DEVELOPMENT OF CLUSTERING AND CLASSIFICATION STRATEGIES FOR THE DETERMINATION OF GEOGRAPHICAL ORIGIN OF HONEY BY USING ATOMIC AND MOLECULAR SPECTROMETRY

A Thesis Submitted to the
Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by
Müşerref YERSEL

July 2011
İZMİR

We approve the thesis of **Müşerref YERSEL**

Assoc.Prof.Dr. Durmuş ÖZDEMİR
Supervisor

Prof.Dr. Ahmet E. EROĞLU
Co-Supervisor

Prof.Dr. Nil ERTAŞ
Committee Member

Prof.Dr. Serdar ÖZÇELİK
Committee Member

Assoc.Prof.Dr. Şerife YALÇIN
Committee Member

Assoc.Prof.Dr. Figen TOKATLI
Committee Member

01 July 2011

Prof.Dr. Serdar ÖZÇELİK
Head of the Department of Chemistry

Prof.Dr. Durmuş Ali DEMİR
Dean of the Graduate School of Engineering and Sciences

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Assoc. Prof Dr. Durmuş ÖZDEMİR for his supports and encouragements throughout of this study. In addition, I am grateful to my co-supervisor Prof. Dr. Ahmet E. EROĞLU for his valuable comments.

I also would like to thank to members of the thesis committee, Assoc. Prof. Dr. Şerife YALÇIN, Assoc. Prof. Dr. Figen KOREL, Assoc. Prof. Dr. Figen TOKATLI, Prof. Dr. Nil ERTAŞ and Pof. Dr. Serdar ÖZÇELİK for their valuable comments and patience.

I am very indebted my friends Aslı ERDEM and Özge TUNUSOĞLU for their patience and encouragement during this thesis.

Special thanks to Arzu ERDEM for her huge support and help in every step of my study.

I am deeply appreciative to the research scientists at the Environmental Research Centre for ICP-MS analyses.

This study would not be possible without the support of The Scientific and Technical Research Council of Turkey (TUBİTAK-Hızlı Destek Project number 209T120).

Finally I dedicate this thesis to my parents, Sevim YERSEL and Mehmet YERSEL.

ABSTRACT

DEVELOPMENT OF CLUSTERING AND CLASSIFICATION STRATEGIES FOR THE DETERMINATION OF GEOGRAPHICAL ORIGIN OF HONEY BY USING ATOMIC AND MOLECULAR SPECTROMETRY

Honey is a natural, nutritious and healthy food produced by honeybees from the nectar of plants. The classification of honey based on geographical origin is of great interest since the quality of honey depends on its chemical composition and geographical origin. In this study, it is aimed to develop classification models using elemental and molecular composition of honey samples via atomic and molecular spectrometry. For this purpose, honey samples from different regions of Turkey were collected from producers and they were scanned with Fourier Transform infrared spectrometer equipped with attenuated total reflectance (FTIR-ATR) accessory, and fluorescence spectrophotometer (synchronous fluorescence mode and 3D excitation emission mode). Afterwards, any clustering of the samples based on their regions was investigated using principal component analysis (PCA) and hierarchical cluster analysis (HCA) and soft independent modeling of class analogies (SIMCA). Finally, inductively coupled plasma mass spectrometry was applied to determine the metal concentrations (Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba) in honey samples and then the same classification methods were performed to compare the results.

In conclusion, molecular spectrometry gave better classification results based on geographical origin compared to the results obtained with atomic spectrometry. Molecular spectrometry is more advantageous for the classification of honey samples in the case of saving time, saving chemicals and ease of usage.

ÖZET

ATOMİK VE MOLEKÜLER SPEKTROMETRİ KULLANILARAK BALIN COĞRAFİK KÖKENİNİN BELİRLENMESİ İÇİN SINIFLANDIRMA VE KÜMELEME STRATEJİLERİNİN GELİŞTİRİLMESİ

Bal, bal arıları tarafından bitkilerin nektarlarından üretilen, besin değeri açısından önemli bir gıda maddesidir. Balların kalitesi, içeriği ve üretildiği bölgeye bağlı olmasından dolayı, sınıflandırılmaları büyük önem taşımaktadır. Bu çalışmada, atomik ve moleküler spektrometri verilerine kemometrik analiz yöntemlerini uygulayarak sınıflandırma modellerinin kurulması amaçlanmıştır. Bu amaçla, Türkiye'nin farklı yörelerinden bal örnekleri toplanmıştır.ve toplanan örnekler, Fourier transform infrared spektrometresinde zayıflatılmış toplam reflektans aparatı (FTIR-ATR) ve floresans spektrometresi (senkronize floresans ve uyarılma-emisyon modu) ile taranmıştır. Sonrasında, bölgelere göre kümelenme olup olmayacağını araştırmak için yönlendirmesiz sınıflandırma (unsupervised classification) metotlarından temel bileşenler analizi (principal component analysis, PCA) ve hiyerarşik kümeleme analizi (hierarchical cluster analysis, HCA) ve yönlendirmeli metotlardan kısmi bağımsız benzeşim modeli (SIMCA) uygulanmıştır. Son olarak, örnekler bozundurularak indüktüf eşleşmiş plazma kütle spektrometre (inductively coupled plasma mass spectrometer-ICPMS) ile metal içerikleri (Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba) belirlenmiş ve karşılaştırma yapmak amacı ile aynı sınıflandırma metotları uygulanmıştır.

Sonuç olarak, balların coğrafik bölgelerine göre sınıflandırılmasında; moleküler spektrometri verilerinin atomik spektrometri verilerine göre daha başarılı olduğu belirlenmiştir. Moleküler spektrometri, analiz için harcanan süre, kimyasal madde kullanımı ve kullanım kolaylığı açısından avantajlıdır.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

1.1. Honey

Honey is a viscous, sweet, substance made by honey bees (*Apis Mellifera*) mostly from nectar of honey plants. Due to its special flavor and sweet taste, honey has been long valued as a desired natural sweetener consumed by humans. Being a valuable constituent of a rough daily diet, honey is also an important ingredient of other food products, beverages and pharmaceuticals. All the properties of honey, including nutritious and medicinal effects, result from its unique chemical composition and a specific processing subjected by bees to the collected sources (Arvanitoyannis et al. 2005). The main parts of honey are simple sugars and water (Figure 1.1). As regards the carbohydrates, the average content of fructose and glucose in honey is 38.2 and 31.3 %, respectively (Pohl et al. 2009a). Other carbohydrates are maltose (7.3 % on average), saccharose (2.4 % on average), and few high sugars (1.5 % on average). The average moisture content is 17.2 %.

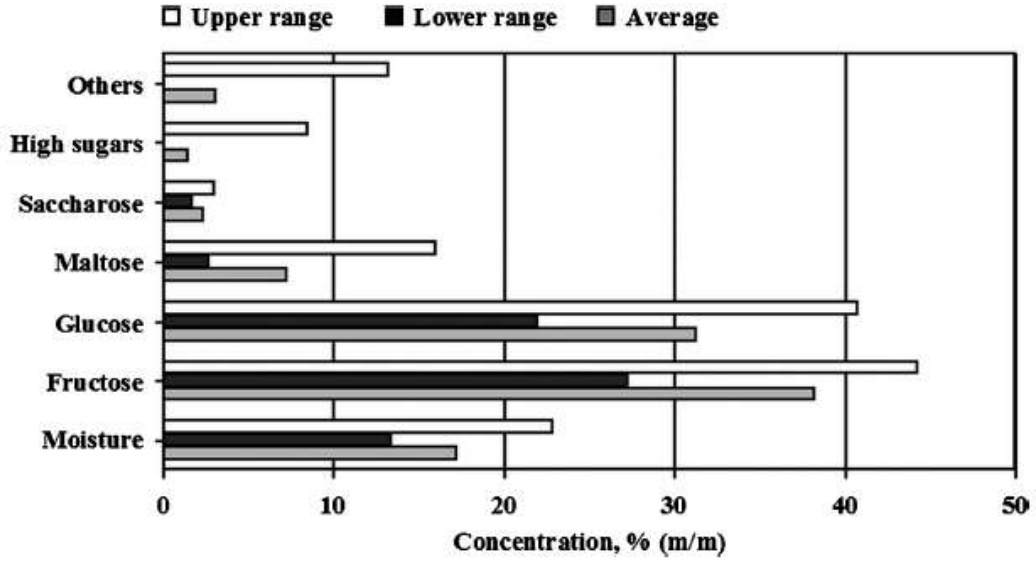

Figure 1.1. Percentage contributions of honey constituents
(Source: Pohl et al. 2009)

Honey additionally contains proteins and enzymes (diastase, invertase, glucose oxidase, and catalase), amino acids, organic acids (acetic, butyric, citric, succinic, lactic, malic, gluconic, and ascorbic), vitamins (riboflavin, niacin, folic acid, panthothenic acid, and B6), flavonoids, and minerals (Belitz et al. 2004). Honey also includes major, minor, and traces metals. Their composition and content are characteristic to the floral type of honey, geochemical and climatic conditions attributed to the forage area, and locality of an apiary. Usually, the total content of all metals in flora honeys contribute to about 0.1–0.2% of their total composition.

1.2. Classification of Honey

Honey is very important food stuff due to its nutrient and therapeutic effects. It is the natural product obtained by honey bees (*Apis Mellifera L.*) from the nectar of flowers or from secretions of living parts of the plants. The botanical origin of the nectars or secretions determines the composition and characteristics of honey. According to different characteristics the classification of honey is possible and there are many studies on this topic.

In literature, several different methods were used for the classification of honey samples. Conti et al. (2007) have characterized three types of Italian honey based on their quality parameters (pH, sugar content, humidity) and mineral content (Na, K, Ca, Mg, Cu, Fe, and Mn). Principal components analysis (PCA) and linear discriminant analysis (LDA) were applied as pattern recognition methods. Thirty two samples of honey from La Pampa (Argentina) have been classified on the basis of their phosphorous, aluminum, iron, calcium, magnesium and sodium contents using inductive coupled plasma optical emission spectrometry (ICP-OES) by Camina et al. (2008). Classification was performed with PCA, cluster analysis (CA) and LDA. Nozal Nalda et al. (2005) have classified a total of 73 different honeys from seven botanical origins ling, heather, rosemary, thyme, honeydew, spike lavender and French lavender by applying discriminant analysis to their metal content data and other common physicochemical parameters. Torres et al. (2005) have determined eleven elements (Zn, P, B, Mn, Mg, Cu, Ca, Ba, Sr, Na and K) by ICP-OES in 40 honey samples from different places of Spain and four different botanical origins: eucalyptus, heather, orange-blossom and Rosemary. Then, PCA, cluster analysis and LDA were used for

differentiation of honeys according to botanical origin. It was found that Zn, Mn, Mg and Na concentrations were strongly dependent on the kind of botanical origin. The characterization of honey samples produced in Cordoba (Argentina) and their classification by geographical origin (North/South) with the use of 15 variables (glucose, pH, free acidity, free amino acids, calcium and zinc) have been reported by Baroni et al. (2009). Chudzinska and Baralkiewicz (2010) have studied the mineral content of fifty five honey samples, which represented three different types of honey: honeydew, buckwheat and rape honey from different areas in Poland. Determination of thirteen elements (Al, B, Ba, Ca, Cd, Cu, K, Mg, Mn, Na, Ni, Pb, and Zn) was performed using inductively coupled plasma-mass spectrometry (ICP-MS) and CA and PCA were applied to classify honey according to mineral content. CA showed three clusters corresponding to the three botanical origins of honey. PCA permitted the reduction of 13 variables to four principal components explaining 77.19% of the total variance. The only work about the classification of Turkish honeys was performed by Senyuva et al. (2009). Seventy authentic honey samples of 9 different floral types (rhododendron, chestnut, honeydew, Anzer (thymus spp.), eucalyptus, gossypium, citrus, sunflower, and multifloral) have been used from 15 different geographical regions of Turkey. The profiles of free amino acids, oligosaccharides, and volatile components together with water activity were used for partial least squares (PLS) followed by linear discriminant analysis (LDA).

Infrared spectroscopy is also used for classification. Fourier transform infrared (FTIR) spectroscopy coupled with partial least squares (PLS) regression analysis, factorial discriminant analysis (FDA), and soft independent modeling of class analogy (SIMCA) were used to verify the origin of honey samples (n=150) from Europe and South America (Hennessy et al. 2008) according to geographical origin. It was found that correct classifications of up to 100% were achieved using SIMCA. Ruoff et al. (2006a) also used FTIR spectroscopy for authentication of the botanical origin and the geographical origin of honey. Principal component analysis and linear discriminant analysis have been applied to evaluate the spectra. Ruoff et al. (2006b) have again used Fourier transform near-infrared spectroscopy (FT-NIR) for discrimination of acacia, chestnut, and fir honeydew honey from the other unifloral and polyfloral honey types studied. Corbella and Cozzolino (2006) have classified floral origin Uruguayan honeys using Moisture (M), pH, electric conductivity (EC) values with LDA. A study on geographical origin of 167 unfiltered honey samples (88 Irish, 54 Mexican, and 25

Spanish) and 125 filtered honey samples (25 Irish, 25 Argentinean, 50 Czech, and 25 Hungarian) has been performed by Woodcock et al. (2007) using near infrared spectroscopy with PCA and SIMCA. Kelly et al. (2006) identified authentic honey and honey adulterated by beet sucrose, dextrose syrups, and partial invert corn syrup by the use of soft independent modeling of class analogy (SIMCA) and partial least-squares (PLS) classification with FTIR-ATR spectra of samples. This combination of spectroscopic technique and chemometric methods was not able to definitely identify adulteration by high-fructose corn syrup or fully inverted beet syrup. In other work, FTIR spectroscopy with ATR accessory has been used to quantify three different adulterants (corn syrup, high fructose corn syrup and inverted sugar) in honeys of four different locations of Mexico (Velazquez et al. 2009). The calibrations for the three adulterants were constructed with partial least squares (PLS) and classification of the Mexican honeys from the four different states was succeeded with soft independent modeling class analogy (SIMCA). SIMCA analysis was capable to classify correctly the origin of the Mexican honeys from four different states. There is not much study on the classification of honey with fluorescence spectrometric information. An example of such study has been performed by Karoui et al. (2006). In this research, front face fluorescence spectroscopy was suggested for seven honey types (acacia, alpine rose, chestnut, rape, honeydew, alpine polyfloral and lowland polyfloral). They used the first 10 principal components (PCs) of the principal component analysis (PCA) extracted from each data set and then analyzed by factorial discriminant analysis (FDA).

In addition, chromatographic methods have been also used in honey analysis for classification. Nozal et al. (2005) presented a study to characterize the botanical origin of honey from a single geographical area, the Province of Soria (Spain), using carbohydrate profiles and canonical discriminant analysis. Fourteen carbohydrates were quantified using high-performance liquid chromatography (HPLC) with pulsed amperometric detection (PAD) in 77 natural honeys, the botanical origins of which were ling, spike lavender, French lavender, thyme, forest, and multifloral. PCA was employed as a first approach to characterize the honey samples analyzed, showing similarities, then they achieved best discrimination with canonical discrimant analysis. Geographical origin of honeys based on volatile compounds profiles was examined by Stanimirova et al. (2010). The volatiles in honeys were analyzed by a head-space solid phase microextraction (SPME) combined with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry (GCxGC–TOFMS). Linear

discriminant analysis (LDA), soft independent modeling of class analogies (SIMCA), discriminant partial least squares (DPLS) were used in this study to discriminate between Corsican and non-Corsican honeys. Head-space solid-phase microextraction (HS-SPME)-based procedure, coupled to comprehensive two dimensional gas chromatography–time-of-flight mass spectrometry (GC×GC–TOF-MS), was employed by Cajka et al. (2009) for classification of honey based on honey volatiles. In total, 374 samples were collected over two production seasons in Corsica (n = 219) and other European countries (n = 155). For chemometric analysis, artificial neural networks with multilayer perceptrons (ANN-MLP) were tested. Thus, the best prediction (94.5%) and classification (96.5%) abilities of the ANN-MLP model were obtained.

High resolution nuclear magnetic resonance (HR-NMR) data was used by Lolli et al. (2008) for the 71 honey samples in order to classify using PCA and general discriminant analysis (GDA). There is also one more study with NMR by Consonni and Cagliani (2008). Forty one honey samples have been analyzed (polyfloral and acacia) from different countries in terms of 1H NMR spectroscopy coupled with multivariate statistical methods. Subsequently, PCA and PLS-DA were used for investigation of geographical origin. An interesting work was developed by Dias et al. (2008). They have used an electronic tongue system based on 20 all-solid-state potentiometric sensors and chemometric data processing (PCA, LDA), with polymeric membranes applied on solid conducting silver-epoxy supports and a Ag-AgCl reference electrode.

1.3. Classification Methods

Most of the published studies in chemometrsics are on pattern recognition. Pattern recognition is used to classify the objects into sets based upon some similarity in properties (Einax et al. 1995). The aim is to classify data (patterns) based on either knowledge or on statistical information. In chemistry, there are many applications using data to determine the patterns. The following examples can be given: wine characterization based on the analysis of the biogenic amine composition using the chromatographic profiles (Garcia-Villar et al. 2007), verifying the geographical origin olive oils by near infrared spectroscopy (Woodcock et al. 2008) and monitoring of water quality using nitrate, sulphate, chloride, turbidity, conductivity, hardness, alkalinity, coliforms and Escherichia coli data (De Luca et al. 2008).

There are many methods for chemical classification. Classification methods in chemometrics are mainly divided into two groups: unsupervised and supervised techniques.

1.3.1 Unsupervised Methods

The main goal of unsupervised methods is to evaluate whether clustering exists in a data set and to find a property of objects using measurements on them. Unsupervised methods do not require any prior knowledge about the group structure in the data, but instead produce the grouping and this type of methods mainly analyzes the data. In some situations the class membership of the samples is known. If the aim is any grouping between samples or any outliers, unsupervised pattern recognition techniques such as principal component analysis (PCA), hierarchical cluster analysis (HCA) can be used. Thus, the class information is known or suspected but is not used initially. (Sharaf et al. 1986)

1.3.1.1 Principal Component Analysis (PCA)

In most studies, many variables are measured on each individual, which result a huge data set consisting of large number of variables. The dimensionality of the data set can often be reduced, without disturbing the main features of the whole data set by Principal Component Analysis (PCA) technique. It is a simple, non-parametric method of extracting relevant information from confusing data sets. PCA represents the relationship among the observations and reveals any deviating observations or groups of observations in the data. The main advantage of PCA is that once you can find the patterns in the data, and you compress the data (Lindsay 2002).

While PCA is performed, the dataset is decomposed into two parts, namely, meaningful information and error (or noise). The transformation is often mathematically described as follows (Brereton 2002).

$$\boldsymbol{X} = \boldsymbol{T}.\boldsymbol{P} + \boldsymbol{E} = \hat{\boldsymbol{X}} + \boldsymbol{E} \tag{1.1}$$

where

• X is the original data

• T is the principal component scores and has as many rows as the original data matrix;

• P is the principal component loadings and has as many columns as the original data matrix;

• E is an error matrix

Many chemometricians use a "hat" notation to indicate a prediction so $\hat{X}$ is the "prediction" of X using the PC model.

In other words, the projection of X down on to a d-dimensional subspace by means of the projection matrix P gives the object coordinates in this plane, T. The columns in T are the score vectors and the rows in P are called loading vectors. Both vectors are orthogonal (Otto 1999).

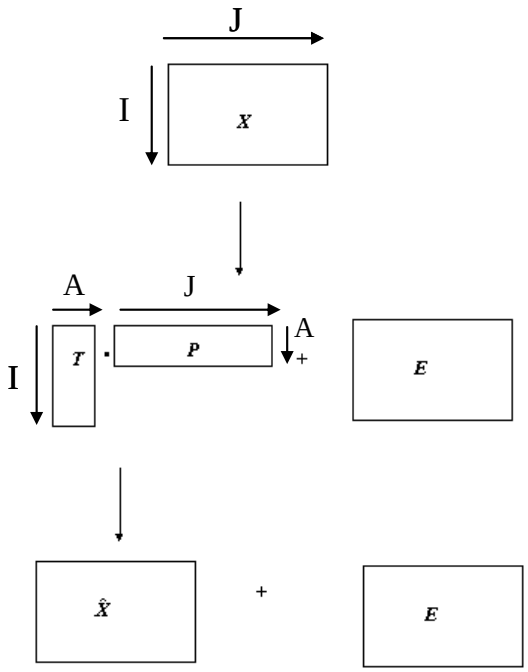

Figure1.2. Principal component analysis
(Source: Brereton 2002)

Each score matrix consists of a series of column vectors and each loading matrix a series of row vectors. These vectors are denoted by ta and pa, where a is the number of principal component (1,2,3,…, A,) The matrices T and P have such vectors one for each principal component. Decomposition of the correlation matrix into eigenvalues and eigenvectors leads to the linear combination with the dataset simplifications. PCA starts with the determination of the number of principal components by the percentage of explained variance, eigenvalues, and cross-validation. Eigenvalue is called as the size of

each component. The most significant component has the largest size. Simple definition of eigenvalue of a principal component is the sum of squares of the scores, so that

$$g_a = \sum_{i=1}^{I} t_{ia}^2 \qquad (1.2)$$

where is the g_a is a^{th} eigenvalue and t_{ia} is score vectors. The eigenvectors represent the linear combination coefficients and the corresponding eigenvalues represent the variance described by each linear combination. As the eigenvalues are in non-decreasing order, the first linear components account for the largest amount of variance. The principal components are determined on the basis of the maximum variance criterion (Figure 1.3). The significance of the each principal component can be tested by cross-validation. In cross-validation, each sample is removed once from the data set and PCA is performed on the remaining samples. Different scores and loadings matrices are obtained depending on removed sample. In this way, all samples are removed once and the remaining sample is predicted.

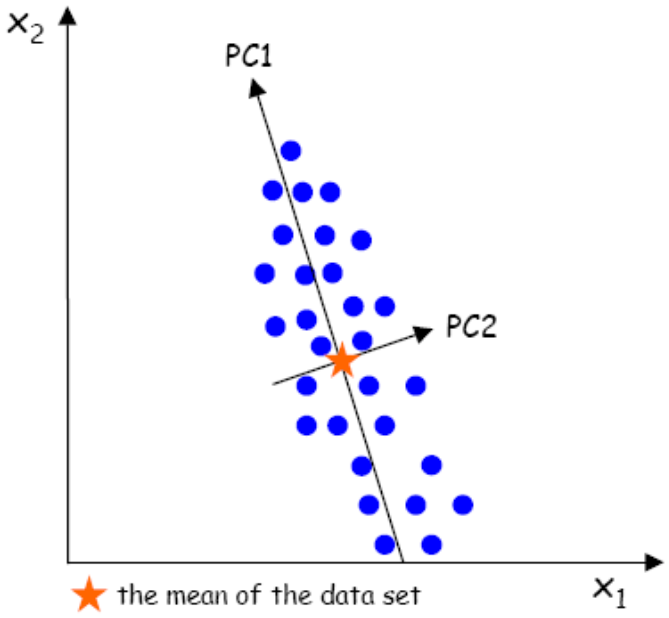

Figure 1.3 A row plot of data in a two-measurement system, with the first two principal component axes.

Each subsequent principal component describes a maximum of variance that is not modeled by the former components. According to this, most of the variance of the data is contained in the first principal component. In the second component, there is

more information than in the third, etc. Finally, as many principal components as are computed, those are needed to explain a preset percentage of the variance.

In conclusion, PCA is mainly used for;

- Visualization of multivariate data by scatter plots
- Transformation of highly correlating x-variables into a smaller set of uncorrelated latent variables that can be used by other methods
- Separation of relevant information (described by a few latent variables) from noise
- Combination of several variables that characterize a chemical-technological process into a single or a few ‘‘characteristic’’ variables (Varmuza and Filzmoser 2008).

1.3.1.2. Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) is an unsupervised technique concerning with forming groups of similar objects based on several measurements of different kinds made on the objects. The main idea is to examine the interpoint distances between all the samples and represents that information in the form of two dimensional plots as a dendrogram. This idea has been applied in many areas including astronomy, archeology, medicine, chemistry, education, psychology, linguistics and sociology.

While constructing a dendrogram, the first step is to determine the similarities between objects. It is possible with measuring the distances between objects. There are many different methods for measuring a distance and the most used ones for hierarchical cluster analysis are as follows:

1) Euclidean distance: The distance between samples k and l is defined by:

$$d_{kl} = \sqrt{\sum_{j=1}^{J} (x_{kj} - x_{lj})^2} \tag{1.3}$$

where there are j measurements and x_{kj} is the j^{th} measurement on sample k.

2) Manhattan distance: This is defined slightly differently to the Euclidean distance and is given by:

$$d_{kl} = \sum_{j=1}^{J} | x_{kj} - x_{lj} | \tag{1.4}$$

3) **Mahalanobis distance:** This method is similar to the Euclidean distance; it takes into account that some variables may be correlated, thus it measures more or less the same properties.

$$d_{kl} = \sqrt{(x_k - x_l)\mathbf{C}^{-1}(x_k - x_l)^{-1}} \tag{1.5}$$

Where C is the covariance matrix. The Mahalanobis distance is as same as with Euclidean distance if the covariance matrix is the identity matrix.

After measurement the distances between the objects, the next step is to link these objects. The most frequent approach is agglomerative clustering where single objects are gradually connected to each other in groups. There are various ways of doing this.

Single linkage: Here the shortest distance between opposite clusters is calculated. Thus, the first cluster is one with two observations that have the shortest distance.

Complete linkage: This is similar to single linkage except that this is based on the maximum distance; minimum distance is not considered. The maximum distance between any two individuals in a cluster represents the smallest (minimum diameter) sphere that can enclose the cluster.

Average linkage: Here the average distance from samples in one cluster to samples in other clusters are used. There are two different way of doing this, according to the size of each group being joined together.

i) Unweighted average linkage: with this method the number of objects in a cluster is used for weighting the cluster distances.

ii) Weighted average linkage: the sizes of clusters and their weights are assumed to be equal.

Centroid (Mean) Method: Euclidean distance is measured between centroids of two clusters.

Ward's method: This method is distinct from all other methods because it uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the Sum of Squares (SS) of any two (hypothetical) clusters that can be formed at each step.

There are also various linkage methods, but it is not needed to use too many combinations of similarity and linkage methods. However, the general way is to check the result of using a combination of approaches.

The next steps consist of ongoing to group the data, until all objects have joined one large group. Since there are four original objects, there will be three steps before this is achieved. At each step, the most similar pair of objects or clusters is identified, and then they are combined into one new cluster, until all objects have been joined Figure 1.4 (Brereton 2002).

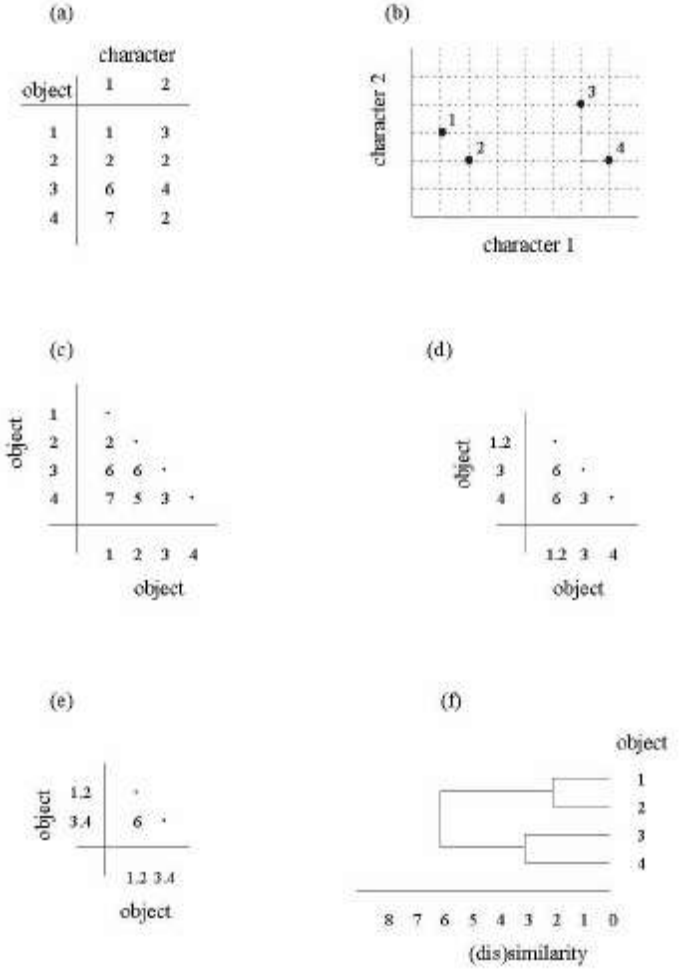

Figure 1.4. Simple example illustrating the protocol for cluster analysis. (a) Data set consisting of four objects, each characterized by two characters, (b) Objects plotted in character space, (c) Similarity matrix showing dissimilarity between objects. (d) and (e) Derived similarity matrices used in successive steps of the clustering process, (f) Dendrogram. (Source: Varmuza and Filzmoser 2008)

1.3.2. Supervised Methods

Supervised methods are used when the goal is to construct a model to be used in the classification of future samples. An example of a supervised model is one developed

for classifying different raw materials. Measurements are taken on samples for each raw material (class) and a model is constructed that the best discriminates the classes. The class of new material is predicted with the constructed model to verify the identity. The K-Nearest Neighbor (KNN), soft independent modeling of class analogies (SIMCA), discriminant analysis are some of the supervised pattern recognition techniques. The K-Nearest Neighbor (KNN) is supervised technique in which models are constructed using the physical closeness of samples in multidimensional space. KNN uses only the physical closeness of samples to construct models while soft independent modeling of class analogies (SIMCA) uses the position and shape of the object formed by the samples in row space. In SIMCA, multidimensional box is constructed for each class using PCA and the classification of future samples is performed by determining within which box the sample belongs.

The purpose of discriminant analysis is to classify objects into one of two or more groups based on a set of features that describe the objects. Thus, in discriminant analysis, the dependent variable (Y) is the group and the independent variables (X) are the object features that might describe the group. The dependent variable is always category (nominal scale) variable while the independent variables can be any measurement scale (i.e. nominal, ordinal, interval or ratio). Discriminant analysis uses continuous variable measurements on different groups of items to highlight aspects that distinguish the groups and to use these measurements to classify new items (Beebe et al. 1998).

In the present study, the unsupervised methods like principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used for classifying honey samples from different regions of Turkey. Afterwards, same samples were analyzed with a supervised technique which was soft independent modeling of class analogies (SIMCA).

1.3.2.1. Soft Independent Modeling of Class Analogies (SIMCA)

In principal component analysis (PCA), the class information is not used in the construction of the model and PCA just describes the overall variation in the data. However, PCA can be coupled with the class information in order to give classification models by means of soft independent modeling of class analogy. SIMCA was first

introduced by Svante Wold in the early 1970s. It is defined as “soft” since no hypothesis on the distribution of variables is made and “independent” since classes are modeled one at a time i.e. each class model is developed independently (Sun 2009).

In SIMCA, a PCA is performed on each class in the data set, and a sufficient number of principal components are kept to explain for most of the variation within each class. Therefore, a principal component model is used to represent each class in the data set. The number of principal components retained for each class is usually different. Thus, the class models may represent lines, planes, boxes or hyper boxes as demonstrated in Figure 1.5.

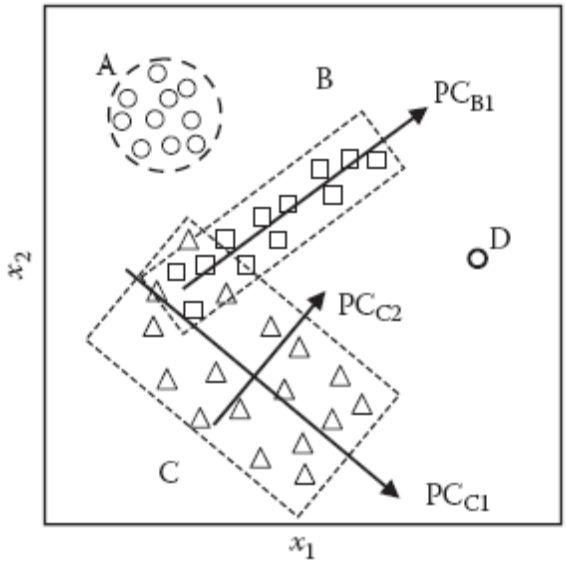

Figure 1.5. Principle of SIMCA modeling and classification
(Source: Varmuza and Filzmose 2008).

Group A can be modeled by a single prototype point (usually the center of the group) and a sphere with an appropriate radius. Group B is distributed along a straight line, and one principal component, PC_{B1}, together with an appropriate radius defines a model with a cylindrical shape. Group C requires two principal components, PC_{C1} and PC_{C2}, and the geometric model is a rectangular box. The single object D would be recognized not to belong to any of the groups A–C.

Deciding on the number of principal components that should be retained for each class is important, as retention of too few components can distort the signal or information content contained in the model about the class, whereas retention of too many principal components diminishes the signal-to-noise. A procedure called cross-

validation ensures that the model size can be determined directly from the data. To perform cross-validation, segments of the data are omitted during the PCA. Using one, two, three, etc., principal components, omitted data are predicted and compared to the actual values. This procedure is repeated until every data element has been kept out once. The principal component model that yields the minimum prediction error for the omitted data is retained (Brereton 2002)

SIMCA develops principal component models for each training class separately and provides information including critical distances which can be calculated as the geometric distance of each object from the principal component models. Following the modeling for classes, each sample is fitted to each model and classification of the sample with corresponding class is achieved. SIMCA results can be graphically visualized. Thus, a plot of the loadings and the scores of the PCA performed on the training set provide information about outliers, sub-groupings and within-class structure. Moreover, a useful tool for the interpretation of SIMCA results is the so-called Cooman's plot (Figure 1.6), which shows the discrimination of two classes.

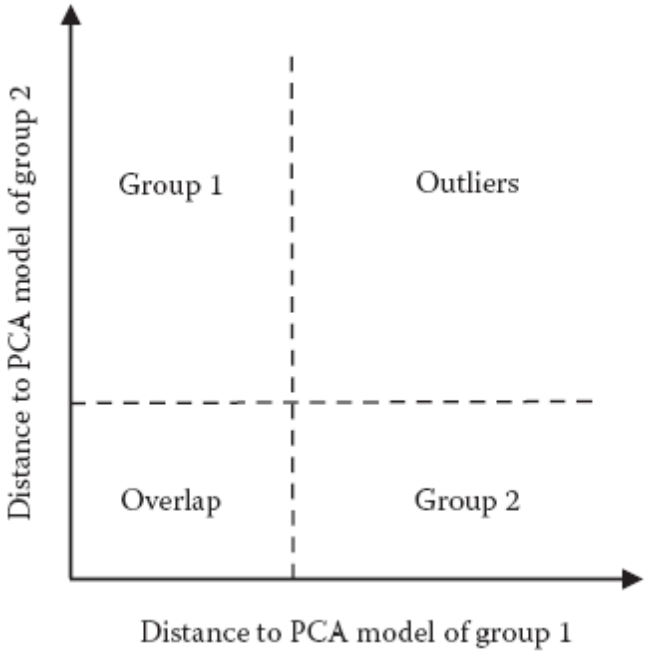

Figure 1.6. The Cooman's plot uses the distances of the objects to the PCA models of two groups. It visualizes whether objects belong to one of the groups, to both, or to none (Source: Varmuza and Filzmose 2008).

The distance from the model for class 1 is plotted against that from model 2. The critical distances (usually at 95% of confidence level) are indicated on both axes.

Consequently, four zones are defined on the plot: class 1, class 2, overlap of classes' 1 and 2, and outlier zone (far from both classes). By plotting objects in this plot it is easy to visualize how certain a classification is (Berrueta et al. 2007). The Cooman's Plot gives information about the class membership of the observations for two models simultaneously.

1.4. Inductively Coupled Plasma Mass Spectrometry

Inductively Coupled Plasma Mass Spectrometry (ICP-MS) is an analytical technique used for trace multielement analysis, often at the part per trillion levels. The technique was commercially introduced in 1983 and has gained general acceptance in many types of laboratories. An ICP-MS combines a high-temperature ICP (Inductively Coupled Plasma) source with a mass spectrometer.

Plasma is collection of electrons, ions and neutral atoms that is electrically neutral. It is an ionized gas, usually argon, which is sustained by a radio frequency (RF) generator. The RF generator applies an electric force through a copper coil that as consequence ionizes the argon gas onto Ar^+ ions and electrons. The argon gas is selected among others because is the noble gas found in more abundance in the atmosphere and has a low ionization energy (15.68 eV) relative with other noble gases with less atomic mass. In the plasma, the temperature can reach 10,000 K and the high temperature of the ICP ensures almost complete decomposition of the sample into its constituent atoms, and the ionization conditions within ICP result in highly efficient ionization of the most elements in the periodic table.

The sample is typically introduced into the ICP plasma as an aerosol, either by aspirating a liquid or dissolved solid sample into a nebulizer or using a laser to directly convert solid samples into an aerosol. Once the sample aerosol is introduced into the ICP torch, it is completely desolvated and the elements in the aerosol are converted first into gaseous atoms and then ionized towards the end of the plasma. Once the elements in the sample are converted into ions, they are then brought into the mass spectrometer via the interface cones. The interface region in the ICP-MS transmits the ions traveling in the argon sample stream at atmospheric pressure (1-2 torr) into the low pressure region of the mass spectrometer (<1 x 10-5 torr). This is done through the intermediate vacuum region created by the two interface cones; the sampler and the skimmer. The

sampler and skimmer cones are metal disks with a small hole (~1mm) in the center. The purpose of these cones is to sample the center portion of the ion beam coming from the ICP torch (Jarvis et al. 1992) (Figure 1.7).

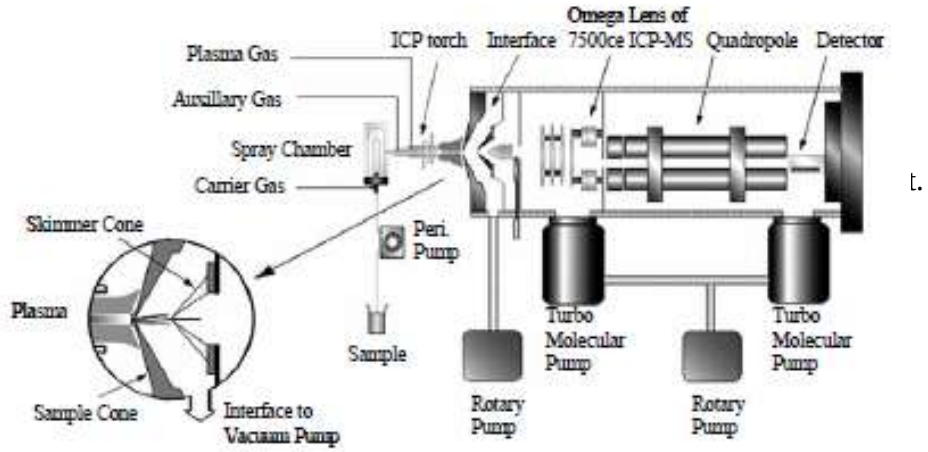

Figure 1.7 Schematic illustration of an ICP-MS instrument with quadrupole analyzer (Source Jarvis et al, 1992).

In the mass spectrometer part, firstly ions are removed from the plasma by a pumped extraction system. An ion beam is produced and focused further into the actual unit. There are several different types of mass analyzers which can be employed to separate isotopes based on their mass to charge ratio. Quadrupole analyzers are compact and easy to use but offer lower resolution when dealing with ions of the same mass to charge (m/z) ratio. Double focusing sector analyzers offer better resolution but are larger and have higher capital cost.

The quadrupole mass filter is made up of four metal rods aligned in a parallel diamond pattern. In this type, 4 rods (approximately 1 cm in diameter and 15-20 cm long) are arranged. In a quadrupole mass filter, alternating AC and DC voltages are applied to opposite pairs of the rods. These voltages are then rapidly switched along with an RF-field. The result is that an electrostatic filter is established that only allows ions of a single mass-to-charge ratio (m/z) pass through the rods to the detector at a given instant in time. So, the quadrupole mass filter is really a sequential filter, with the settings being change for each specific m/z at a time. However, the voltages on the rods can be switched at a very rapid rate. Many combinations of voltages are chosen which allows an array of different m/z ratio ions to be detected.

The most common type of ion detector found in an ICP-MS system is the channeltron electron multiplier. This cone or horn shaped tube has a high voltage applied to it opposite in charge to that of the ions being detected. Ions leaving the quadrupole are attracted to the interior cone surface. When they strike the surface additional secondary electrons are emitted which move farther into the tube emitting additional secondary electrons. As the process continues even more electrons are formed, resulting in as many as 108 electrons at the other end of the tube after one ion strikes at the entrance of the cone (Jarvis et al. 1992).

1.5. Fourier Transform Infrared (FTIR) Spectrometry

The region starts from 4000 cm^{-1} and ends at 400 cm^{-1} in the electromagnetic spectrum assigns the middle infrared region. Infrared radiation is not sufficient to cause the transitions between the electronic states. The vibrational levels and infrared spectra are generated by the characteristic twisting, bending, rotating and vibrational motions of atoms in a molecule. All of the motions can be described in terms of two types of molecular vibrations. One type of vibration, a stretch, produces a change of bond length. A stretch is a rhythmic movement along the line between the atoms so that the interatomic distance is either increasing or decreasing. The second type of vibration, a bend, results in a change in bond angle. These are also called scissoring, rocking or wigwag motions. Each of these two main types of vibration can have variations. A stretch can be symmetric or asymmetric (Figure 1.8).

In a Fourier Transform Infrared Spectrometer, a continuum source of light is used to produce light over a broad range of infrared wavelengths. Light coming from this continuum source is split into two paths using a half-silvered mirror; this light is then reflected from two mirrors back onto the beam splitter, where it is recombined. Because the path that one beam travels is a fixed length and the other is constantly changing as its mirror moves, the signal which exits the interferometer is the result of these two beams “interfering” with each other. The resulting signal is called an interferogram which has the unique property that every data point (a function of the moving mirror position) which makes up the signal has information about every infrared frequency which comes from the source.

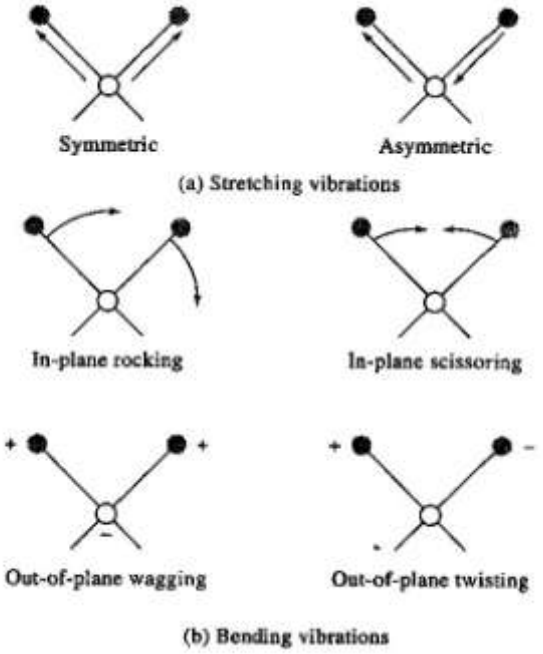

Figure 1.8. Types of molecular vibrations. + indicates motion from the page toward the reader; - indicates the motion away from the reader. (Source: Skoog et al. 1998)

In infrared instruments, Nernst glower, globar, tungsten filament, mercury arc or CO_2 laser are used as a source. Due to the heat property of sources, the detectors should be resistant to the heat. Thermocouples, bolometer, photoconducting tubes or pyroelectrics are generally used detectors in infrared spectrometers and also the mostly used one as an interferometer is the Michelson interferometer. Figure 1.9 shows the optical diagram of an infrared instrument.

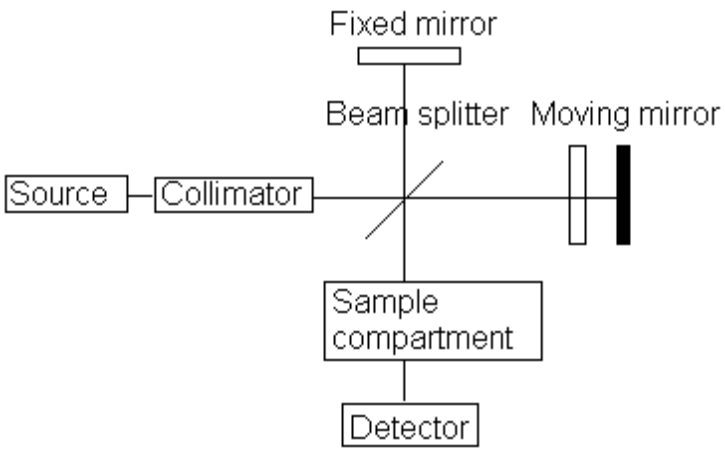

Figure 1.9. Optical diagram of Fourier transform infrared spectrometer (Source: wikipedia.com 2011)

The analysis of aqueous solutions is complicated by the solubility of the NaCl cell window in water. One way to obtaining infrared spectra on aqueous solutions is to use attenuated total reflectance (ATR) instead of transmission. Figure 1.10 shows a diagram of a typical ATR sampler, consisting of an IR-transparent crystal of high refractive index, such as ZnSe, surrounded by a sample of lower-refractive index. Radiation from the source enters the ATR crystal, where it goes through a series of total internal reflections before exiting the crystal. During each reflection, the radiation penetrates into the sample to a depth of a few microns. The result is a selective attenuation of the radiation at those wavelengths at which the sample absorbs.

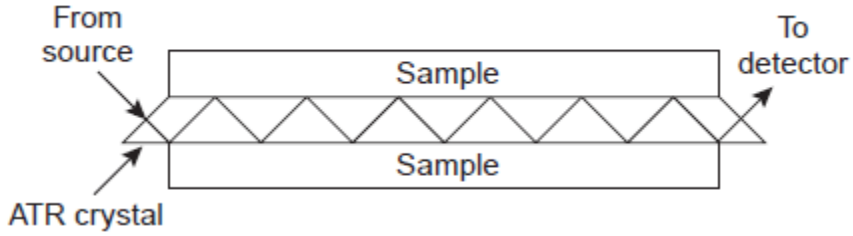

Figure 1.10. Attenuated total reflectance (ATR) cell used in infrared spectroscopy. (Source: Harvey 2000)

Solid samples also can be analyzed by means of reflectance. The ATR sampler described for the analysis of aqueous solutions can be used for the analysis of solid samples, provided that the solid can be brought into contact with the ATR crystal.

1.6. Fluorescence Spectrometry

Photoluminescence is divided into two categories: fluorescence and phosphorescence. Absorption of an ultraviolet or visible photon promotes a valence electron from its ground state to an excited state with conservation of the electron's spin. For example, a pair of electrons occupying the same electronic ground state has opposite spins (Figure 1.11) and are said to be in a singlet spin state. Absorbing a photon promotes one of the electrons to a singlet excited state. Emission of a photon from a singlet excited state to a singlet ground state is called fluorescence. Fluorescence decays rapidly after the excitation source is removed. In some cases an electron in a singlet excited state is transformed to a triplet excited state in which its spin is no longer

paired with that of the ground state. Emission between a triplet excited state and a singlet ground state is called phosphorescence (Skoog et al. 1998).

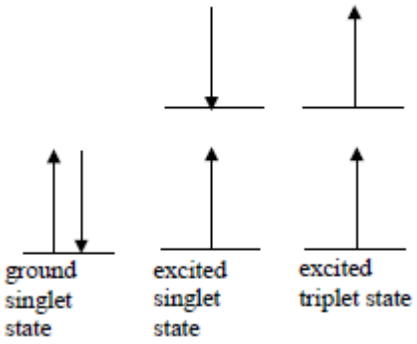

Figure 1.11. Singlet/Triplet excited states

The process which occurs between the absorption and emission of light are usually illustrated by a Jabloński diagram (Figure 1.12.). Jabloński diagram typically explains the fluorescence and phosphorescence in terms of energy level. Three non-radiational processes are also explained here. These are internal conversion (IC), intersystem crossing (ISC), and vibrational relaxation. Internal conversion is the transition between energy states of the same spin state. The transition between the different spin states called as intersystem crossing. The last non-radiational process vibrational relaxation occurs in a molecule which is in excited vibrational and rotational sates.

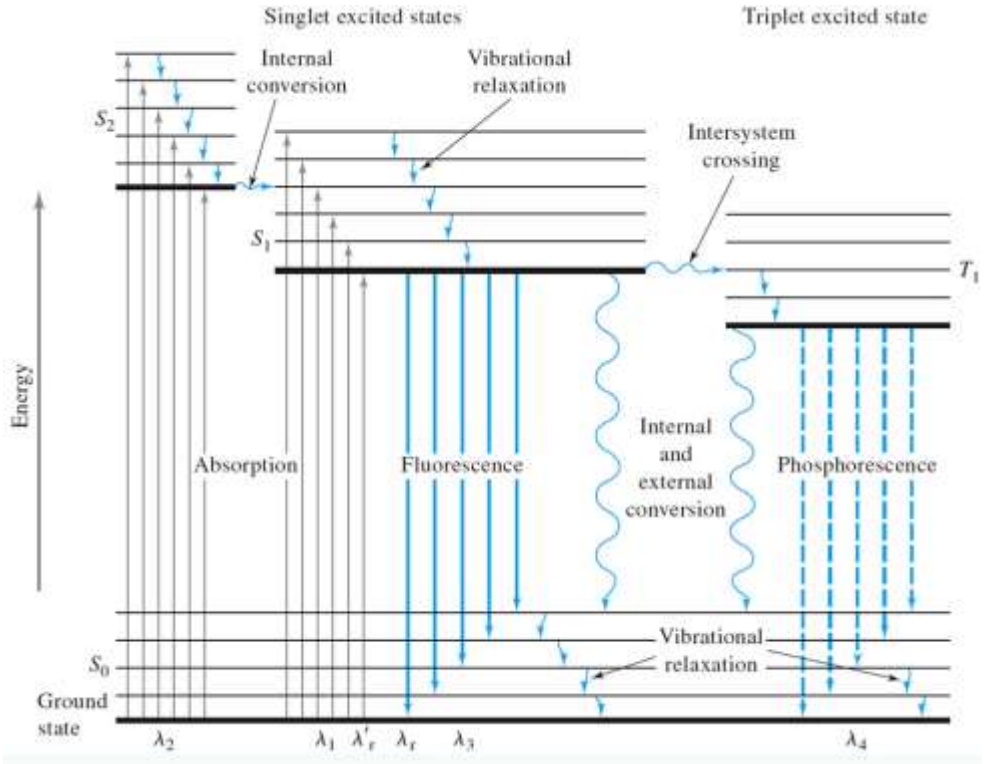

Figure 1.12. Partial energy diagram for a photoluminescent system
(Source: Skoog et al. 2002)

The basic design of instrumentation for monitoring molecular fluorescence and molecular phosphorescence is similar to that found for other spectroscopies. A typical instrumental block diagram for molecular fluorescence is shown in Figure 1.13. In contrast to instruments for absorption spectroscopy, the optical paths for the source and detector are usually positioned at an angle of 90°. Two basic instrumental designs are used for measuring molecular fluorescence. In a fluorometer, the excitation and emission wavelengths are selected with absorption or interference filters. The excitation source for a fluorometer is usually a low pressure mercury vapor lamp that provides intense emission lines distributed throughout the ultraviolet and visible region. When a monochromator is used to select the excitation and emission wavelengths, the instrument is called a spectrofluorometer. With a monochromator, the excitation source is usually a high-pressure Xe arc lamp, which has a continuum emission spectrum. Either instrumental design is appropriate for quantitative work, although only a spectrofluorometer can be used to record an excitation or emission spectrum. Generally, the fluorescence signal is a low intensity, therefore photomultiplier tubes are the most

common transducer in fluorescence instruments. Diode-array and charge transfer detectors have been also proposed for spectrofluorometers (Skoog et al. 1998).

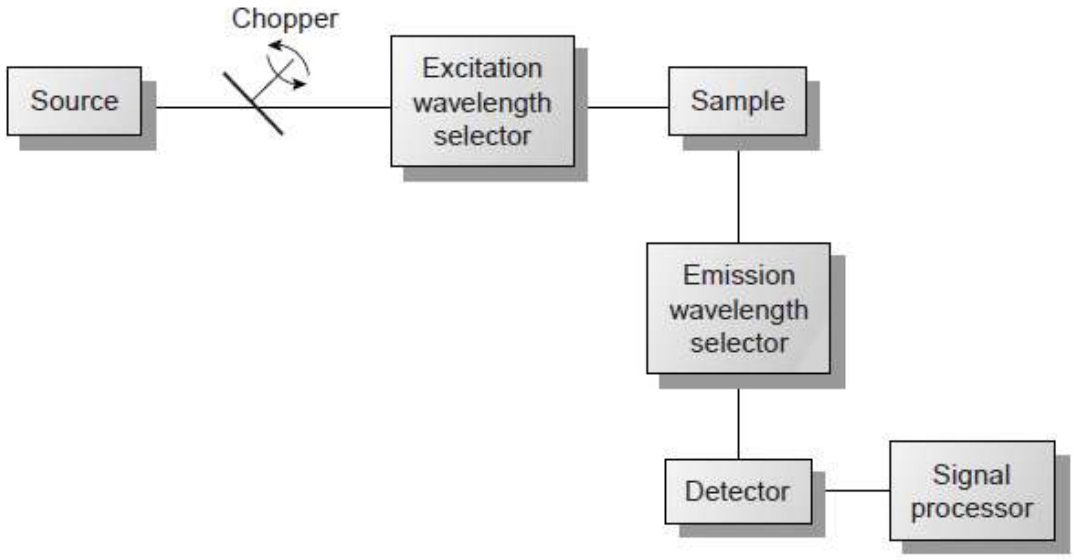

Figure 1.13. Block diagram for molecular fluorescence spectrometer.

1.7. Aim of This Work

Honey is considered as natural and healthy product that has a highly concentrated solution of a complex mixture of carbohydrates. The quality of a honey depends on its chemical composition and botanical origin. Therefore, classification of honey according to geographical origin is of great interest. In this study, it is aimed to develop classification models of honey produced in Turkey based on geographical origin via atomic and molecular spectrometry. The honey samples were taken from different regions of Turkey and then they were scanned with molecular spectrometric methods. Afterwards, unsupervised and supervised methods were used for the classification of honey samples. Finally, the metal content of honey samples was investigated through ICP-MS measurements and the same methods were again applied for classification based on geographical origin.

CHAPTER 2

EXPERIMENTAL

2.1. Chemicals and Reagents

Nitric acid (Merck) and hydrogen peroxide (Riedel de Haën) were of analytical grade. Calibration standards were prepared from ICP multielement standard (Merck). Glassware and plastic ware used in digestion analysis were cleaned by soaking in 10 % (v/v) nitric acid and rinsed with distilled water prior to use. 18 MΩ ultra pure water was used throughout studies.

2.2. Samples

Totally 145 honey samples which were from different regions of Turkey were collected from locally bee keepers, Altıparmak Gıda San. ve Tic. Koll. Şti and Doğa Arıcılık. The samples stored in dark and at room temperature until they were analyzed. During the examination of classification methods (PCA, HCA and SIMCA) for honey samples, all of the samples were not used. Since the collected honey samples were received in different geographic regions of the TURKEY from east to west, it was worth to investigate if there is any clustering of the samples based on their regions. The collected samples were scanned with middle infrared spectrometer equipped with attenuated total reflectance (MIDIR-ATR), and fluorescence spectrophotometer (synchronous fluorescence mode and 3D excitation emission mode). In addition, the collected samples were digested and the metal concentrations were determined with inductively coupled plasma mass spectrometer (ICP-MS). The sampling regions, collected numbers of samples and the labels given are listed in Table 2.1.

Table 2.1. The sampling region and number of honey samples

Sampling region	Number of samples	Label given
İzmir	35	I
Denizli	6	DN
Aydın	6	AY
Datça	6	D
Marmaris	11	M
Antalya	1	AN
Adana	3	AD
Kayseri	9	K
Malatya	2	ML
Sivas	3	S
Kırşehir	3	KI
Afyon	1	AF
Eskişehir	1	E
Balıkesir	1	BL
Bozcaada	1	BO
Trakya	12	T
Safranbolu	1	SF
Giresun	1	G
Diyarbakır	11	DY
Muş	10	MU
Bingöl	2	BN
Bitlis	2	BT
Urfa	7	U
Yüksekova	7	Y
Ardahan	1	AR
Kars	1	KR

2.3. Measurements Using Fourier Transform Infrared-Attenuated Total Reflectance (FTIR-ATR)

Fourier Transform infrared spectra of the honey samples were collected at room temperature on Perkin Elmer Spectrum 100 FTIR Spectrometer (Waltham, MA, USA) between 630 and 4000 cm^{-1}. Since honey is a viscous liquid attenuated total reflectance with diamond was used for measurements. The spectra were saved as log 1/R and the resolution was 8 cm^{-1}. Background spectrum was obtained empty and dry ATR cell. Before and after each sample analyses background was collected to reduce the contaminations that would come from the ATR crystal. ATR crystal was cleaned with pure ethanol and allowed to dry.

2.4. Measurements Using Fluorescence Spectrophotometer

The honey samples were diluted with 1/5 (w/v) ratio with distilled water as the honey is viscous liquid. The fluorescence spectra of diluted samples were obtained using a Varian Cary Eclipse spectrophotometer (Varian, Inc. Hansen Way, Palo Alto, CA). The instrument is equipped with a xenon flash lamp and spectra were collected in 2 modes: excitation emission 3D scan mode (EEF) and synchronous fluorescence (SF) mode. The EEF spectra were collected between 400 and 700 nm emission wavelength by exciting the samples with a wavelength increment of 10 nm ($\Delta\lambda$) from 330 to 380 nm. The slit widths were 10 nm for excitation and 5 nm for emission monochromators. The SF spectra of the samples were recorded between 250 and 600 nm with a $\Delta\lambda$ of 15 nm. The slit widths of excitation and emission monochromators were set to 5 nm in SF modes.

2.5. Measurements Using Inductively Coupled Plasma Mass Spectrometer

An Agilent 7500ce quadrupole (Tokyo, Japan) inductively coupled plasma mass spectrometer (ICP-MS) with a high solid nebulizer was used to determine the concentration of Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba, Pb in honey samples. The operating conditions of the instrument are given in Table 2.2.

Table 2.2. ICP-MS operating conditions.

Forward power	1500 W
Reflected power	1W
Coolant gas flow rate	15 L min^{-1}
Auxilary flow rate	0.90 L min^{-1}
Sample uptake time	25 sec
Integration time	100 msec

Before metal analysis the honey samples were needed to transfer to the aqueous media by digestion. Two different types of digestion procedures were tested with brown bread certified reference material (CRM,BCR 191) in order to decide the digestion procedure before ICP-MS analysis: microwave digestion and wet digestion without applying heat.

A CEM Mars 5 Plus (Matthews, North Carolina, USA) microwave digestion system was used. Approximately 0.5 g of honey was placed in a PTFE vessel and 6 mL of concentrated HNO_3 (65%,m/v) and 2 mL of H_2O_2 (30%, m/v) were added. The vessels were capped, tightened and placed in the rotor of the microwave oven. The digestion was carried out with the following digestion program: 1000 W/10 min up to 170 °C; then 1000 W/15 min at 170 °C. Ventilation was performed for 20 min after the end of the second step. Finally the vessels were cooled, opened and the contents quantitatively transferred to falcon tubes then completed to 25 mL with ultrapure water. Finally the samples were diluted in 1/10 ratio with ultrapure water for ICP-MS analysis.

For wet digestion without applying heat, just about 0.5 g of honey samples was weighed and the mixture of 6 mL HNO_3 and 2 mL H_2O_2 was added to sample. After that the mixture was homogenized with shaking and then waited one night for digestion

by itself. The mixture was completed to 25 mL with ultrapure water. Finally the solution is diluted in 1/10 ratio before ICP-MS analysis.

2.6. Statistical Classification Studies

The multivariate unsupervised classification analyses (principal component analysis, PCA and hierarchical cluster analysis, HCA) were carried out by Minitab 15 (Minitab Inc.) and the supervised classification analysis (SIMCA) were performed by SIMCA-P v.10.5 (Umetrics, Umea, Sweden). Data obtained from analyses were put in a matrix with the rows relating to the honey varieties and geographical origins for and the columns relating to the individual absorbance or intensity values (k variables). Prior to multivariate analysis, the data were pre-processed by the standard procedure. This procedure includes mean-centering (the mean value of each variable is calculated and subtracted from the data) and normalization.

The models were developed for classification of honey samples according to geographical origin. PCA results were illustrated on the plot of the firs component vs the second component and meanwhile HCA results were shown on dendrograms. In SIMCA models on principal components, the distance from the model for class 1 was plotted against that from model 2. The discrimination of each class was shown in the Cooman's plots of the class models.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Classification Studies with Molecular Spectrometric Data

3.1.1. Results of Unsupervised Methods

It is worth to investigate the clustering of the collected honey samples based on their regions as they are received in different geographic regions of TURKEY from east to west. For this purpose, four different scenarios were tested and they were decided according to their sample size and the regions. The first group was based on the samples from Trakya, Kayseri, Malatya and Sivas and the second one was established on the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. The third group was formed with the honey samples collected from İzmir, Datça and Marmaris and finally, the fourth group was figured out with samples from Trakya, Muş Bingöl and Bitlis. The sample names are coded according to city from where they are collected. The sample codes are illustrated in Table 2.1.

The mentioned four groups were scanned with the spectroscopic methods, such as FTIR, synchronous fluorescence and excitation-emission fluorescence and then analyzed with PCA and HCA.

3.1.1.1 FTIR-ATR Results

Fourier Transform infrared spectrometer is used for classifying the honey samples based on their spectral features. The spectrometer is equipped with attenuated total reflectance (FTIR-ATR) accessory that carries a dimond-ZnSe crystal plate. The samples are scanned between 4000 and 600 cm^{-1} and the collected spectra are shown in Figure 3.1.

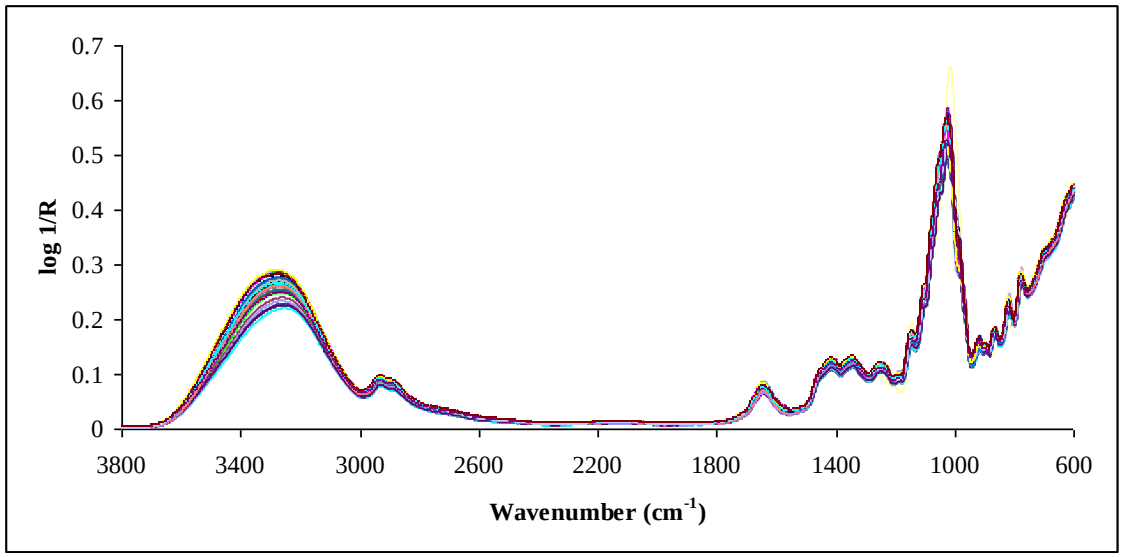

Figure 3.1. The FTIR-ATR spectra of honey samples.

The absorption region 750 and 1500 cm^{-1} in Figure 3.1 corresponds to monosaccharides of honey such as fructose and glucose and disaccharides such as sucrose. In the region 750–900 cm^{-1}, the signals are matched with the anomeric region and are characteristic of the saccharide configuration. The wave numbers of C–O and C–C stretching modes are in the region of 900-1150 cm^{-1} and around 1200–1480 cm^{-1} are due to the bending modes of O–C–H, C–C–H and C–O–H. The bands at 2800–3000 cm^{-1} are corresponding to the stretching mode of C–H groups and O–H of carboxylic groups. Bands around 3600 and 1600 cm^{-1} are related to O–H stretching mode and to residual water (Bertelli et al. 2007).

After scanning the honey samples with FTIR-ATR spectrometer, the collected spectra were used for PCA and HCA by Minitab software. As it is known PCA is an unsupervised classification method and is generally used to obtain a lower dimensional graphical representation which describes the maximum variation in a data set. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible (Beebe et al. 1998).

The first combination is made with two groups which consist of honey samples from Trakya, Kayseri, Malatya and Sivas. The score plot of the first component versus the second component is demonstrated in Figure 3.2. The first and second principal components explained 73 % of the variation of the data.

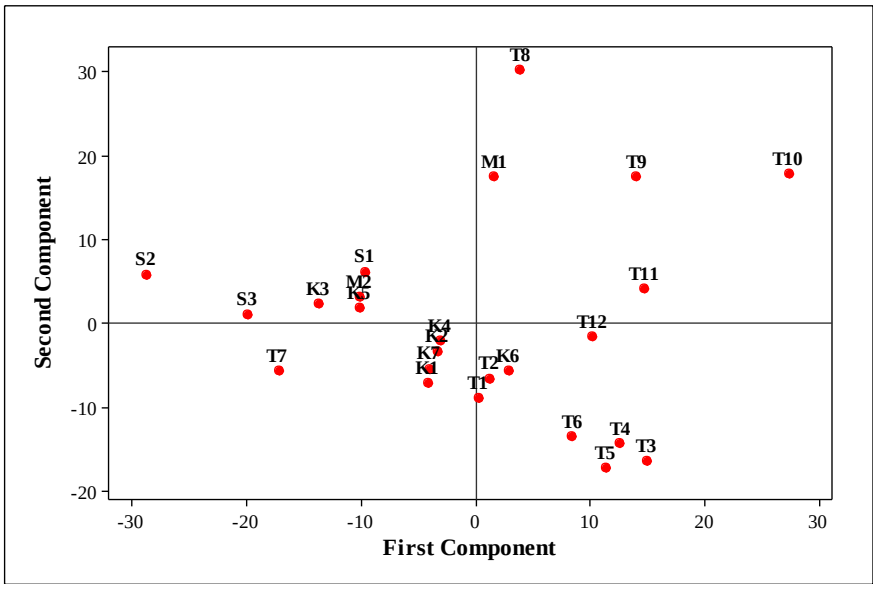

Figure 3.2. The score plot of the first component versus the second component for honey samples from Trakya, Kayseri, Malatya and Sivas using FTIR spectra.

When the score plot of the samples is examined, it is seen that all of the Trakya samples except T7 were characterized with positive value of the first component; nevertheless, all of the Kayseri, Malatya and Sivas samples, except K6 and M1, were classified on the negative side of the first component. Since Trakya is located on the northwest and Kayseri, Sivas and Malatya are located nearly central of the Turkey, the samples from these regions are classified separately.

Following the PCA analysis, another unsupervised classification method which is also commonly used to demonstrate the similarities between the samples is applied to the same spectroscopic data set. This method is called hierarchical cluster analysis (HCA) and it generates rectangular tables of variables and objects that are called dendrograms. The aim of HCA is to find out the grouping of the objects (samples) and variables (features) in addition to similarities possibly, in terms of a hierarchy of embedded groups. Briefly, two main steps are repeated. The first step is to investigate the distance matrix for the two closest objects (or variables) whereas the second one is used to consider this pair of objects as a single individual and to recompute the distances between this new element and the rest of the objects. (Devillers et al. 2002). As it can be directly applied to raw data set, it is also possible to apply HCA to the PCA score vectors and loading vectors. In fact, when the original data contains too many variables (e.g. spectroscopic data contains several absorbance values at corresponding wavelengths or wave numbers), it is better to preprocess the data with PCA so that the

dimensionality of the original data (either normalized or not) can be reduced to a few most important PC's. After PCA analysis, the resulting significant score and loading vectors can be used to cluster the objects and variables, respectively. If there are only a few original variables in the data set, HCA can be directly applied to the original data. In the present study, the FTIR spectra of the honey samples have contained around 1800 individual wave numbers and therefore, the HCA analysis has to be applied to PCA score and loading vectors. Figure 3.3 depicts the dendrogram of honey samples from Trakya, Kayseri, Malatya and Sivas obtained with HCA. The first three PC score vectors that are accounted 85 % of the total variability in the original spectral data are used in the distance calculation.

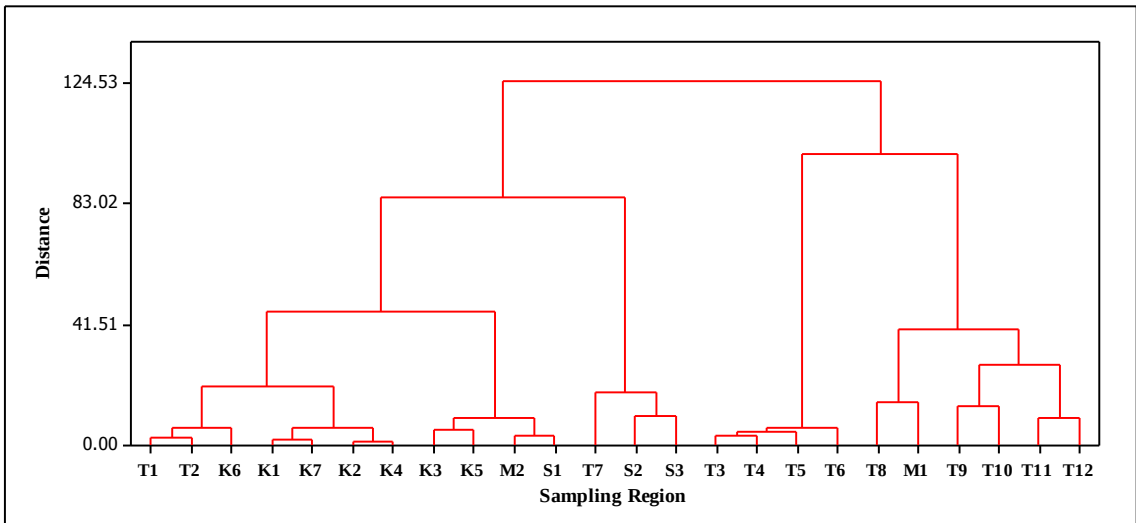

Figure 3.3. Dendrogram for honey samples from Trakya, Kayseri, Malatya and Sivas using FTIR spectra.

Altough the samples in dendrogram were not separated in two main classes according to their sampling regions, the clusters contained the samples from the same city; for example, Trakya samples were clustered together and Kayseri samples were classified separately. Dendrogram also shows the closeness of the samples.

The next combination was made up with the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. It is important to investigate the classification since these regions are very close to each other on the map of the Turkey. The score plot of the first component versus the second component is presented in Figure 3.4. The two PC's explain approximately 71 % of the total variance of the data.

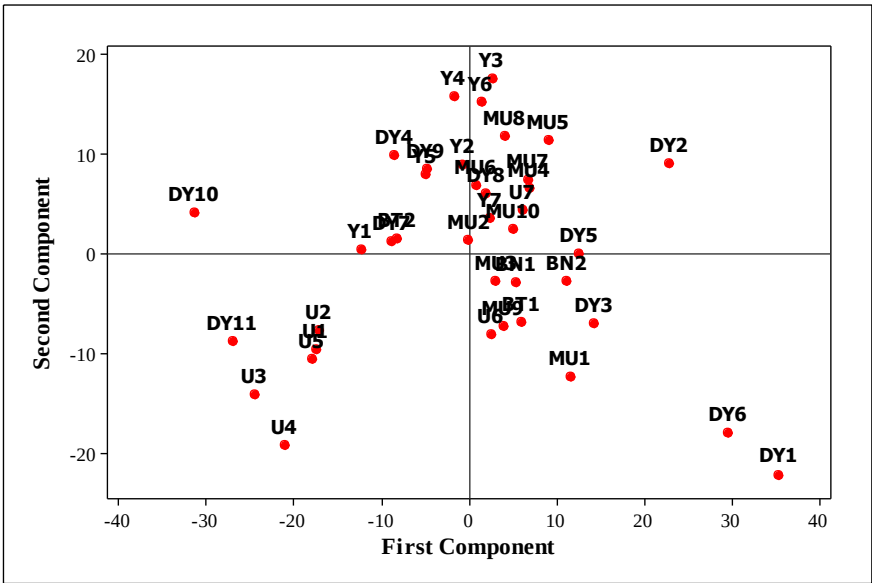

Figure 3.4. The score plot of the first component versus the second component for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using FTIR spectra.

As can be seen from the Figure 3.4, the samples from Urfa were classified in the negative region of both components whereas Yüksekova samples were characterized with the positive side of the second component and Muş samples with the positive side of the first component. Nonetheless, the honey samples from Diyarbakır were not classified and scattered on four regions.

In order to see the closeness of the honey samples, HCA dendrogram is applied by using five principal components which are explained 92 % variation of the original data and the dendogram is depicted in Figure 3.5. It is clearly seen that there are two main classes, and one of the main cluster which is located on the right side contains Urfa samples. Other samples are not clustered according to sampling city and it can be concluded that FTIR spectra is not sufficient enough to classify the samples in which the regions are very close.

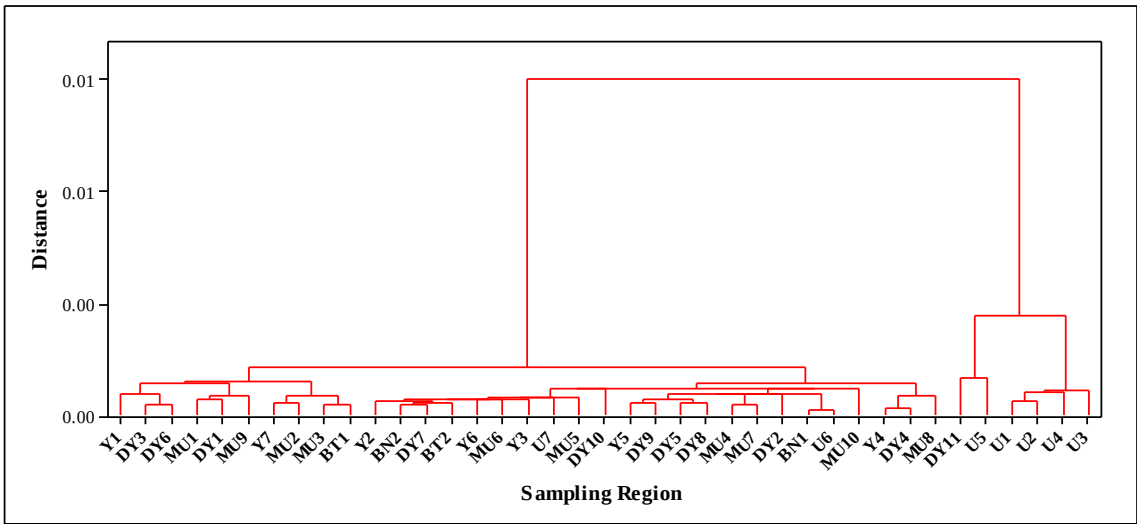

Figure 3.5. Dendrogram for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using FTIR spectra.

The following trial for geographical grouping was again performed with the honey samples of close regions which were İzmir, Datça, Marmaris and one sample from Antalya. PCA result is illustrated in Figure 3.6. The first and the second components cover 77 % of the total variance of the data. PCA result explains that there is no grouping of the honey samples according to their geographical origins as the samples from all regions are overlapped on the PC graph.

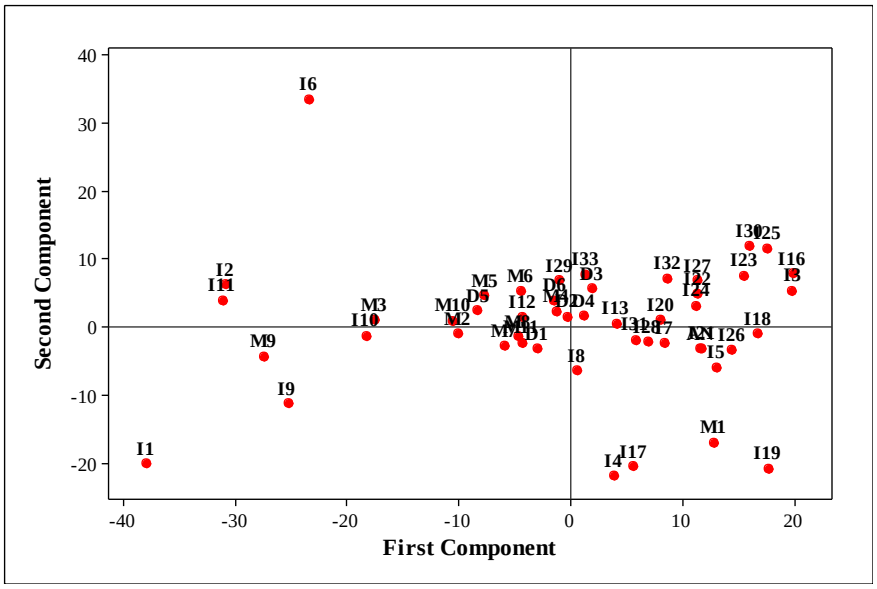

Figure 3.6. The score plot of the first component versus the second component for honey samples from İzmir, Datça, Marmaris and Antalya using FTIR spectra.

Furthermore, HCA also demonstrates that similar samples are clustered in the same region (Figure 3.7). The number of PCs for HCA is 3 which explained about 85 % of the variation in the data.

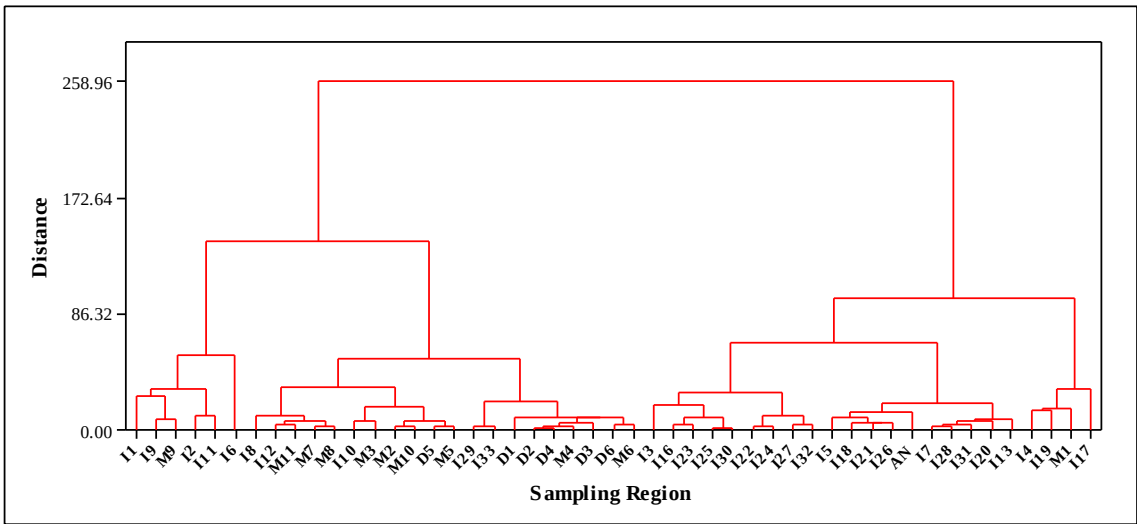

Figure 3.7. Dendrogram for honey samples from İzmir, Datça, Marmaris and Antalya using FTIR spectra.

As it can be concluded from Figure 3.7, most of the honey samples from İzmir are clustered at right cluster and the others at the left cluster. Although the samples were from very close neighbor regions they were clustered mainly with some exceptions.

The last combination was constructed with honey samples from Trakya, Muş, Bingöl and Bitlis. The result of classification analysis with PCA is shown in Figure 3.8 and 73 % of the total variance of the data is explained with the first two components.

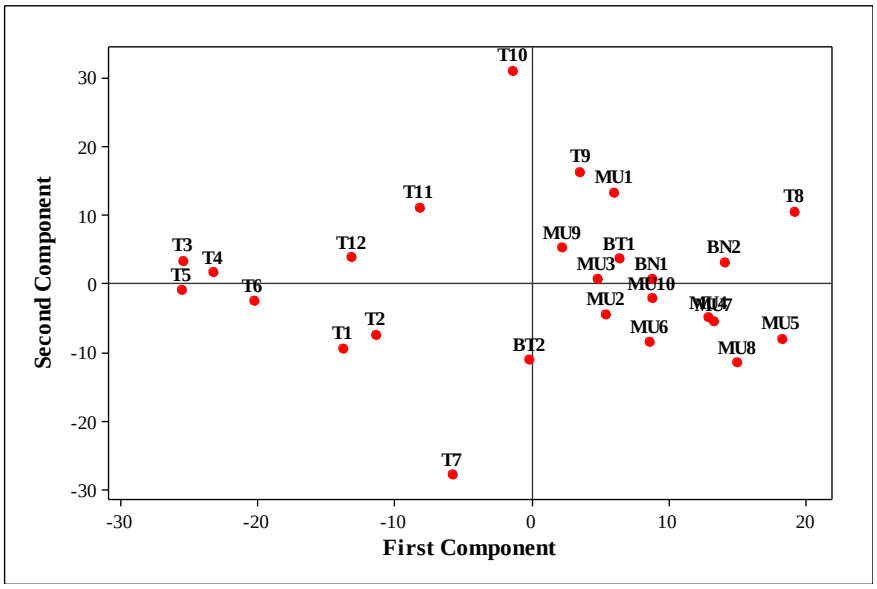

Figure 3.8. The score plot of the first component versus the second component for honey samples from Trakya, Muş, Bingöl and Bitlis using FTIR spectra.

The score plot clearly represents that all of the honey samples from Trakya are separated from the others except the samples labeled T8 and T9. Since Bitlis, Muş and Bingöl are neighbor cities, their samples are classified positive part of the first component. In order to investigate the similarities between the honey samples, HCA was performed with six principal components. The dendrogram is illustrated in Figure 3.9 and the same conclusions can be done. As, Trakya (North West) samples were clustered at the right side, the honey samples from east are clustered together. The samples labeled T8, T9 and T10 are clustered in the second main class.

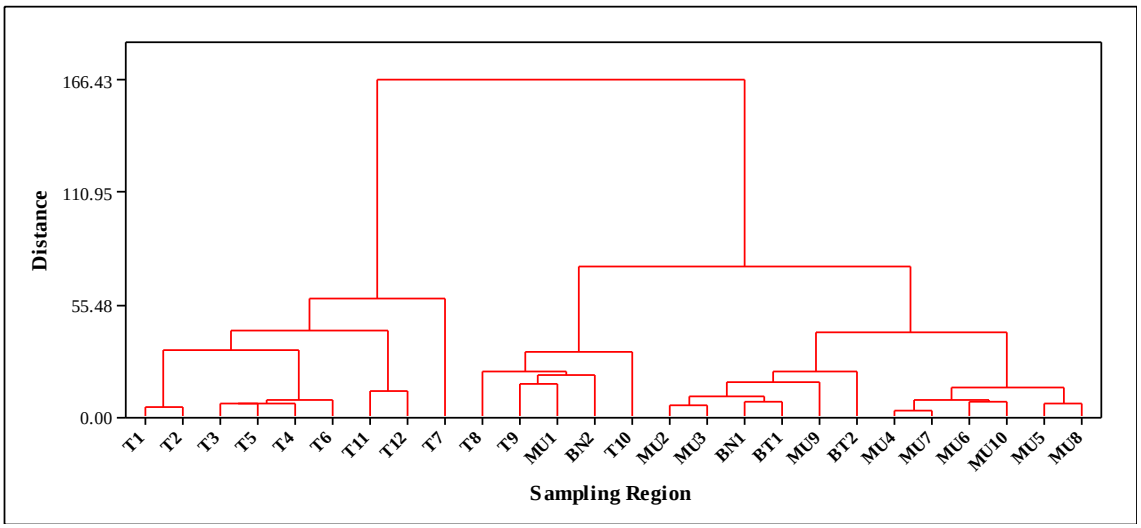

Figure 3.9. Dendrogram for honey samples from Trakya, Muş, Bingöl and Bitlis using FTIR spectra.

3.1.1.2. Synchronous Fluorescence Results

Synchronous fluorescence (SF) spectrum was recorded by scanning simultaneously both the excitation and emission monochromators at the same speed with a fixed wavelength interval (Δλ) between the excitation and emission wavelengths for each honey samples. SF spectra has the advantage of absorption as well as the emission characteristics of a given compound, so sharper spectral features representing different components of a multi component system can be observed.

In order to compare the classification results using different molecular spectrometric methods, the honey samples were scanned with fluorescence spectrometer using the synchronous mode between 250 and 600 nm with a Δλ of 15 nm. The synchronous fluorescence spectra of samples are represented in Figure 3.10.

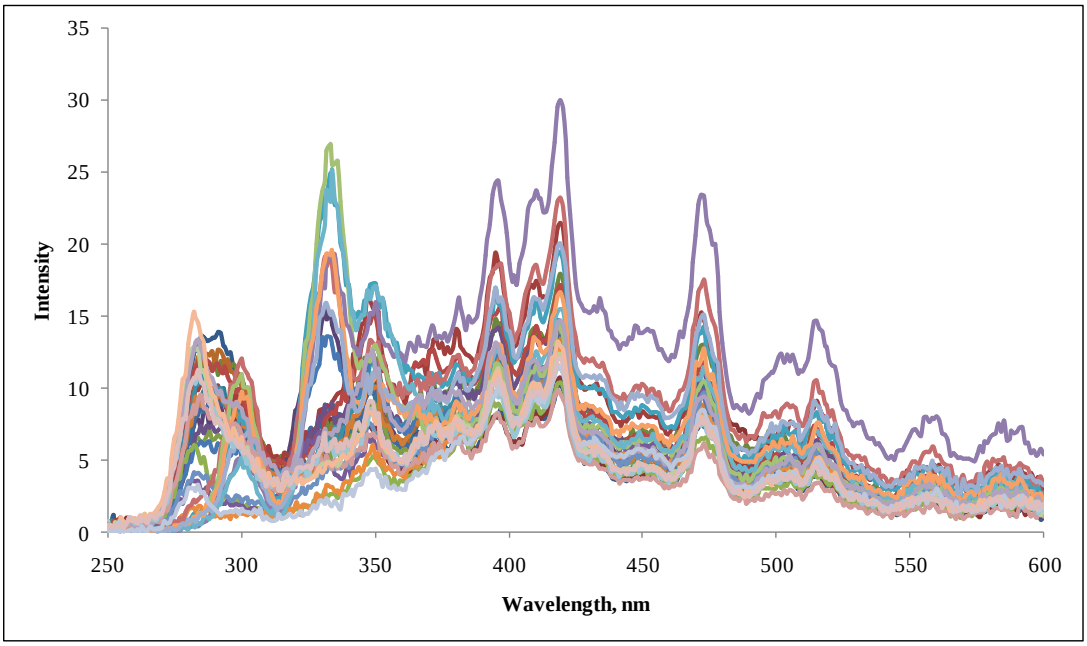

Figure 3.10. Synchronous fluorescence emission spectra of 61 honey samples between 250 and 600 nm (Δλ=15 nm).

The same combinations of the samples that are given in FTIR were used for PCA and PCA-HCA analysis in order to compare the results. The first grouping had included Trakya, Kayseri, Malatya and Sivas honey samples. The PCA result using synchronous fluorescence is shown in Figure 3.11.

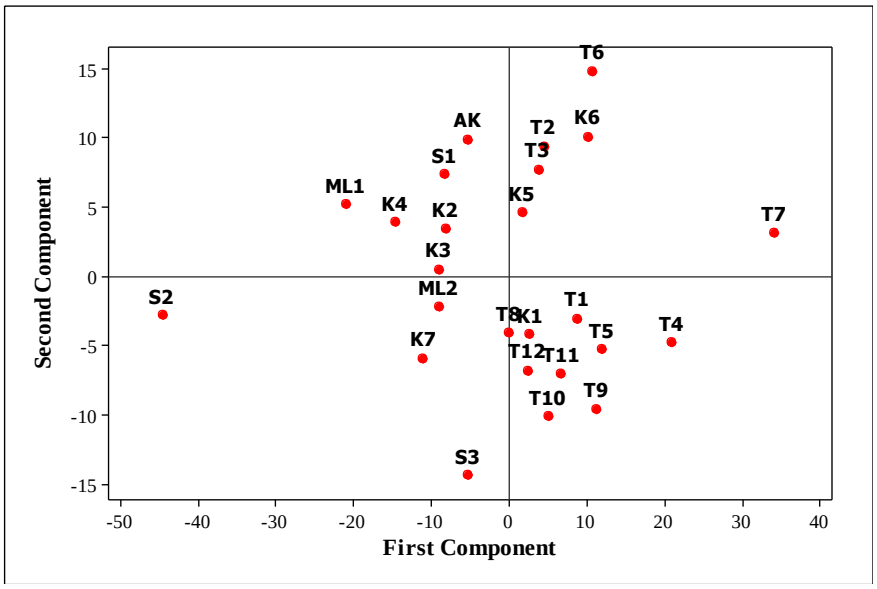

Figure 3.11. The score plot of the first component versus the second component for honey samples from Trakya, Kayseri, Malatya and Sivas using synchronous fluorescence spectra.

As seen from the plot, all Trakya samples were located on the positive side whereas many of Kayseri, Malatya and Sivas samples were placed on the negative side of the first component on score plot. It can also be observed that the sample labeled T7 which was from Trakya was located far away from the others. Hence, it can be concluded that the first component can characterize the honey samples based on the geographical origin using synchronous fluorescence spectra. The two principal components which were used to construct this score plot were able to cover about 80% the variation in the spectra of honey samples.

PCA-HCA results using six of the first principal components are depicted in Figure 3.12 and these three PC's explained 90% of the variability within the data set.

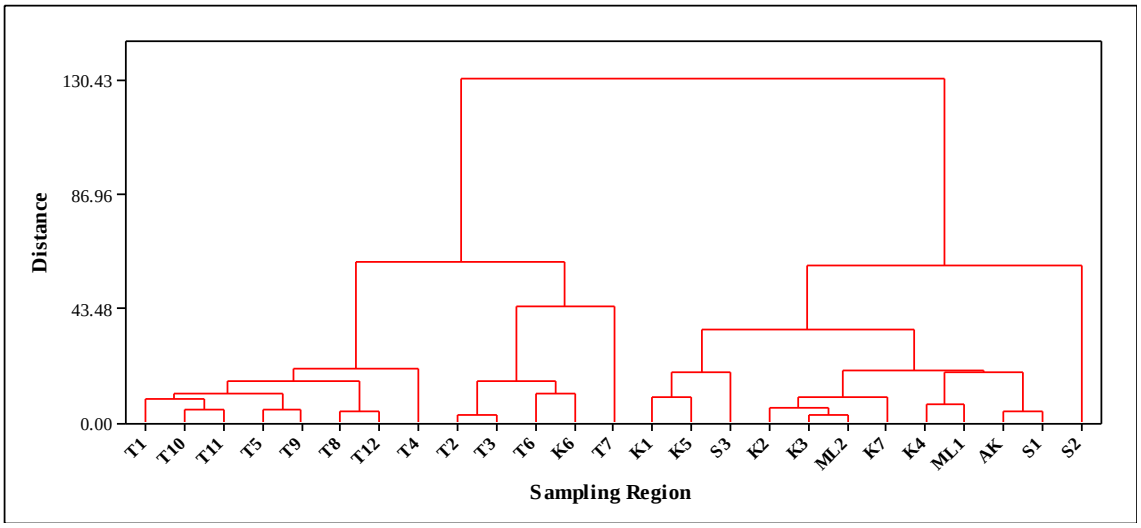

Figure 3.12. Dendrogram for honey samples from Trakya, Kayseri, Malatya and Sivas using synchronous fluorescence spectra.

As seen from Figure 3.12, the differentiation of honey samples was obviously performed according to sampling region; however, the samples labeled K6 and S2 were misclassified.

The second combination of the samples consists of Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. The score plot is given in Figure 3.13 and the resulted two-component graphs explain 88 % of total variance.

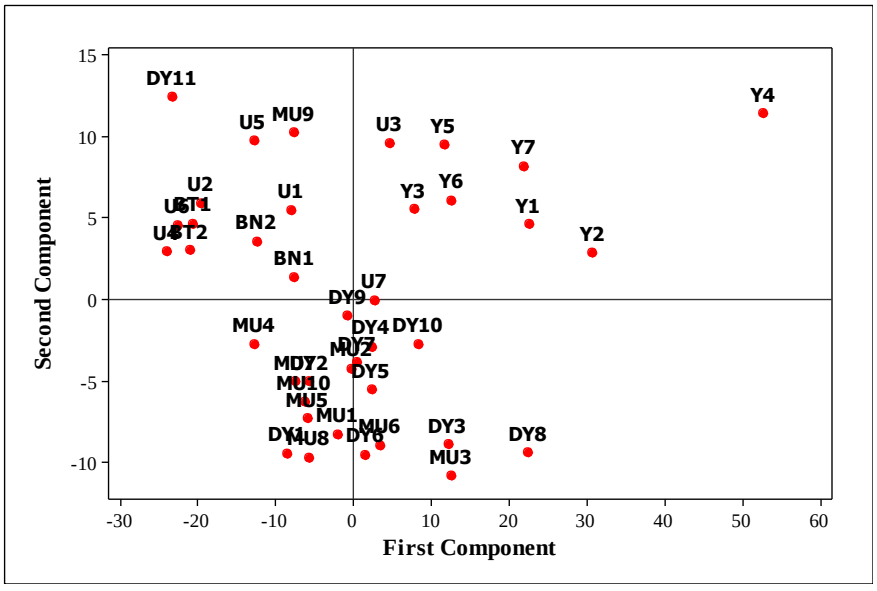

Figure 3.13. The score plot of the first component versus the second component for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using synchronous fluorescence spectra.

Although the sampling regions of the honey samples were very close to each other, classification of the samples was again achieved interestingly. For example, all Yüksekova samples were separated from the others and characterized on both positive sides of the principal components. In addition, all Urfa samples were located on the positive region of the second component. Nevertheless, the rest of the samples were overlapped on the plot.

Figure 3.14 represents the HCA dendrogram obtained with three principal components covering again 93 % of the variability in the data set. This dendrogram shows similarity with the result obtained by PCA and shows the closeness of the samples. As can be seen, most of the samples from cities formed small clusters with respect to each other and samples from Yüksekova were separated from others as observed in PCA plot.

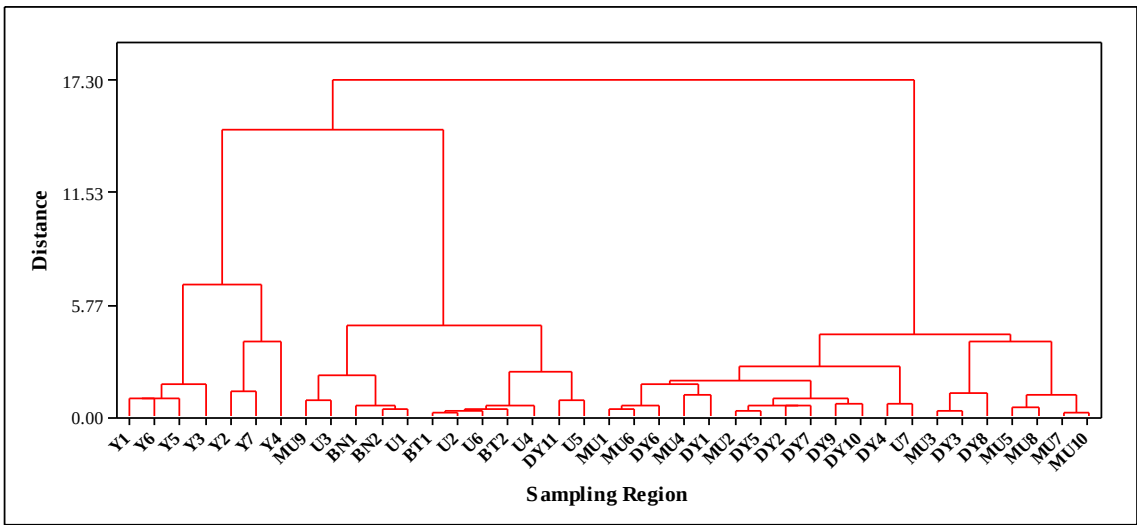

Figure 3.14. Dendrogram for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using synchronous fluorescence spectra.

Honey samples from İzmir, Datça, Marmaris and Antalya were analyzed and the PCA plot in Figure 3.15 was obtained. The plot showed that there was no clear differentiation between the classes, but most of Datça and Marmaris samples were placed on the positive part of the fist component

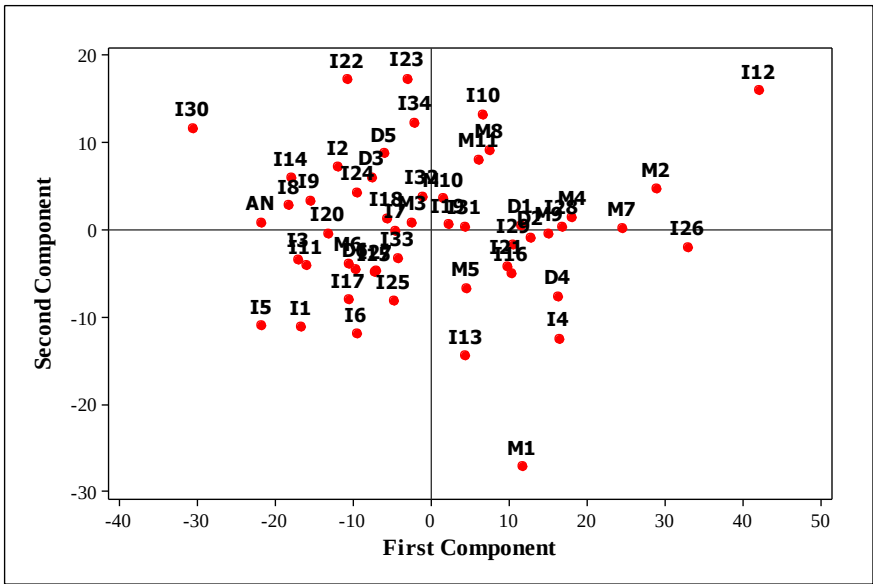

Figure 3.15. The score plot of the first component versus the second component for honey samples from İzmir, Datça, Marmaris and Antalya using synchronous fluorescence spectra.

In addition, the samples from İzmir were also scattered on the PC plot. PCA-HCA dendrogram shown in Figure 3.16 was generated with three PC's (90 % variation) to investigate the similarities between the samples. There were two main classes in dendrogram and most of the samples from Datça and Marmaris were clustered separately.

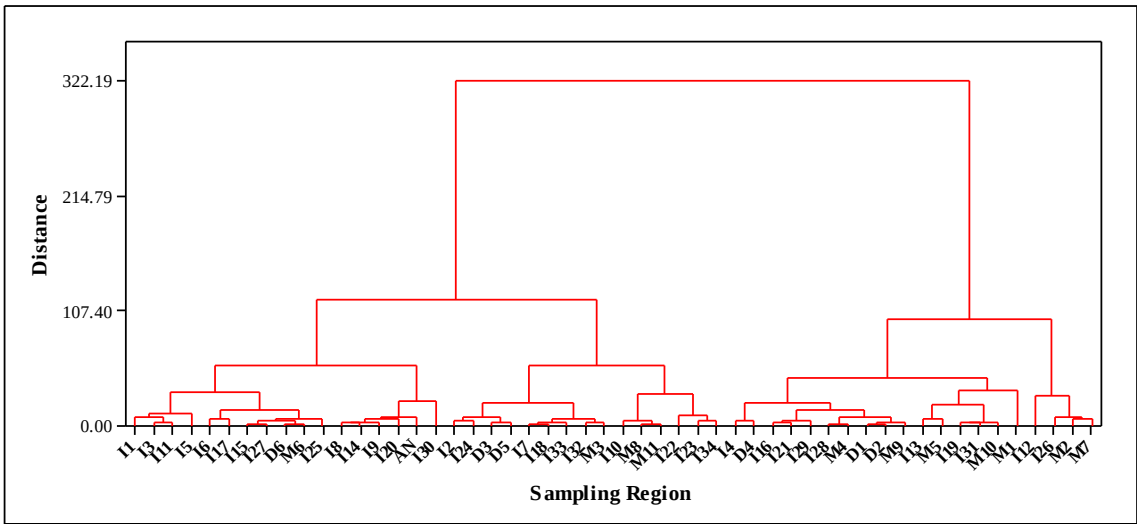

Figure 3.16. Dendrogram for honey samples from İzmir, Datça, Marmaris and Antalya using synchronous fluorescence spectra.

The last trial was performed with honey samples from Trakya, Muş, Bingöl and Bitlis and the resulting PC plot obtained with Minitab is presented in Figure 3.17. ın the

figure, obvious classification of the samples was observed even for all cities and gave better results compared to FTIR data. There were three different classes: the first one included Trakya samples and located on the positive region of the first component, the second class is constructed by samples from Bitlis and Bingöl and placed on left bottom of the plot; finally, the third class corresponded to the samples from Muş and it was on left upper region of the plot.

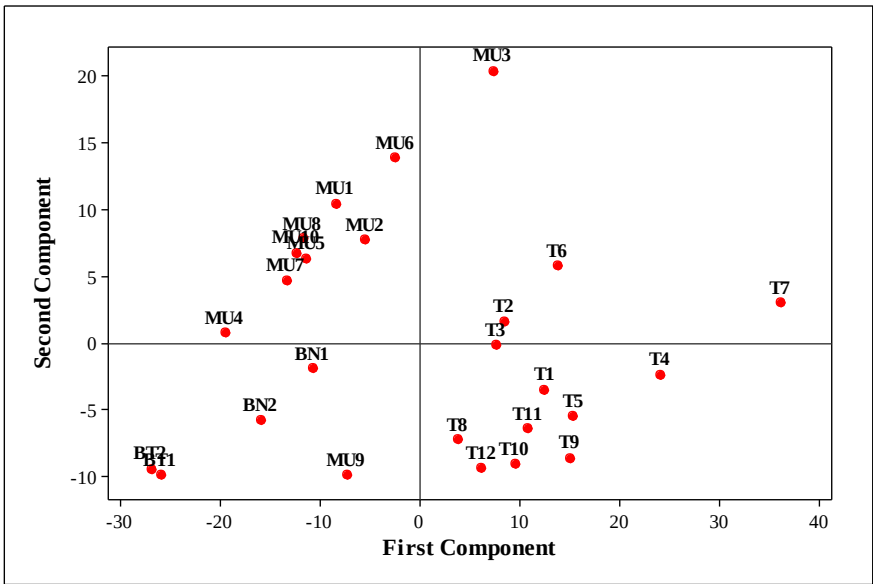

Figure 3.17. The score plot of the first component versus the second component for honey samples from Trakya, Muş, Bingöl and Bitlis using synchronous fluorescence spectra.

PCA-HCA results using three of the first principal components covering 93 % of the variability in the data set is demonstrated in Figure 3. 18. The dendrogram showed up the same results with the principal component analysis, however, only the sample coded as T7 was out of the clusters as indicated with FTIR results. In the present plot, even Muş samples constructed separate class from Bingöl and Bitlis samples. In conclusion, synchronous florescence data were quite successful in the classification of honey samples according to geographical origin even for very close neighbor samples.

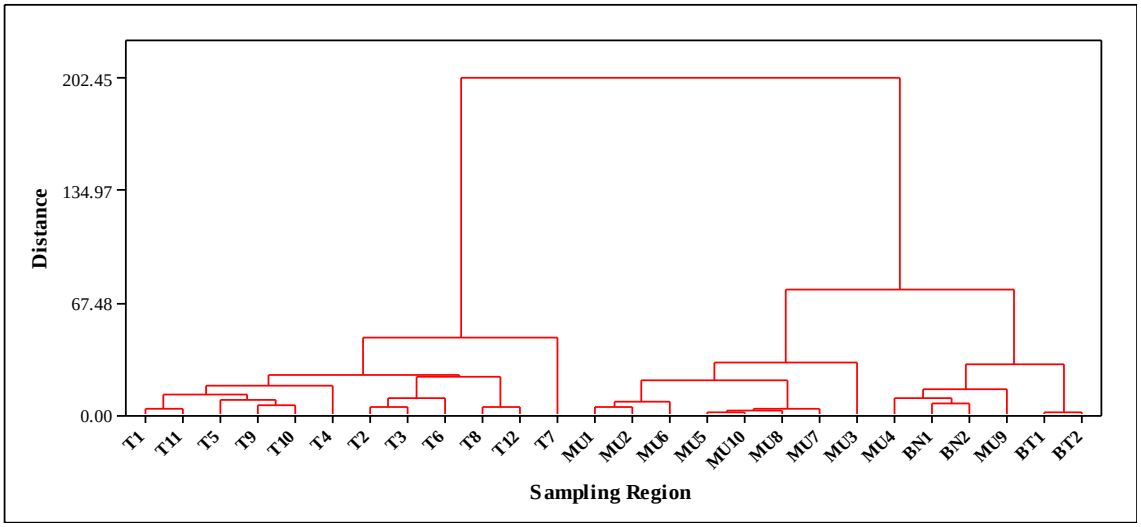

Figure 3.18. Dendrogram for honey samples from Trakya, Muş, Bingöl and Bitlis using synchronous fluorescence spectra.

3.1.1.3. Excitation-Emission Fluorescence Results

Spectra with excitation-emission fluorescence mode were collected by changing excitation wavelengths with 10 nm intervals between 330 to 380 nm while emission intensities were scanned in the range of 400 and 700 nm. Figure 3.19 illustrates a three dimensional excitation-emission spectrum of a honey sample. As can be seen from the figure, the honey sample gave the maximum emission intensity at approximately 500 nm.

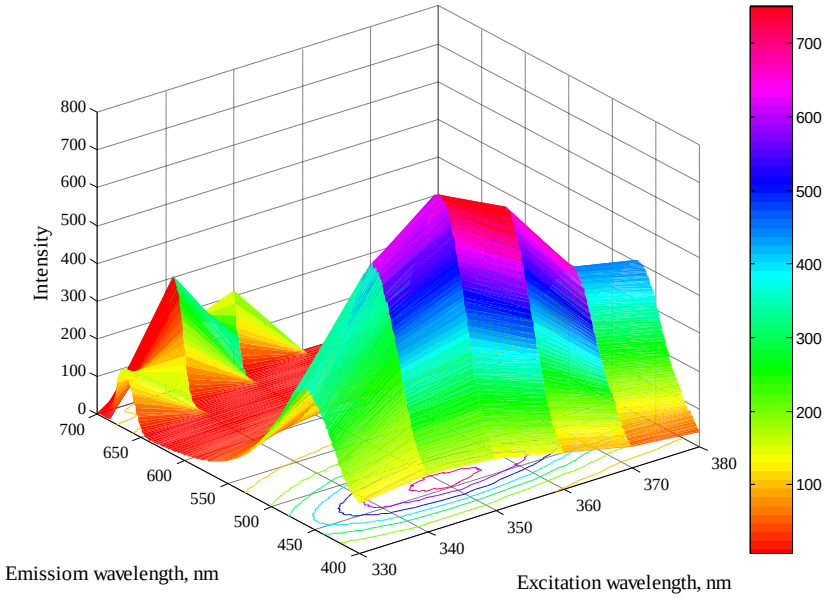

Figure 3.19. Excitation and emission fluorescence spectra of a honey sample between 400 and 700 nm emission wavelengths and between 330 and 380 nm excitation wavelengths.

After collecting three dimensional spectra of all honey samples, PCA and HCA analysis were performed with the same combinations of regions mentioned above in order to compare molecular spectrometric methods for the classification of honey samples based on the geographical region and the score plot obtained is represented in Figure 3.20. The explained total variation in the result of PCA was 81 % with the honey samples from Trakya, Kayseri, Malatya and Sivas. West and central samples were classified separately as in the case of the other spectrometric methods, however, there were some exceptions like the samples labeled as T5, T12 and K6.

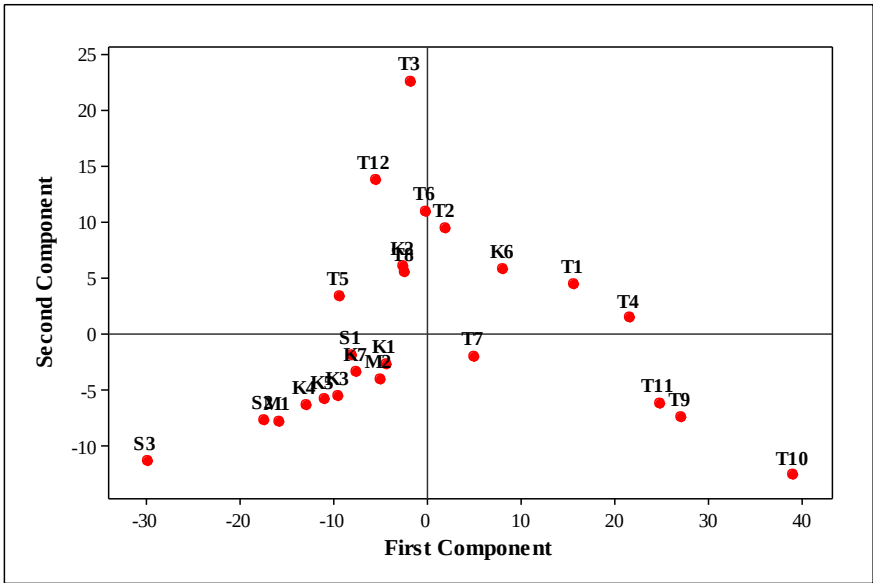

Figure 3.20. The score plot of the first component versus the second component for honey samples from Trakya, Kayseri, Malatya and Sivas using excitation-emission fluorescence (EEF) spectra.

Finally, HCA was processed with 3 principal components that were explained by 93 % variation of the data and the resulting dendrogram is represented in Figure 3.21. As observed from the figure, the samples were clustered effectively based on the sampling region, but there were some exceptions in samples labeled as K6 and K1.

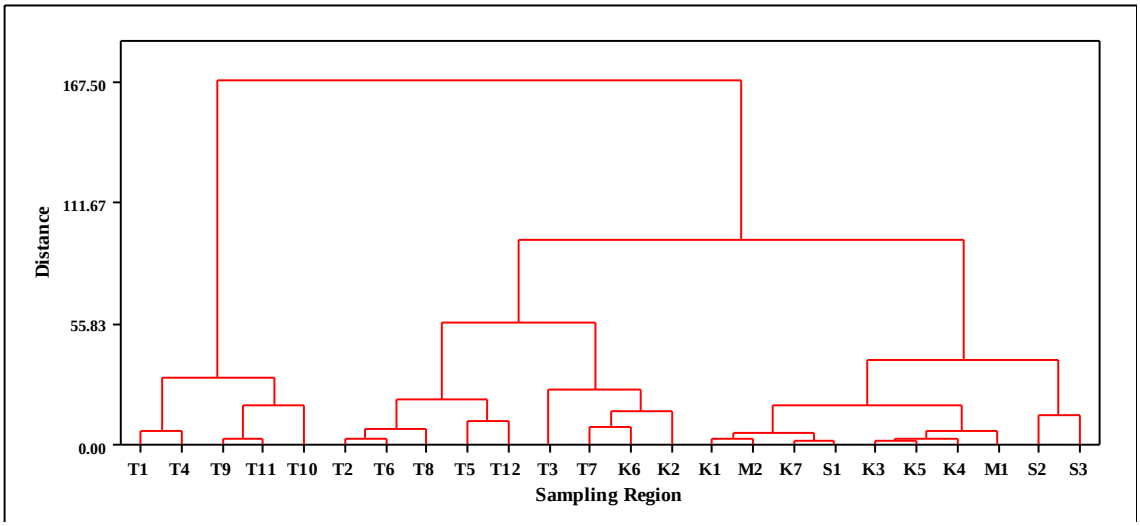

Figure 3.21. Dendrogram for honey samples from Trakya, Kayseri, Malatya and Sivas using excitation-emission fluorescence (EEF) spectra.

The results of principal component analysis for the honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa is demonstrated in Figure 3.22. As can be seen from the figure, the samples from Diyarbakır and Yüksekova formed different classes, nevertheless, Muş, Bingöl and Bitlis samples were overlapped and combined to form a unique class. The sample labeled as DY2 was from Diyarbakır and was placed far away from the others and could be said that it didn't show similar fluorescence characteristics when compared with the others.

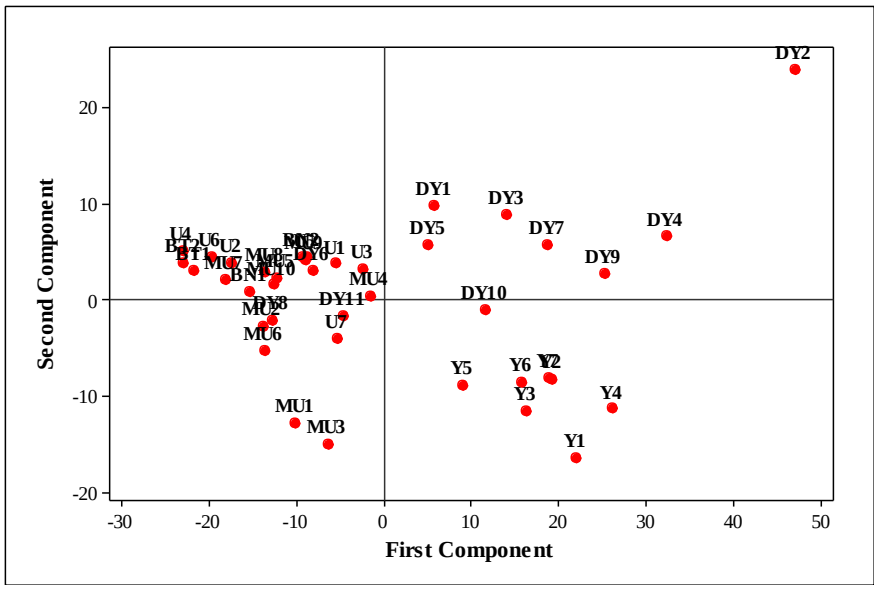

Figure 3.22. The score plot of the first component versus the second component for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using excitation-emission fluorescence (EEF) spectra.

After the PCA analysis, the HCA method is applied and the result of dendrogram is depicted in Figure 3.23. As can be seen, there was obvious separation of Yüksekova samples as mentioned previously, Diyarbakır samples were also clustered together and seemed to have similar characteristics to Yüksekova samples. In conclusion, the PCA results were in consistent with the dendrogram.

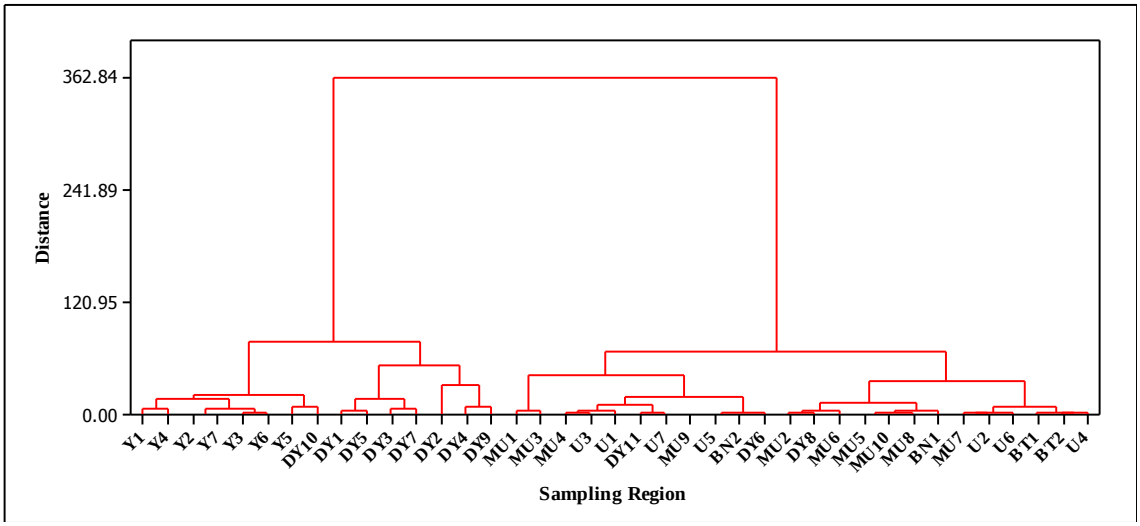

Figure 3.23. Dendrogram for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using excitation-emission fluorescence (EEF) spectra.

The next trial was constructed with the samples from İzmir, Datça, Marmaris and Antalya and the plot of the first two components which were accounted for 88 % variation are illustrated in Figure 3.24. All Datça and Marmaris samples were located at the positive side of the fist component as observed with the other two spectrometric methods. However, İzmir samples were not distinguished and spread over the plot.

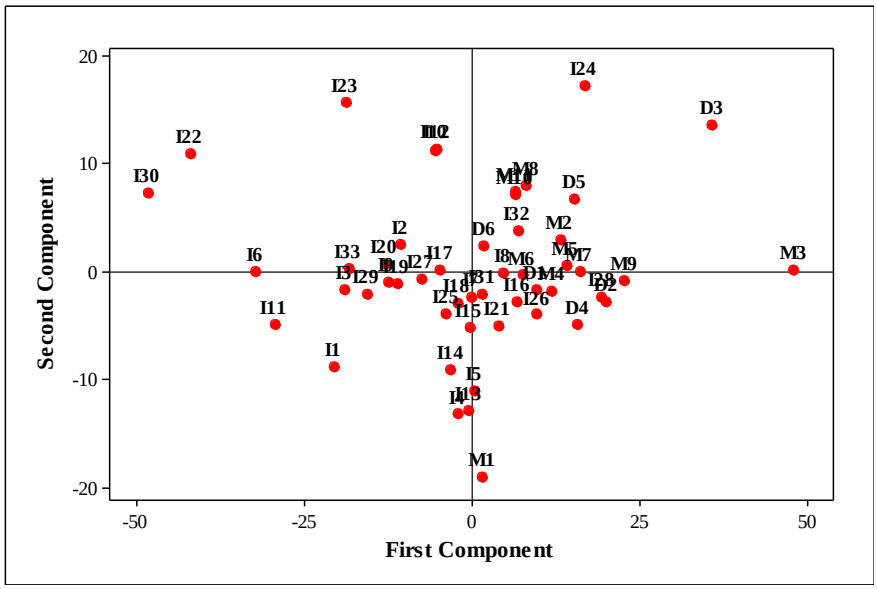

Figure 3.24. The score plot of the first component versus the second component for honey samples from İzmir, Datça, Marmaris and Antalya using excitation-emission fluorescence (EEF) spectra.

Two PCs (89% variation) were used for figuring the HCA dendrogram presented in Figure 3.25. Surprisingly, although samples were from close neighbor regions, they were differentiated in different clusters. İzmir samples were also grouped together and constructed with two main clusters at the left side. Finally, Datça and Marmaris clusters were located as neighbors.

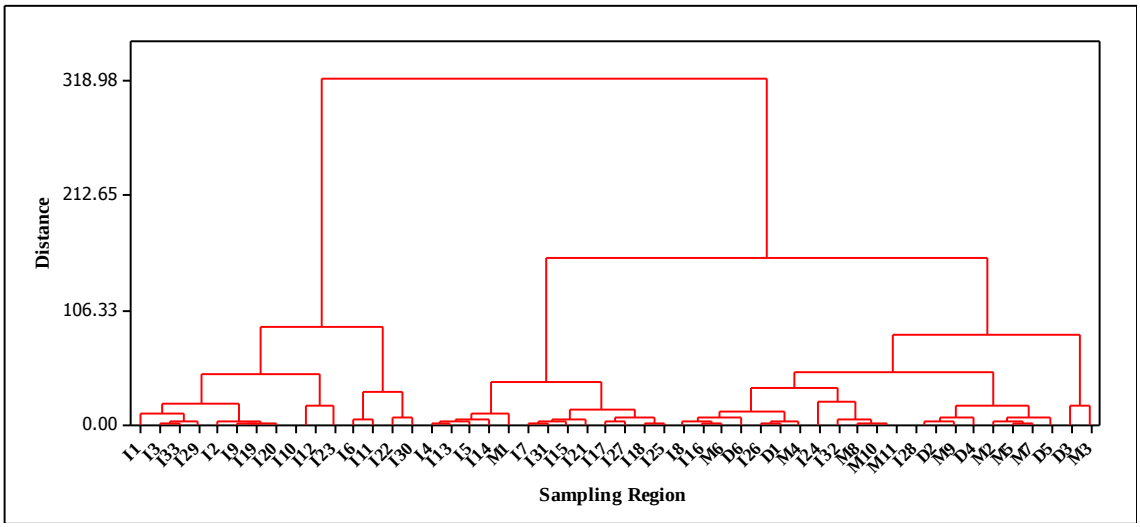

Figure 3.25. Dendrogram for honey samples from İzmir, Datça, Marmaris and Antalya using excitation-emission fluorescence (EEF) spectra.

The last combination was figured out with the honey samples from Trakya, Muş, Bingöl and Bitlis. The score plot of the first component versus the second component is shown in Figure 3.26. The two PC's explain approximately 83% of the total variance of the data. The plot showed us a clear differentiation between the honey samples from west to east. The positive side of the first component included Trakya samples whereas the negative part of the first component included east samples.

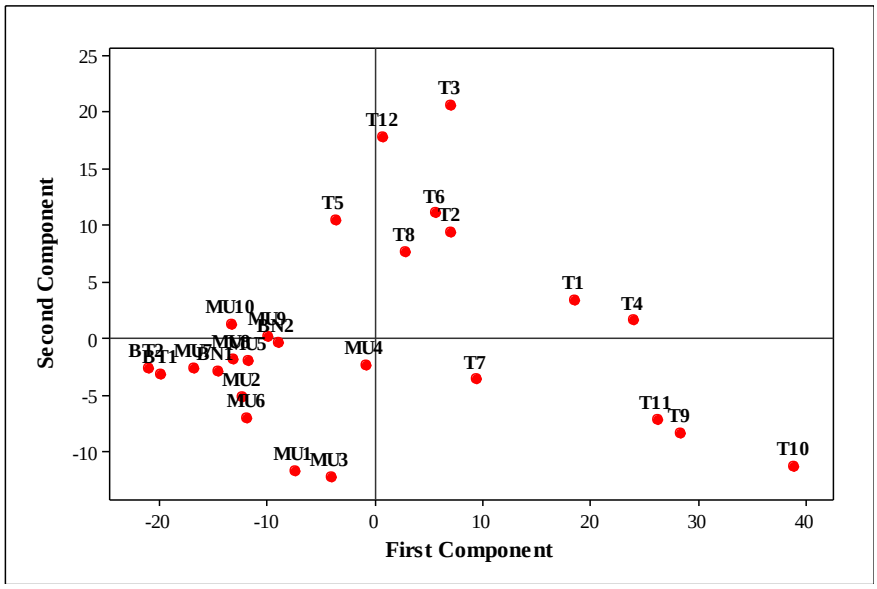

Figure 3.26. The score plot of the first component versus the second component for honey samples from Trakya, Muş, Bingöl and Bitlis using excitation-emission fluorescence (EEF) spectra.

To see the closeness of the honey samples, HCA dendrogram is drawn using three principal components and the dendrogram is demonstrated in Figure 3.27. According to the dendrogram, the results of the principal component was supported by drawing two main clusters, but only one sample from Trakya (T7) was beside from other Trakya samples.

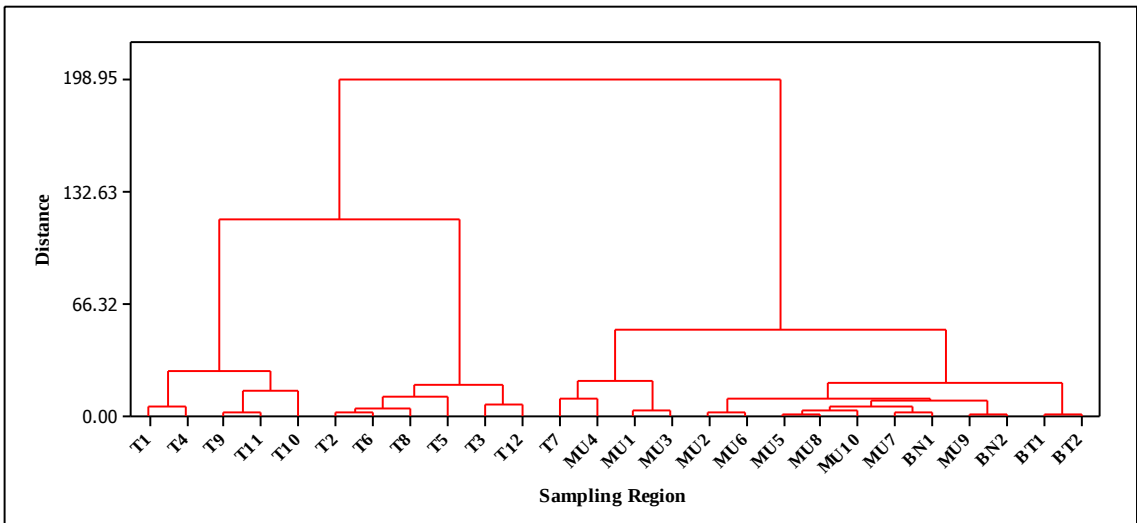

Figure 3.27. Dendrogram for honey samples from Trakya, Muş, Bingöl and Bitlis using excitation-emission fluorescence (EEF) spectra.

3.1.2. Classification Studies with Supervised Methods

After the classification studies with unsupervised methods, the same combinations were tested with a supervised method SIMCA. In principal component analysis (PCA), the class information is not used in the construction of the model and PCA just describes the overall variation in the data. However, PCA can be coupled with the class information to give classification models by means of soft independent modeling of class analogy. In SIMCA, a PCA is performed on each class in the data set, and a sufficient number of principal components are kept to explain for most of the variation within each class. Therefore, a principal component model is used to represent each class in the data set. The number of principal components retained for each class is usually different. SIMCA develops principal component models for each training class separately and provides information including critical distances which can be calculated as the geometric distance of each object from the principal component models. SIMCA models on principal components were developed for classification of honey samples according to geographical origin. The distance from the model for class 1 was plotted against that from model 2. The discrimination of each class was shown in the Cooman's plots of the class models.

3.1.2.1 FTIR-ATR Results

The spectra collected were used for SIMCA software. Prior to multivariate analysis, the data were pre-processed by the standard procedure. This procedure includes mean-centering (the mean value of each variable is calculated and subtracted from the data) and normalization. As it is known SIMCA does not use all of the principal components, but the principal components are selected by the cross validation. The calculated principal components of honey sample classes according to regions of the model and their explained variances are given in the Table 3.1. The distance from the model for class 1 was plotted against that from model 2. The discrimination of each class was shown in the Cooman's plots of the class models.

Table 3.1. Number of PCs used and general statistics of each honey sample classes for FTIR spectral data

Honey sample classes	Number of PCs	% explained variance
Trakya (n=12)	5	96
Kayseri, Malatya,Sivas (n=13)	9	93
Yüksekova (n=7)	3	88
Muş, Bingöl, Bitlis (n=14)	6	94
İzmir (n=35)	16	100
Datça, Marmaris, Antalya (n=18)	8	97

The same scenarios were also investigated and the first combination was figured out with two groups which consisted of honey samples from Trakya, Kayseri, Malatya and Sivas. The Cooman's plot for honey samples is exhibited in Figure 3.28. When Cooman's plot is examined, Kayseri, Malatya and Sivas samples were designed as a class and successful differentiation was achieved. All of the samples except the sample labeled as T4 were placed in their class regions on the plot, T4 is in outlier region.

The next combination was made with the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. These groups of samples were from neighbor cities and since SIMCA differentiate two classes of samples, the samples from Yüksekova were introduced as class 1 whereas the samples from Muş, Bingöl, and Bitlis were as class 2. The others were not defined as classes in order to see the behavior of the honey samples. The results obtained from SIMCA are displayed in Figure 3.29. As can be seen from the figure, most of the honey samples were discriminated effectively according to geographical origin although they are neighbors and the samples which were not defined as class (Diyarbakır, Urfa) were located in the outlier region.

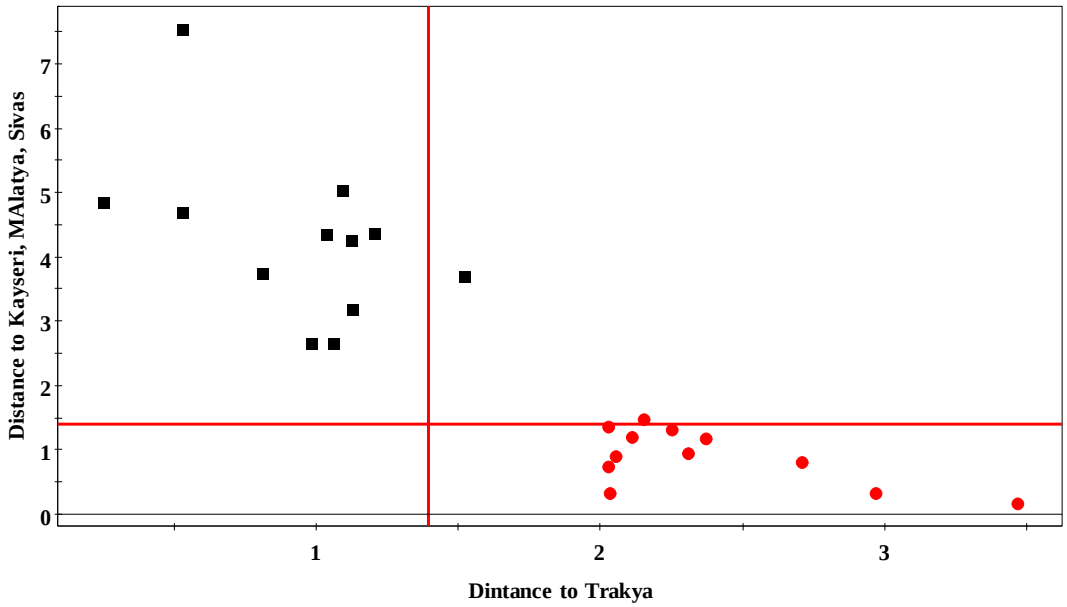

Figure 3.28. Cooman's plot for classification of honey samples from Trakya, Kayseri, Malatya, Sivas using FTIR spectra ((■)Trakya and (●) Kayseri, Malatya, Sivas)

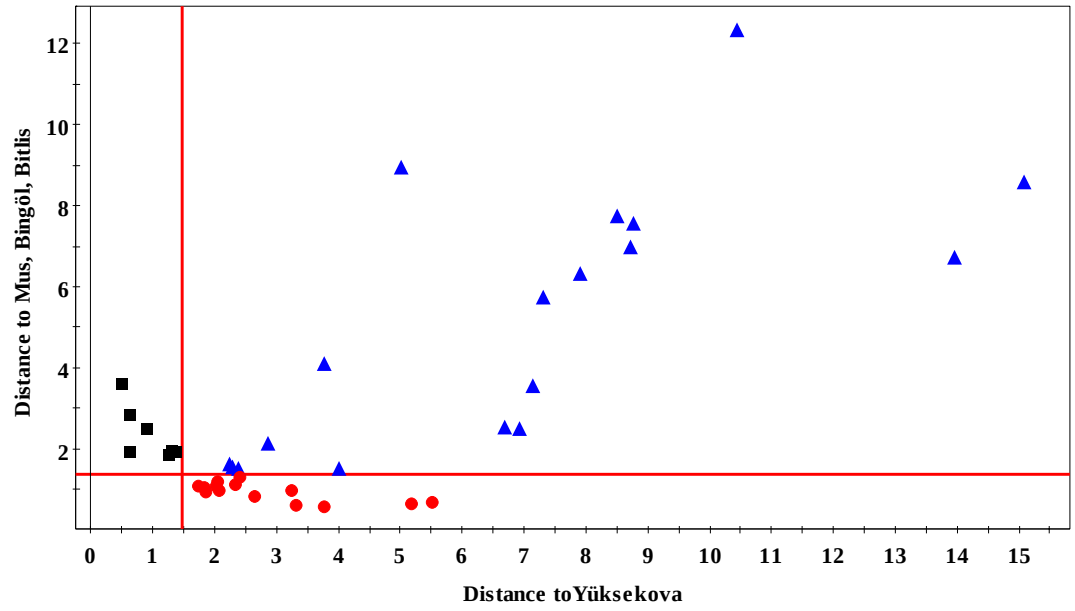

Figure 3.29. Cooman's for classification of honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using FTIR spectra ((■)Yüksekova, (●) Muş, Bingöl, Bitlis and (▲) Diyarbakır, Urfa).

The following trial was performed with the honey samples from İzmir, Datça, Marmaris and Antalya for their geographical grouping. Figure 3.30 represents the Cooman's plot for honey samples from those regions. The differentiation between honey samples from İzmir, Datça, Marmaris and Antalya seemed clearly even the

region so close to each other, only a sample from İzmir labeled as I10 was misclassified and it was located in the overlap region.

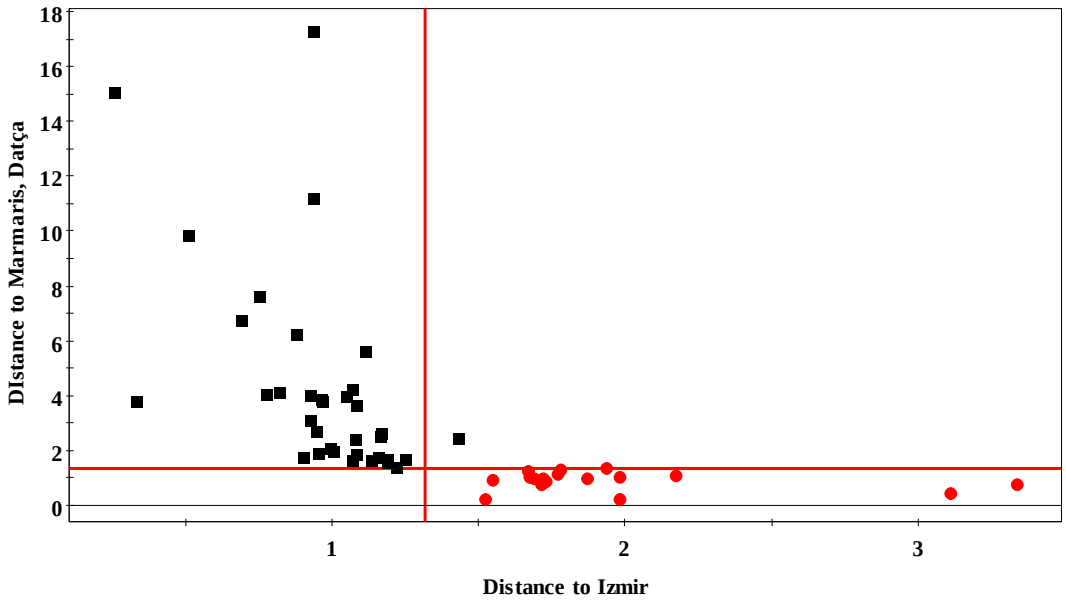

Figure 3.30. Cooman's for classification of honey samples from İzmir, Datça, Marmaris and Antalya using FTIR spectra ((■) İzmir, (●) Marmaris, Datça, Antalya).

The last combination included the samples from Trakya, Muş, Bingöl and Bitlis and was compared with FTIR results. The resulted Cooman's plot is depicted in Figure 3.31. As can be seen from the figure, all honey samples from that region were classified but there was only one exception (T4). The sample labeled as T4 was in outlier region when Trakya samples were matched with Kayseri and Malatya samples and it was not seemed belong to its class.

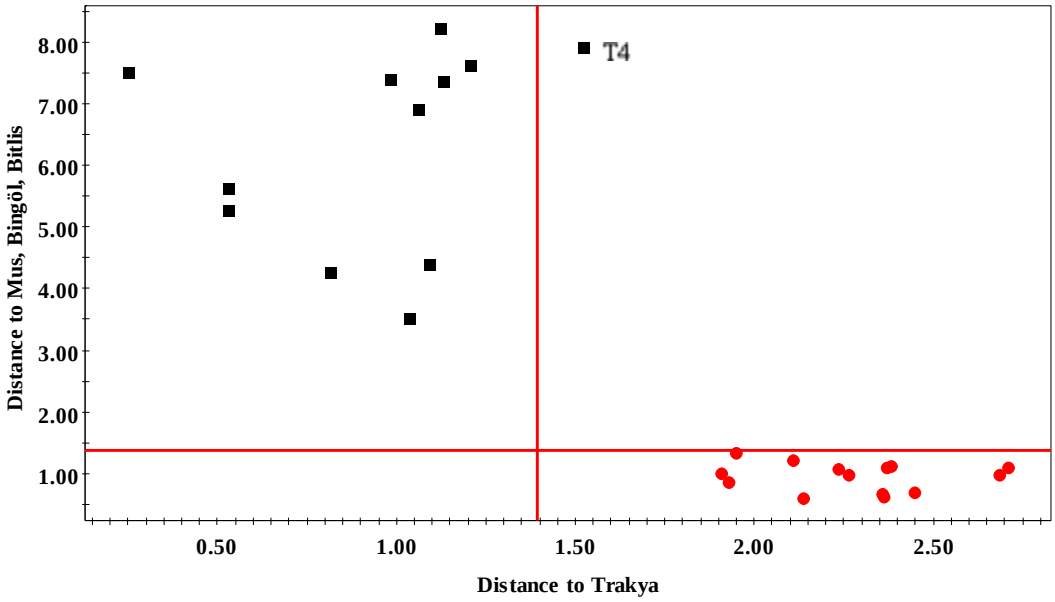

Figure 3.31. Cooman's for classification of honey samples from Trakya, Muş, Bingöl and Bitlis using FTIR spectra (■) Trakya, (●) Muş, Bingöl and Bitlis).

3.1.2.2. Synchronous Fluorescence Results

In order to compare the classification results using different molecular spectrometric methods, the honey samples were scanned with fluorescence spectrometer via synchronous mode. Afterwards, the same combinations of the samples that were given in FTIR were used for SIMCA analysis in order to compare the results. Table 3.2 represents the calculated principal components of the model for honey sample classes according to regions and their total variances.

Table 3.2. Number of PCs used and general statistics of each honey sample classes for synchronous florescence spectral data

Honey sample classes	Number of PCs	% explained variance
Trakya (n=12)	5	79
Kayseri, Malatya, Sivas (n=13)	6	96
Yüksekova (n=7)	3	94
Muş, Bingöl, Bitlis (n=14)	6	97
İzmir (n=35)	16	98
Datça, Marmaris, Antalya (n=18)	8	97

The first grouping had included Trakya, Kayseri, Malatya and Sivas samples. The classification results with SIMCA are indicated in Figure 3.32 with the help of Cooman's plot. As seen from the figure, all of the samples were classified according to their geographical origin and there was no sample in neither overlap nor outlier regions. Distance of sample which was coded as S2 was larger than others ant it was located far away from the members of the class.

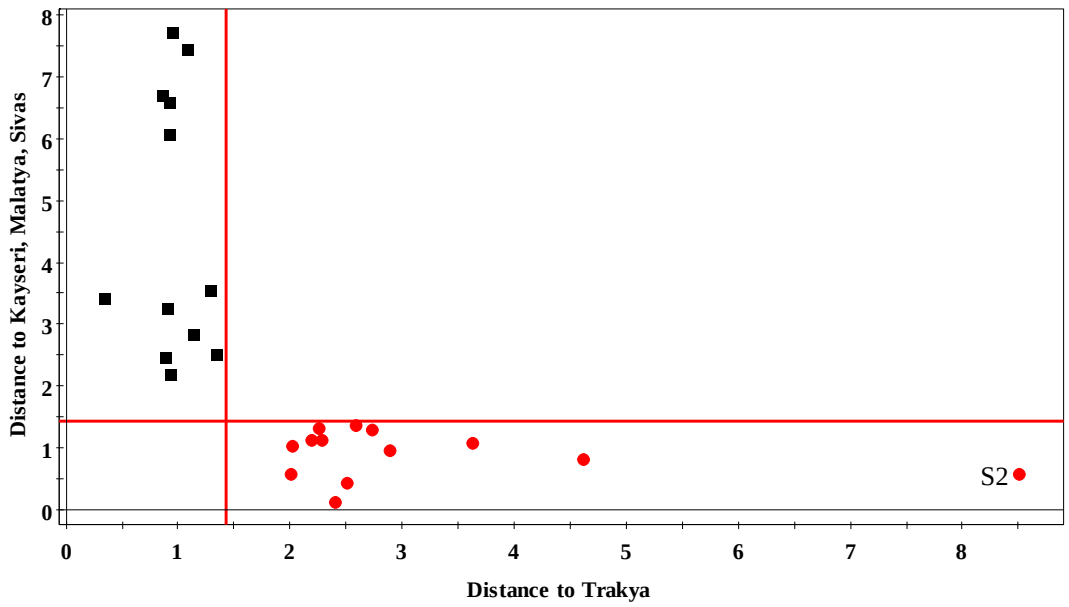

Figure 3.32. Coomans' plot for classification of honey samples from (■) Trakya, (●) Kayseri, Malatya, Sivas using synchronous fluorescence spectra.

When honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa samples were analyzed, the Cooman's plot in Figure 3.33 was obtained. The plot showed that the honey samples in Yüksekova, Muş, Bingöl, Bitlis were differentiated clearly according to synchronous spectral data and most of Diyarbakır and Urfa samples were separated in outlier region. Five of Urfa samples and one of Diyarbakır samples were located on the critical line. It is understood that those samples were close to Yüksekova samples.

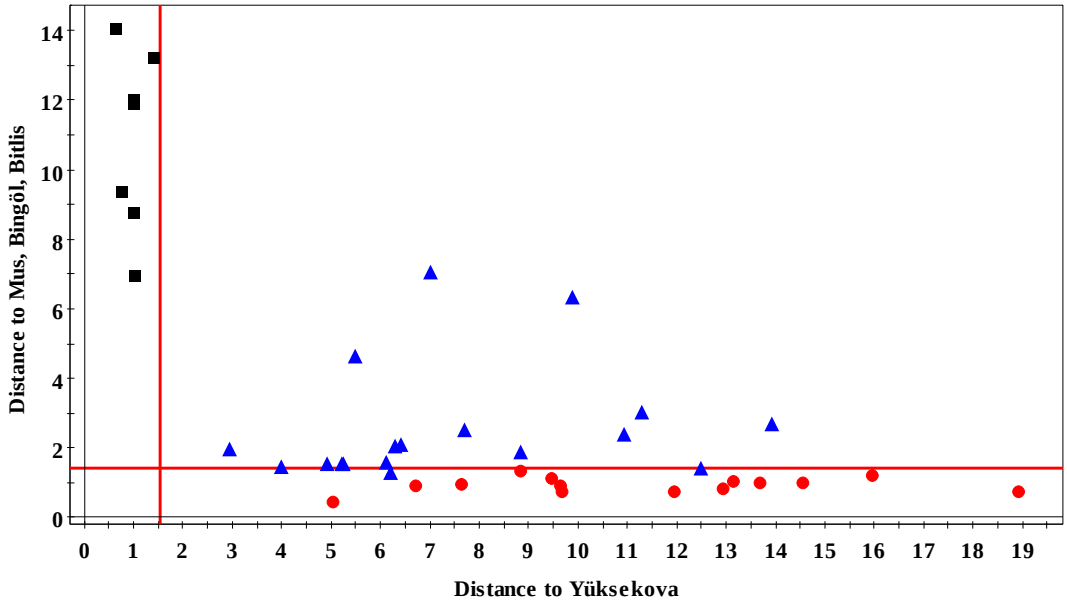

Figure 3.33. Cooman's for classification of honey samples from (■) Yüksekova, (●) Muş, Bingöl, Bitlis and (▲) Diyarbakır and Urfa using synchronous fluorescence spectra.

The next trial was performed with honey samples from İzmir, Datça, Marmaris and Antalya. The resulting plot obtained with SIMCA is illustrated in Figure 3.34. When a comparison was done with the FTIR results, a similar trend was not obtained and it was unsuccessful to classify the honey samples based on geographical region. Some İzmir samples and the half of Datça, Marmaris samples were located in the overlap region.

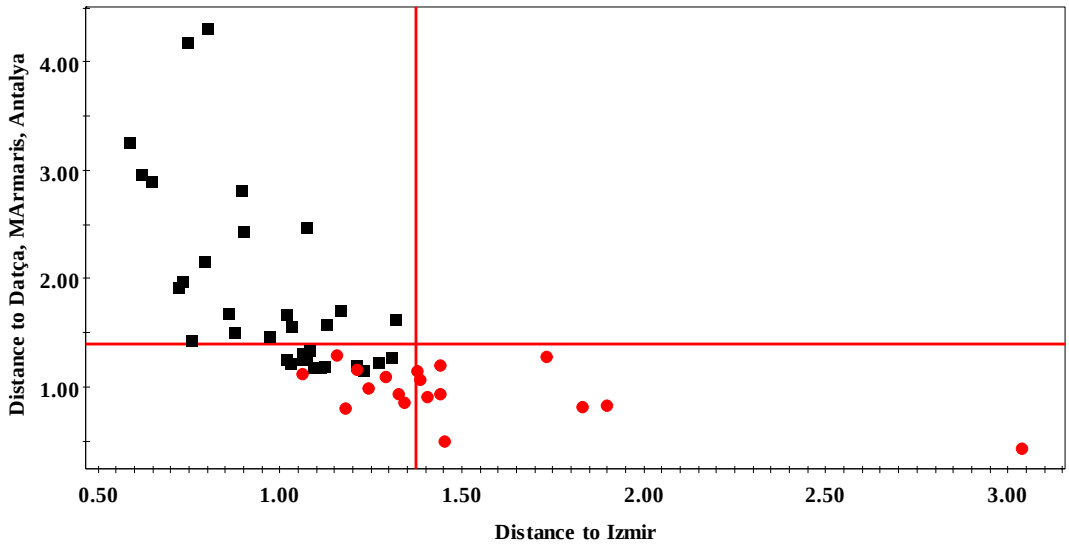

Figure 3.34. Cooman's for classification of honey samples from (■) İzmir, (●) Datça, Marmaris and Antalya using synchronous fluorescence spectra.

The last examination for geographical region was constructed with honey samples from Trakya, Muş, Bingöl and Bitlis. Since Trakya is North West whereas Muş, Bingöl and Bitlis are in the east of Turkey, the classification of these groups was expected. The Cooman's plot was obtained after SIMCA analysis and it was demonstrated in Figure 3.35. As can be concluded from the figure, all of the samples were discriminated according to their classes as expected and there were no overlapping and outlier samples.

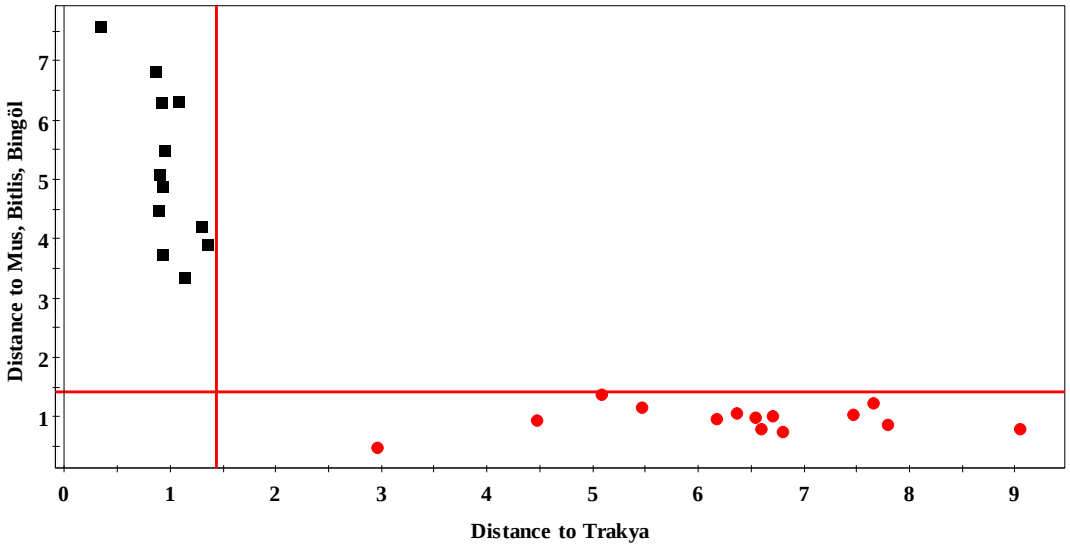

Figure 3.35. Cooman's for classification of honey samples from (■) Trakya, (●) Muş, and Bitlis using synchronous fluorescence spectra.

3.2.3. Excitation-Emission Fluorescence Results

The same honey samples were scanned with excitation-emission fluorescence mode, and then the collected three dimensional spectra of honey samples were used in SIMCA analysis with the same combinations of regions mentioned before in order to compare molecular spectrometric methods for classification according to geographical region. The calculated principal components of the model for honey sample classes according to regions and their total variances during SIMCA analysis are indicated in Table 3.3.

Table 3.3. Number of PCs used and general statistics of each honey sample classes for excitation-emission fluorescence spectral data.

Honey sample classes	**Number of PCs**	**% explained variance**
Trakya (n=12)	5	100
Kayseri, Malatya,Sivas (n=13)	5	100
Yüksekova (n=7)	3	92
Muş, Bingöl, Bitlis (n=14)	6	100
İzmir (n=35)	5	99
Datça, Marmaris, Antalya (n=18)	8	99

Figure 3.36 shows the Cooman's plot resulted after the analysis of honey samples from Trakya, Kayseri, Malatya, Sivas using excitation-emission spectral data. When compared with the other molecular spectrometric methods, excitation-emission fluorescence data gave also the same differentiation of the honey samples from Trakya, Kayseri, Malatya, Sivas. All of the samples except sample T12 were in the right class and T12 was observed as outlier.

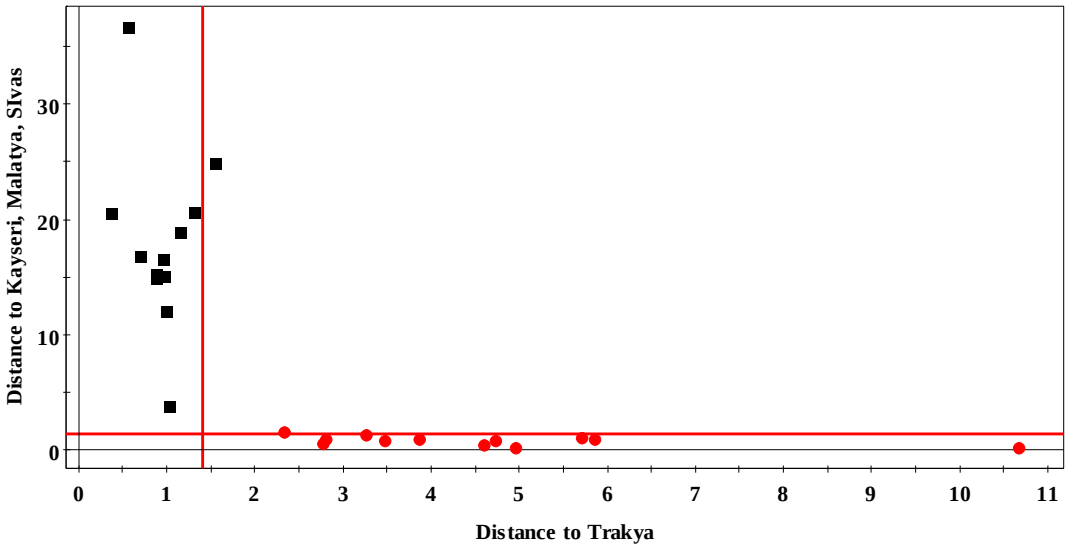

Figure 3.36. Coomans' plot for classification of honey samples from (■) Trakya, (•) Kayseri, Malatya, Sivas using excitation-emission fluorescence spectra

The following trial was with the honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa and the Cooman's plot in Figure 3.37 represents that the samples were mainly classified according to their geographical origin. There was only one Yüksekova sample observed in outlier region, nevertheless, there were no samples in the overlap region. As observed with other molecular spectrometric data, some of Urfa samples were very close to the class of Muş, Bingöl, Bitlis while they were located on the critical line.

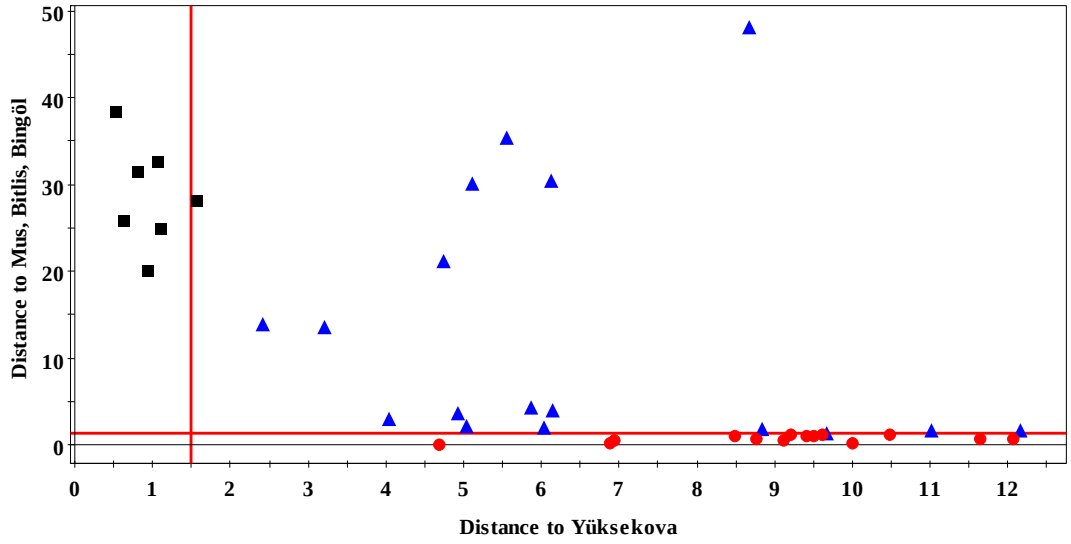

Figure 3.37. Cooman's for classification of honey samples from (■) Yüksekova, (•) Muş, Bingöl, Bitlis, (▲)Diyarbakır and Urfa using excitation-emission fluorescence spectra.

The analysis result of honey samples from İzmir, Datça, Marmaris and Antalya is displayed in Figure 3.38. When Cooman's plot is examined, a good differentiation of the samples was not clearly observed according to their region. Only one misclassification was seen that belonged to sample from İzmir in the outlier region, nonetheless, half of the Datça, Marmaris samples were located in overlap region.

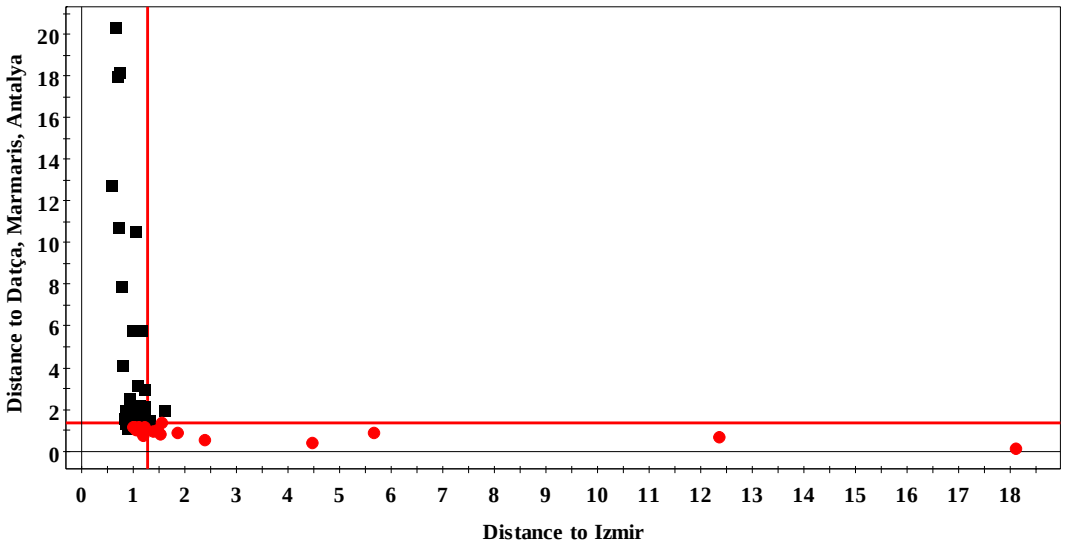

Figure 3.38. Cooman's for classification of honey samples from (■) İzmir, (●) Datça, Marmaris and Antalya using excitation-emission fluorescence spectra.

The last combination was constructed with honey samples from Trakya, Muş, Bingöl and Bitlis as seen in Figure 3.39. The previously obtained differentiation was also observed using excitation-emission profiles. There was only one sample (T12) that exceeded the critical limit and the samples were classified according to their geographical origin.

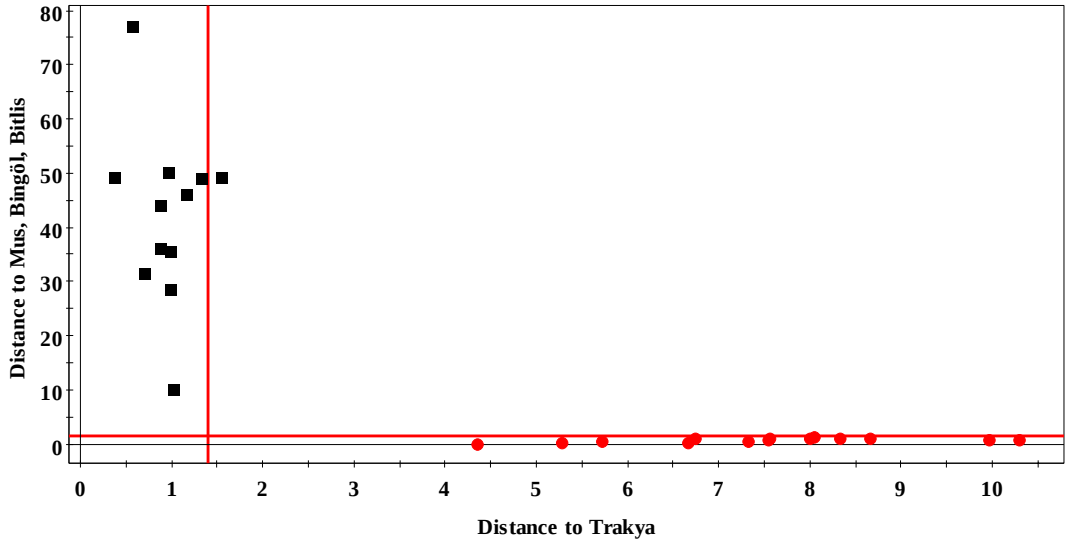

Figure 3.39. Cooman's for classification of honey samples from (■)Trakya, (●) Muş, Bingöl and Bitlis using excitation-emission fluorescence spectra.

3.2. Classification Studies with Atomic Spectrometric Data

3.2.1. Calibration Plots and Detection Limits of Investigated Elements

In order to compare the classification results obtained with molecular spectroscopy, the honey samples were also analyzed with atomic spectroscopic data. For this purpose, the collected samples were firstly digested and then the metal concentrations were determined by inductively coupled plasma mass spectrometer. The metal ions detected with ICP-MS are Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba and Pb. The metal that has the highest concentration is Mg and the metals that have the lowest concentrations are Pb and Co. The aqueous standard solutions were prepared from multielement standard and the calibration plots of metals for ICP-MS readings were constructed according to concentration ranges in digested aqueous honey samples. Figure 3.40 and Figure 3.41 represents the calibration plots obtained with ICP-MS measurements for the studied elements.

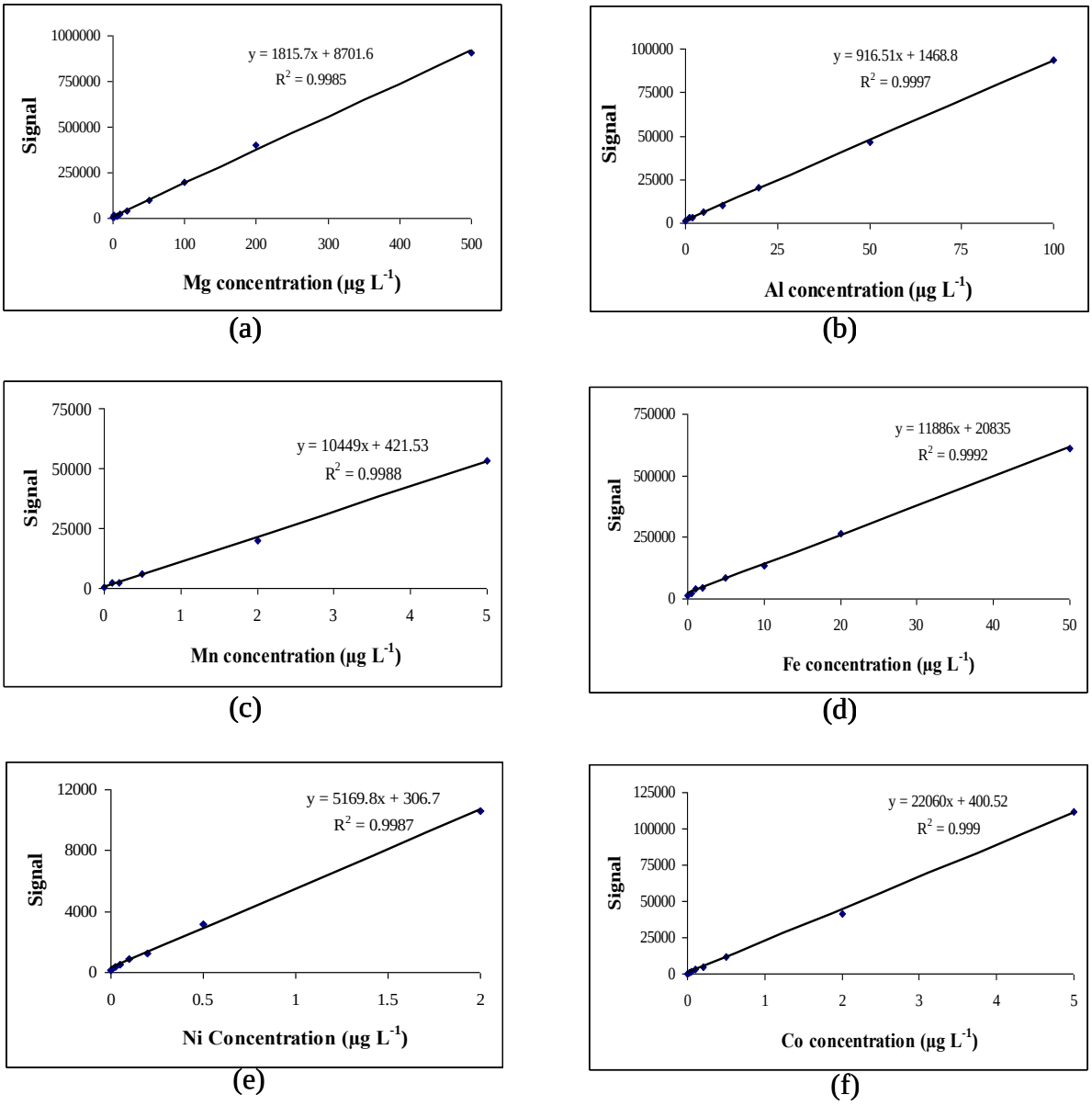

Figure 3.40. Calibration plots of **(a)** Mg (y = 1815.7x + 8701.6; R^2 = 0.9985), **(b)** Al (y = 916.51x + 1468.8; R^2 = 0.9997, **(c)** Mn (y = 10449x + 421.53; R^2 = 0.9988), **(d)** Fe (y = 11886x + 20835; R^2 = 0.9992), **(e)** Ni (y = 5169.8x + 306.7 R^2 = 0.9987), **(f)** Co (y = 22060x + 400.52; R^2 = 0.999).

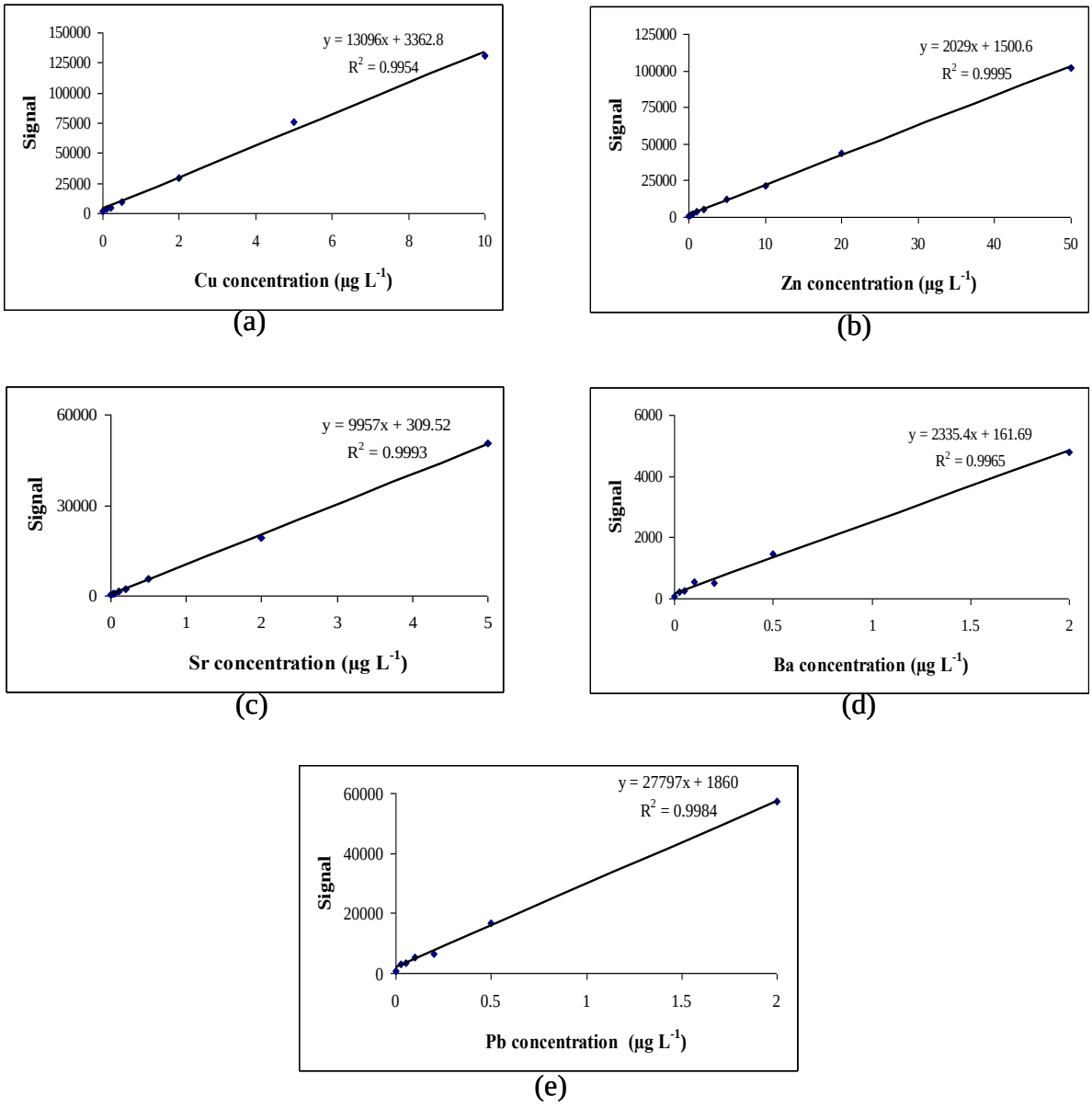

Figure 3.41. Calibration plots of **(a)** Cu (y = 13096x + 3362.8; R^2 = 0.9954), **(b)** Zn (y = 2029x + 1500.6; R^2 = 0.9995, **(c)** Sr (y = 9957x + 309.52; R^2 = 0.9993), **(d)** Ba (y = 2335.4x + 161.69; R^2 = 0.9965), **(e)** Pb (y = 27797x + 1860; R^2 = 0.9984).

The detection limits were also calculated for each metal based on 3σ of the blank and the results are given in Table 3.4.

Table 3.4. Detection Limits for Metals by ICP-MS

Metals	Limit of Detection (LOD) (µg $L^{-1)}$
Mg	10.3
Al	0.49
Mn	0.008
Fe	1.38
Co	0.022
Ni	0.037
Cu	0.04
Zn	0.26
Sr	0.024
Ba	0.009
Pb	0.014

3.2.2. Method Validation for Digestion Procedures

Samples are in the form of a solution in most of the measurement techniques. The sample should not contain high amounts of insoluble particles and its viscosity should not be too high. These requirements affect the choice of the pre-treatment method. Even though honey is soluble in water**,** it can be pre-treated just dissolving in water or acidic solutions for metal determination**,** samples are usually decomposed before measurements. Finally, the mineral components are transferred to the resulting digests. This treatment decomposes the organic matrix of honey and extracts metal species. It also prevents variations in physico-chemical properties of sample solutions and eliminates accumulation or deposition carbonaceous residues in the burner heads or the nebulizers of spectrometers used (Pohl 2009b). Honey is treated with different digestion procedures before the metal analysis. Generally, dry ashing (Silici et al. 2008, Terrab et al. 2003, Osman et al. 2007, Downey et al. 2005), wet acid digestion (Torres et al. 2005, Rashed and Soltan 2004, Devillers et al. 2002, Tuzen et al. 2007) and microwave digestion (Stankovska et al. 2008, Tuzen et al. 2007, Frazzoli et al. 2007,

Ajtony et al. 2007) have been used to disrupt the organic substances and release metals from the complex sample matrix into solution.

In this study, two types of digestion procedures were tried and compared according to characteristics such as easy in application, digestion without loosing metals and time saving. These methods were microwave digestion, and wet digestion without applying heat. A certified reference material was used in order to determine the best digestion procedure prior to ICP-MS analysis. Additionally, accuracy of the method was tested through the use of Brown Bread BCR CRM 191, one of the reference materials with a high content of carbohydrate and therefore closes enough to the complex matrix of honey. Table 3.5 shows the certified metal concentrations in BCR CRM 191.

Table 3.5. Certified values of elements in CRM BCR 191.

CRM BCR-191		
Elements	**Certified value**	**Uncertainity**
Cd	28.4 µg/kg	1.4 µg/kg
Pb	187 µg/kg	14 µg/kg
Cu	2.63 mg/kg	0.07 mg/kg
Zn	19.5 mg/kg	0.5 mg/kg
Fe	40.7 mg/kg	2.3 mg/kg
Mn	20.3 mg/kg	0.7 mg/kg

Reference material was weighed as 0.5 g and digested with the two digestion procedures as explained in experimental section. Subsequently, the digested samples were analyzed with ICP-MS for the metals demonstrated in Table 3.5. The determined concentration values and recoveries are indicated in Table 3.6 and Table 3.7, respectively, for two digestion procedures.

Tablo 3.6. Recovery results of CRM BCR 191 for microwave digestion method

Elements	Certified value (µg/kg)	Found (µg/kg)	% Recovery
Mn	20300	23518 ± 1990	116
Fe	40700	41556 ± 3223	102
Cu	2630	2683 ± 220	102
Zn	19500	18655 ± 1375	96
Cd	28.4	15.0 ± 2	54
Pb	187	198 ± 47	106

Tablo 3.7. Recovery results of CRM BCR 191 for wet digestion without heating method.

Elements	Certified value (µg/kg)	Found (µg/kg)	% Recovery
Mn	20300	23012 ± 999	113
Fe	40700	39495 ± 3800	97
Cu	2630	2772 ± 330	105
Zn	19500	18524 ± 2030	95
Cd	28.4	22 ± 2	78
Pb	187	207 ± 21	111

Once the recovery values were compared for microwave digestion and wet digestion without heat procedures, it was seen that there were no such differences between recovery results and generally values were near 100 %. However, there was one exception with Cd recovery results for both digestion procedures and the results showed the loss of Cd during digestion process. Nevertheless, Cd concentration was always below the detection limit while performing the preliminary studies of honey. Therefore, the wet digestion without applying heat was decided as the digestion procedure, in addition, for every analysis, one honey sample was digested with microwave and this sample was used as a control sample for checking whether there was any difference between readings.

3.2.3. Classification Studies with Unsupervised Methods

In order to compare the classification results using molecular spectroscopic data with atomic spectroscopic data, the collected samples were digested and metal concentrations were determined with inductively coupled plasma mass spectrometer. The metals which of the concentrations were determined are Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba. The metal concentrations of the collected honey samples are shown in Table 3.8.

The metal Pb was also determined but the concentration values nearly all samples were below the detection limit, and the others have concentration values close to detection limit. For the reason that the relative standard deviation (RSD) values were so high, hence the results can not be reliable. As can be seen from Table 3.8, the metal that has the highest concentration was Mg and the metal that has the lowest concentration was Co. Mg concentrations in honey samples were changing between 10 and 120 mg/kg. The highest concentration value for Co in honey samples was 26 µg/kg and in some samples Co concentrations could not be detected because of very low values.

Afterwards the same ICP-MS data are processed with the principal component analysis (PCA), then with the hierarchical cluster analysis (HCA) in order to see whether there is grouping and to observe their closeness. The fact that the element content ranges were changing such as Mg concentration was at ppm level on the other hand Co was in ppb level, it was a necessity to mean center and normalize the data. After this preprocessing PCA was applied the first two PC's were chosen to examine the data set.

Tablo 3.8. The concentrations (µg/kg) of the elements determined by ICP-MS in honey samples (BDL: Below the Detecetion Limit) (N=3)

	DY1	DY2	DY3	DY4	DY5	DY6	DY7	DY8	DY9	DY10
Mg	10199 ± 1117	15025 ± 1697	9321 ± 361	8154 ± 246	9046 ± 528	12733 ± 541	16108 ± 358	23858 ± 5594	12836 ± 1258	15495 ± 178
Al	567 ± 105	1158 ± 175	569 ± 27	723 ± 34	618 ± 18	691 ± 152	1031 ± 90	675 ± 85	596 ± 33	874 ± 62
Mn	104 ± 10	137 ± 13	119 ± 7	110 ± 10	94 ± 10	97 ± 3	133 ± 2	298 ± 39	408 ± 183	131 ± 9
Fe	2286 ± 332	3202 ± 267	2269 ± 142	2359 ± 272	3159 ± 97	2261 ± 130	3307 ± 452	5995 ± 767	3572 ± 172	3034 ± 262
Co	BDL	BDL	BDL	BDL	BDL	3.8 ± 0.2	3.2 ± 0.7	4.5 ± 0.8	32.8 ± 2.9	1.3 ± 0.1
Ni	BDL	24 ± 5	22 ± 3	BDL	BDL	5 ± 2	12 ± 3	16 ± 9	2594 ± 193	31 ± 7
Cu	214 ± 24	370 ± 70	187 ± 14	189 ± 24	193 ± 13	56 ± 2	152 ± 5	152 ± 59	194 ± 90	111 ± 9
Zn	2147 ± 228	3944 ± 521	1359 ± 95	1591 ± 96	3641 ± 203	6809 ± 261	3109 ± 109	6332 ± 592	6144 ± 11	2476 ± 313
Sr	103 ± 15	206 ± 20	57 ± 3	96 ± 9	153 ± 13	224 ± 4	184 ± 7	305 ± 29	202 ± 20	157 ± 6
Ba	62 ± 4	75 ± 7	43 ± 9	60 ± 0.5	72 ± 30	15 ± 3	31 ± 1	31 ± 2	BDL	BDL

	MU1	MU2	U1	MU3	ML1	Y1	ML2	S1	Y2	Y3
Mg	18137 ± 638	19915 ± 5321	17647 ± 246	21634 ± 9044	9747 ± 706	30012 ± 2742	20964 ± 669	28430 ± 538	11906 ± 239	12972 ± 262
Al	419 ± 32	643 ± 23	759 ± 29	872 ± 148	837 ± 48	3779 ± 208	975 ± 8	738 ± 101	3326 ± 265	2210 ± 47
Mn	429 ± 19	220 ± 15	183 ± 16	327 ± 74	552 ± 37	816 ± 59	381 ± 13	460 ± 32	402 ± 8	340 ± 12
Fe	1787 ± 264	3240 ± 750	3714 ± 203	1699 ± 397	1652 ± 180	6892 ± 974	8243 ± 155	3117 ± 138	6459 ± 352	6107 ± 827
Co	1.3 ± 0.1	BDL	BDL	1.6 ± 0.7	6.1 ± 0.7	8.3 ± 1.9	2.0 ± 0.2	BDL	0.8 ± 0.2	1.7 ± 0.4
Ni	25 ± 4	18 ± 9	17 ± 1	20 ± 9	41 ± 4	82 ± 8	24 ± 6	59 ± 3	15 ± 4	11 ± 2
Cu	40 ± 12	1171 ± 135	95 ± 14	67 ± 22	BDL	336 ± 78	BDL	BDL	33 ± 7	365 ± 3
Zn	4759 ± 4	3667 ± 42	5397 ± 533	3277 ± 709	1483 ± 65	742 ± 144	2613 ± 29	4640 ± 97	812 ± 3	749 ± 5
Sr	387 ± 7	180 ± 8	318 ± 55	211 ± 44	188 ± 14	255 ± 44	429 ± 11	740 ± 21	173 ± 6	168 ± 6
Ba	BDL	109 ± 23	17 ± 3	31 ± 17	117 ± 3	179 ± 19	135 ± 6	235 ± 14	111 ± 20	138 ± 4

	Y4	Y5	Y6	Y7	MU4	MU5	MU6	MU7	MU8	MU9
Mg	14700 ± 1560	23686 ± 5467	13264 ± 1956	20611 ± 666	22852 ± 1116	13255 ± 1915	14225 ± 2772	20204 ± 790	10369 ± 1258	17079 ± 129
Al	2595 ± 91	3234 ± 495	3415 ± 236	2529 ± 61	1233 ± 95	1313 ± 195	1119 ± 210	1345 ± 22	1348 ± 97	1444 ± 143
Mn	377 ± 37	1529 ± 20	442 ± 38	480 ± 40	254 ± 17	750 ± 88	345 ± 63	1053 ± 21	594 ± 44	621 ± 52
Fe	4135 ± 105	8960 ± 383	5422 ± 455	5628 ± 486	2224 ± 374	3229 ± 452	1738 ± 449	2470 ± 182	1568 ± 1	7641 ± 858
Co	8.6 ± 2.4	36.6 ± 1.8	28.0 ± 2.5	25.7 ± 3.2	16.6 ± 1.3	10.4 ± 0.3	13.1 ± 1.4	12.6 ± 1.2	11.4 ± 0.7	19.8 ± 0.6
Ni	12 ± 4	112 ± 11	40 ± 2	25 ± 8	55 ± 3	62 ± 12	33 ± 8	94 ± 7	47 ± 8	40 ± 2
Cu	469 ± 76	432 ± 74	417 ± 18	446 ± 33	152 ± 31	129 ± 2	189 ± 56	66 ± 4	92 ± 22	236 ± 51
Zn	645 ± 11	4214 ± 68	975 ± 35	469 ± 162	2624 ± 309	670 ± 27	1816 ± 197	2420 ± 107	483 ± 83	1116 ± 118
Sr	172 ± 23	313 ± 5	219 ± 19	253 ± 23	292 ± 17	200 ± 3	102 ± 20	243 ± 10	229 ± 22	122 ± 19
Ba	185 ± 10	178 ± 4	193 ± 50	215 ± 15	170 ± 1	175 ± 62	157 ± 1	165 ± 11	119 ± 6	233 ± 27

(cont. on the next page)

Tablo 3.8.(cont.)

	MU10	K1	K2	K3	K4	K5	K6	K7	T1	T2
Mg	15485 ± 259	33007 ± 6610	19039 ± 532	17026 ± 1269	22743 ± 2010	20221 ± 1197	33395 ± 1469	21577 ± 1434	61349 ± 999	54791 ± 3244
Al	1222 ± 77	4111 ± 972	798 ± 61	439 ± 35	429 ± 35	436 ± 22	1316 ± 13	770 ± 33	48 ± 3	140 ± 5
Mn	707 ± 8	848 ± 172	279 ± 20	324 ± 14	336 ± 3	327 ± 3	262 ± 9	374 ± 18	1157 ± 22	305 ± 7
Fe	2164 ± 123	3971 ± 512	2682 ± 144	2619 ± 342	2903 ± 11	3546 ± 405	4497 ± 195	3628 ± 321	2094 ± 70	3411 ± 355
Co	10.4 ± 0.5	24.9 ± 3.7	12.2 ± 1.9	BDL	BDL	BDL	BDL	BDL	BDL	BDL
Ni	60 ± 2	97 ± 4	58 ± 13	36 ± 1	34 ± 2	51 ± 1	59 ± 5	56 ± 3	132 ± 9	83 ± 2
Cu	42 ± 11	401 ± 60	412 ± 25	258 ± 37	447 ± 42	380 ± 12	756 ± 23	375 ± 49	346 ± 18	350 ± 1
Zn	666 ± 66	449 ± 125	1379 ± 222	1850 ± 111	3142 ± 385	2070 ± 281	2543 ± 357	2395 ± 33	1882 ± 98	2274 ± 100
Sr	188 ± 8	170 ± 36	262 ± 16	185 ± 20	368 ± 29	150 ± 41	347 ± 6	290 ± 18	378 ± 61	265 ± 12
Ba	92 ± 4	281 ± 3	162 ± 27	89 ± 6	80 ± 3	72 ± 3	75 ± 5	115 ± 3	213 ± 9	97 ± 10

	T3	T4	T5	T6	D1	D2	D3	D4	M1	M2
Mg	45066 ± 4469	98571 ± 3592	65358 ± 2324	53176 ± 403	71799 ± 721	63577 ± 3118	30037 ± 587	72800 ± 1068	73019 ± 832	87214 ± 1523
Al	170 ± 3	577 ± 18	410 ± 10	252 ± 41	1847 ± 233	3692 ± 284	700 ± 60	3048 ± 92	2618 ± 96	2677 ± 160
Mn	254 ± 4	1419 ± 72	675 ± 71	430 ± 1	874 ± 31	1051 ± 47	647 ± 59	905 ± 50	1041 ± 18	1527 ± 37
Fe	1437 ± 155	5131 ± 118	10557 ± 170	2876 ± 111	3911 ± 594	4215 ± 396	2729 ± 632	3561 ± 136	7234 ± 35	12502 ± 3971
Co	BDL	BDL	BDL	BDL	9.6 ± 1.8	8.4 ± 0.4	6.1 ± 0.4	8.3 ± 0.9	11.7 ± 0.5	11.5 ± 0.9
Ni	71 ± 12	167 ± 0	132 ± 9	96 ± 2	476 ± 57	169 ± 5	47 ± 2	359 ± 4	227 ± 68	276 ± 18
Cu	419 ± 23	1234 ± 107	508 ± 17	391 ± 14	746 ± 78	795 ± 4	366 ± 22	771 ± 76	872 ± 8	2679 ± 237
Zn	3931 ± 738	4223 ± 694	4998 ± 132	2105 ± 116	2215 ± 179	1761 ± 299	1323 ± 382	1046 ± 42	3509 ± 394	2820 ± 711
Sr	232 ± 20	441 ± 57	233 ± 10	266 ± 28	191 ± 62	548 ± 8	2301 ± 38	143 ± 1	100 ± 18	130 ± 25
Ba	127 ± 8	365 ± 19	90 ± 13	113 ± 13	205 ± 92	81 ± 8	53 ± 3	116 ± 12	67 ± 5	90 ± 4

	D5	D6	M3	M4	M5	M6	G	M7	M8	AD
Mg	39965 ± 804	49033 ± 1358	79118 ± 1208	79029 ± 1574	40311 ± 3262	24576 ± 2811	26818 ± 414	95894 ± 5025	90153 ± 2866	87909 ± 6727
Al	2142 ± 48	2254 ± 131	2352 ± 43	3792 ± 350	2614 ± 102	2521 ± 495	2318 ± 105	2014 ± 7	2152 ± 319	1947 ± 104
Mn	587 ± 25	884 ± 23	951 ± 5	1141 ± 87	542 ± 35	305 ± 37	7530 ± 399	745 ± 40	871 ± 34	1044 ± 76
Fe	5084 ± 425	3663 ± 242	4413 ± 268	3541 ± 143	3353 ± 199	4472 ± 292	14386 ± 45	5506 ± 750	9784 ± 209	9012 ± 568
Co	8.2 ± 0.9	7.6 ± 0.7	11.7 ± 1.1	11.3 ± 5.3	BDL	BDL	5.2 ± 0.5	3.1 ± 0.1	1.2 ± 0.5	2.7 ± 0.2
Ni	212 ± 22	161 ± 3	135 ± 12	172 ± 12	140 ± 19	114 ± 15	66 ± 6	965 ± 57	157 ± 5	267 ± 15
Cu	558 ± 84	699 ± 26	915 ± 19	1256 ± 67	682 ± 66	375 ± 23	341 ± 34	1644 ± 42	1118 ± 50	1801 ± 14
Zn	3514 ± 282	2235 ± 216	1800 ± 271	4040 ± 1354	2240 ± 167	2550 ± 348	1421 ± 150	1940 ± 427	2245 ± 768	1347 ± 353
Sr	164 ± 12	217 ± 15	88 ± 6	193 ± 2	172 ± 13	306 ± 51	317 ± 29	137 ± 8	420 ± 6	113 ± 13
Ba	58 ± 7	70 ± 10	46 ± 1	159 ± 19	127 ± 16	185 ± 5	870 ± 8	41 ± 14	145 ± 11	77 ± 3

(cont. on the next page)

Tablo 3.8.(cont.)

	M9	T7	M10	M11	U2	U3	DY11	U4	BL	U5
Mg	77736 ± 2959	69353 ± 3664	61205 ± 4846	77183 ± 1355	54265 ± 4830	38617 ± 2438	29371 ± 323	54728 ± 2921	34357 ± 2300	67845 ± 3643
Al	3486 ± 157	2896 ± 56	19348 ± 1475	3556 ± 154	1441 ± 169	1344 ± 94	1179 ± 42	2629 ± 80	1137 ± 66	1342 ± 80
Mn	945 ± 28	769 ± 46	810 ± 51	987 ± 15	653 ± 27	433 ± 9	389 ± 14	667 ± 26	488 ± 12	715 ± 40
Fe	8293 ± 330	5393 ± 234	9793 ± 684	13505 ± 1973	3098 ± 318	2668 ± 242	3255 ± 381	8765 ± 973	3518 ± 349	3482 ± 359
Co	25.7 ± 5.7	17.1 ± 0.8	26.8 ± 1.8	19.3 ± 0.4	11.9 ± 1.2	10.5 ± 0.7	5.8 ± 0.8	8.6 ± 1.5	3.0 ± 0.2	6.1 ± 0.2
Ni	320 ± 17	319 ± 21	414 ± 1	324 ± 19	54 ± 3	78 ± 8	54 ± 3	178 ± 5	27 ± 3	43 ± 5
Cu	1184 ± 47	1044 ± 111	1055 ± 125	1168 ± 31	598 ± 125	348 ± 39	176 ± 59	389 ± 98	294 ± 67	219 ± 32
Zn	83 ± 7	1429 ± 49	4668 ± 154	3274 ± 558	4852 ± 105	7026 ± 726	1864 ± 95	4103 ± 540	2523 ± 597	4398 ± 284
Sr	228 ± 22	241 ± 9	352 ± 34	464 ± 49	1425 ± 96	847 ± 17	757 ± 63	705 ± 78	767 ± 0	1852 ± 113
Ba	236 ± 26	189 ± 16	219 ± 16	253 ± 21	337 ± 15	255 ± 16	242 ± 9	188 ± 19	304 ± 12	294 ± 3

	U6	I1	AY1	I2	I3	I4	I5	I6	I7	I8
Mg	50640 ± 9569	28310 ± 566	45032 ± 920	43535 ± 1945	13732 ± 862	53699 ± 1785	55270 ± 3594	28295 ± 733	33420 ± 213	32184 ± 1016
Al	2811 ± 153	24841 ± 1129	7392 ± 378	2578 ± 62	2119 ± 720	3575 ± 217	3449 ± 316	3984 ± 526	3616 ± 432	1077 ± 108
Mn	561 ± 10	915 ± 120	908 ± 21	952 ± 145	74 ± 18	831 ± 36	728 ± 80	333 ± 112	342 ± 74	311 ± 41
Fe	5206 ± 363	6535 ± 461	3103 ± 470	14262 ± 2177	843 ± 135	261329 ± 43767	5201 ± 486	5955 ± 752	4716 ± 641	3434 ± 181
Co	8.2 ± 1.2	BDL	BDL	BDL	BDL	BDL	BDL	BDL	BDL	5.2 ± 0.2
Ni	75 ± 3	197 ± 10	152 ± 1	214 ± 34	51 ± 8	133 ± 9	383 ± 13	83 ± 3	177 ± 18	95 ± 3
Cu	450 ± 23	591 ± 2	695 ± 2	518 ± 13	128 ± 12	1112 ± 46	820 ± 69	396 ± 25	553 ± 17	325 ± 14
Zn	31994 ± 65	2300 ± 235	1213 ± 87	9421 ± 782	1017 ± 145	7231 ± 879	5551 ± 838	8288 ± 953	1372 ± 349	1071 ± 241
Sr	1273 ± 84	249 ± 6	143 ± 3	694 ± 148	299 ± 247	194 ± 15	184 ± 67	247 ± 31	164 ± 3	420 ± 17
Ba	294 ± 13	126 ± 4	164 ± 10	271 ± 48	111 ± 9	155 ± 24	158 ± 31	172 ± 5	87 ± 10	124 ± 11

	I9	I10	BO	SF	AF	AN	T8	KI1	BN1	M12
Mg	8238 ± 524	32945 ± 343	56941 ± 1761	29358 ± 300	35384 ± 2811	15271 ± 810	57026 ± 3640	22496 ± 550	13193 ± 536	79561 ± 588
Al	522 ± 25	2025 ± 213	2941 ± 282	2662 ± 604	34795 ± 2660	1003 ± 143	1539 ± 857	1133 ± 249	1698 ± 558	2036 ± 3
Mn	150 ± 10	341 ± 16	682 ± 25	306 ± 5	982 ± 99	189 ± 43	416 ± 6	401 ± 32	263 ± 8	659 ± 1
Fe	4287 ± 362	3258 ± 355	12628 ± 1386	5379 ± 0	4682 ± 968	6882 ± 699	1796 ± 510	4292 ± 317	1849 ± 479	4849 ± 7
Co	BDL	4.3 ± 0.4	7.8 ± 0.5	3.8 ± 0.0	20.3 ± 0.4	0.5 ± 0.0	1.0 ± 0.2	3.6 ± 0.5	2.7 ± 0.2	13.1 ± 0.2
Ni	BDL	89 ± 10	233 ± 4	62 ± 4	203 ± 5	28 ± 3	152 ± 13	22 ± 2	31 ± 2	136 ± 0
Cu	176 ± 12	386 ± 10	849 ± 11	358 ± 21	661 ± 35	129 ± 12	232 ± 1	287 ± 8	178 ± 24	580 ± 11
Zn	1991 ± 251	892 ± 24	1378 ± 124	881 ± 151	1382 ± 196	1813 ± 33	2575 ± 402	2957 ± 891	3377 ± 133	677 ± 1
Sr	23 ± 5	373 ± 31	262 ± 20	86 ± 19	315 ± 34	114 ± 12	165 ± 17	498 ± 31	185 ± 17	285 ± 0
Ba	62 ± 2	89 ± 4	82 ± 6	65 ± 17	134 ± 3	43 ± 5	123 ± 22	141 ± 7	87 ± 10	87 ± 0

(cont. on the next page)

Tablo 3.8.(cont.)

	KI2	KI3	T9	BN2	BN3	BT1	BT2	BL	T10	S2
Mg	41731 ± 582	34868 ± 960	67074 ± 2648	10838 ± 75	12210 ± 505	10664 ± 281	9487 ± 47	61478 ± 1269	66079 ± 5094	27441 ± 1222
Al	9179 ± 615	2304 ± 364	359 ± 64	684 ± 192	1729 ± 466	519 ± 40	533 ± 103	764 ± 119	571 ± 63	3494 ± 65
Mn	321 ± 30	527 ± 20	3441 ± 177	369 ± 22	492 ± 38	220 ± 2	209 ± 17	1793 ± 106	4231 ± 74	1141 ± 47
Fe	2182 ± 7	2976 ± 113	2427 ± 63	3655 ± 7	2827 ± 615	1575 ± 235	1337 ± 298	2902 ± 100	7496 ± 444	5600 ± 920
Co	17.3 ± 0.9	13.0 ± 0.5	6.7 ± 0.5	2.6 ± 2.8	4.4 ± 0.5	3.8 ± 0.6	BDL	BDL	BDL	24.8 ± 2.6
Ni	44 ± 2	61 ± 4	159 ± 10	36 ± 4	33 ± 1	24 ± 4	15 ± 2	73 ± 7	90 ± 9	99 ± 7
Cu	518 ± 115	461 ± 22	243 ± 27	128 ± 11	90 ± 3	128 ± 11	120 ± 4	241 ± 29	252 ± 25	213 ± 8
Zn	2447 ± 430	14302 ± 16414	1860 ± 42	1229 ± 126	1336 ± 56	5374 ± 168	2471 ± 410	3574 ± 38	10703 ± 1175	1867 ± 178
Sr	221 ± 31	295 ± 20	350 ± 12	150 ± 12	243 ± 42	138 ± 37	153 ± 7	218 ± 19	349 ± 53	252 ± 18
Ba	63 ± 7	142 ± 67	339 ± 44	146 ± 34	86 ± 8	105 ± 69	70 ± 7	262 ± 18	291 ± 28	350 ± 27

	T11	E	I10	I11	I12	U7
Mg	51972 ± 982	36243 ± 1934	22618 ± 716	64362 ± 2752	27298 ± 1652	24062 ± 1593
Al	623 ± 42	1106 ± 27	1581 ± 50	5336 ± 338	853 ± 210	1398 ± 553
Mn	2356 ± 55	357 ± 11	536 ± 17	1063 ± 201	1005 ± 112	544 ± 12
Fe	11636 ± 880	2215 ± 144	14520 ± 460	14315 ± 2415	8897 ± 1826	1833 ± 80
Co	BDL	BDL	5.0 ± 0.2	18.7 ± 1.5	8.7 ± 0.0	BDL
Ni	65 ± 2	62 ± 5	93 ± 3	2021 ± 336	85 ± 15	30 ± 3
Cu	202 ± 4	307 ± 3	295 ± 9	1782 ± 560	212 ± 39	101 ± 9
Zn	3841 ± 129	2520 ± 366	3868 ± 123	2646 ± 319	2496 ± 210	2089 ± 224
Sr	361 ± 45	251 ± 16	131 ± 4	92 ± 15	268 ± 12	721 ± 36
Ba	190 ± 10	173 ± 2	153 ± 5	53 ± 21	181 ± 19	139 ± 11

As mentioned earlier, since the collected honey samples were received in different geographic regions of the Turkey from east to west, any clustering of the samples based on their regions was investigated. For this purpose, the same four scenarios as mentioned before were tested. The first combination was made with two groups which consist of honey samples from Trakya, Kayseri, Malatya and Sivas. The score plot of the PC1 versus PC2 is shown in Figure 3.42. The first and second principal components explained 58 % of the variation of the data.

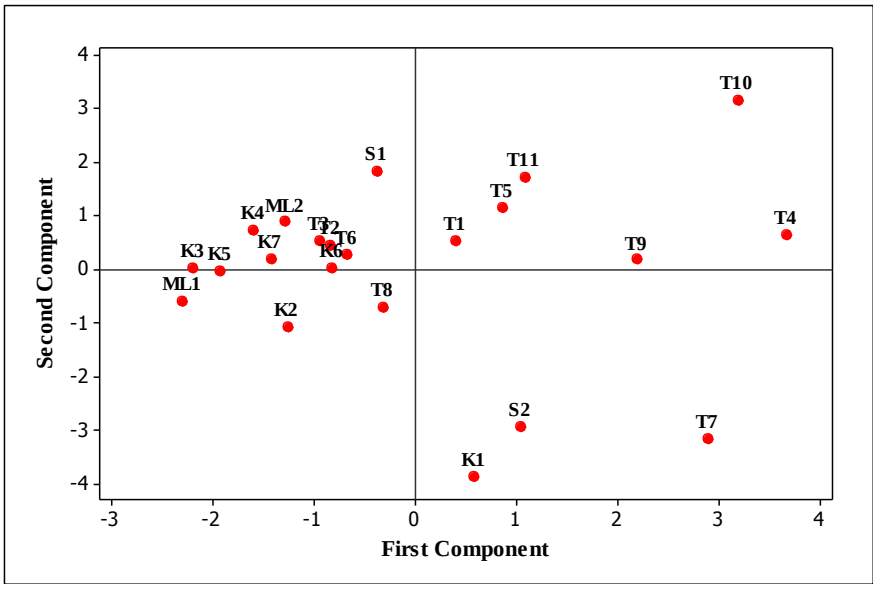

Figure 3.42. The score plot of the first component versus the second component for honey samples from Trakya, Kayseri, Malatya and Sivas using metal concentration data.

When PC score plot is examined, most of the Trakya samples (seven of eleven) were characterized based on the first component. Likely samples from east of Turkey except S2 and K1 were differentiated according to first component and located on the negative part. HCA method was also applied to same samples in order to see the relationship between them. (Figure 3.43).

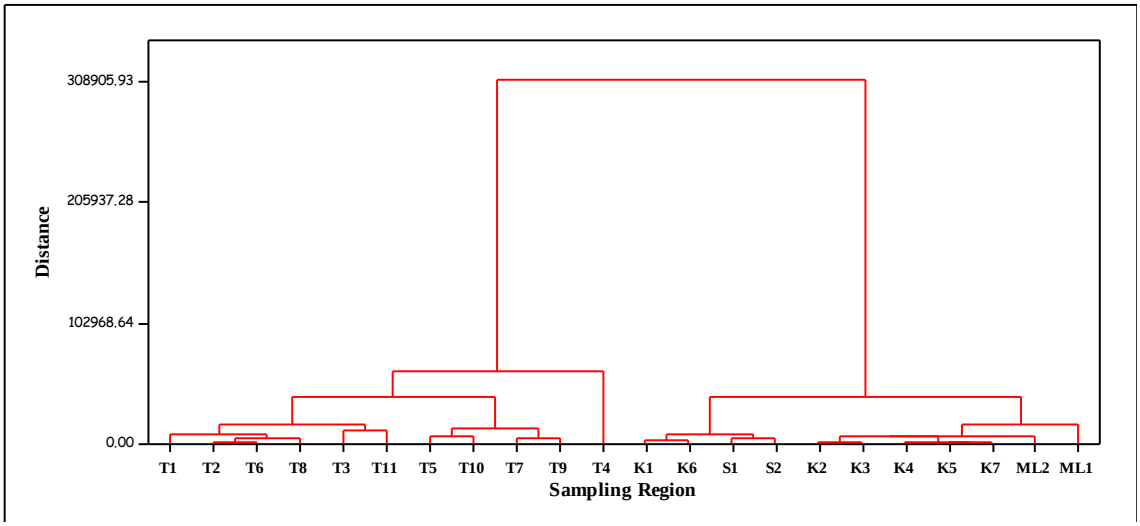

Figure 3.43. Dendrogram for honey samples from Trakya, Kayseri, Malatya and Sivas using metal concentration data.

The dendrogram supported strongly PCA result, two main clusters were observed and the classes were clustered separately as east and west samples. In addition the sample coded as T4 were characterized separately but in its region class. We can say that metal concentration data were successful in separating samples from east and west.

The next combination was made with the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. The score plot of PC1 versus PC2 is shown in Figure 3.44. The two PC's explain approximately 58 % of the total variance of the data. According to the plot, the exact grouping according to the geographical origin cannot be observed; however the samples from same cities were located very close and also Urfa samples were separated from others.

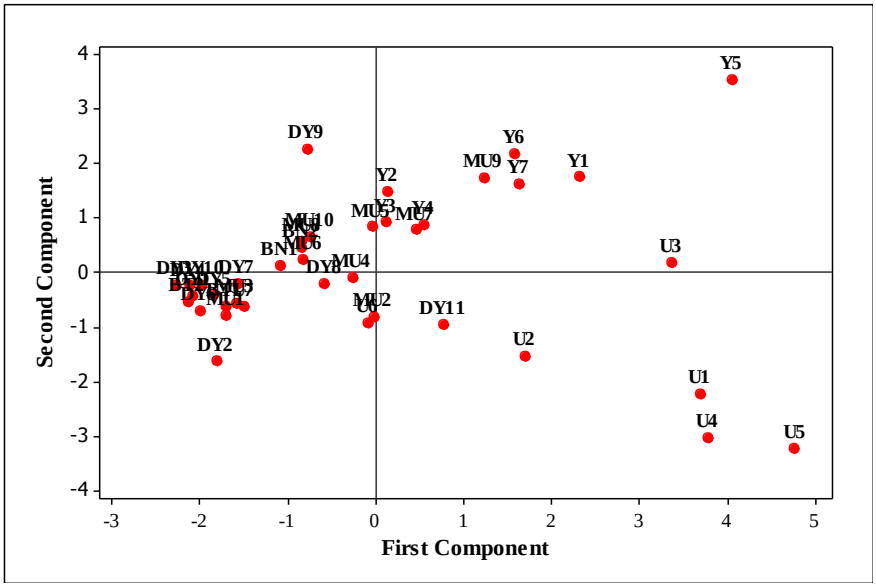

Figure 3.44. The score plot of the first component versus the second component for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using metal concentration data.

In order to see the closeness of the honey samples, HCA dendrogram was drawn using the metal concentration data directly (Figure 3.45).

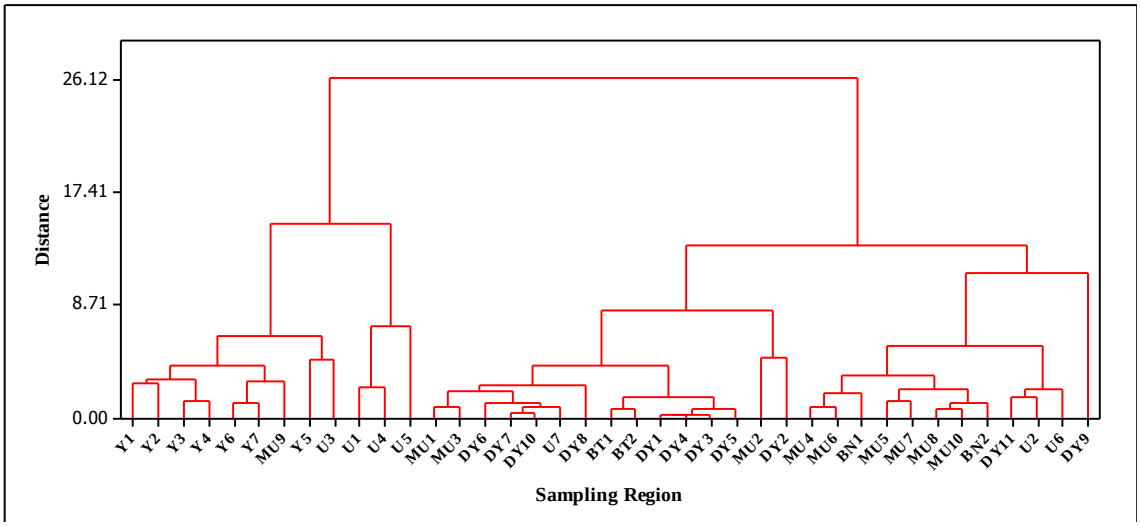

Figure 3.45. Dendrogram for honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using metal concentration data.

HCA result confirmed that the samples constructed separate clusters according to sampling cities. The differentiation of Yüksekova samples were very obvious from others. In addition Diyarbakır samples formed small separate clusters also. The

differentiation was not fully clear but it can be said that that classification was not failed with metal concentration data.

The third trial for geographical grouping was again performed with the honey samples which were İzmir, Datça, Marmaris and one sample from Antalya as mentioned previously. All of İzmir samples were not digested and thirteen of them were selected for analysis with ICP-MS because of limited time and in order to equalize the number of samples in groups. The plot for the first two principal components which explained about 50% of the variation in the data is illustrated in Figure 3.46.

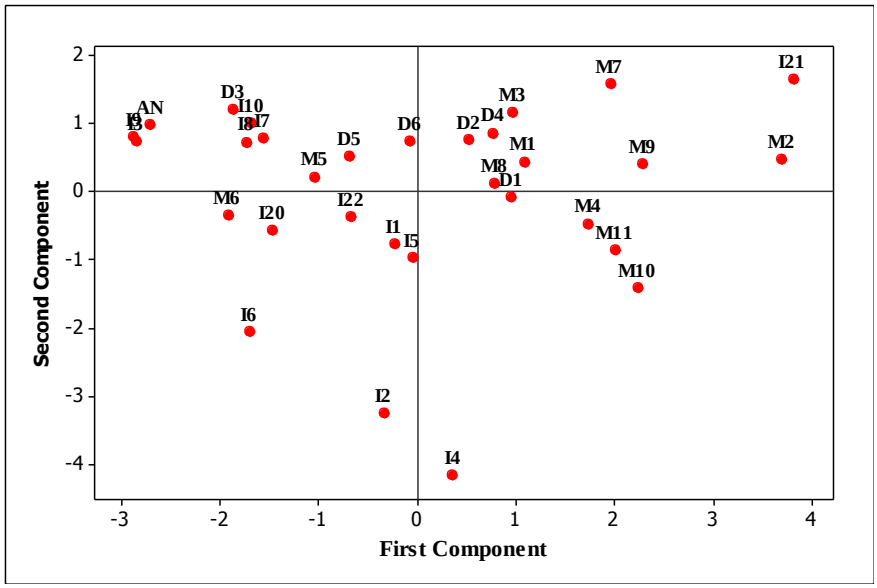

Figure 3.46. The score plot of the first component versus the second component for honey samples from İzmir, Datça, Marmaris and Antalya using metal concentration data.

The plot showed us that İzmir samples were differentiated according to first component and located negative part. On the other hand most of Datça and Marmaris samples were charactherized based on the second component and placed on the positive part of it. Also, HCA dendrogram shown in Figure 3.47 was generated to investigate the similarities between the samples.

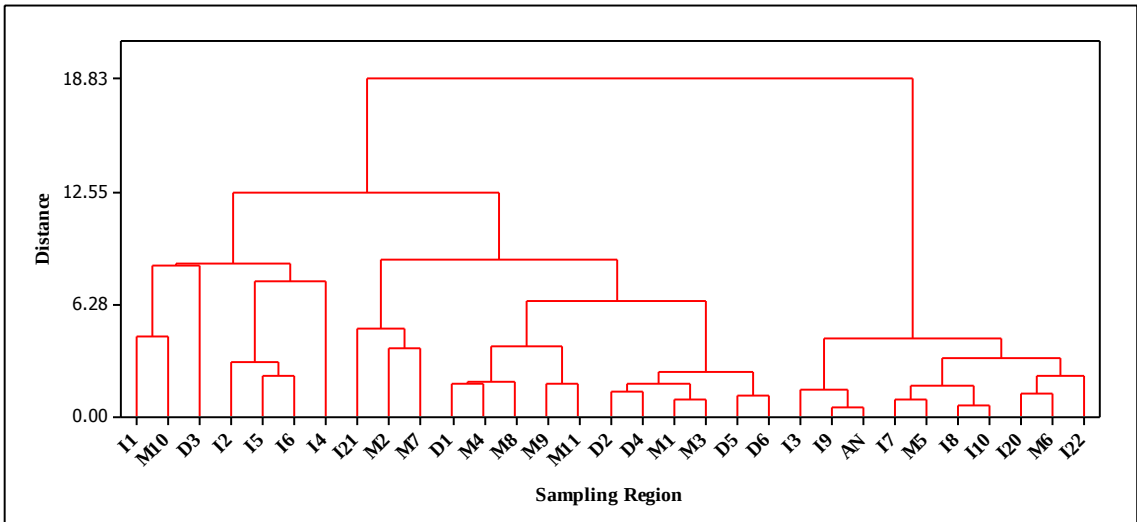

Figure 3.47. Dendrogram for honey samples from İzmir, Datça, Marmaris and Antalya using metal concentration data.

As can be seen from the dendrogram the samples constructed two main clusters and some of İzmir samples were classified in the cluster on the right side. Datça and Marmaris samples were not differentiated since they were located very close to each other on the map.

Lastly the classification of honey samples from Trakya, Muş, Bingöl and Bitlis was investigated and the PCA plot is depicted in Figure 3.48. The resulted two-component graphs explain 62 % of total variance.

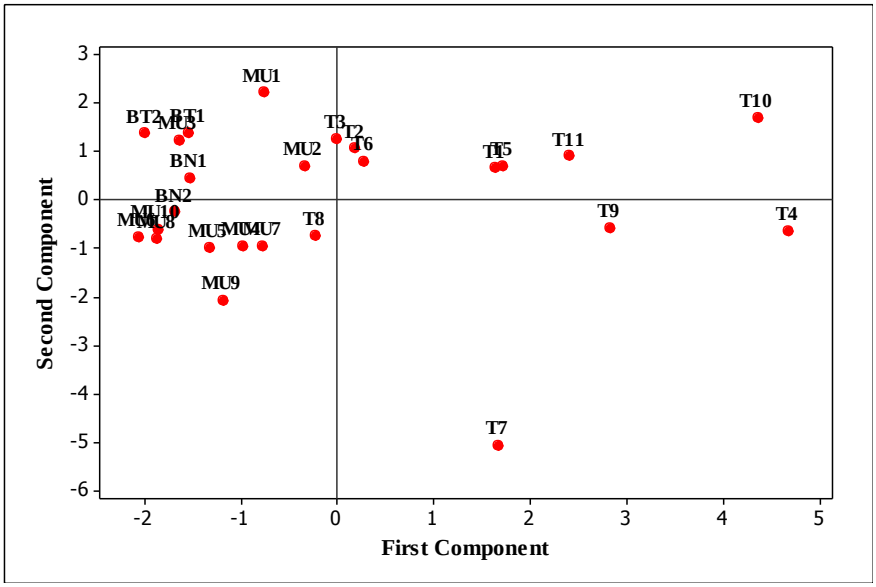

Figure 3.48. The score plot of the first component versus the second component for honey samples from Trakya, Muş, Bingöl and Bitlis using metal concentration data.

The metal content data is also succesful in differentiate the samples from east and west of the country. Again the first component distinguishes two classes. The result of cluster analysis was illustrated in dendrogram in Figure 3.49.

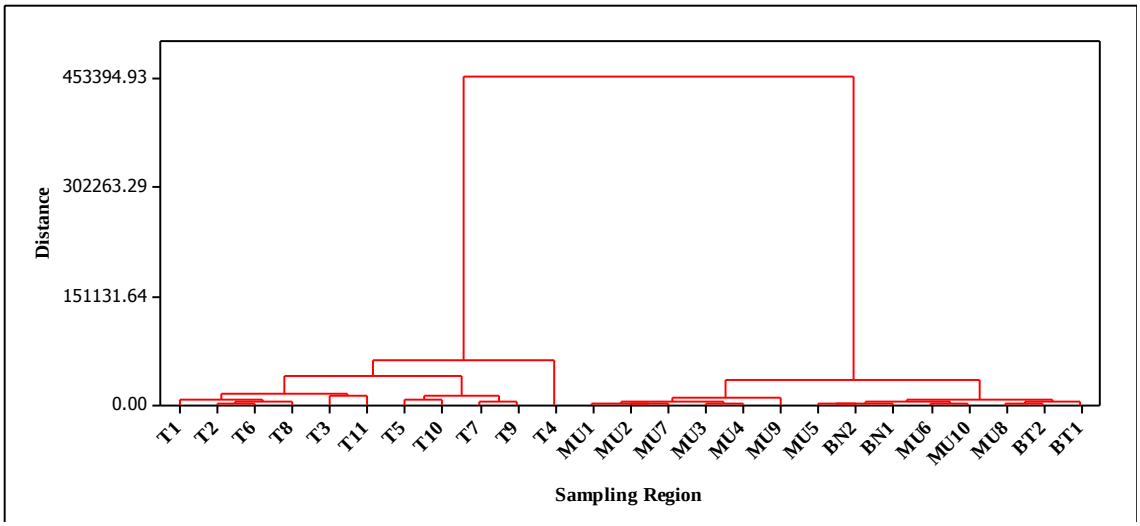

Figure 3.49. Dendrogram for honey samples from Trakya, Muş, Bingöl and Bitlis using metal concentration data.

Also the dendrogram supports discrimination of classes and the samples formed perfect two main clusters as east and west. In conclusion metal concentration data were also successful in differentiation samples from different regions.

3.2.4. Classification Studies with Supervised Methods

Following the classification studies with PCA and HCA, SIMCA is also used for metal concentration data. The same combinations of grouping were investigated and the first grouping was figured out with the samples from Trakya, Kayseri, Malatya and Sivas. The Cooman's plot for honey samples is exhibited in Figure 3.50.

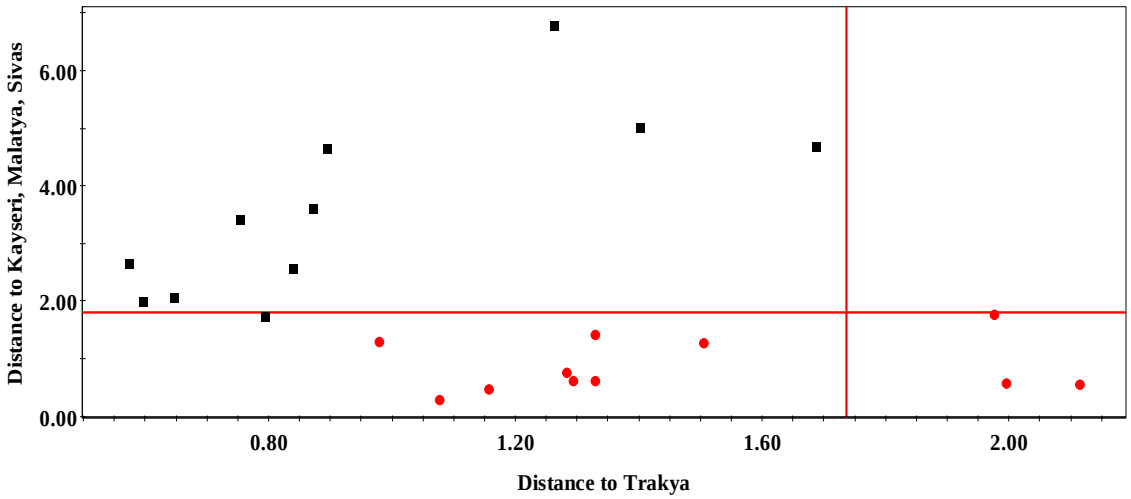

Figure 3.50. Cooman's plot for classification of honey samples from Trakya, Kayseri, Malatya, Sivas using metal concentration data ((■)Trakya and (•) Kayseri, Malatya, Sivas)

When Cooman's plot is examined, all of Trakya samples were on their own class region but the samples from Kayseri Malatya and Sivas samples were overlap region. This results means the samples were differentiated other than the samples were similar. The next combination was made with the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. The results obtained from SIMCA are on the plot in Figure 3.51.

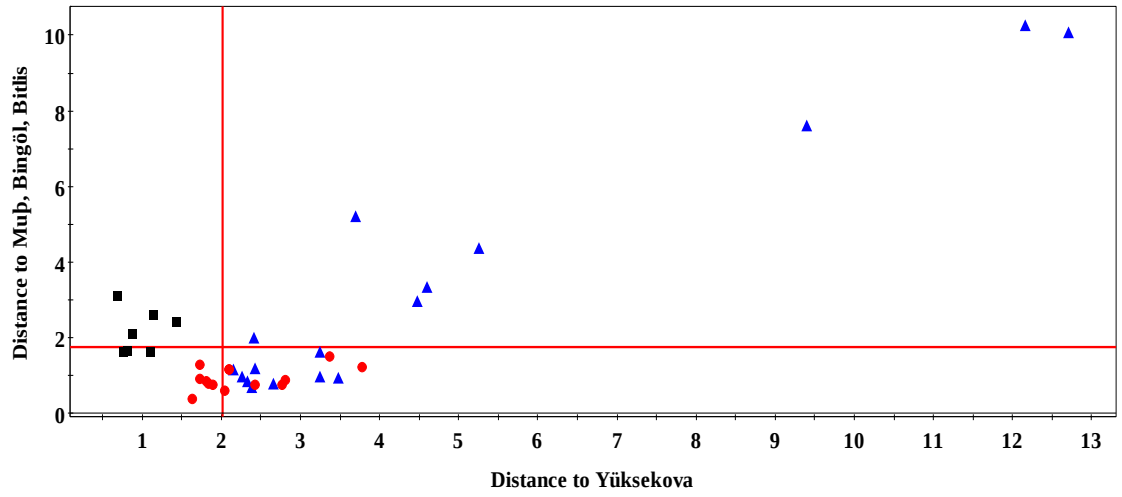

Figure 3.51. Cooman's plot for classification of honey samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa using metal concentration data (■:Yüksekova ; ●: Muş, Bingöl, Bitlis; ▲: Diyarbakır, Urfa).

The plot showed us that most of Yüksekova samples were differentiated obviously according to metal concentration data and the half of Diyarbakır and Urfa samples were separated in outlier region, on the contrary the other were classified with Muş, Bingöl Bitlis samples. The next trial was formed with honey samples from İzmir, Datça, Marmaris and Antalya and the resulting plot by SIMCA is illustrated in Figure 3.52.

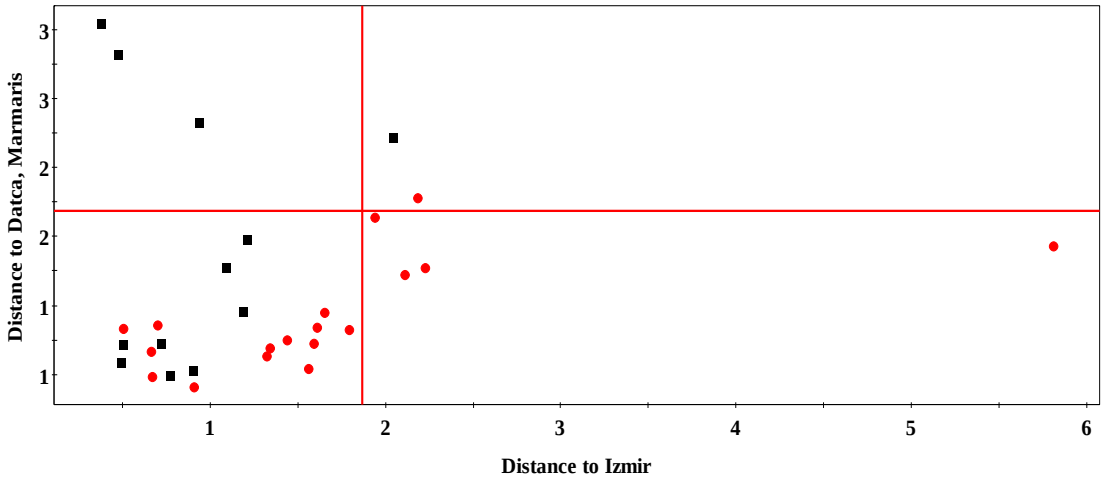

Figure 3.52. Cooman's for classification of honey samples from İzmir, Datça, Marmaris and Antalya using metal concentration data ((■) İzmir, (•) Marmaris, Datça, Antalya).

As can be seen from the Figure 3.52. most of the honey samples were not discriminated according to geographical origin, all samples were in overlap region meaning the software identified samples as a class. The last examination for geographical examination was constructed with honey samples from Trakya, Muş, Bingöl and Bitlis. The plot in Figure 3.53 was obtained after analysis by SIMCA.

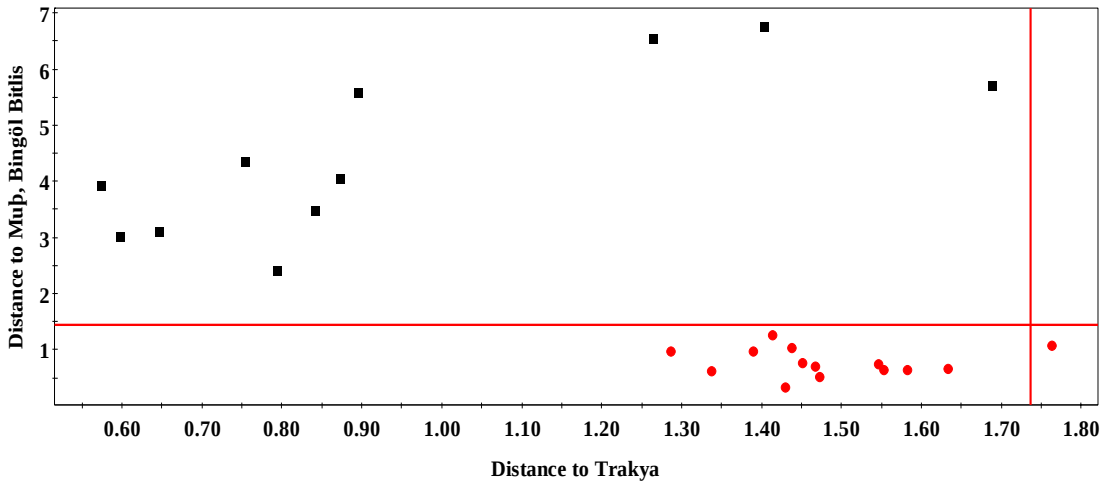

Figure 3.53. Cooman's for classification of honey samples from Trakya, Muş, Bingöl and Bitlis using metal concentration data (■) Trakya, (●) Muş, Bingöl and Bitlis).

As shown in Cooman's plot mainly, all honey samples from Trakya were classified, conversely all of Muş, Bingöl and Bitlis samples were in overlap region. In conclusion metal concentration data was not successful in classification using supervised method such as SIMCA like molecular spectrometric data. The metal concentration data was not sufficient for SIMCA to separate honey samples according to sampling regions.

CHAPTER 4

CONCLUSION

In this thesis, it is aimed to develop classification models of honey produced in Turkey based on geographical origin via atomic and molecular spectrometry. The honey samples were taken from different regions of Turkey and then they were scanned with molecular spectrometric methods (FTIR-ATR, and fluorescence spectrophotometer with synchronous fluorescence mode and 3D excitation emission mode). Afterwards, unsupervised (PCA, HCA) and supervised (SIMCA) methods were used for the classification of honey samples. Finally, the metal content of honey samples was investigated through ICP-MS measurements and the same methods were again applied for classification based on geographical origin.

Differentiation of honey samples was examined based on their geographical origins. For this purpose, four different scenarios were tested. The first one was based on the samples from Trakya, Kayseri, Malatya and Sivas and the second one was established on the samples from Yüksekova, Muş, Bingöl, Bitlis, Diyarbakır and Urfa. The third group was formed with the honey samples collected from İzmir, Datça and Marmaris, lastly, fourth group was figured with samples from Trakya, Muş Bingöl and Bitlis. Successful differentiations were obtained even with samples from neighbor regions (İzmir and Datça, Marmaris) by processing molecular spectrometric data. The most winning separation according to geographical origin was obtained with the samples Trakya, Muş Bingöl and Bitlis because of their far locations. Three of the molecular spectrometric data discriminated the honey samples with four combinations.

In addition, the honey samples were digested and the metal concentrations (Mg, Al, Mn, Fe, Co, Ni, Cu, Zn, Sr, Ba) were determined in order to process and use for classification. The differentiation with honey samples from far regions were achieved using unsupervised techniques.

In conclusion, molecular spectrometry gave better classification results based on geographical origin compared to the results obtained with atomic spectrometry. Molecular spectrometry is more advantageous for the classification of honey samples in the case of saving time, saving chemicals and ease of usage.

REFERENCES

Ajtony, Z., Bencs, L., Haraszi, R., Szigeti, J., Szoboszlai, N. 2007. Study on the simultaneous determination of some essential and toxic trace elements in honey by multi-element graphite furnace atomic absorption spectrometry. *Talanta* 71(2): 683-690.

Arvanitoyannis, I.S., Chalhoub, C., Gotsiou, P., Lydakis-Simantiris, N., Kefalas, P. 2005. Novel Quality Control Methods in Conjuction with Chemometrics (Multivariate Analysis) for Detecting Honey Authenticity. *Critical Reviews in Food Science and Nutrition* 45: 193-203.

Baroni, M.V., Arrua, C., Nores, M.L., Faye, P., Díaz, M.P., Chiabrando, G.A, Wunderlin, D.A. 2009. Composition of honey from Córdoba (Argentina): Assessment of North/South provenance by chemometrics. *Food Chemistry* 114: 727-733.

Beebe, K.R., Pell, R.J., Seasholtz, M.B. 1998. *Chemometrics, a practical guide.* (Wiley-Interscience: John Wiley & Sons, Inc.).

Berrueta, L.A, Alonso-Salces, R.M., Heberger, K., 2007. Supervised pattern recognition in food analysis. *Journal of Chromatography A* 1158(1-2): 196-214.

Bertelli, D., Plessi, M., Sabatini, A.G., Lolli, M., Grillenzoni, F. 2007. Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS). *Food Chemistry* 101: 1565-1570.

Brereton R.G. 2003. *Chemometrics, data analysis for the laboratory and chemical plant*, (Wiley: John Wiley & Sons Ltd.).

Cajka, T., Hajslova, J., Pudil, F., Riddellova, K. 2009. Traceability of honey origin based on volatiles pattern processing by artificial neural networks. *J. of Chromatog. A* 1216(9): 1458-1462.

Camina, J.M., Cantarelli, M.A., Lozano, V.A., Boeris, M.S., Irimia, M.E., Gil, R.A., Marchevsky, E.J. 2008. Characterization of Honey Produced in La Pampa Argentina, from Their Elemental Content, Using Inductively Coupled Plasma Optical Emission Spectrometry. *Journal of Agricultural Research* 47(2): 102-107.

Chudzinska, M., Baralkiewicz, D. 2010. Estimation of honey authenticity by multielement characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics. *Food and Chemical Toxicology* 48(1): 284-290.

Conti, M.E., Stripeikis, J., Campanella, L., Cucina, D., Tudino, M.B., 2007. Characterization of Italian honeys (Marche Region) on the basis of their mineral content and some typical quality parameters. *Chemistry Central Journal*, 1(14): 1-10.

Corbella, E., Cozzolino, D. 2006. Classification of the floral origin of Uruguayan honeys by chemical and physical characteristics combined with chemometrics. *Food Science and Technology* 39: 534-539.

De Luca, M., Oliverio, F., Ioelea, D., Husson, G.P., Ragno, G. 2008. Monitoring of water quality in South Paris district by clustering and SIMCA. *Intern. J. Environ. Anal. Chem.* 88(15): 1087-1105.

Devillers, J., Dor, J.C., Marenco, M., Poirier-Duchne, F., Galand, N., Viel, C., 2002. Chemometrical analysis of 18 metallic and nonmetallic elements found in honeys sold in France. *J. Agric. Food Chem.* 50(21): 5998-6007.

Dias, L.A., Peres, A.M., Boas, M.V., Rocha, M.A., Estevinho, L., Machado, A.S.C. 2008. An electronic tongue for honey classification. *Microchim Acta* 163(1-2): 97-102.

Downey, G., Hussey, K., Kelly, J.D., Walshe, T.F., Martin, P.G. 2005. Preliminary contribution to the characterization of artisanal honey produced on the island of Ireland by palynological and physico-chemical data. *Food Chem.* 91: 347-354.

Einax, J., Hutzinger, O., Christy, A.A., 1995. *Chemometrics in Environmental Chemistry: Applications*, (Springer-Verlag Berlin Heidelberg).

Frazzoli, C., D'Ilio, S., Bocca, B. 2007. Determination of Cd and Pb in honey by SF-ICP-MS: Validation figures and uncertainty of results. *Analytical Letters* 40(10):1992-2004.

Garcia-Villar, N., Hernandez-Cassou, S., Javier, S. 2007. Characterization of Wines through the Biogenic Amine Contents Using Chromatographic Tecniques and Chemometric Data Analysis. *J. Agric. Food Chem.* 55(18): 7453-7461.

Harvey, D. 2000. *Modern Analytical Chemistry*, (The McGraw-Hill Companies, USA),

Hennessy, S., Downey, G., O'Donnell C. 2008. Multivariate Analysis of Attenuated Total Reflection–Fourier Transform Infrared Spectroscopic Data to Confirm the Origin of Honeys. *Applied Spectroscopy* 62: 1115-1123.

http://en.wikipedia.org/wiki/FTIR (accessed May 15, 2011).

Jarvis, K.E., Gray, A.L., Houk, R.S. 1992. *Handbook of Inductively Coupled Plasma Mass Spectrometry*, (Blackie, Chapman and Hall, New York).

Karoui, R., Dufour, E., Bosset, J.O., Baerdemaeker, J.De, 2007. The use of front face fluorescence spectroscopy to classify the botanical origin of honey samples produced in Switzerland. *Food Chemistry* 101(1): 314-323.

Kelly, J.D., Petisco, C., Downey, G. 2006. Application of Fourier Transform Midinfrared Spectroscopy to the Discrimination between Irish Artisanal Honey and

Such Honey Adulterated with Various Sugar Syrups. *J. Agric. Food Chem.* 54: 6166-6171.

Lolli, M., Bertelli, D., Plessi, M., Sabatini, A.G., Restani, C. 2008. Classification of Italian Honeys by 2D HR-NMR. *J. Agric. Food Chem*., 56: 1298-1304.

Nozal Nalda, M.J., Bernal Yagüe, J.L., Diego Calva, J., Martin Gomez, M. T. 2005. Classifying honeys from the Soria Province of Spain via multivariate analysis. *Anal Bioanal Chem* 382: 311-319.

Osman, K.A., Al-Doghairi, M.A., Al-Rehiayani, S., Helal, M.I.D. 2007. Mineral contents and physicochemical properties of natural honey produced in Al-Qassim region, Saudi Arabia. *International Journal of Food, Agriculture and Environment* 5: 142-146.

Otto, M., 1999. *Chemometrics statistics and computer Application in Analytical Chemistry*, (John Wiley & Sons Inc.).

Pohl, P., Sergiel, I., Stecka, H. 2009a. Determination and Fractionation of Metals in Honey. *Critical Reviews in Analytical Chemistry* 39(4): 276-288.

Pohl, P. 2009b. Determination of metal content in honey by atomic absorption and emission spectrometries. *Trends in Analytical Chemistry* 28(1): 117-128.

Rashed, M.N., Soltan, M.E., 2004. Major and trace elements in different types of Egyptian mono-floral and non-floral bee honeys. *J. Food Compos. Anal.* 17(6): 725-735.

Ruoff, K., Luginbül, W., Künzli, R., Iglesias, M. T., Bogdanov, S., Bosset, J.O., Von Der Ohe, K., Von Der Ohe, W., Amado R., 2006a, Authentication of the Botanical and Geographical Origin of Honey by Mid-Infrared Spectroscopy. *J. Agric. Food Chem.* 54: 6873-6880.

Ruoff, K., Luginbül, W., Bogdanov, S., Bosset, J.O., Estermann, B., Ziolko, T., Amado, R. 2006b. Authentication of the Botanical Origin of Honey by Near-Infrared Spectroscopy. *J. Agric. Food Chem.* 54(18): 6867-6872.

Senyuva, H.Z., Gilbert, J., Silici, S., Charlton, A., Dal, C., Gürel, N., Cimen, D. 2009. Profiling Turkish Honeys to Determine Authenticity Using Physical and Chemical Characteristics. *J. Agric. Food Chem.* 57(9): 3911-3919.

Sharaf, M.A., Illma, D.L., Kowalski, B.R. 1986. *Chemometrics*, (John Wiley & Sons Inc.), p.

Silici, S., Uluözlü, O.D., Tüzen, M., Soylak, M. 2008. Assessment of trace element levels in Rhododendron honeys of Black Sea Region, Turkey. *J. Hazard. Mater*. 156(1-3): 612-618.

Skoog, D.A., Holler, F.J., Nieman, T.A. 1998. *Principles of instrumental analysis*, (Philedelphia: Saunders College Publishing, Harcourt Brace College Publishers), p. 357, 383.

Smith, L.I. 2002. A tutorial on Principal Components Analysis. p 1-26.

Stanimiroval, I., Charlton, A.J., Wrobel, M.S. 2010. Multivariate discrimination of wines with respect to their grape varieties and vintages. *European Food Research and Technology* 231(5): 733-743.

Stankovska, E., Stafilov, T., Sajn, R. 2008. Monitoring of trace elements in honey from the Rebuplic of Macedonia by atomic absorption spectrometry. *Environmental Monitoring and Assessment* 142(1-3): 117-126.

Sun, D.W. 2009. *Infrared Spectroscopy for Food Quality Analysis and Control*, (Elsevier).

Terrab, A., Gonzalez, A.G., Diez, M.J., Heredia, F.J. 2003. Mineral Content and electrical conductivity of the honeys produced in Northwest Morocco and their contribution to the characterization of unifloral honeys. *J. Sci. Food Agric.* 83(7): 637-643.

Torres, R.F., Bernal, J.L.P., Bello-Lopez, M.A., Moch´on, M.C., Sanchez, J.C.J., Perez, A.G. 2005. Mineral content and botanical origin of Spanish honeys. *Talanta* 65: 686-691.

Tuzen, M., Silici, S., Mendil, D., Soylak, M. 2007. Trace element levels in honeys from different regions of Turkey, *Food Chemistry* 103(2): 325-330.

Varmuza, K., Filzmose, P. 2008. *Introduction to multivariate statistical analysis in chemometrics*, (Taylor and Francis.).

Velazquez, T.G., Revilla, G.O., de Loa, M.Z., Espinoza, Y.R. 2009. Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. *Food Research International* 42: 313-331.

Woodcock, T., Downey, G., Kelly, J. D., O'Donnell, C. 2007. Geographical Classification of Honey Samples by Near-Infrared Spectroscopy: A Feasibility Study. *J. Agric. Food Chem.* 55: 9128-9134.

VITA

PERSONAL INFORMATION

Surname, Name: YERSEL MÜŞERREF

Nationality: Turkish (TC)

Date and Place of Birth: 31.05.1980, İzmir

Marital Status: Single

Phone: +90 (232) 452 96 57

E-mail: myersel2@hotmail.com

EDUCATION

Degree Institution Year of Graduation

MS İzmir Institute of Technology 2005

BS İzmir Institute of Technology 2002

High School Şirinyer Lisesi 1998

WORK EXPERIENCE

Year Place Enrollment

2002-2011 İzmir Institute of Technology Research Assistant

1997 Hıfsızsıhha Enstitüsü Trainee

1997 DYO Inks Trainee

FOREIGN LANGUAGES

English (Advanced)

PUBLICATIONS

Müşerref Yersel, Aslı Erdem, Ahmet E. Eroğlu, Talal Shahwan, Separation of trace antimony and arsenic prior to hydride generation atomic absorption spectrometric determination, Anal. Chim. Acta, 2005, 534:293-300.