

## Comparison of Principal Component Analysis and Multidimensional Scaling Methods for Clustering Some Honey Bee Genotypes

<sup>1</sup>Irfan Ozturk, <sup>2</sup>Hikmet Orhan and <sup>1</sup>Zeki Dogan

<sup>1</sup>Department of Animal Science, Faculty of Agriculture, Harran University, Sanliurfa, Turkey

<sup>2</sup>Department of Animal Science, Faculty of Agriculture, S. Demirel University, Isparta, Turkey

**Abstract:** The aim of this study, was to make use of Cluster Analysis, Principal Component Analysis and Multidimensional Scaling for clustering and appointments of the units to the true clusters. The honey bee genotypes (*Apis mellifera* L.) belonging to 30 provinces of Turkey were clustered according to Cluster Methods cited below. From these methods, McQuitty, Single Linkage, Complete Linkage, Average Linkage and Centroid Linkage showed the similar results and the results were in good agreement for separation graphics obtained by Principal Component Analysis and Multidimensional Scaling Methods while different results were found from median, centroid linkage and k-means methods.

**Key words:** Multivariate analyses, clustering, honey bee, *Apis mellifera* L., k-means methods

### INTRODUCTION

Regarding the similarities and differences in their characteristics, living organisms constitute a multidimensional model of variations in the nature (Steffes, 2007). The variations formed by both ecologic and genetic effects between the subsections of species bring up the structure of population pertaining to the species.

For classification, population and population groups compared with each other and tried to arrange them in defined rules. The most important thing in a classification is to grouping the organisms and ranking them in a hierarchic category. Systematic zoology, used differences and similarities between organisms, put population into species or higher taxons; also, put forward an idea about biological reasons for similarities and differences of taxons (Dufrene and Legendre, 1997).

Today, the data obtained from all the research and experiments are being evaluated by various statistical methods. Similarly, for the correct classification of plants and animals as well, appropriate statistical analyses have to be used. For this aim a number of methods for cluster analysis have been developed. However, uncompromising results obtained from different methods leave the researchers in a contradiction. To overcome this difficulty, the distribution of the data at hand and the application of the method relevant to these data should be known by researcher. This is not an easy task. Therefore, in this study, by using the methods such as Multidimensional Scaling (MDS)

and Principal Component Analysis (PCA), we have investigated whether they are of help in determining the clusters.

In hierarchical clustering methods, if an element is determined to go into a cluster, it cannot go into another cluster at all. In case of some extreme values, this situation can lead to a formation of wrong clusters. It has been argued that, by way of obtaining first 2 principal component scores, the approximate scatter diagrams can be drawn and they could give to the researcher an idea about the correctness of cluster analysis results and a preview about its inference (Anderberg, 1973).

In MDS, using the matrix of lengths between n elements, with as minimum dimension as possible, the scatter of elements is aimed to be determined as close as to the original form. From the scatter diagram, the distribution of the elements in the multidimensional space is obtained. Observing this scatter diagram, it is possible to decide which clustering methods can better cluster the elements without any divergence (Johnson and Wichern, 1992; Dufrene and Legendre, 1997).

### MATERIALS AND METHODS

The data of this study, were consisted of honeybees (*Apis mellifera* L.) genotypes values obtained from 30 different locations of Turkey (Ozturk *et al.*, 1992). From each of the honeybee genotypes, 12 measures pertaining to morphological characteristics have been taken, via distance values based on these measurements, bee genotypes were tried to be cast to appropriate clusters.

Measuring scale and variable types should be correctly determined in order to have an appropriate result from applied methods. Since, the morphological characteristics of honeybees depend on measuring, in statistical analysis ratio scaling is used. However, due to having different properties and measuring units, these morphological characteristics are standardized; consequently, a standard Z-matrix is obtained (Dueck *et al.*, 2005). Using this standard Z-matrix, the distances between the paired honeybee genotypes are calculated by the use of Euclidian distances, probably the most commonly used distance measure and the most familiar, where the distance between points i and j denoted by  $d_{ij}$  is defined as:

$$d_{ij} \Rightarrow d_2(x_i, x_j) = \left[ \sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right]^{1/2}$$

where:

- $X_{ik}$  = The value of the kth variable for the ith entity
- $d_{ij}$  = The distance between observations i
- j and p = The number of variables (Everitt, 1979; Sandoz, 2006)

**Cluster methods:** Although, there are many clustering methods for the classification of data, in this study, we have applied the methods, which have a common application and take place in many software packages like SPSS (2006), Minitab Inc. (2003) and Chirpaz *et al.* (2004). These methods can be named as Single Linkage, Complete Linkage, Average Linkage, Centroid Linkage, Median Linkage, McQuitty, Ward Clustering and k-means.

**Principal Components Analysis (PCA):** Although, it is easy to make analysis in a multivariable case, inference pertaining to their results is not an easy task. In cluster analysis, there are many distance measures and methods based on these measures. Depending on either distance measure or selected method, the results of cluster analysis could be different and this can lead researcher into an uncertainty. That is why, in recent years, in cluster analysis Principal Components is mostly used. By this way, on the one hand, the number of variables is reduced; on the other hand, the correlation pattern between variables, which is negatively affecting the multi variable analysis methods, can be removed. Furthermore, it is possible to derive detailed information from the plot of observations over the first 2 principal components. The resulting diagram can give the researcher an idea about the correctness and inference of cluster analysis results (Bensmail *et al.*, 1997).

The reasons of the Principal Components could be given as follows; let us think about the position in the p-dimensional space of an X data matrix, which contains n individual for each of p variables. The data matrix can be expressed as a cloud consisting of many points, each of which are determined by an individual. Due to the absence, of an exact independence between variables, the axes of cloud-shaped geometric form will not be orthogonal to each other and on this account there will not be any definition of it. In fact, if these points are taken into ellipsoid orthogonal axes, we can derive more detailed and explanatory information. For this aim, in the transformation applied, without changing the total variance of points along the first axes, the new axes are brought into orthogonality.

In practice, since the measuring units of variables are different from each other, we do not use the  $X_{p \times n}$  dimensional raw data matrix, we will make use of  $Z_{p \times n}$  dimensional standard matrix, which is consisted of standardized values of X variables. In this case,  $T_{p \times p}$  being a transformation matrix, we have the following relation. After transformation, we obtain unrelated  $Y_{ij}$  values with each other from the related  $Z_{ij}$  values.

$$Y_{p \times n} = T'_{p \times p} \cdot Z_{p \times n}$$

In principal component analysis to account for the variance-covariance structure, instead of p variables, k principal components, which are, linear combination of them is used. The method of PCA brings about a basic structure for the ensuing multivariable statistical analysis. The mentioned analyses are multivariable regression, cluster and factor analysis (Sharma, 1996).

In PCA, the basic instrument used for the analysis is either the variance-covariance matrix ( $\Sigma$ ) or correlation matrix. Let our vector pertaining to the variables be in the form of:

$$X': \{X_1, X_2, \dots, X_p\}$$

Eigen values related to these variables are:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$$

and eigen vectors are:

$$e_i' : \{e_1, e_2, e_3, \dots, e_p\}$$

To define in brief the aforementioned eigen values and eigen vectors, let A be a pth order square matrix. Having  $\lambda$  as a constant, the characteristic equation of this matrix is formed like the following equality.

$$|A - \lambda \cdot I_n| = 0 \text{ or } \lambda^p + b_{p-1}\lambda^{p-1} + \dots + b_1\lambda + b_0 = 0$$

p roots obtained from the above p-th order polynomial are called the eigen values (or characteristic roots) of matrix A. For a matrix A to have real eigen values, it should hold the property of being a symmetrical matrix that is  $A = A^T$ , the original matrix and its transposed matrix should be equal. Otherwise its eigen values will be complex numbers.

On the other hand, if  $\lambda_1, \lambda_2, \dots, \lambda_p$  are eigen values of A, then

$$|A| = \prod_{j=1}^p \lambda_j$$

In this case, if a matrix A of order p has a rank less than p, then at least one of the eigen values of this matrix is zero and its determinant value will be zero as well.

Let, A be a matrix of order  $p \times p$  and  $\lambda$  be a constant, if a non zero vector of order  $p \times 1$  provides the following condition then this vector is said to be an eigen vector of A related with the eigen value  $\lambda$ .

$$(A - \lambda I) e = 0$$

The eigen vectors of a symmetrical matrix are orthogonal to each other and this property provides the very basis of many multivariable analysis (Chang, 1983).

Using this property, the principal components can be found via following equality:

$$Y_i = e_i'x = e_{i1}x_1 + e_{i2}x_2 + \dots + e_{ip}x_p \quad i = 1, 2, \dots, p$$

**Multidimensional Scaling (MDS) methods:** This method tries to determine the scatter of n individuals according to p variables in k-dimensional ( $k < p$ ) space.

Cluster analysis is a method that brings about the relation between n individuals depending on p variables and makes use of dendograms formed by calculated distances. However, here especially in hierarchical cluster methods an individual assigned to any cluster cannot be assigned again in later stages to a more suitable cluster taking it out from the first cluster to which it is assigned before. This point can cause wrong formations of cluster structures for some data. Therefore, to compare the structure of formed clusters it is necessary to consider the scatter of original variables in k-dimensional space (Malhotra, 1987).

As in the case of Cluster Analysis, MDS also makes use of the distances between individuals. These original distances are taken as absolute distances and put into the

Table 1: Kruskal's stress statistics

Stress coefficient	Degrees of fitness
0.20-1.00	Unfitting
0.10-0.20	Low fit
0.05-0.10	Good fit
0.025-0.05	Excellent fit
0.00-0.025	Complete fit

process. According to the mentioned distances; indicator distances, which are quite close to the original ones, are tried to be produced on the coordinate system. The goodness of fit between this indicator distances and original ones are determined by the stress measure as follows:

$$\text{Stress } 1 = [\sum_{i < k} (d_{ij} - td_{ij})^2 / \sum_{i < k} (td_{ij})^2]^{1/2}$$

The goodness of fit between resulting scatter diagrams and that of real variables in space can be most reliably tested by stress coefficient. For this reason, the fitness of determined form to the original one is possible to evaluate by the stress values reported by Ozdamar (1999) (Table 1).

Another statistical advantage of MDS method is that it does not require an assumption of a specific distribution about population. Since our data are in the rational scale, predictor configuration distances ( $td_{ij}$ ) are calculated by linear regression equations:

$$d_{ij} = a + b\delta_{ij} + e$$

These predictor distances are called inequalities (differences) and inequality values of  $td_{ij}$  are distance data that are quite close to  $d_{ij}$  values and representing  $\delta_{ij}$  values (Ozdamar, 1999).

**Hotelling-lawley trace statistics:** This method is found to be the most appropriate criterion for determining the number of cluster (Milligan and Cooper, 1985; Dincer and Ozdamar, 1993). In this method, B and W being, respectively between groups and within groups sum of squares matrices, we show the rth eigen value as  $\lambda_r$  for the product matrix of  $BW^{-1}$ . The trace of the product matrix of  $(BW^{-1})$  defined as the sum of diagonal elements is  $[\text{tr}(BW^{-1})]$ , which is equal to the sum of eigen values, gives the Hotelling- $T^2$  statistics:

$$T_o^2 = \sum_{r=1}^p \lambda_r$$

is being used Giurcaneanu and Tabus (2004). A big value of  $T_o^2$  implies the rejection of  $H_0$  hypothesis. Provided with sufficiently big values of n, using the approximation of:

$$HL = nT_0^2 \sim X_{p(k-1), \alpha}^2$$

the table values of chi-square ( $\chi^2$ ) distribution with  $p(k-1)$  degrees of freedom are used as critical value. Furthermore, having done the transformation

$$F = \frac{p(n-p-1)+2}{(k-1)p^2} \cdot T_0^2$$

the F table values with  $v_1 = (k-1)p^2$  and  $v_2 = p(n-p-1)+2$  degrees of freedom can be used as well, as a critical value (Everitt, 1979).

### RESULTS

According to the aforementioned clustering methods honeybees belonging to the 30 regions have been clustered so as to form 2, 3, 4 and 5 clusters. The results of most significant cluster structure are reported in Table 2.

Having applied PCA to the variables, the distribution of bee genotypes have been obtained according to the resulting 2 principal components as in Fig. 1. With the help of this graphic a visual prediction about the scatter

and structure of clusters has been obtained and it is shown that the bee genotypes can be divided into 4 or 5 clusters.

In a similar way, for honeybee genotypes, the MDS analysis and its scatter in the multi-dimensional space have been obtained as in Fig. 2. The analysis has been separately evaluated for 2, 3 and 4 dimensions. The related stress values obtained according to Kruskal's stress statistics are calculated, respectively as 0.07995; 0.04079 and 0.02119. Naturally, when having the scatter of genotypes evaluated as being 4-dimensioned, the stress value proved to be a minimum. However, small is the stress value of diagram obtained in multidimensional scaling analysis, the number of dimensions also is required to be a minimum. That is because when the number of dimensions increases graphical illustrations become far from being easily understandable. The preferred solution is 3 or less ( $3 \geq k$ ) dimensional one. For this reason, three-dimensional scaling of the data has been decided, resulting scale values and loci of genotypes determined to be as in Fig. 2. Moreover, the stress value derived from three dimensions is (0.04079) and the significance attached to this value is 0.025-0.05, that is the goodness of fit of configuration distances to the original ones are excellent.

Table 2: Clusters obtained according to different cluster methods

Regions	Average linkage	Ward	Complete linkage	McQuitty	Singel linkage	Median	Centroid linkage	k-Means
Boyalı (Beypazari)	1	1	1	1	1	1	1	2
Eldiven (Cankiri)	1	1	1	1	1	1	1	2
Evyaba (Erbaa)	1	1	1	1	1	1	1	2
Cakir (Erbaa)	1	1	1	1	1	1	1	3
Yazibasi (Zile)	1	1	1	1	1	1	1	2
Kocak (Zile)	1	1	1	1	1	1	1	2
Almus	1	1	1	1	1	1	1	3
Ardahan	1	1	1	1	1	1	1	3
Gundes (Hanak)	1	1	1	1	1	1	1	3
Ovundu (Cildir)	1	1	1	1	1	1	1	3
Merkez (Posof)	1	1	1	1	1	1	1	3
Yurtbasi (Posof)	1	1	1	1	1	1	1	3
Uzunali (Kirsehir)	2	2	2	2	2	2	2	1
Oymaagac (Beypazari)	2	2	2	2	2	2	2	1
Beypazari 2	2	2	2	2	2	2	2	1
Sabanozu (Cankiri)	2	2	2	2	2	2	2	1
Cifteler (Eskişehir)	2	2	2	2	2	2	2	1
Edirne	3	3	3	3	3	1	1	4
Balikesir	3	3	3	3	3	1	1	4
Canakkale	3	3	3	3	3	1	1	4
Denizli	3	3	3	3	3	1	1	4
Mugla 1	3	3	3	3	3	1	1	4
Mugla 2	3	3	3	3	3	1	1	4
Mugla 3	3	3	3	3	3	1	1	4
Simav (Kutahya)	3	3	3	3	3	3	3	4
Menemen (Izmir)	3	3	3	3	3	1	1	4
Sanliurfa	4	4	4	4	4	4	4	1
Orta Anadolu Genel	1	1	1	1	1	1	1	2
Karadeniz Gecit	1	1	1	1	1	1	1	2
Ardahan	1	1	1	1	1	1	1	3

Table 3: The tests of cluster structures by hotelling lawley trace statistics and their significance levels

No. clusters	Type of cluster	Clustering methods	Hottelling lawley trace scale	F-test statistics	Level of significance (P)
3	1	McQuitty, Ward, Single, Complete, Average and Central Linkage	107.34	67.1	0.000
	2	Medyan	61.83	38.6	0.000
	3	Ward and k-Average	137.24	85.8	0.000
4	1	McQuitty, Ward, Single, Complete, Average Linkage	256.86	9.5	0.000
	2	Central Linkage and Median	111.99	42.5	0.000
	3	k-Average	160.15	60.8	0.000
5	1	McQuitty, Median, Single, Complete, Average and Central Linkage	311.56	81.1	0.000
	2	Ward	280.06	72.9	0.000
	3	k-Average	193.04	50.3	0.000

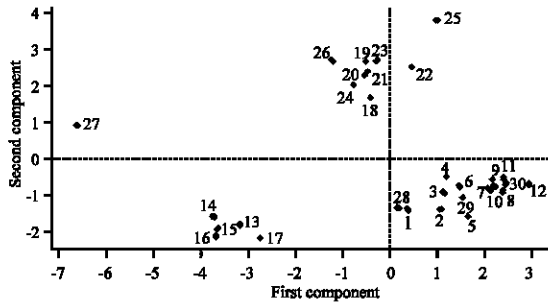


Fig. 1: Scatter diagram of the honeybee genotypes obtained from first 2 PCs

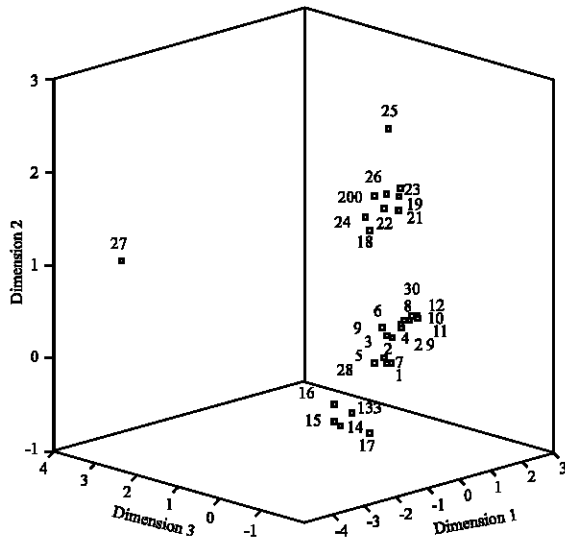


Fig. 2: Three-dimensional scatter diagram of honeybee genotypes using MDS method

Having considered Fig. 2, it is observed that the honeybee genotypes for 30 regions could be assigned to 4 or 5 clusters. These results have come out to be similar with those scatters obtained from PCA and it is proven to be no significant differences across the methods. As a matter of fact, both PCA and MDS methods have given results supporting each other for the scatter of honeybee genotypes.

In case of the number of cluster to be 4, the methods of Single linkage, Complete Linkage, mean Linkage, McQuitty and Ward come out with the best similar cluster structures. In case of the number of cluster to be 5, the methods of Single linkage, Complete Linkage, Mean Linkage, McQuitty, Central Linkage and Median come out with the best similar cluster structures. To make this point clearer, the statistics of Hotelling Lawley Trace test is applied to each of the resulting cluster structures and test results have been reported in Table 3.

### DISCUSSION

Having done the analyses and tests, many cluster structures have been eliminated. Hotelling Lawley trace test applied to some clustering methods and structures seeming most significant in Table 3 have all ended up with significant p-values. However, at the  $p < 0.001$  significance level, we have observed that the quadruple cluster provides the best fitting cluster structure formed by methods such as McQuitty, Single Linkage, Complete Linkage, Average Linkage, Central Linkage and Ward with highest F-value ( $F = 97.5$ ) and producing graphics complying with those of PCA and MDS analysis (Table 3).

Furthermore, when the genotypes of honeybees are considered, Sanliurfa bees, which take place within the location number 27 and forming the cluster 4 by itself, do not show any similarities with those of other regions. This result confirms the findings of Ruttner (1988), regarding the bees of Southeast Anatolian region to be one of the 6 ecotypes of Iran bees (*A.m. meda*).

### CONCLUSION

Using only cluster analysis by itself is not sufficient in assigning the individuals into the right clusters and there so it can be said that use should be made of other aforementioned methods. In this study, however, though they have given sufficient explanatory preliminary information in determining the number of clusters, it is

established that in different applications they do not always provide the same explanatory information. In this situation, when the subject matter is the clustering of individuals, it can be said that the following the process below could prove to be useful.

- First of all one should bring about a profile of individuals by a cluster analysis
- In the next stage, by Multidimensional Scaling or Principal Components Analysis, the number of clusters and suitable clustering methods for assigning individuals to the correct clusters has to be determined
- In the third stage, by making use of statistics such as Hotelling Lawley trace value, the best feasible cluster structure and significance level (P) should be determined

## REFERENCES

- Anderberg, M.R., 1973. Cluster Analysis for Applications. 1st Edn. Academic Press, New York, pp: 10-24. ISBN: 0-12-057650-3. <http://adsabs.harvard.edu/abs/1973caa..book.....A>.
- Bensmail, H., G. Celeux, A.E. Raftery and C.P. Robert, 1997. Inference in model-based cluster analysis 1. *Statistics and Computing*, 7 (1): 1-10. DOI: 10.1023/A:1018510926151. <http://portal.acm.org/citation.cfm?id=599227.599262>.
- Chang, W.C., 1983. On using Principal Components before separating a mixture of 2 multivariate normal Populations. *J. Applied Stat.*, 32: 267-275. DOI: 10.1177/096228029200100105. <http://smm.sagepub.com/cgi/content/abstract/1/1/69>.
- Chirpaz, E., M. Colonna and J.F. Viel, 2004. Cluster analysis in geographical epidemiology: The use of several statistical methods and comparison of their results. *Epidemiol Sante Publique*, 52 (2): 139-149. PMID: 15138393. <http://www.cababstractsplus.org/google/abstract.asp?AcNo=20043085514>.
- Dincer, K.S. and K. ve Ozdamar, 1993. Comparison of appropriate cluster criteria in cluster analyze. *J. Sci. Eng.* [Kumeleme cozumlemesinde uygun kumeleme olcutlerinin karsilastirilmesi. Hacettepe Univ. Fen ve Muhendislik Bilimleri Dergisi], 14: 17-33. [http://www.ulakbim.gov.tr/cabim/vt/uvvt/muh/uvvt\\_journal\\_index.php?the\\_db=3](http://www.ulakbim.gov.tr/cabim/vt/uvvt/muh/uvvt_journal_index.php?the_db=3).
- Dueck, D., Q.D. Morris and B.J. Frey, 2005. Multy-way clustering of microarray data using prob-abilistic sparse matrix factorization. *Bioinformatics*, 21 (Suppl-1): 144-151. DOI: 10.1093/bioinformatics/bti1041. PMID: 15961451. [http://bioinformatics.oxfordjournals.org/cgi/content/abstract/21/suppl\\_1/i144?ikey=y6wMylQyQzAJpB&keytype=ref](http://bioinformatics.oxfordjournals.org/cgi/content/abstract/21/suppl_1/i144?ikey=y6wMylQyQzAJpB&keytype=ref).
- Dufrene, M. and P. Legendre, 1997. Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach. *Ecological Monographs*, 67 (3): 345-366. [http://www.ufz.de/data/Dufrene\\_1997\\_3161.pdf](http://www.ufz.de/data/Dufrene_1997_3161.pdf).
- Everitt, B.S., 1979. Unresolved problems in cluster analysis. *Biometrics*, 35 (1): 169-181. <http://www.jstor.org/stable/2529943>.
- Giurcaneanu, C.D. and I. Tabus, 2004. Cluster structure inference based on clustering stability with applications to microarray data analysis. *EURASIP J. Applied Signal Process.*, (1): 64-80. DOI: 10.1155/S1110865704309078. <http://portal.acm.org/citation.cfm?id=1289345>.
- Johnson, R.A. and D.W. Wichern, 1992. Applied Multivariate Statistical Analysis. 5th Edn. Prentice Hall International Inc., New Jersey. ISBN-10: 0130925535. [http://www.pearsonhighered.com/academic/product/0,0130925535,00%2Ben-USS\\_01DBC.html](http://www.pearsonhighered.com/academic/product/0,0130925535,00%2Ben-USS_01DBC.html).
- Minitab Inc., 2003. Minitab release 14 Statistical Software, Minitab Inc. USA.
- Malhotra, N.K., 1987. Validity and structural reliability of multidimensional scaling. *J. Marketing Res.*, 24: 164-173. <http://web.ebscohost.com/ehost/pdf?vid=2&hid=12&sid=b14d2b65-0bb4-4d85-a8def64191ae1e5%40sessionmgr109>.
- Milligan, G.W. and M.C. Cooper, 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50 (2): 159-179. DOI: 10.1007/BF02294245. <http://www.springerlink.com/content/ap740680558kq8u4>.
- Ozdamar, K., 1999. Data analyze with statistical packaged software-2 (MANOVA). 2nd Edn. [Paket programlar ile istatistiksel veri analizi-2 (Cok Degiskenli Analiz)], Kaan Kitabevi, Eskisehir-Turkey, pp: 479-486. ISBN: 9756787015. [http://kutuphane.gyte.edu.tr/fini/xsl/yordam/y1\\_detay.xml?-db=YordamBT&%23050+082+S%C4%B1n%C4%B1flama=%C3%9693&-sortfield.1=%23008+Yay%C4%B1n+Tarihi&-sortorder.1=ascend&-token.aranankelime=%C3%9693&-token.tk=&-token.gt=0&-token.d=1&-skip=21&-find&-token.yzym=zym](http://kutuphane.gyte.edu.tr/fini/xsl/yordam/y1_detay.xml?-db=YordamBT&%23050+082+S%C4%B1n%C4%B1flama=%C3%9693&-sortfield.1=%23008+Yay%C4%B1n+Tarihi&-sortorder.1=ascend&-token.aranankelime=%C3%9693&-token.tk=&-token.gt=0&-token.d=1&-skip=21&-find&-token.yzym=zym).
- Ozturk, A.I., I. Alatas, A. Settar, Y. Boduroglu, F.B. Uyguner and M. Bozkurt, 1992. Determination of some morphological characteristics in Aegean honeybee population. *J. Aegean Agric. Res. Institute* [Ege bolgesi ari populasyonlarında bazı morfolojik özelliklerin saptanması. *Ege Tar. Aras. Ens. Dergisi*], Menemen-Izmir. <http://www.etae.gov.tr/bolumler.php?op=sayfa&id=4>.

- Ruttner, F., 1988. Biogeography and taxonomy of honeybee. 1st Edn. Springer-Verlag, Berlin, pp: 85-98. ISBN: 13: 978-0387177816. <http://orton.catie.ac.cr/cgi-bin/wxis.exe/?IsisScript=SIBE01.xis&method=post&formato=2&cantidad=1&expresion=mfn=007858>.
- SPSS Inc., 2006. SPSS for Windows version 15.0 Statistical Software, SPSS Inc. Chicago.
- Sandoz, J.C., 2006. Odour-evoked responses to queen pheromone components and to plant odours using optical imaging in the antennal lobe of the honey bee drone *Apis mellifera* L. *J. Exp. Biol.*, 209: 3587-3598. DOI: 10.1242/jeb.02423. PMID: 16943499. <http://jeb.biologists.org/cgi/content/full/209/18/3587>.
- Sharma, S., 1996. Applied Multivariate Techniques. 1st Edn. John Wiley and Sons. Inc. New York, pp: 185-195. ISBN: 978-0-471-31064-8. [http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471310646\\_descCd-tableOfContents.html](http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471310646_descCd-tableOfContents.html).
- Steffes, D.M., 2007. Panpsychic organicism: Sewall Wright's philosophy for understanding complex genetic systems. *J. Hist. Biol.*, 40 (2): 327-361. PMID: 18175605. DOI: 10.1007/s10739-006-9105-5. <http://hs7nx8de4e.scholar.serialssolutions.com/?sid=google&auinit=DM&aualast=Steffes&atitle=Panpsychic+Organicism:+Sewall+Wright%E2%80%99s+Philosophy+for+Understanding+Complex+Genetic+Systems&id=doi:10.1007/s10739-006-9105-5>.