Nereis. Revista Iberoamericana Interdisciplinar de Métodos, Modelización y Simulación	8	23-33	Universidad Católica de Valencia San Vicente Mártir	Valencia (España)	ISSN 1888-8550

Classification of Food Spices by Proximate Content: Principal Component, Cluster, Meta-Analyses

Fecha de recepción y aceptación: 4 de octubre de 2015 y 6 de noviembre de 2015

Francisco Torrens Zaragozá[1*]

[1] Institut Universitari de Ciència Molecular.
[*] Correspondencia: Universitat de València. Edifici d'Instituts de Paterna. P. O. Box 22085. 46071 Valencia. España. E-mail: francisco.torrens@uv.es

ABSTRACT

Proximate composition of six food spices commonly used in South-East Nigeria are classified by principal component analysis (PCAs) of constituents and spices cluster analysis (CAs). Samples are grouped into two classes. Compositional PCA and spice CA permit classificating them and group the similar ones. The first PCA axis explains 61% of the variance; first two, 93%; first three, 99; *etc*. Different behaviour of species depends on *ash*, *fibre*, *fat*, *moisture*, *etc*. Macronutrients (*protein*, *carbohydrate*, *fat*) contents are adequate. *Carbohydrate* amounts are high. *Fat* quantities are moderate. *Fat* is closer to *protein* than to *carbohydrate*.

KEYWORDS: *chemotaxonomic analysis, meta-analysis, principal component analysis, cluster analysis, distribution, proximate, phytochemical, spice, nutrition, ethnomedicinal.*

RESUMEN

Se clasifica la composición de los constituyentes principales de seis especias alimenticias comúnmente usadas en el sudeste de Nigeria por análisis de componentes principales (ACPs) de los constituyentes y análisis de agregados (AAs) de especias. Las muestras se agrupan en dos clases. El ACP de composición y AA de especias permiten clasificarlas y agruparlas según se asemejen. El primer eje ACP explica el 61% de la varianza, los dos primeros, el 93%, los tres primeros, el 99%, *etc*. El diferente comportamiento de las especias depende de la *ceniza*, *fibra*, *grasa*, *humedad*, *etc*. Los contenidos de macronutrientes (*proteína*, *carbohidrato*, *grasa*) son adecuados. Las cantidades de *carbohidrato* son altas. Las cantidades de *grasa* son moderadas. La *grasa* está más cerca de la *proteína* que del *carbohidrato*.

PALABRAS CLAVE: *análisis quimiotaxonómico, metaanálisis, análisis de componentes principales, análisis de agregados, distribución, constituyente principal, fitoquímico, especia, nutrición, etnomedicinal.*

INTRODUCTION

Plants synthesize chemical compounds sorted into primary and secondary metabolites [1]. The latter are responsible for therapeutic potentialities [2]. Cravings were a factor in progress, and did much to change history and geography [3]. Spices play a role in many industries (perfume, soaps, incense, as dyes) [4]. They are therapeutically useful in several ailments (convulsion, leprosy, stomach ache, inflammation, rheumatoid pains, cough, loss of appetite, gastrointestinal disorder) [5,6]. They are used for preparing soups for mothers to prevent post-partum contraction and aid lactation [7,8]. Proximate and nutrient analysis of medicinal plants, edible fruits and vegetables play a role in assessing nutritional significance [9]. Olubunmi [10] reported proximate (*cf.* Table 1) and phytochemical content of six commonly used food spices in South-East Nigeria [11]. The proximate analysis shows that *Monodora myristica* had the highest *protein* and *fat* content. *Chrysobalanus icaco* had the least *protein* content and *Tetrapleura tetraptera* had the least *fat*. *Tetrapleura tetraptera* had the highest *ash* content and the highest *crude fibre* content. *Chrysobalanus icaco* had the highest *carbohydrate* content. *Afromomum danielli* had the least *ash* content while *M. myristica* had the least *crude fibre* content. *Monodora myristica* had the least *carbohydrate* content. *Afromomum danielli* had the highest *moisture* content with *T. tetraptera* having the least. The high *carbohydrate* content in all the spices ranks them as *carbohydrate*-rich food sources. The moderate *fat* content indicates that the spices are not sources of lipid accumulation, which causes arteriosclerosis and ageing. Macronutrients (*carbohydrate*, *protein*, *fat*) are needed by the body in large amounts for growth, metabolism and body functions. Regular use of plant foods rich in *protein* makes a valuable addition to a diet. People need *protein* for growth, tissue repair, immune function, making essential hormones and enzymes, energy when *carbohydrate* is not available, preserving lean muscle mass, *etc*. *Fats* insulate and protect body organs. They are essential for normal growth and development, energy, absorbing and transporting vitamins, maintaining cell membranes, providing taste, consistency and stability to foods.

Table 1. Proximate composition (% dry weight). Proximates: i_1, ash; i_2, protein; i_3, crude fibre; i_4, carbohydrate; i_5, fat; i_6, moisture

Sample	Ash	Protein	Crude fibre	Carbohydrate	Fat	Moisture
1. *Tetrapleura tetraptera*	9.42	7.80	18.56	54.66	4.45	5.11
2. *Xylopia aethopica*	6.97	13.89	9.28	52.94	7.81	9.59
3. *Monodora myristica*	6.49	21.69	7.38	41.02	13.24	9.72
4. *Syzygium aromaticum*	5.94	11.65	9.06	54.26	9.21	9.88
5. *Chrysobalanus icaco*	5.89	6.78	11.24	61.36	4.79	9.94
6. *Afromomum danielli*	3.35	10.94	13.00	53.48	7.16	11.63

Earlier publications in *Nereis* classified yams [12] and fruits [13] by principal component (PCA), cluster (CA) and meta-analyses. The main aim of the present report is to develop code learning potentialities and, since spices proximates are more naturally described *via* varying size-structured representation, find general approaches to structured information processing. In view of the spices' nutritional benefits the objective was to categorize them with PCA and CA, which differentiated proximates. The next section shows the method. Following that, two sections illustrate and discuss the results. Finally, the last section summarizes my conclusions.

COMPUTATIONAL METHOD

Principal components' analysis (PCA) is a dimension reduction technique [14–19]. From original variables X_j, PCA builds orthogonal variables $\tilde{P}_j$, linear combinations of mean-centred ones $\tilde{X}_j = X_j - \overline{X}_j$ corresponding to eigenvectors of sample co-variance matrix $S = 1/(n-1)\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})$. For every loading vector $\tilde{P}_j$, matching eigenvalue $\tilde{l}_j$ of S tells how much data variability is explained: $\tilde{l}_j = \mathrm{Var}(\tilde{P}_j)$. Loading vectors are sorted in decaying eigenvalues. First k PCs explain most variability. After selecting k, one projects p-dimensional data on to subspace spanned by k loading vectors and computes co-ordinates *vs.* $\tilde{P}_j$, yielding scores:

$$\tilde{\mathbf{t}}_i = \tilde{P}'(\mathbf{x}_i - \overline{\mathbf{x}}) \tag{1}$$

for every i = 1, …, n having trivially zero mean. With respect to original co-ordinate system, projected data point is computed fitting:

$$\hat{\mathbf{x}}_i = \bar{\mathbf{x}} + \tilde{P}\tilde{\mathbf{t}}_i \tag{2}$$

Loading matrix $\tilde{P}$ (pxk) contains loadings column-wise and diagonal one $\tilde{L} = \left(\tilde{l}_j\right)_j$ (kxk), eigenvalues. Loadings k explain variation:

$$\left(\sum_{j=1}^{k}\tilde{l}_j\right) \Big/ \left(\sum_{j=1}^{p}\tilde{l}_j\right) \geq 80\% \tag{3}$$

Cluster analysis (CA) encompasses different classification algorithms [20,21]. Starting point is nxp data matrix **X** containing p components measured in n samples. One assumes data were pre-processed to remove artefacts, and missing values, imputed. The CA organizes samples into small number of clusters so that samples within cluster are similar. Distances l_q between samples $x,x' \in \Re^p$ are:

$$\|x - x'\|_q = \left(\sum_{i=1}^{p}|x_i - x'_i|^q\right)^{1/q} \tag{4}$$

(*e.g.*, Euclidean l_2, Manhattan l_1 distances). Comparing samples, *Pearson's correlation coefficient* (PCC) is advantageous:

$$r(x - x') = \frac{\sum_{i=1}^{p}(x_i - \bar{x})(x'_i - \bar{x}')}{\left[\sum_{i=1}^{p}(x_i - \bar{x})^2 \sum_{i=1}^{p}(x'_i - \bar{x}')^2\right]^{1/2}} \tag{5}$$

where $\bar{x} = \left(\sum_{i=1}^{p} x_i\right)/p$ is measure mean value for sample x [22–28].

CALCULATION RESULTS

Six spices taken from Olubunmi were used as data. The PCC matrix **R** was computed between seeds; upper triangle is:

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.955 & 0.833 & 0.955 & 0.977 & 0.965 \\ & 1.000 & 0.947 & 0.998 & 0.987 & 0.988 \\ & & 1.000 & 0.937 & 0.886 & 0.911 \\ & & & 1.000 & 0.991 & 0.991 \\ & & & & 1.000 & 0.992 \\ & & & & & 1.000 \end{pmatrix}$$

Correlations between spices are high, *e.g.*, *Xylopia aethopica–Syzygium aromaticum* R_{24} = 0.998. They are illustrated in the partial correlation diagram (PCD) that could contain high ($r \geq 0.75$), medium ($0.5 \leq r < 0.75$), low ($0.25 \leq r < 0.5$) and *zero* ($r < 0.25$) partial correlations. The PCD contains 15 high partial correlations (*cf.* Fig. 1, *red*), which show similar compositions. The corresponding interpretation is the similar proximates in all six spices.

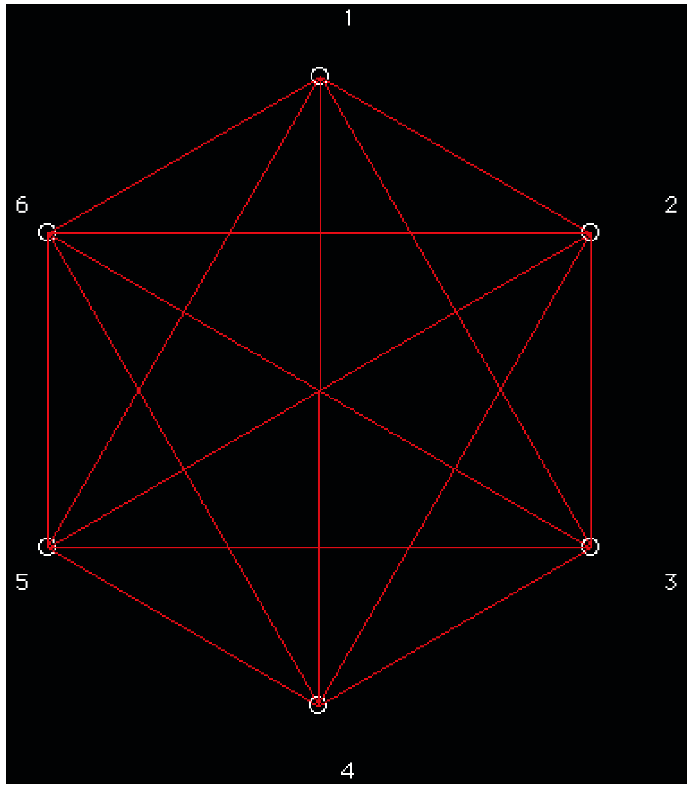

Figure 1. Partial correlation diagram showing 15 high partial correlations (*red*).

The dendrogram of spices according to proximates (*cf.* Fig. 2) shows different behaviour depending on *ash*, *crude fibre*, *fat* and *moisture*. Two clusters are clearly recognized:

(1,3)(2,4,5,6)

Class 1 is separated from spices in grouping 2: *T. tetraptera* and *M. myristica* present high *ash*, *etc*. and are clustered into class 1; *X. aethopica*, *S. aromaticum*, *C. icaco* and *A. danielli* show high *carbohydrate* and *moisture*, low *ash*, *protein*, *etc*. and are grouped into cluster 2. The spices in both classes appear highly correlated in PCD (Fig. 1). However, the results should be taken with care because the branching with a unique individual (*M. myristica*) is a possible outlier.

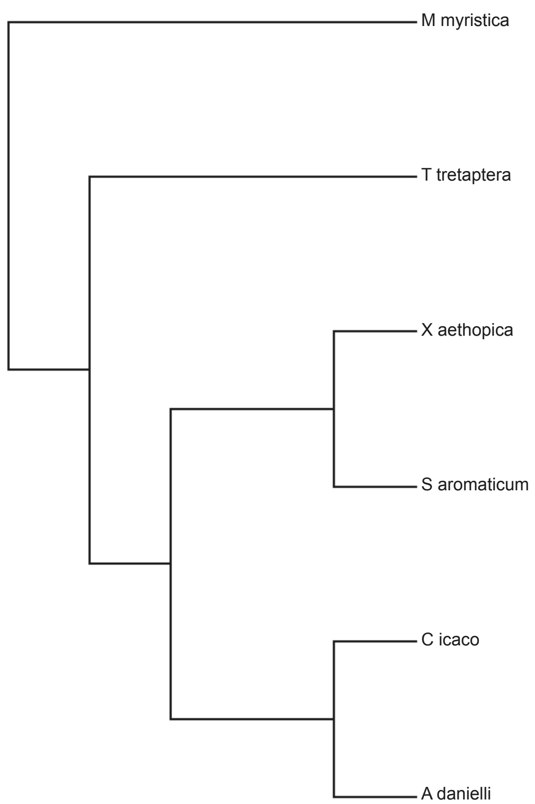

Figure 2. Dendrogram of spices according to proximate composition.

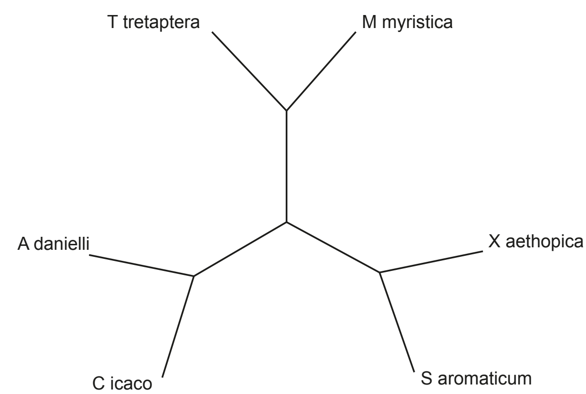

Figure 3. Radial tree of spices according to proximate composition.

The radial tree (*cf*. Fig. 3) shows different behaviour of spices depending on *ash*, *etc*. The same classes above are clearly recognized in agreement with PCD and dendrogram (Figs. 1 and 2). Again, spices with high *ash*, *etc*. are grouped into cluster 1, *etc*.

The splits graph for six spices in Table 1 (*cf*. Fig. 4) reveals conflicting relations in class 2 because of interdependences [29]. It indicates spurious relations between spices resulting from base-composition effects. It illustrates different behaviour of spices depending on *ash*, *etc*. It is in qualitative agreement with PCD and binary/radial trees (Figs. 1-3).

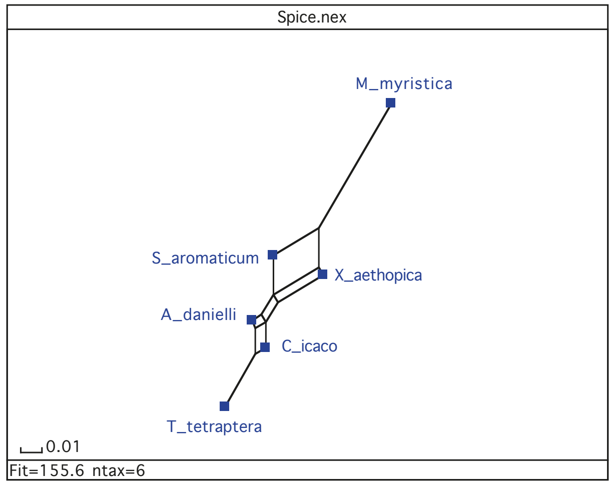

Figure 4. Splits graph of spices according to proximate composition.

Principal components (PCs) *analysis* (PCA) allows *summarizing* information contained in **X**-matrix. It decomposes **X**-matrix as a product of matrices **P** and **T**. *Loading matrix* **P** with information about variables contains a few vectors: PCs that are obtained as linear combinations of original *X*-variables. In *score matrix* **T** with information about objects, every object is described in terms of projections on PCs instead of the original variables: $\mathbf{X} = \mathbf{TP}' + \mathbf{E}$, where ' denotes transposed matrix. Information not contained in matrices remains *unexplained* X-*variance* in *residual matrix* **E**. Every PC$_i$ is a new co-ordinate expressed as linear combination of the old x_j: $\mathrm{PC}_i = \Sigma_j b_{ij} x_j$. New co-ordinates PC$_i$ are *scores* (*factors*) while coefficients b_{ij} are *loadings*. Scores are ordered according to information content *vs*. total variance among objects. *Score–score plots* show positions of compounds in new co-ordinate system, while *loading–loading plots* display location of features that represent compounds in new co-ordination. The PCs show properties: (1) they are extracted by decaying importance; (2) every PC is orthogonal to each other. A PCA was performed for spices. Importance of PCA factors $F_{1\text{–}6}$ for proximates (*cf.* Table 2) shows that first factor F_1 explains 61% variance (39% error), first two factors $F_{1/2}$, 93% variance (7% error), first three factors $F_{1\text{–}3}$, 99% variance (1% error), *etc*. For F_1 variable i_5 shows greatest weight; however, F_1 cannot be reduced to two variables $\{i_2,i_5\}$ without 53% error. For F_2 variable i_1 presents greatest weight; notwithstanding, F_2 cannot be reduced to two variables $\{i_1,i_6\}$ without 30% error. For F_3 variable i_3 assigns greatest weight; nevertheless, F_3 cannot be reduced to two variables $\{i_1,i_3\}$ without 21% error, *etc*.

Table 2. Importance of principal component analysis factors for proximate composition of spices

Factor	Eigenvalue	Percentage of variance	Cumulative percentage of variance
F_1	3.68662289	61.44	61.44
F_2	1.88354048	31.40	92.84
F_3	0.38799897	6.46	99.30
F_4	0.04139353	0.69	99.99
F_5	0.00044413	0.01	100.00
F_6	0.00000000	0.00	100.00

Scores plot of PCA F_2–F_1 for spices (*cf.* Fig. 5) illustrates different behaviour depending on *ash*, *etc*. Two clusters are clearly distinguished: class 1 with two spices ($0 \approx F_1 > F_2$, *bottom*) and grouping 2 with four seeds ($0 \approx F_1 < F_2$, *top*). The diagram is in qualitative agreement with PCD, binary/radial trees and split graph (Figs. 1-4).

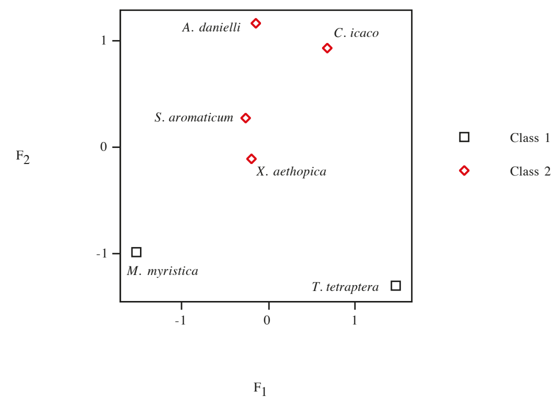

Figure 5. PCA scores plot of spices according to proximate composition.

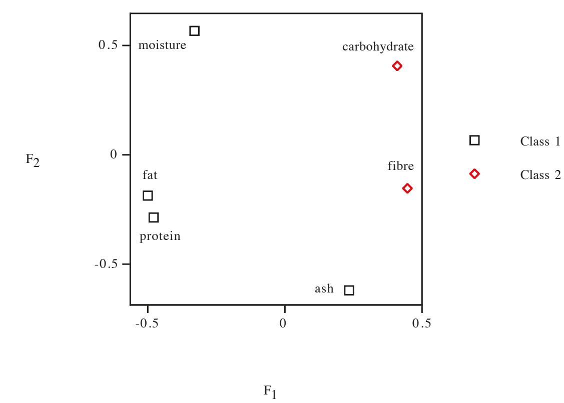

Figure 6. PCA loadings plot according to proximate composition.

From PCA factors loading of spices, F_2–F_1 *loadings plot* (*cf.* Fig. 6) depicts six proximates (Table 1). Two clusters are clearly distinguished: class 1 with four main constituents {1,2,5,6} ($F_1 < F_2 \approx 0$, *bottom*) and grouping 2 with two components {3,4} ($F_1 > F_2 \approx 0$, *top*). Macronutrients are relatively separated: *protein* and *fat* result included in class 1 while *carbohydrate* appears in grouping 2; *fat* is closer to *protein* than to *carbohydrate*. In addition as a complement to scores diagram for loadings it is confirmed that spices in class 1, located in the bottom, present a more pronounced contribution from proximates in grouping 1 located in the same position in Fig. 5. Plants in cluster 2 at the top show a contribution from components in class 2 found in the same location in Fig. 5.

Instead of six spices in space $\Re^6$ of six proximates consider six components in space $\Re^6$ of six plants. Matrix **R** upper triangle is:

$$\mathbf{R} = \begin{pmatrix} 1.000 & -0.088 & 0.417 & -0.037 & -0.214 & -0.938 \\ & 1.000 & -0.682 & -0.948 & 0.955 & 0.276 \\ & & 1.000 & 0.429 & -0.769 & -0.686 \\ & & & 1.000 & -0.885 & -0.066 \\ & & & & 1.000 & 0.388 \\ & & & & & 1.000 \end{pmatrix}$$

Low PCC correlations result between *ash* and other proximates $|R_{1,i}| \sim 0.3$ except *moisture* $R_{16} = -0.938$. High correlations appear between pairs of macronutrients (*protein*, *carbohydrate*, *fat*) $|R_{24}|$, R_{25}, $|R_{45}| \sim 0.9$. Again *fat* is closer to *protein* than to *carbohydrate* in agreement with PCA loadings plot (Fig. 6). The dendrogram for six proximates of spices (*cf.* Fig. 7) separates the same two classes above in agreement with PCA loadings plot (Fig. 6). One more time *protein* and *fat* result included in class 1 and *carbohydrate* in grouping 2. However, the results should be taken with care because the branching with only one component (*ash*) is a possible outlier.

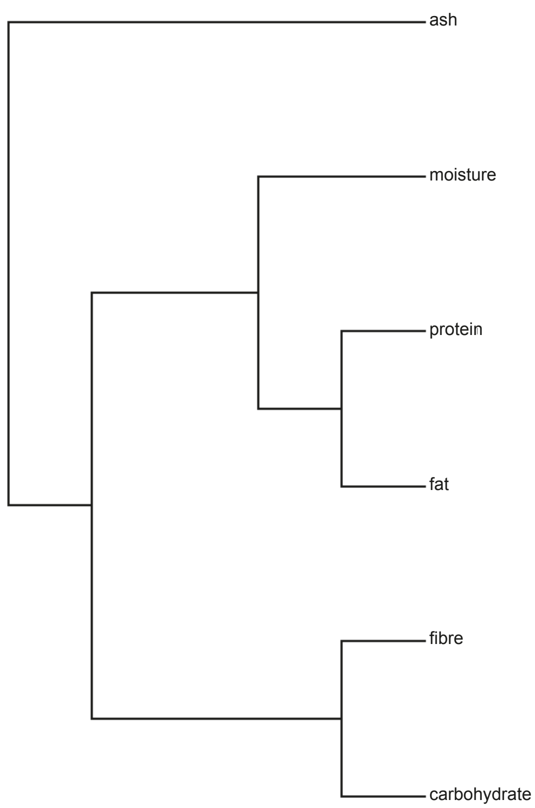

Figure 7. Dendrogram of proximate composition for spices.

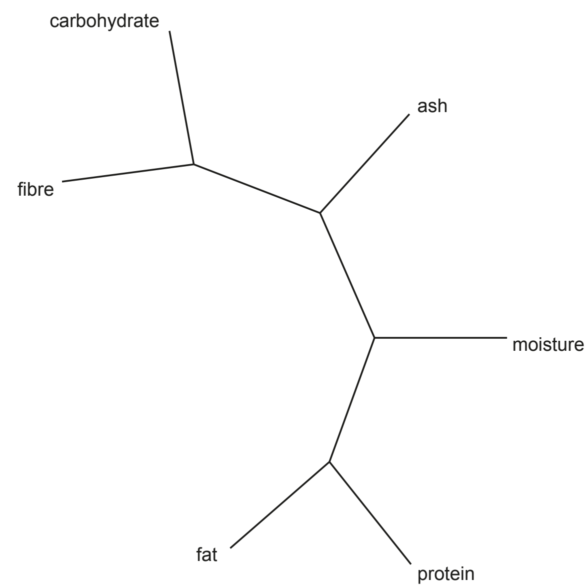

Figure 8. Radial tree of proximate composition for spices.

The radial tree for six proximates of spices (*cf.* Fig. 8) separates the same two classes above, in agreement with PCA loadings plot and dendrogram (Figs. 6 and 7). Once more, *protein* and *fat* result included in class 1 and *carbohydrate* in grouping 2.

Splits graph for six proximates of spices (*cf.* Fig. 9) reveals conflicting relations between classes because of interdependences. It separates the same two classes above in agreement with PCA loadings plot and binary/radial trees (Figs. 6-8). Again, *protein* and *fat* result included in class 1 and *carbohydrate* in grouping 2.

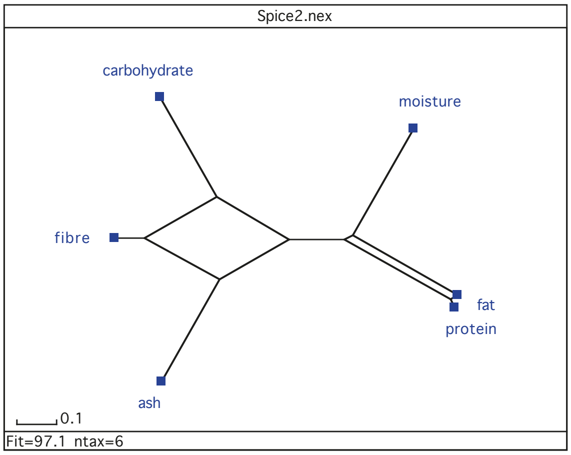

Figure 9. Splits graph of proximate composition for spices.

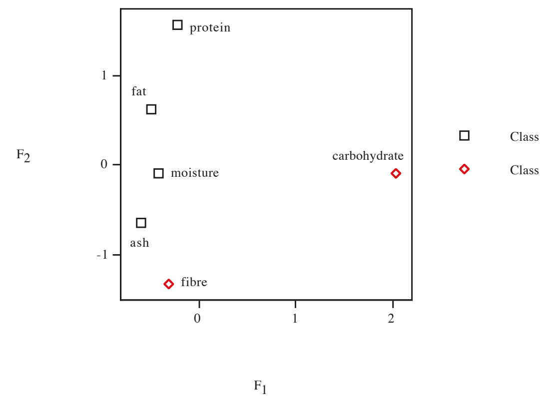

Figure 10. PCA scores plot of proximate composition for spices.

A PCA was performed for proximates. Factor F_1 explains 96% variance (4% error), $F_{1/2}$, 99.4% variance (0.6% error), F_{1-3}, 99.8% variance (0.2% error), *etc*. Scores plot of PCA F_2–F_1 for proximates (*cf.* Fig. 10) shows that two clusters are clearly distinguished: class 1 with four components ($F_1 < F_2 \approx 0$, *left*) and grouping 2 with two constituents ($F_1 >> F_2$, *right*). Again, *protein* and *fat* result included in class 1 and *carbohydrate* in cluster 2; macronutrients are separated: *fat* is closer to *protein* than to *carbohydrate*. The diagram separates the same classes above in qualitative agreement with PCA loadings plot, binary/radial trees and splits graph (Figs. 6-9).

DISCUSSION

The Bible offers descriptions of *ca*. 30 healing plants. Frankincense and myrrh enjoyed their status of great worth because of medicinal properties. Reported to present antiseptic properties, they were employed as mouthwashes. The spices are a good source of nutrients and natural antioxidants. Adequate macronutrients (*protein*, *carbohydrate*, and *fat*) of the seeds are beneficial to the body.

There is a need to search for medicinal plants with the aim of validating the ethnomedicinal, phytochemical, antioxidant, and antinutrient uses, and isolation and characterization of compounds, which will be added to the potential list of drugs. Phytochemical analysis revealed flavonoids, anthraquinones, saponins, phenols, tannins, alkaloids, cardiac glycosides, and terpenoids, which indicate that spices add nutritional and ethnomedicinal value to the diet. The seeds are sources of antioxidants. The potent antioxidant activity of flavonoids (phytoestrogens) reveals their ability to scavenge hydroxyl radicals $OH^{\bullet}$, superoxide anions $O_2^{\bullet-}$ and lipid peroxy radicals. Phenols are strong antioxidants that prevent oxidative damage to biomolecules, *e.g.*, deoxyribonucleic acid (DNA), lipids, and proteins, which play a role in chronic diseases, *e.g.*, cancer and cardiovascular disease. Plant phenols interfere with all stages of cancer, potentially resulting in a reduction of risk. The high chemical content of the sows supports the benefits the consumer may derive.

Cultures around the world rely on herbs and spices to add flavour and zest to food. Many spices, however, contain high numbers of bacteria, making them a potent source for food spoilage and pathogens. The combination of garlic and clove is able to kill 99% of *Escherichia coli* in salami.

CONCLUSIONS

From the discussion of the present results the following conclusions can be drawn.

1. Follow three advices: (a) Never blindly trust what you get. (b) Mathematical and statistical models are not the panacea. (c) Use this knowledge in data analysis to guide your investigation or experimentation, not as an end in itself.
2. Follow two advices on designing and improving experiments: (a) How do you design your experiments? (b) How do you improve your experiments?
3. Cluster segmentation's tool was used not only to segment the heterogeneous dataset into one smaller with distinct groups but also to extract knowledge from the dataset to analyse it. Data clustering was extensively used to understand the characteristics of homogeneous spice groups. This study developed different spice segment models and analysed every segment based on spice proximate composition.
4. Branchings with only one item, *e.g.*, spice *Monodora myristica* and proximate *ash*, are possible outliers.
5. Different behaviour of spices depends on proximate composition: *ash*, *fibre*, *fat*, *moisture*, *etc*. Macronutrients' (*protein*, *carbohydrate,* and *fat*) content are adequate. *Carbohydrate* amounts are high. *Fat* quantities are moderate. *Fat* is closer to *protein* than to *carbohydrate*.
6. The results in the technical literature suggest that the spices present a good antioxidant capacity, which could be accounted for their flavonoids and phenols content. Equally important are the results from proximate macronutrient content. For these reasons, the seeds may be used as a functional food, which could be useful in hypertension prevention and treatment.

REFERENCES

[1] D. Mukherjee, M. M. Laloraya, Keto acids and free amino acids during leaf growth in *Bauhinia purpurea*, *Experientia* **33** (1977) 304-305.

[2] G. K. Sharanabasappa, M. K. Santosh, D. Shaila, Y. N. Seetharam, I. Sanjeevarao, Phytochemical studies on *Bauhinia racemosa* Lam. *Bauhinia purpurea* Linn. and *Hardwickia binata* Roxb., *E-J. Chem*. **4** (2007) 21-31.

[3] A. A. Akindahunsi, S. O. Salawu, Phytochemical screening and nutrient-antinutrient composition of selected tropical green leafy vegetables, *Afr. J. Biotechnol*. **4** (2005) 497-501.

[4] I. Onyesom, P. N. Okoh, Qualitative analysis of nitrate and nitrite contents in vegetables commonly consumed in Delta State, Nigeria, *Br. J. Nutr*. **96** (2006) 902-905.

[5] D. E. OKWU, Phytochemicals and vitamin content of indigenous spices of Southeastern Nigeria, *J. Sustain. Agric. Environ*. **6** (2004) 30-37.

[6] M. Valko, D. Leibfritz, J. Moncol, M. T. Cronin, M. Mazur, J. Telser, Free radicals and antioxidants in normal physiological functions and human disease, *Int. J. Biochem. Cell Biol*. **39** (2007) 44-84.

[7] D. E. Okwu, Flavoring properties of spices on cassava fufu, *Afr. J. Root Tuber Crops* **3** (1999) 19-21.

[8] D. E. Okwu, Evaluation of the chemical composition of indigenous spices and flavoring agents. *Global J. Pure Appl. Sci*. **7** (2001) 455-459.

[9] M. A. B. Pandey, S. Abidi, R. P. Singh, Nutritional evaluation of leafy vegetables paratha, *J. Hum. Ecol*. **19** (2006) 155-156.

[10] J. O. Olubunmi, Proximate analysis and phytochemical evaluation of some commonly used food spices in South-East Nigeria. *1st International Conference on Science and Education Development Potentials*, Akure, Ondo, Nigeria, 2013.

[11] G. Smith, M. Cleeg, C. Keen, L. Grivetti, Mineral values of selected plant foods common to Southern Burkina Faso and to Niamey, Niger West Afr., *Int. J. Food Sci. Nutr*. **47** (1996) 41-53.

[12] F. Torrens-Zaragozá, Molecular categorization of yams by principal component and cluster analyses, *Nereis* (5) (2013) 41-51.

[13] F. Torrens-Zaragozá, Classification of fruits proximate and mineral content: Principal component, cluster, meta-analyses, *Nereis* (7) (2015) 39-50.

[14] H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* **24** (1933) 417-441.
[15] R. Kramer, *Chemometric techniques for quantitative analysis*, Marcel Dekker, New York, 1998.
[16] S. K. Patra, A. K. Mandal, M. K. J. Pal, *Photochem. Photobiol., Sect. A* **122** (1999) 23-31.
[17] I. T. Jolliffe, *Principal component analysis*, Springer, New York, 2002.
[18] J. Xu, A. Hagler, *Molecules* 7 (2002) 566-600.
[19] P. J. A. Shaw, *Multivariate statistics for the environmental sciences*, Hodder-Arnold, New York, 2003.
[20] IMSL, *Integrated Mathematical Statistical Library (IMSL)*, IMSL, Houston, 1989.
[21] R. C. J. Tryon, *Chronic Dis.* **20** (1939) 511-524.
[22] I. Priness, O. Maimon, I. Ben-Gal, Evaluation of gene-expression clustering via mutual information distance measure, *BMC Bioinformatics* **8** (2007) 111-112.
[23] R. Steuer, J. Kurths, C. O. Daub, J. Weise, J. Selbig, The mutual information: Detecting and evaluating dependencies between variables, *Bioinformatics* **18**(Suppl. 2) (2002) S231-S240.
[24] P. D'Haeseleer, S. Liang, R. Somogyi, Genetic network inference: From coexpression clustering to reverse engineering, *Bioinformatics* **16** (2000) 707-726.
[25] C. M. Perou, T. Sørlie, M. B. van Eisen, M. de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A. L. Børresen-Dale, P. O. Brown, D. Botstein, Molecular portraits of human breast tumours, *Nature (London)* **406** (2000) 747-752.
[26] R. A. Jarvis, E. A Patrick, Clustering using a similarity measure based on shared nearest neighbors, *IEEE Trans. Comput.* **C22** (1973) 1025-1034.
[27] R. D. M. Page, *Program TreeView*, Universiy of Glasgow, UK, 2000.
[28] M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. U. S. A.* **95** (1998) 14863-14868.
[29] D. H. Huson, SplitsTree: Analyzing and visualizing evolutionary data, *Bioinformatics* **14** (1998) 68-73.

ACKNOWLEDGEMENT

The author thanks support from the Spanish Ministerio de Economía y Competitividad (Project No. BFU2013-41648-P) and EU ERDF.