998. 348 pp. ISBN 978-0-471-12451-1.

[5] Corbella, E. – Cozzolino, D.: Classification of the floral origin of Uruguayan honeys by chemical and physical characteristics combined with chemometrics. Lebensmittel Wissenschaft und Technologie – Food Science and technology, 39, 2006, pp. 534–539.

[6] Erdtman, G. (1960). Investigation of honey pollen. The acetolysis method. Svensk Botanisk Tidskrift, (54), 561-564.

[7] Kristina K., Jolana K., Zlatica K., Katarina B., (2008) .Authentification of honey by multivariate analysis of its physico-chemical parameters. Journal of food and Nutrition Research.Vol 47 (4)Pp 170-180.

[8] Kuchla , M. , Araujo, M.D.M., Soares, A.F., Quinaia, S.P., Felsner, M.L. (2015) Classification of wild honeys of different Mesoregions from Parana State, Brazil, by Principal Component analysis. Revista Virtual de Quimica. 7(6) Pp 2301-2313. ISSN 1984-6835.DOI 10.5935/1984-6835.20150137.

[9] Poderoso, J. C. M. and Maria, E. C., Carlos da Paz, L. (2012). Botanical preferences of Africanised bees Apis mellifera on the Coast and in the Atlantic forest of Sergipe, Brazil.

[10] Samir, K.M., Sudha, G., A.P. Das, Subir, B. (2007). The beekeeping potential of Sub Himalayan West Bengal, India: A palynological assessment of honey.Journal of Apicultural research and Bee world. 46(3): 164-177.