We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Classification and Ordination Methods as a Tool for Analyzing of Plant Communities

Mohammad Ali Zare Chahouki

Additional information is available at the end of the chapter

1. Introduction

Community ecologists aim at understanding the occurrence and abundance of taxa (usully species) in space and time and the goal of all studies in plant ecology, is finding spatial and temporal interactions add to the complexity of vegetation systems. Hence for this purpose, it is necessary to imply best statistical methods (Causton, 1988)

In this study, some important classification and ordination methods such as cluster analysis (CA), Two way Indicator Species Analysis (TWINSPAN), Polar Ordination (PO), Nonmetric Multidimensional Scaling (NMS), Principal component analysis (PCA), Detrended Correspondence Analysis (DCA), Canonical correspondence analysis (CCA), Redundancy analysis (RDA) will be explained briefly.

Ordination (or inertia) methods, like principal component and correspondence analysis,and clustering and classification methods are currently used in many ecological studies (Anderson, 1971; Gauch et aL, I982a; Orloci, 1978; Whittaker et al, 1967; Legendre & Legendre, 1998).

The choice of the mathematical method of analysis is mainly determined by availability rather than an accurate knowledge of the properties and limitations of the possible different methods (Legendre & Legendre, 1998).

This study aims to explain these methods as tool for analyzing of plant Communities. The use of multivariate analysis has been extended much more widely over the past 20 years. Much more is included on techniques such as Canonical Correspondence Analysis (CCA) and Non-metric Multidimensional Scaling (NMS), Principal component analysis (PCA) and another technique to include plant communication and plant-environment relationships (Kent, 2006). It is a main objective in data analysis to distinguish random from deterministic components. Therefore spatial and temporal interactions add to the complexity of vegetation systems (Wildi, 2010).

INTECH
open science | open minds

Some basic knowledge of Classification and Ordination methods that influence vegetation ecology might be needed to understand the examples presented in this study.

Studying the vegetation distribution pattern is a basic aspect of the design and management (Zhang et al., 2006). Quantitative separation was studied by previous scholars to investigate the contribution of environmental factors to the whole or different layers of plant community distribution pattern. (Zhang et al., 2004). Actually, natural plant communities are distributed continuously, and they are composed of plant communities at different succession stages which response to environmental factors differently.

2. Data

Commonly, data interpreted using Classification and ordination, are collected in a species by sample data matrix, similar to the matrixes presented below.

Species abundances as main data matrix will also use the standardized set of no redundant environmental variables for use with clustering and indicator species analysis. Will be not need a second matrix, although Cluster analysis will produce one for use during this exercise. For explaining the issue, using data from Study area that is located in the North-East of the Semnan province in center of Iran (35º 53′ N, 54º 24′ E to 35º50′ N, 53º43′ E) (Fig 1).

270	plots					
9	Species					
		Q	Q	Q	Q	Q	Q
		Ar.si	Se.ro	Eu.ce	St.ba	Zy.er	...
1	10	0.5	0.5	0.5	0.5	...
2	1.75	3.75	0.5	0.5	0.5	...
3	1	0.5	3.75	1	0.5	...
4	3.75	0.5	3.75	0.5	0.5	...
5	6.25	0.5	0.5	1.75	0.5	...
6	1	0.5	3.75	1	0.5	...
7	10	0.5	1	1	0.5	...
8	3.75	0.5	1	0.5	0.5	...
9	3.75	0.5	1	0.5	0.5	...
10	6.25	0.5	1.75	0.5	0.5	...
11	6.25	1.75	3.75	0.5	0.5	...
12	3.75	0.5	0.5	0.5	1	...
13	10	0.5	15	1	0.5	...
14	3.75	0.5	1.75	0.5	0.5	...

Table 1. Data matrix using in Classification (using ordinal scale of Van-der-Marrel)
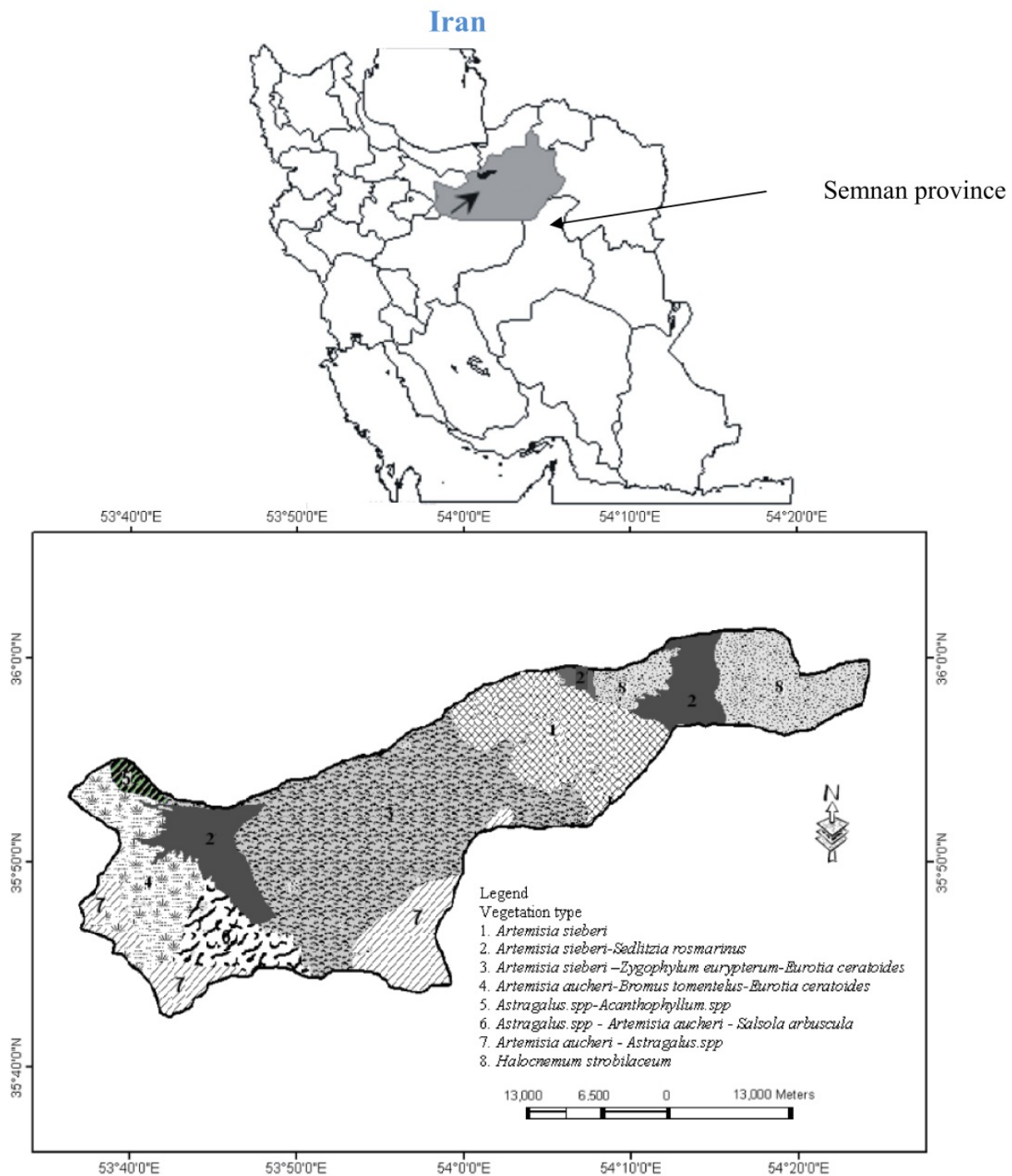
Figure 1. Location of study area and the distribution of the vegetation types.

The below is a relatively simple data set. However, it is easy to imagine that a true data set may encounter dozens of species over 270 of samples. Complex sample by species matrices represent dozens to 270 of dimensions which are impossible to visualize or interpret. Even graphed, species response curves of large community data sets can be nearly impossible to interpret.

A quantitative survey of the vegetation is carried out during 2009-2010. In each of the studied types, soil and vegetative attributes were described within quadrates located along three 150m transverse transects. Quadrate size was determined for each vegetation type using the minimal area method. Considering variation of vegetation and environmental factors, forty five quadrates with a distance of 50m from each other were established in each vegetation type. Sampling method was randomized systematic. Floristic list, density and canopy cover percentage were determined in each quadrate. Vegetation cover data were recorded using ordinal scale of Van-der-Marrel (1979).

In fact, the cover data transformed using an eight-point scale ((0–1=0.5, 1–2.5=1.75, 2.5–5=3.75, 5–7.5=6.25, 7.5–12.5=10, 12.5–17.5=15, 17.5–22.5=20, 22.5–27.5=25, >27.5=30)

Sample data may include measures of density, biomass, frequency, importance values, presence/absence, or any number of abundance measures.

Ordination can help us find structure in these complicated data sets. By using various mathematical calculations, ordination techniques will identify similarity between species and samples. Results are then projected onto two dimensions in such a way that species and samples most similar to one another will be close together, and species and samples most dissimilar from one another will appear farther apart (as shown at this study).

6	type				
22	factor				
	Q	Q	Q	Q	Q
	gr1	gr2	clay1	clay2	...
A.sieberi-E.ceratoides	28.2016	45.6333	22.1667	21	...
H.strobilaceum	8.04E+00	2.83667	26.8	29.3333	...
A.sieberi-Z.eurypterum	35.5167	50.0333	17.5	16	...
Z.eurypterum-A.sieberi	27.5933	36.44	16.6667	23.6667	...
A.au-As.ssp-B.tomentelus	28.48	47.6433	26.4533	33.1667	...
S.rosmarinus	28.15	37.475	22.8333	20.6667	...

Table 2. Data matrix using in Ordination

Data analysis was performed on the species, averaging all plots per site. All numerical analyses were done with the PC-ORD, V. 4 package (McCune and Mefford, 1999).

3. Methods of classification analysis

Classification method is an act of putting things in groups. Most commonly in community ecology, the "things" are samples or communities. Classification can be completely subjective, or it can be objective and computer-assisted (even if arbitrary). Hierarchical classification means that the groups are nested within other groups. There are two general kinds of hierarchical classification: divisive and agglomerative. A Divisive method starts with the entire set of samples, and progressively divides it into smaller and smaller groups. An agglomerative method starts with small groups of few samples, and progressively

groups them into larger and larger clusters, until the entire data set is sampled (Pielou, 1984).

Cluster analysis, on the other hand, seeks to divide the n quadrates into groups of high internal similarity with respect to species or characters used. In the classical approach of Williams & Lambert (1959), the so-called Association-Analysis, communities are defined by the presence or absence of single species. This is highly dependent on the vagaries of sampling; many workers have felt the method may result in botanical over simplification, so that nowadays polythetic methods are more usually applied.

From the above discution, it can be seen that ordination and cluster analysis are not competing approaches and provided the ecologist is cautious in making inferences, both can reasonably be applied in the examination of multivariate samples (Pritchard & Anderson, 1971).

In classification of species the basic idea is that a characteristic species combination (or at least a group of differentiated species) should gather samples containing these species into clusters of similar samples (Tavili & Jafari, 2009).

In fact, Classification assumes from the outset that the species assemblages fall into discontinuous group, whereas ordination starts from the idea that such assemblages very gradually

3.1. Cluster analysis

Clustering, sometimes simply a synonym of classification, but more usually referring to agglomerative classification.

Clustering is a straightforward method to show association data, however, the confidence of the nodes are highly dependent on data quality, and levels of similarity for cluster nodes is dependent on the similarity index used. Krebs (1999) shows that mean linkage is superior to single and complete linkage methods for ecological purposes because the other two are extremes, either producing long or tight, compact clusters respectively. There are, however, no guidelines as to which mean-linkage method is the best (Swan, 1970).

The objective of Cluster Analysis is to graphically show the relationship between cluster analyses and your individual data points.

The resulting graph makes it easy to see similarities and differences between rows in the same group, rows in different groups, columns in the same group, and columns in different groups. Groups of rows and columns relate to each other, could be seen graphically. Two-way clustering refers to doing a cluster analysis on both the rows and columns of your matrix, followed by graphing the two dendrograms simultaneously, adjacent to a representation of your main matrix. Rows and columns of your main matrix are re-ordered to match the order of items in your dendrogram (Mucina, 1997).

Fig 1 showed dendrogram of Cluster analysis (study area: North East of Semnan rangelands, Iran). Grouping was performed using Euclidean distance and the Ward method. Species with less than 2 entries in the matrix were deleted from the analysis.
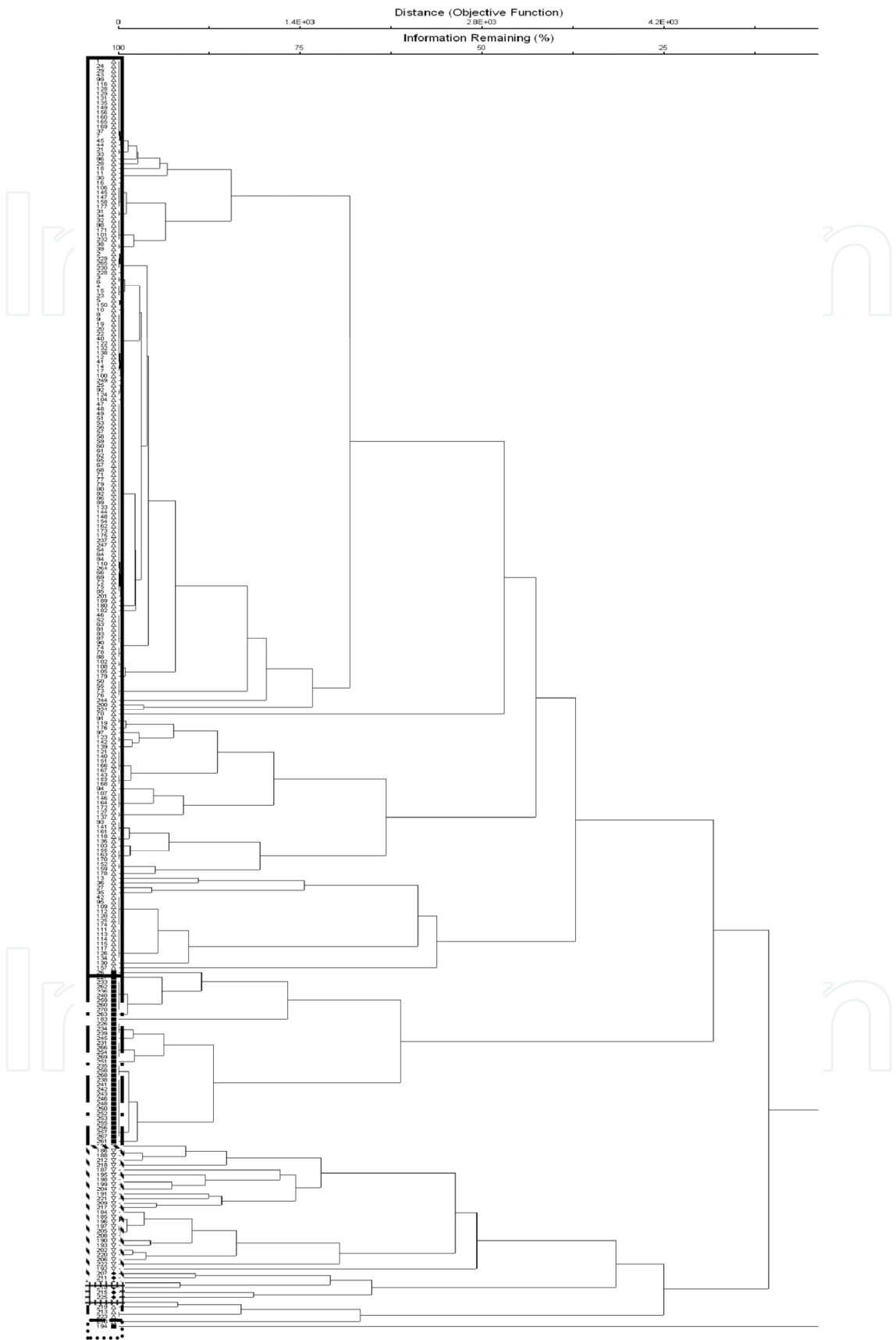
Figure 2. Dendrogram of the cluster grouping of the study sites

Cluster analysis can be performed using either presence–absence or quantitative data. Each pair of sites is evaluated on the degree of similarity, and then combined sequentially into clusters to form a dendrogram with the branching point representing the measure of similarity.

In fact, the aim is to form a hierarchical classification (i.e. groups, containing subgroups) which is usually displayed by a dendrogram (as shown in above). The groups are formed from the most similar objects are first joined to form the first cluster, which is then considered an object, and the joining continues until all the objects are joined in the final cluster, containing all the objects (fig 2).

The procedure has two basic steps: in the first step, the similarity matrix is calculated for all the pairs of the objects (the matrix is symmetric, and on the diagonal there are either zeroes – for dissimilarity – or the maximum possible similarity values). In the second step, the objects are clustered (joined, amalgamated) so that after each amalgamation, the newly formed group is considered to be an object, and the similarities of the remaining objects to the newly formed one are recalculated. The individual procedures (algorithms) differ in the way they recalculate the similarities (Leps & Smilauer, 2003).

Major types of hierarchical, agglomerative, polythetic clustering strategies followed:

1. Nearest Neighbor
2. Farthest Neighbor
3. Median
4. Group Average
5. Centroid: It (weighted) mean of a multivariate data set. Can be represented by a vector. For many ordination techniques, the centroid is a vector of zeros (that is, the scores are centered and standardized). In a direct gradient analysis, a categorical variable is often best represented by a centroid in the ordination diagram.
6. Ward's Method (Ward's is also know as Orloci's and Minimum Variance Method)
7. Flexible Beta
8. McQuitty's Method

This analysis of the vegetation – environment relations and the classification of the Semnan rangelands, is also relevant for the rangelands of arid and semi arid in Iran, and provides a base line for other studies intended to conserve and restore this ecosystem.

Although clustering is an agglomerative classification technique and TWINSPAN is divisive, both produced comparable results. In addition, TWINSPAN provided indicator species.

In addition, to identify species with particular diagnostic value and to confirm clustering results, the floristic data were classified with the two way indicator species analysis (TWINSPAN) (Hill, 1979).

3.2. TWINSPAN

The TWINSPAN method is one of the more popular classification programs used in plant community ecology (Hill 1979; Hill et al. 1975). The two approaches differ between two

classification methods is that, TWINSPAN creates groups and also finds indicator species for those groups, while Cluster analysis requires a before-the-fact assignment of group membership as input. In this case, will be used hierarchical clustering to identify groups for vegetation classification. TWINSPAN produces no graphical output. The biggest volume of the result is the description of each division. For each division, TWINSPAN identifies the indicator pseudo species and their signs (positive or negative for one end of the ordination or the other) and lists the samples assigned to each subgroup. Two popular agglomerative polythetic techniques are Group Average and Flexible. McCune et al. (2002) recommend Ward's method in addition. Gauch (1982a) preferred to use divisive polythetic techniques such as TWINSPAN.

This method works with qualitative data only. In order not to lose the information about the species abundances, the concepts of pseudo-species and pseudo-species cut levels were introduced. Each species can be represented by several pseudo-species, depending on its quantity in the sample. A pseudo-species is present if the species quantity exceeds the corresponding cut level.

TWINSPAN is a program for classifying species and samples, producing an ordered two-way table of their occurrence. The process of classification is hierarchical; samples are successively divided into categories, and species are then divided into categories on the basis of the sample classification. TWINSPAN, like DECORANA, has been widely used by ecologists.

For example, TWINSPAN was performed for vegetation analysis in 270 plots using ordinal scale of Van-der-Marrel (1979). The end of results file is the two-way ordred table summarizing the classification (Fig3). The table has species (not pesudo species) as rows and samples as columns.The results of TWINSPAN classification are presented in Fig.4. According to the above-mentioned table, figure, and also eigenvalue of each division, vegetation of the study area was classified in to six main types. Each type differs from the other in terms of it's environmental needs.

These types are as follows:

1. *Artemisia sieberi-Eurotia ceratoides*
2. *Artemisia aucheri, Astragalus spp., Bormus tomentellus*
3. *Artemisia sieberi–Zygophylom eurypterum*
4. *Zygophylom eurypterum- Artemisia sieberi*
5. *Seidlitzia rosmarinus*
6. *Halocnemum strobilaceum*

4. Methods of ordination analysis

Ordination serves to summarize community data (such as species abundance data) by producing a low-dimensional ordination space in which similar species and samples are plotted close together, and dissimilar species and samples are placed far apart (Peet, 1980)
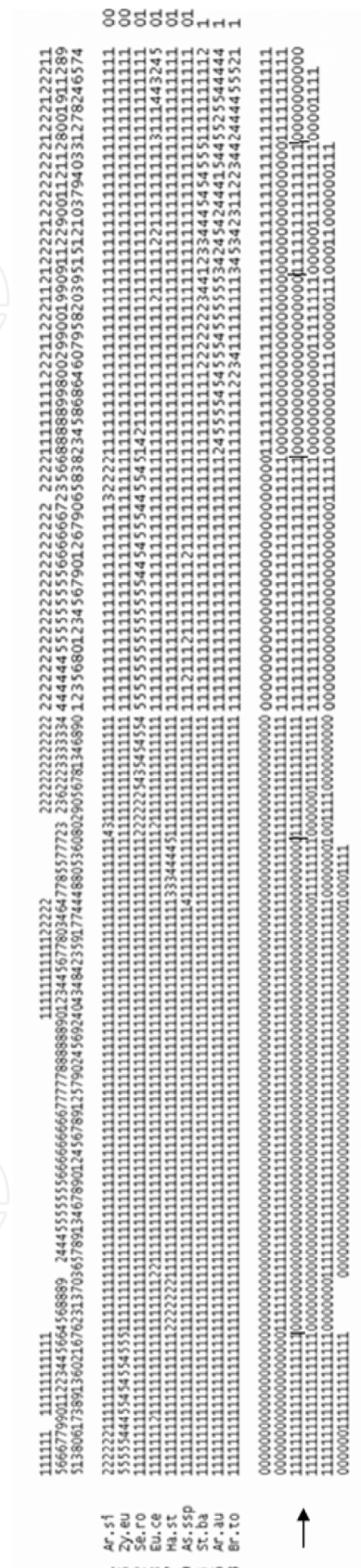
Figure 3. TWINSPAN of the vegetation cover in 270 quadrates and 9 species
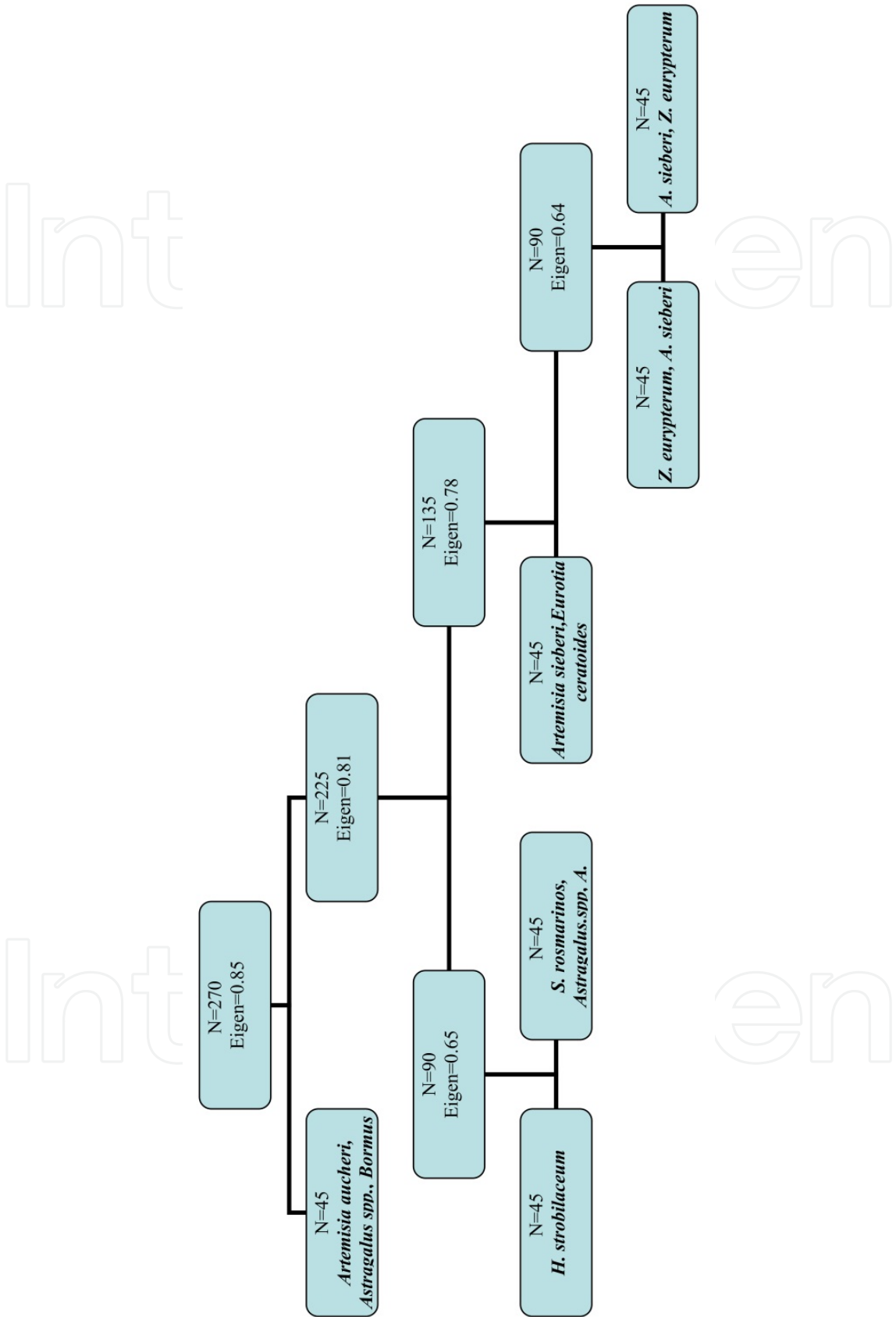
Figure 4. Schematic comparison of Ordination techniques

Ordination methods can be divided in two main groups, direct and indirect methods. Direct methods use species and environment data in a single, integrated analysis. Indirect methods use the species data only (Fig 5). Finally, ordination techniques are used to describe relationships between species composition patterns and the underlying environmental gradients which influence these patterns. Although community ecology is a fairly young science, the application of quantitative methods began fairly early (McIntosh,. 1985).
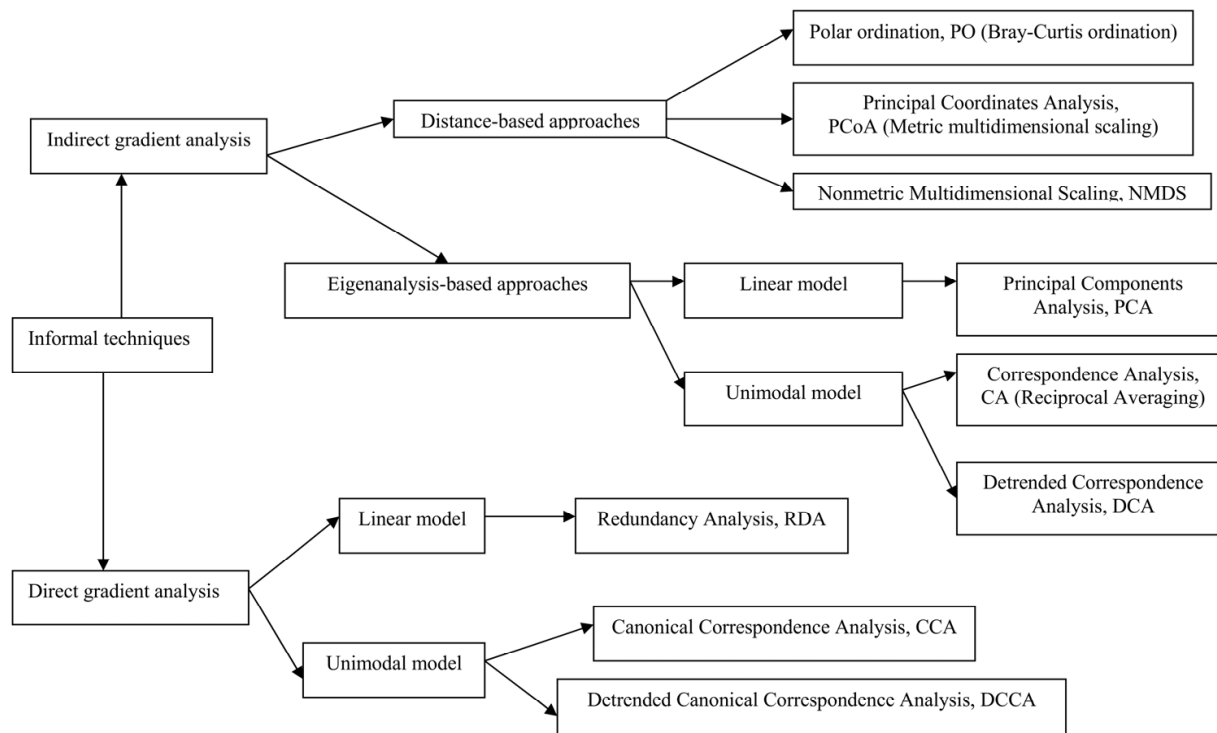

Figure 5. Schematic comparison of Ordination techniques

In 1930, began to use informal ordination techniques for vegetation. Such informal and largely subjective methods became widespread in the early 1950's (Whittaker 1967). In 1951, Curtis and McIntosh developed the 'continuum index', which later lead to conceptual links between species responses to gradients and multivariate methods. Shortly thereafter, Goodall (1954) introduced the term 'ordination' in an ecological context for Principal Components Analysis.

Each method was applied to data from a North east of Semnan (In Iran). If objective of study is examining the distribution patterns of six plant type in the rangelands, ordination could be used to determine which species are commonly found associated with one another, and how the species composition of the community changes with increase and decrease in each environment factor (Zare Chahouki et al, 2010). The objective of this method was to establish a monitoring system that may serve to identify and predict future vegetation changes and to assess impacts of conservation and management practices.

There are several different ordination techniques, all of which differ slightly, in the mathematical approach used to calculate species and sample similarity/dissimiarity. Rather

than reinventing the wheel by discussing each of these techniques. Our example study illustrates the most frequent use of ordination methods in community ecology, we will offer only a brief description of the most commonly used methods here. Further details can be found in the following.

Polar Ordination (PO)

Bray and Curtis (1957) developed polar ordination, which became the first widely-used ordination technique in ecology.

Polar Ordination arranges samples with respect to poles (also termed end points or reference points) according to a distance matrix (Bray and Curtis 1957). These endpoints are two samples with the highest ecological distance between them, or two samples suspected of being at opposite ends of an important gradient. This method is especially useful for investigating ecological change (e.g., succession, recovery).

For example, Fig 6 shows ordination diagram for vegetation types and soil variables by Bray-Curtis analysis.

Endpoints for axis 1 was *Halocnemum strobilaceum*, *Artemisia aucheri-Astragalus spp-Bromus tomentellus.* Distances (ordination scores) are from *Halocnemum strobilaceum* Sum of squares of non-redundant distances in original matrix was .199621E+12. Axis 1 extracted 100.00% of the original distance matrix. Sum of squares of residual distances remaining is .672048E+05. Regression coefficient for this axis was -6.40 and Variance in distances from the first endpoint was 0.65.

Endpoints for axis 2: *Artemisia sieberi-Zygophylum eurypterum, Ar.au-As.spp-Br.to* distances (ordination scores) were from *Artemisia siberi-Zygophylum eurypterum.* Regression coefficient for this axis was -3.53. Variance in distances from the first endpoint was 0.0.

Axis 2 extracted 1.87% of the original distance matrix, Cumulative was 98.15%. Sum of squares of residual distances remaining was .948501E-01.

Polar ordination has strengths and weaknesses. The advantage of this method is that: (Beals 1984).

1. It is Simple, easy to understand geometric method, easily taught.
2. It is Ideal for evaluating problems with discrete endpoints. Polar Ordination ideal for testing specific hypotheses (e.g., reference condition or experimental design) by subjectively selecting the end points

The weaknesses of Polar Ordination method is that: (Beals 1984).

1. Axes are not orthogonal. With large data sets, it may be difficult to get a consistent ordination.
2. Not completely objective won't always get the same answer. However, this is a function of the decision regarding reference stands, and is really amounts to viewing the ordination from different angles, although the problem of nonorthogonal axes can cause considerable distortion to the ordination space.

Some of this problem can be overcome by using rules to define the reference stands.
3. Distances are not metric (i.e., they are relative only)
4. No explicit statement of underlying model.

In the earliest versions of PO, these endpoints were the two samples with the highest ecological distance between them, or two samples which are suspected of being at opposite ends of an important gradient (thus introducing a degree of subjectivity).

Beals (1984) extended Bray-Curtis ordination and discussed its variants, and is thus a useful reference. The polar ordination, simplest method is to choose the pair of samples, not including the previous endpoints, with the maximum distance of separation.
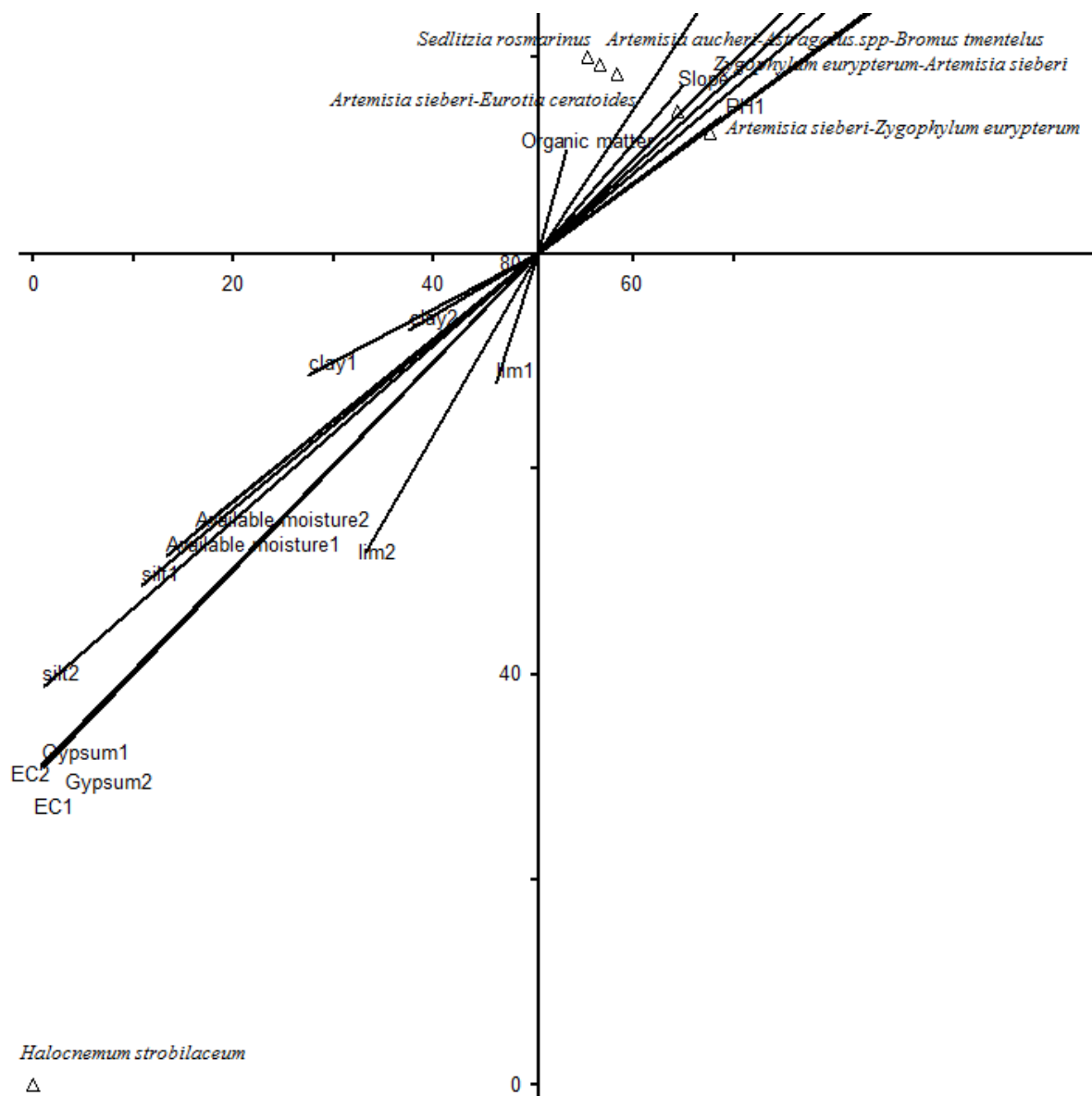

Figure 6. Bray-Curtis–ordination diagram of the environmental data. For vegetation types and variables abbreviations. (Δ) is the representative of the vegetation types.

These patterns are consistent with others in the literature (cited and reanalyzed in Palmer 1986).

Principal Components Analysis (PCA)

Principal Components Analysis (PCA) was one of the earliest ordination techniques applied to ecological data. PCA uses a rigid rotation to derive orthogonal axes, which maximize the variance in the data set. Both species and sample ordinations result from a single analysis. Computationally, Principal components analysis is the basic eigen analysis technique. It maximizes the variance explained by each successive axis.

The sum of the eigenvalues will equal the sum of the variance of all variables in the data set. PCA is relatively objective and provides a reasonable but crude indication of relationships.

PCA was invented in 1901 by Karl Pearson (Dunn,et al,1987) Now it is mostly used as a tool in exploratory data analysis and for making predictive models.

PCA is a method that reduces data dimensionality by performing a covariance analysis between factors (Feoli and Orl¢ci. 1992).

This method is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components.

The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components (ter Braak and S ˘ milauer, 1998).

PCA method was used to determine the association between plant communities and environmental variables, i.e. in an indirect non-canonical way (ter Braak and Loomans, 1987).

For example to determine the most effective variables on the separation of vegetation types, PCA was performed for 22 factors in six vegetation types. The results of the PCA ordination are presented in Table 3 and Fig.5. Broken-stick eigenvalues for data set indicate that the first two principal components (PC1 and PC2) resolutely captured more variance than expected by chance. The first two principal components together accounted for 86% of the total variance in data set. Therefore, 61% and 25% variance were accounted for by the first and second principal components, respectively. This means that the first principal component is by far the most important for representing the variation of the six vegetation types.

Considering the characteristics of solidarity with the components, the first component includes silt and gravel in 20-80 depth, Available moisture in 0-20 depth, sand, gypsum and EC of both the depths. The second component consists of clay in 0-20 depth and lime in both depths.

AXIS	Eigenvalue	% of Variance	Cum.% of Var.	Broken-stick Eigenvalue
1	13.494	61.335	61.335	3.691
2	5.512	25.053	86.388	2.691
3	1.460	6.636	93.024	2.191
4	0.968	4.398	97.422	1.857
5	0.567	2.578	100.000	1.607
6	0.000	0.000	100.000	1.407
7	0.000	0.000	100.000	1.241
8	0.000	0.000	100.000	1.098
9	0.000	0.000	100.000	0.973
10	0.000	0.000	100.000	0.862

Factor	1	2	3	4	5	6
gr_1	-0.2636	0.0012	-0.0447	-0.0562	0.3161	0.1371
gr_2	-0.2589	0.0904	0.0166	-0.1657	0.2022	0.0355
$clay_1$	0.1792	0.3148	0.1002	-0.0093	0.1005	-0.1242
$clay_2$	0.1504	0.2595	-0.3168	-0.3208	-0.3702	-0.2055
$silt_1$	0.2476	0.0278	-0.1910	0.3450	0.0191	0.1166
$silt_2$	0.2691	0.0624	-0.0028	0.0323	0.0133	-0.0807
$sand_1$	-0.2437	-0.1583	0.0828	-0.2235	-0.0573	-0.0706
$sand_2$	-0.2356	-0.1862	0.1819	0.0264	0.1395	0.0824
lim_1	0.0828	-0.3939	-0.0644	-0.0424	0.2794	0.0946
lim_2	0.1606	-0.3190	0.0101	-0.1881	0.3162	0.0212
$O.M_1$	-0.0253	0.3944	-0.0388	-0.0561	0.4768	0.0649
$O.M_2$	-0.0768	0.2109	0.2962	0.3680	0.0688	-0.0525
$A.W_1$	0.2440	0.1148	-0.2414	0.1038	0.2249	0.1069
$A.W_2$	0.2353	0.1306	-0.2399	0.0725	0.3501	0.1342
gyp_1	0.2662	-0.0688	0.0925	-0.0716	0.0125	-0.1236
gyp_2	0.2662	-0.0688	0.0925	-0.0716	0.0125	-0.1257
EC_1	0.2662	-0.0693	0.0957	-0.0628	0.0188	-0.1148
EC_2	0.2653	-0.0729	0.1017	-0.0773	0.0127	-0.1281
pH_1	-0.1360	-0.1130	-0.6739	0.0644	-0.1513	0.2438
pH_2	-0.2205	-0.1334	-0.2747	0.3324	0.2260	-0.8329
elevat	-0.1945	0.2594	0.0252	0.3383	-0.1141	0.0904
sl	-0.1345	0.2559	-0.1863	-0.5878	0.1327	-0.1505

*Non-trivial principal component as based on broken-stick eigenvalues

Table 3. PCA applied to the correlation matrix of the environmental factors in the study area

In the study area, environmental conditions in *Halocnemum strobilaceum* type differ from the others. With attention to the position of this type in the four quarter of the diagram, it has a high correlation with the first axis. Therefore, this type has the most relation with variables of the first axis.

Because of the bigger distance of *H. strobilaceum* type from the second axis, this type has a weak relation with factors such as clay and lime. *Artemisia sieberi-Eurotia ceratoides* and *Seidlitzia rosmarinus* types have inverse relation with indicator environmental characteristics of the first and second axes except for clay, sand and gravel. *A. aucheri–Astragalus. spp.-Bromus tomentellus* type has more relation with indicator characteristics of the first and second axes.

Indicator environmental factors of the first and second axes in *A. sieberi–Zygophylom eurypterum* and *Z. eurypterum-A. sieberi* types are approximately similar. *A. sieberi–Z. eurypterum* type has a direct relationship with gravel and sand, and an inverse relationship with EC, silt, available moisture and gypsum. While *A. aucheri-As. spp.-B. tomentellus* type has a direct relationship with clay and inversely related to lime.
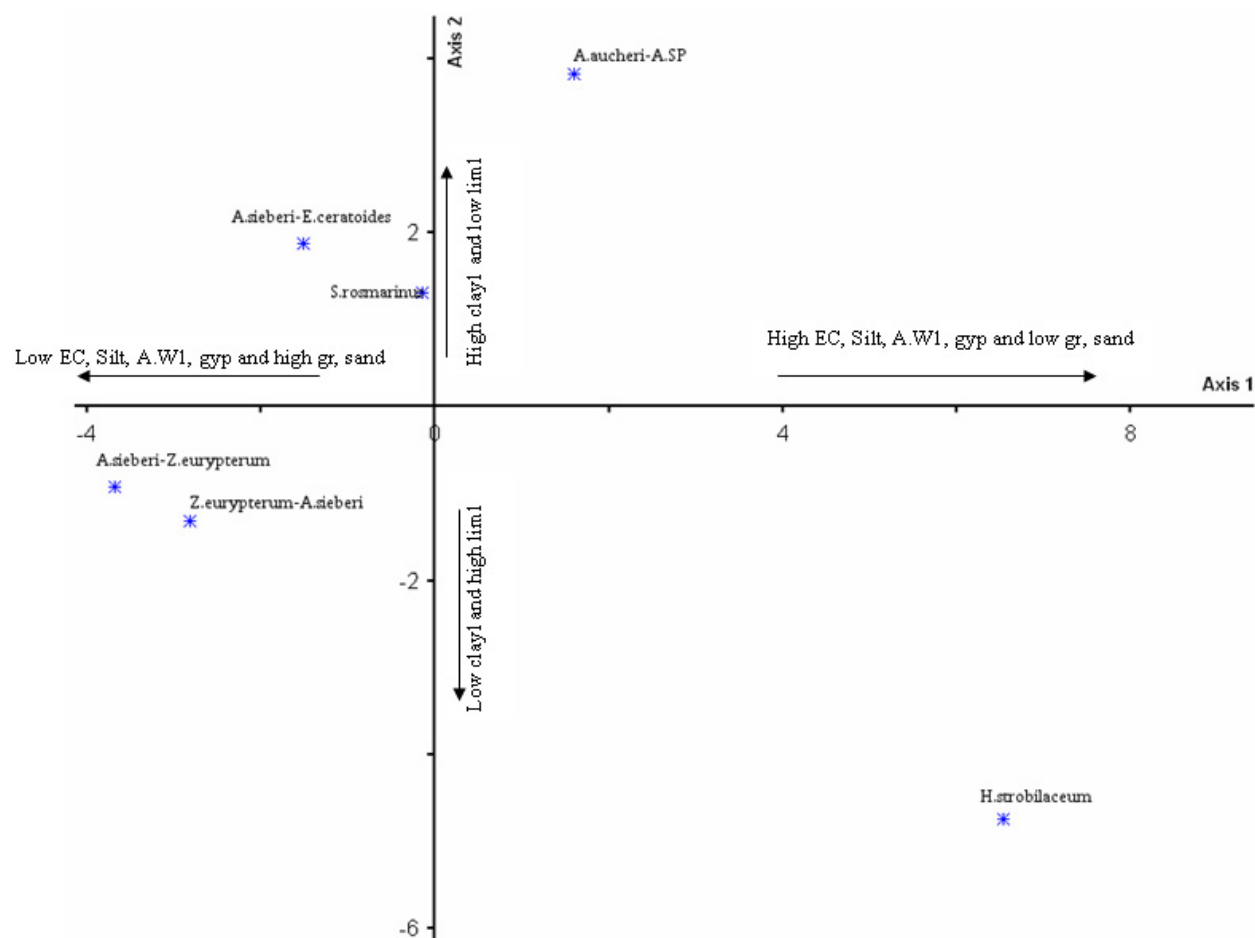

Figure 7. PCA–ordination diagram of the vegetation types related to the environmental factors in the study area. For vegetation types abbreviations, see Appendix A.

PCA operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data

The one advantage of PCA is that once you have found patterns in the data, and you compress the data, ie by reducing the number of dimensions, without much loss of information and While PCA finds the mathematically optimal method (as in minimizing the squared error), it is sensitive to outliers in the data that produce large errors PCA tries to avoid. It therefore is common practice to remove outliers before computing PCA.

However, in some contexts, outliers can be difficult to identify. For example in data mining algorithms like correlation clustering, the assignment of points to clusters and outliers is not known beforehand.

A recently proposed generalization of PCA based on Weighted PCA increases robustness by assigning different weights to data objects based on their estimated relevancy.

Although it has severe faults with many community data sets, it is probably the best technique to use when a data set approximates multivariate normality. PCA is usually a poor method for community data, but it is the best method for many other kinds of multivariate (Bakus, 2007).

In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. This gives you the components in order of significance. Now, if you like, you can decide to ignore the components of lesser significance. You do lose some information, but if the eigenvalues are small, you don't lose much. If you leave out some components, the final data set will have less dimensions than the original.

To be precise, if you originally have dimensions in your data, and so you calculate eigenvectors and eigenvalues, and then you choose only the first eigenvectors, then the final data set has only dimensions. What needs to be done now is you need to form a feature vector, which is just a fancy name for a matrix of vectors. This is constructed by taking the eigenvectors that you want to keep from the list of eigenvectors, and forming a matrix with these eigenvectors in the columns.

Deriving the new data set is the final step in PCA, and is also the easiest. Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

In the case of keeping both eigenvectors for the transformation, we get the data and the plot found in Figure 5. This plot is basically the original data, rotated so that the eigenvectors are the axes. This is understandable since we have lost no information in this decomposition.

In figure 5 showed sample of PCA–ordination diagram of the vegetation types related to the environmental factors.

In contrast to Correspondence Analysis and related methods (see below), species are represented by arrows. This implies that the abundance of the species is continuously increasing in the direction of the arrow, and decreasing in the opposite direction.

Canonical correspondence analysis (CCA)

Canonical correspondence analysis (CCA) is a direct gradient analysis that displays the variation of vegetation in relation to the included environmental factors by using environmental data to order samples (Kent & Coker, 1992). This method combines multiple regression techniques together with various forms of correspondence analysis or reciprocal averaging (Ter Braak, 1986, 1987). The statistical significance of the relationship between the species and the whole set of environmental variables was evaluated using Monte Carlo permutation tests.

The CCA analysis method Ordination is a combination of conventional linear Environment variables with the highest value of dispersion Species shows. In other words, the best weight for CCA describes environment variables with the first axis shows. Species information structure using a reply CCA Nonlinear with the linear combination of variables will consider environmental characteristics of acceptable behavior characteristics of species with environment shows. CCA analysis combined with non-linear species and environmental factors shows the most important environmental variable in connection with the axes shows.

In ecology studies, the ordination of samples and species is constrained by their relationships to environmental variables.

The adventag of CCA Analysis is that: (Palmer, 1993)

1. Patterns result from the combination of several explanatory variables. And many extensions of multiple regressions (e.g. stepwise analysis and partial analysis) also apply to CCA.
2. It is possible to test hypotheses (though in CCA, hypothesis testing is based on randomization procedures rather than distributional assumptions).
3. Another advantage of CCA lies in the intuitive nature of its ordination diagram, or triplot. It is called a triplot because it simultaneously displays three pieces of information: samples as points, species as points, and environmental variables as arrows (or points).
   If data sets are few, CCA triplots can get very crowded then should be separate the parts of the triplot into biplots or scatterplots (e.g. plotting the arrows in a different panel of the same figure) or rescaling the arrows so that the species and sample scores are more spread out. And we can only plotting the most abundant species (but by all means, keep the rare species in the analysis).
4. When species responses are unimodal, and by measuring the important underlying environmental variables, CCA is most likely to be useful.

And one of limitations to CCA is that correlation does not imply causation, and a variable that appears to be strong may merely be related to an unmeasured but 'true' gradient. As with any technique, results should be interpreted in light of these limitations (McCune 1999).

It was used to examine the relationships between the measured variables and the distribution of plant communities (Ter Braak, 1986). CCA expresses species relationships as linear combinations of environmental variables and combines the features of CA with canonical correlation analysis (Green, 1989). This provides a graphical representation of the relationships between species and environmental factors.

Canonical Correlation Analysis is presented as the standard method to relate two sets of variables (Gittins, 1985). However, the latter method is useless if there are many species compared to sites, as in many ecological studies, because its ordination axes are very unstable in such cases.

The best weight for CCA describes environment variables with the first axis shows. Species information structure using a reply CCA Nonlinear with the linear combination of variables will consider environmental characteristics of acceptable behavior characteristics of species with environment shows. CCA analysis combined with non-linear species and environmental factors shows the most important environmental variable in connection with the axes shows.

In Canonical Correspondence Analysis, the sample scores are constrained to be linear combinations of explanatory variables. CCA focuses more on species composition, i.e. relative abundance.

When a combination of environmental variables is highly related to species composition, this method, will create an axis from these variables that makes the species response curves most distinct. The second and higher axes will also maximize the dispersion of species, subject to the constraints that these higher axes are linear combinations of the explanatory variables, and that they are orthogonal to all previous axis.

Monte Carlo permutation tests were subsequently used within canonical correspondence analysis (CCA) to determine the significance of relations between species composition and environmental variables (ter Braak, 1987)

The outcome of CCA is highly dependent on the scaling of the explanatory variables. Unfortunately, we cannot know a priori what the best transformation of the data will be, and it would be arrogant to assume that our measurement scale is the same scale used by plants and animals. Nevertheless, we must make intelligent guesses (Bakus, 2007).

It is probably obvious that the choice of variables in CCA is crucial for the output. Meaningless variables will produce meaningless results. However, a meaningful variable that is not necessarily related to the most important gradient may still yield meaningful results (Palmer 1988).

Explanatory variables need not be continuous in CCA. Indeed, dummy variables representing a categorical variable are very useful. A dummy variable takes the value 1 if the sample belongs to that category and 0 otherwise. Dummy variables are useful if you have discrete experimental treatments, year effects, different bedrock types, or in the case of the bryophyte example, host tree species (Bakus, 2007).

If many variables are included in an analysis, much of the inertia becomes 'explained'. Any linear transformation of variables (e.g. kilograms to grams, meters to inches, Fahrenheit to Centigrade) will not affect the outcome of CCA whatsoever.

There are as many constrained axes as there are explanatory variables. The total 'explained inertia' is the sum of the eigenvalues of the constrained axes. The remaining axes are unconstrained, and can be considered 'residual'. The total inertia in the species data is the sum of eigenvalues of the constrained and the unconstrained axes, and is equivalent to the sum of eigenvalues, or total inertia, of CA. Thus, explained inertia, compared to total inertia, can be used as a measure of how well species composition is explained by the variables. Unfortunately, a strict measure of 'goodness of fit' for CCA is elusive, because the arch effect itself has some inertia associated with it (Bakus, 2007).

The ordination diagrams of canonical correlation analysis and redundancy analysis display the same data tables; the difference lies in the precise weighing of the species (ter Braak, 1987, 1990; ter Braak & Looman, 1994). Recent, good ecological examples of canonical correlations analysis, with many more sites than species, are Van der Meer (1991) and Varis (1991).

For example, according to Tables 4 and 5, first axis (Eigenvalue=0.869) accounted for 98.7% variation in environmental factors data. Correlation between the first axis and species–environmental variables was 0.99 and Monte Carlo permutation test for the first axis was highly significant (P=0.01). The second axis (Eigenvalue=0.182) explained 0.4% variation in data set. Correlation between the second axis and species–environmental variables was 0.92. In addition, the Monte Carlo test for the second axis was highly significant (P=0.02).

	Axis 1	Axis 2	Axis 3
Eigenvalue	0.869	0.003	0.003
Variance in species data			
% of variance explained	98.7	0.4	0.3
Cumulative % explained	98.7	99.1	99.4
Pearson Correlation, Spp-Envt*	0.998	0.920	0.959
Kendall (Rank) Corr., Spp-Envt	0.481	0.706	0.584

* Correlation between sample scores for an axis derived from the species data and the sample scores that are linear combinations of the environmental variables. Set to 0.000 if axis is not canonical.

Table 4. Canonical correspondence analysis for environmental data.

Axis	Spp-Envt Corr.	Mean	Minimum	Maximum	p
1	0.998	0.838	0.195	0.996	0.0100
2	0.920	0.607	0.072	0.935	0.0200
3	0.959	0.342	0.032	0.709	0.0100

p = proportion of randomized runs with species-environment correlation greater than or equal to the observed Species-environment correlation; i.e., p = (1 + no. permutations >= observed)/(1 + no. permutations)

Table 5. Mont Carlo test result –Speacies-Enviroment

Species responses to environmental conditions cannot be inferred in a causal way from multivariate analysis or any other statistical method; however, these techniques are useful to identify spatial distribution patterns and to assess which of the included environmental variables contribute most to species variability and which factors should be experimentally tested (D í ez et al, 2003).

The results of CCA ordination are presented in Fig.8. Each environmental factor is an indicator of the specific habitat. *Artemisia sieberi-Eurotia ceratoides*, *A. sieberi–Zygophylum eurypterum* and *Zygophylom eurypterum- A. sieberi types* have nonlinear relation with gravel, sand, silt, clay, lime, organic matter and available moisture. Relation power depends on the relative distance between indicator points of soil characteristics and vegetation types. *H. strobilaceum* type has non linear relation with gypsum and EC in both layers that is, EC and gypsum are indicator of habitat of this type. *A. sieberi–Z. eurypterum* and *Z. eurypterum- A. sieberi* types have non linear relation with them while *A.aucheri-As.sp.* and *S. rosmarinus* types are different from each other and they have less non linear relation with ecological factors.
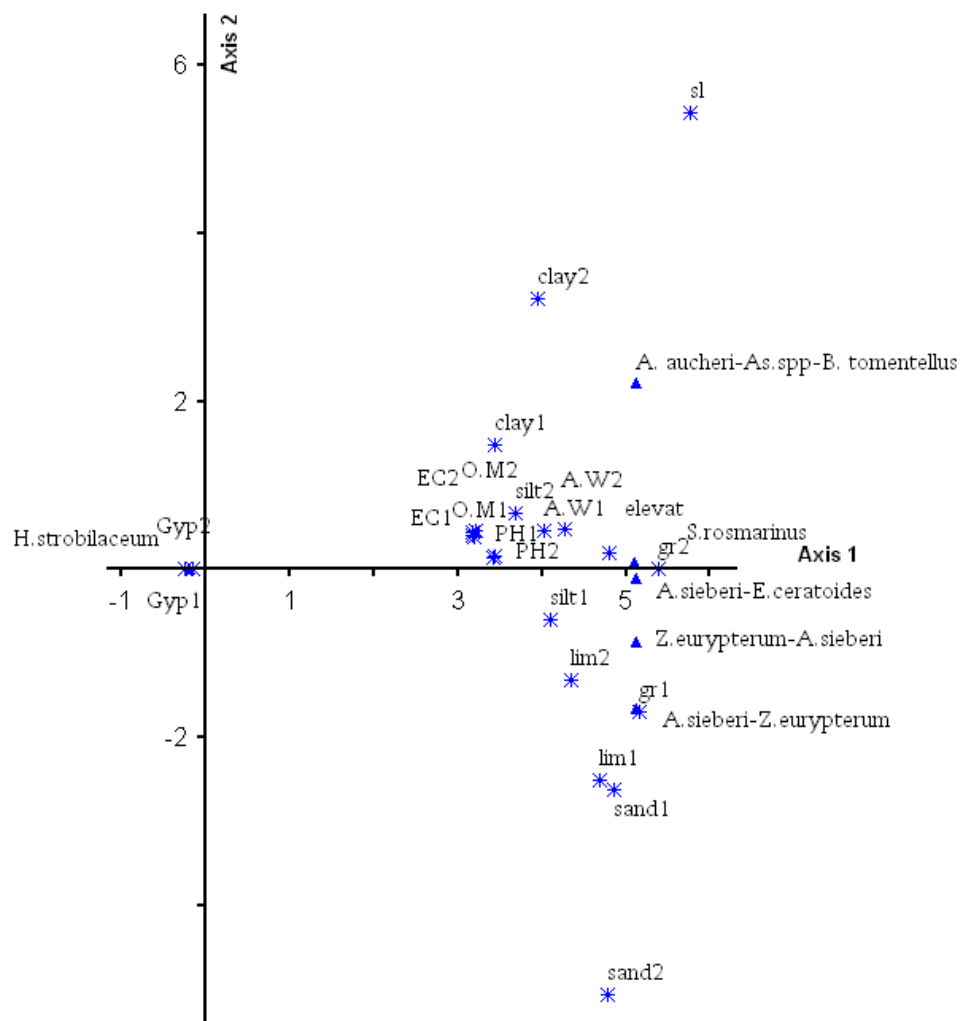

Figure 8. CCA–ordination diagram of the environmental data. For vegetation types and variables abbreviations, see Appendix A. (Δ) is the representative of the vegetation types. (*) is the representative of the environmental factors.

Reciprocal Averaging (RA) - Correspondence Analysis

RA is an ordination technique related conceptually to weighted averages. Because one algorithm for finding the solution involves the repeated averaging of sample scores and species scores (citations), Correspondence Analysis (CA) is also known as reciprocal averaging (Gittins, 1985).

RA places sampling units and species on the same gradients, and maximizes variation between species and sample scores using a correlation coefficient. It serves as a relatively objective analysis of community data.

CA is a graphical display ordination technique which simultaneously displays the rows (sites) and columns (species) of a data matrix in low dimensional space (Gittins, 1985). Row identifiers (species) plotted close together are similar in their relative profiles, and column identifiers plotted close together are correlated, enabling one to interpret not only which of the taxa are clustered, but also why they are clustered (Zhang et al,2005). Reciprocal analysis and canonical correlation analysis are linear methods. So, if well produced, their ordination diagrams are biplots or the superposition of biplots (a triplot). For illustration I use the Dune Meadow Data from Jongman et al. (1987). Reciprocal averaging is performed in PC-ORD by selecting options in program. Reciprocal averaging (RA) yields both normal and transpose ordinations automatically. Like DCA, RA ordinates both species and samples simultaneously. RA is the new technique that selects the linear combination of environmental variables that maximizes the description of the species scores. This gives the first RA axis. In RA, composite gradients are linear combinations of environmental variables, giving a much simpler analysis and the non-linearity enters the model through a unimodal model for a few composite gradients, taken care of in RA by weighted averaging. It provides a summary of the species-environment relations. This method is an ordination technique related conceptually to weighted averages. Results are generally superior to the results from PCA. However, RA axis ends are compressed relative to the middle, and the second axis is often a distortion of the first axis, resulting in an arched effect.

For example the analysis of variance showed in table.4 that there was a significant correlation among species and soil axis. The eigenvalues represent the variance in the sample scores. RA axis 1 has an eigenvalue of 0.86. RA axis 2 with an eigenvalue of 0.017 is less important. Table 6 shows the score classified site. Total variance (inertia) in the species data is 0.8887.

The results of RA ordination are presented in Fig 6. Six group sites were determined in relation to the environmental factors. Sites were determined in relation to the environmental factors.

The eigenvalue of the CA axis is equivalent to the correlation coefficient between species scores and sample scores (Gauch 1982b, Pielou 1984). It is not possible to arrange rows and/or columns in such a way that makes the correlation higher. The second and higher axes also maximize the correlation between species scores and sample scores, but they are constrained to be uncorrelated with (orthogonal to) the previous axes.

Since CA is a unimodal model, species are represented by a point rather than an arrow (Figure 7). This is (under some choices of scaling; see ter Braak and Šmilauer 1998) the weighted average of the samples in which that species occurs. With some simplifying assumptions (ter Braak and Looman 1987), the species score can be considered an estimate of the location of the peak of the species response curve (Figure 7).
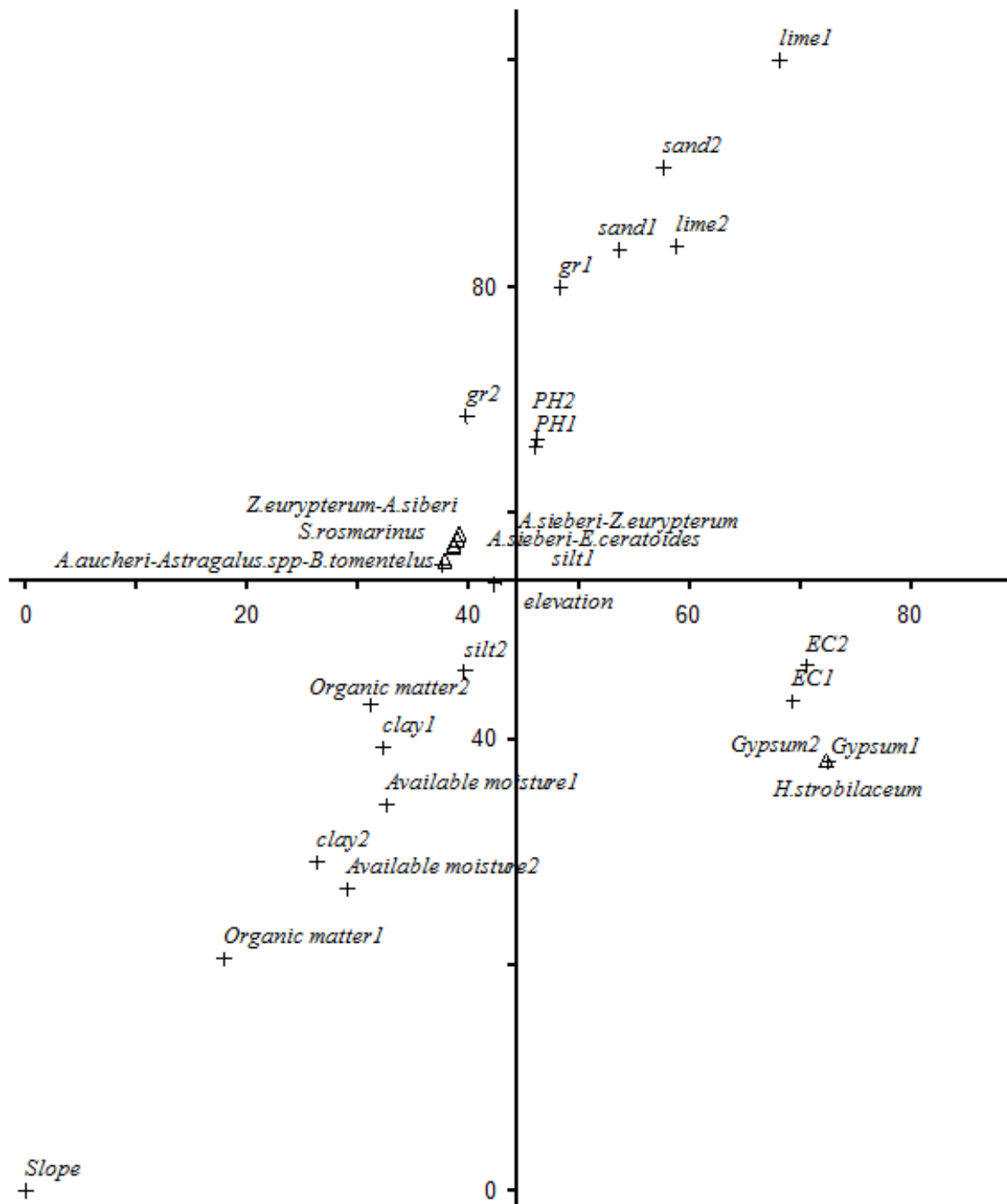

Figure 9. RA–ordination diagram of the environmental data. For vegetation types and variables abbreviations. (Δ) is the representative of the vegetation types. (+) is the representative of the environmental factors.

However, RA axis ends are compressed relative to the middle, and the second axis is often a distortion of the first axis, resulting in an arched effect.

N NAME	AX1	AX2	AX3	RANKED 1		RANKED 2	
				EIG=0.861		EIG=0.017	
1 Ar.si-Er.ce	2443	55	-97	1 Ar.si-Er.ce	2443	5 Ar.au-As.spp-B.to	206
2 Ha.sp	-25	0	0	3 Ar.si-Zy.eu	2441	2Ha.st	55
3 Ar.si-Zy.eu	2441	-73	-72	5Ar.au-As.spp-B.to	2435	1 Ar.si-Er.ce	0
4 Zy.eu-Ar.si	2421	-69	-25	4Zy.eu-A.si	2421	4Zy.eu-A.si	69
5 Ar.au-As.spp-B.to	2435	206	76	6Se.ro	2399	3 Ar.si-Zy.eu	73
6 Se.ro	2399	-161	131	2Ha.st	-25	6Se.ro	161

Table 6. Sample scores - which are weighted mean species scores

Row identifiers (species) plotted close together are similar in their relative profiles, and column identifiers plotted close together are correlated, enabling one to interpret not only which of the taxa are clustered, but also why they are clustered (Bakus,2007).

Reciprocal averaging (RA) yields both normal and transpose ordinations automatically. Like DCA, RA ordinates both species and samples simultaneously. Instead of maximizing 'variance explained', CA maximizes the correspondence between species scores and sample scores.

If species scores are standardized to zero mean and unit variance, the eigenvalues also represent the variance in the sample scores (but not, as is often misunderstood, the variance in species abundance).

The CA distortion is called the arch effect, which is not as serious as the horseshoe effect of PCA because the ends of the gradients are not incurved. Nevertheless, the distortion is prominent enough to seriously impair ecological interpretation (Bakus, 2007).

In other words, the spacing of samples along an axis may not affect true differences in species composition. The problems of gradient compression and the arch effect led to the development of Detrended Correspondence Analysis.

Detrended Correspondence Analysis (DCA)

Detrended correspondence analysis (DCA), an ordination technique used to describe patterns in complex data sets, and produced the following sequence of ordination axis scores (ter Braak,1986).

DCA is an eigenvector ordination technique based on Reciprocal Averaging, correcting for the arch effect produced from RA. Hill and Gauch (1980) report DCA results are superior to those of RA. Other ecologists criticize the detrending process of DCA. DCA is widely used

for the analysis of community data along gradients. DCA ordinates samples and species simultaneously. It is not appropriate for the analysis of a matrix of similarity values between community data (Gauch, 1982b).

Detrended Correspondence Analysis (DCA) eliminates the arch effect by detrending (Hill and Gauch 1982). There are two basic approaches to detrending: by polynomials and by segments (ter Braak and Šmilauer 1998). Detrending by polynomials is the more elegant of the two: a regression is performed in which the second axis is a polynomial function of the first axis, after which the second axis is replaced by the residuals from this regression. Similar procedures are followed for the third and higher axes. Unfortunately, results of detrending by polynomials can be unsatisfactory and hence detrending by segments is preferred. To detrend the second axis by segments, the first axis is divided up into segments, and the samples within each segment are centered to have a zero mean for the second axis (see illustrations in Gauch 1982). The procedure is repeated for different 'starting points' of the segments. Although results in some cases are sensitive to the number of segments (Jackson and Somers 1991), the default of 26 segments is usually satisfactory. Detrending of higher axes proceeds by a similar process.

One way to determine this relationship is to analyze the species data first by detrended correspondence analysis (DCA) and to examine the length of the maximum gradient. If the gradient exceeds 3 sd (sd¼standard deviation) (most of the species are replaced along the gradient), the data show unimodal response (Hill & Gauch, 1980). For example, in North East rangeland of Semnan, DCA axis 1 has an eigenvalue of 0.86 and a gradient length of 15.44. DCA axis 2 with an eigenvalue of 0.016 and a gradient length of 0.39 is less important. Fig 8 shows ordination diagram for vegetation types and soil variables. Table 5 shows the score classified site.

N NAME	AX1	AX2	AX3	RANKED 1		RANKED 2	
				EIG=0.861		EIG=0.017	
1 Ar.si-Er.ce	1714	23	10	1 Ar.si-Er.ce	1714	5 Ar.au-As.spp-B.to	39
2 Ha.sp	0	27	12	3 Ar.si-Zy.eu	1713	2Ha.st	27
3 Ar.si-Zy.eu	1713	8	0	5Ar.au-As.spp-B.to	1710	1 Ar.si-Er.ce	23
4 Zy.eu-Ar.si	1704	9	14	4Zy.eu-A.si	1704	4Zy.eu-A.si	9
5 Ar.au-As.spp-B.to	1710	12	12	6Se.ro	1694	3 Ar.si-Zy.eu	8
6 Se.ro	1694	0	15	2Ha.st	0	6Se.ro	

Table 7. Sample Scores- Weighted are weighted mean species scores (FIRST 6 EIGENVECTORS)

Figure 8 is an example of ordination plots showing the sites plotted on two axes. The ordination was a detrended correspondence analysis, and the sites with the same treatment level are outline for clarity.

One additional note, the different plots illustrate another common approach when using ordination: including only data on certain species thought to be more important as indicator species. This allows for different runs of the test to detect similarities or differences in composition based on a particular group.
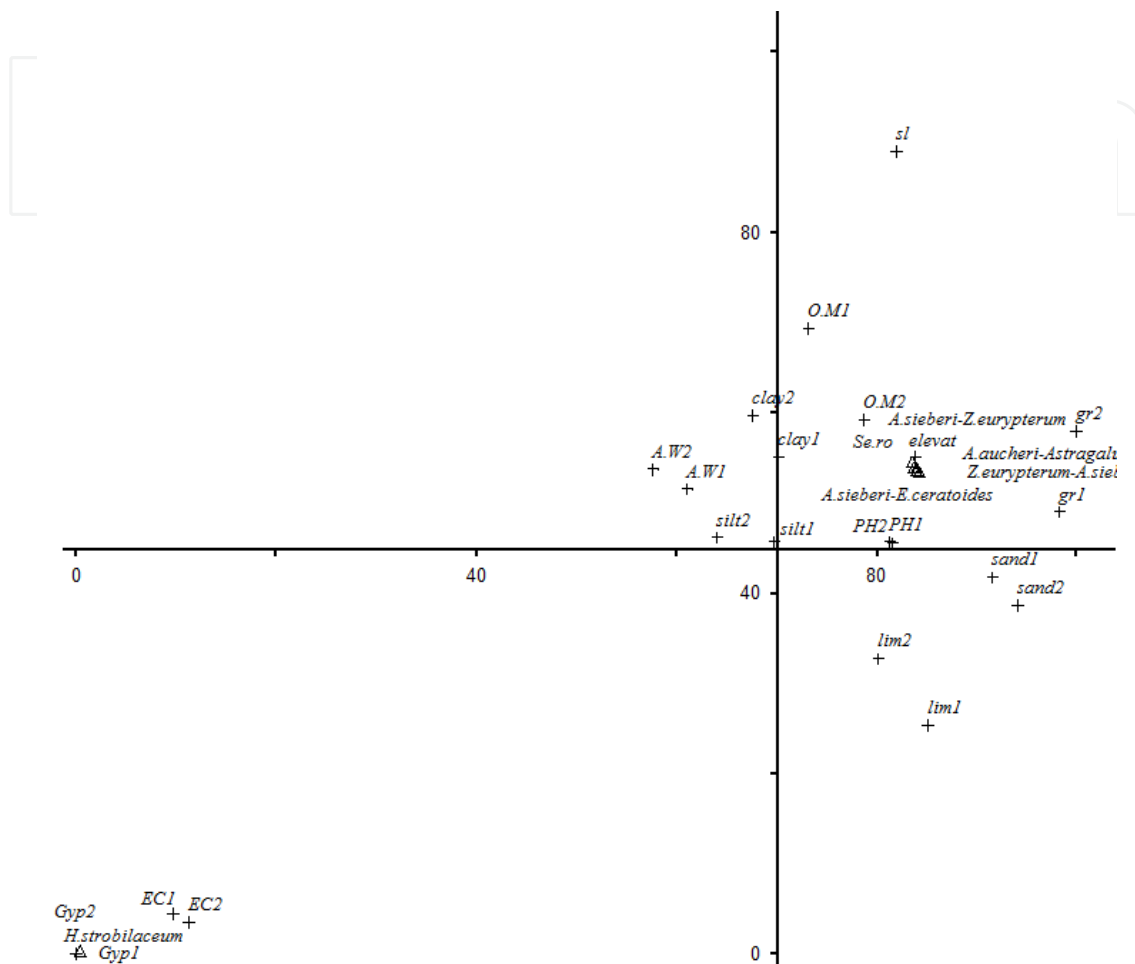
(Δ) is the representative of the vegetation types. (+) is the representative of the environmental factors.

Figure 10. DCA–ordination diagram of the environmental data. For vegetation types and variables abbreviations.

Nonmetric Multidimensional Scaling (NMS)

NMS actually refers to an entire related family of ordination techniques. These techniques use rank order information to identify similarity in a data set. NMS is a truly nonparametric ordination method which seeks to best reduce space portrayal of relationships. The verdict is still out on this type of ordination. Gauch (1982b) claims NMS is not worth the extra computational effort and that it gives effective results only for easy data sets with low diversity. Others hold NMS is extremely effective (Kenkel and Orloci, 1986, Bradfield and Kenkel, 1987).

DCA and NMDS are the two most popular methods for indirect gradient analysis. The reason they have remained side-by-side for so long is because, in part, they have different

strengths and weaknesses. While the choice between the two is not always straightforward, it is worthwhile outlining a few of the key differences.

Some of the issues are relatively minor: for example, computation time is rarely an important consideration, except for the hugest data sets. Some issues are not entirely resolved: the degree to which noise affects NMDS, and the degree to which NMDS finds local rather than global options still need to be determined (Bakus, 2007).

Since NMDS is a distance-based method, all information about species identities is hidden once the distance matrix is created. For many, this is the biggest disadvantage of NMDS (Bakus, 2007).
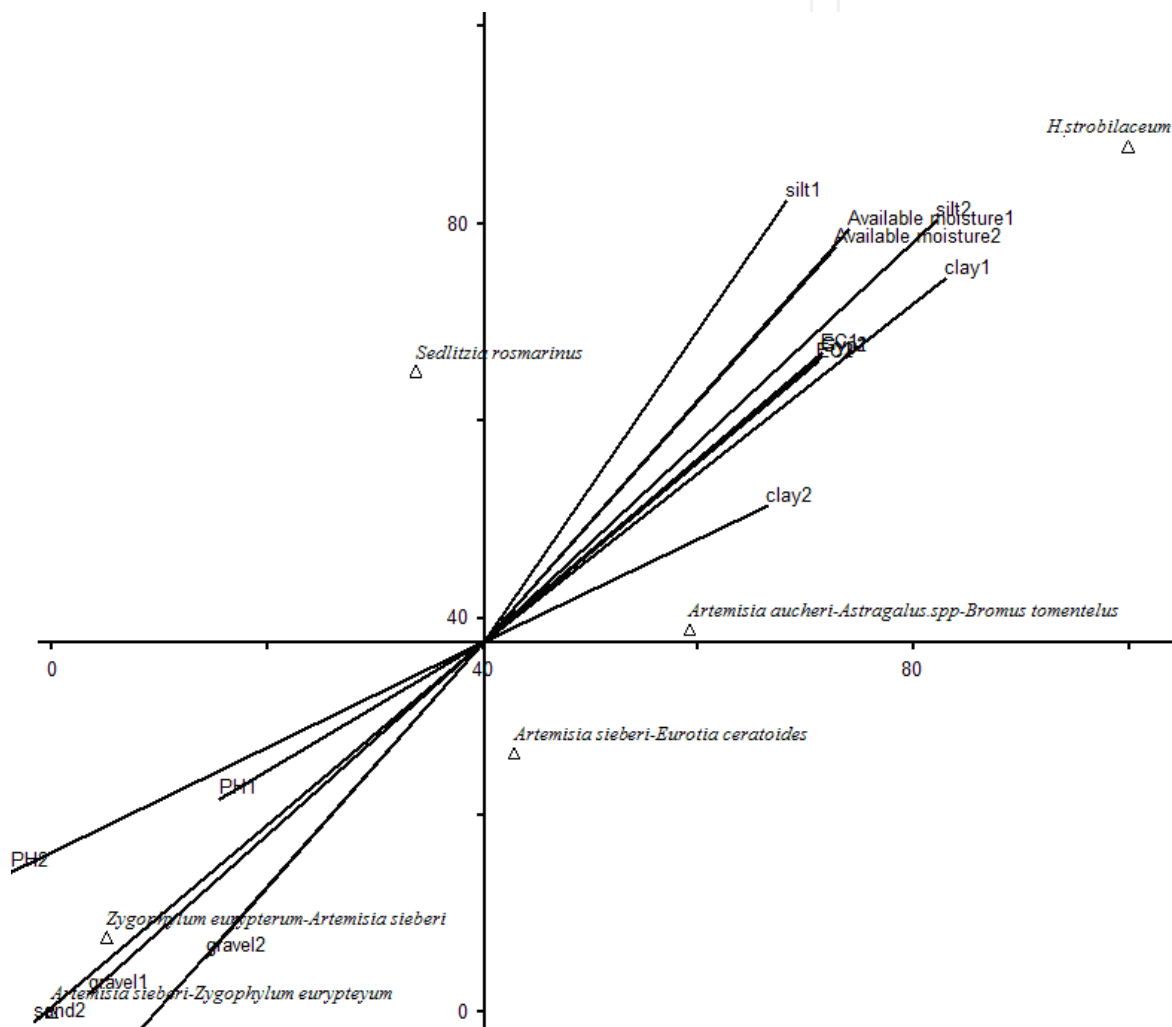

Figure 11. NMS ordination of plant species and environmental factors in along the rangelands of Semnan in Iran

DCA is based on an underlying model of species distributions, the unimodal model, while NMDS is not. Thus, DCA is closer to a theory of community ecology. However, NMDS may be a method of choice if species composition is determined by factors other than position along a gradient: For example, the species present on islands may have more to do with vicariance biogeography and chance extinction events than with environmental preferences

– and for such a system, NMDS would be a better *a priori* choice. As De'ath (1999) points out, there are two classes of ordination methods - 'species composition restoration' (e.g. NMDS) and 'gradient analysis' (e.g. DCA). The choice between the methods should ultimately be governed by this philosophical distinction.

Non-metric multidimensional scaling (NMS) (PC-ORD v. 4.25, 1999) was used to identify environmental variables correlated with plant species composition. A random starting location and Sorensen's distance measurement were used with the NMS autopilot slow and thorough method. Stepwise multiple linear regression (S-PLUS, 2000) was used to select models correlating vegetation cover and structure with environmental factors. Environmental explanatory factors that were not significant contributors (as determined from using stepwise selection at $\alpha = 0.05$) were excluded from the final model (Davies et al, 2007).

A Monte Carlo test of 30 runs with randomized data indicated the minimum stress of the 2 axes NMS ordination were lower than would be expected by chance ($p = 0.0968$). The final stress and instability of the 2-D solution were 23.71 and 0.00001, respectively. The first ordination axis (NMS1) captured 41.9% of the variability in the dataset and the second (NMS2) captured 31.8%, leading a cumulative 73.7% of variance in dataset explained (Fig.11).

5. Conclusion

Multivariate statistical analysis techniques were used to establish the relationships between plant diversity, Topography and soil factors. Plant community, structure and biodiversity have been shown to have a high degree of spatial variability that is controlled by both abiotic and biotic factors (Fu et al, 2004).

CCA is the constrained form of CA, and therefore is preferred for most ecological data sets (since unimodality is common). CCA also is appropriate under a linear model, as long as one is interested in species composition rather than absolute abundances (ter Braak and Šmilauer 1998). Correspondence analysis (CA) and canonical correspondence analysis (CCA) are widely used to obtain unconstrained unconstrained or constrained ordinations of species abundance data tables and the corresponding biplots or triplots which are extremely useful for ecological interpretation CA provided a good approximation for species with unimodal distributions along a single environmental gradient. There is a problem with this metric, however: a difference between abundance values for a common species contributes less to the distance than the same difference for a rare species, so that rare species may have an unduly large influence on the analysis (Greig-Smith 1983; ter Braak and Smilauer 1998; Legendre and Legendre 1998).

The most other general ordination technique, nonmetric multidimensional scaling (NMDS), which is based on the rankings of distances between points (Shepard, 1962), circumvents the linearity assumption of metric ordination methods. This method, used in ecological investigations (Kenkel and Orloci, 1986), Comparative studies of ordination techniques have, moreover, demonstrated the superiority of NMDS, and some authors have re commended its use, notwithstanding the computational burden.

The NMDS approach can in fact be tested each time measures of re semblance or dissimilarity are used to classify OTUs, whatever the causes and origins of arrangements found (Guiller et al, 1998).

In the biplots, where only the first two axes were used, all methods based upon PCA gave a fair representation of the relative numerical importance of the rare species. The weights in CCA are given by a diagonal matrix containing the square roots of the row sums of the species data table. This means that a site where many individuals have been observed contributes more to the regression than a site with few individuals. CCA should only be used when the sites have approximately the same number of individuals, or when one explicitly wants to give high weight to the richest sites. This problem of CCA was one of our incentives for looking for alternative methods for canonical ordination of community composition data.

For the analysis of sites representing short gradients, PCA may be suitable. For longer gradients, many species are replaced by others along the gradient and this generates many zeros in the species data table. Community ecologists have repeatedly argued that the Euclidean distance (and thus PCA) is inappropriate for raw species abundance data involving null abundances (e.g. Orlóci 1978; Wolda 1981; Legendre and Legendre 1998). For that reason, CCA is often the method favoured by researchers who are analysing compositional data, despite the problem posed by rare species.

De-trended correspondence analysis (DCA) is perhaps the most widely used method of indirect vegetation ordination. But direct ordination of vegetation and environment is achieved with canonical correspondence analysis (CCA). CCA is a relatively new method in which the axes of a vegetative ordination are restricted to linear groups of environmental variables (Zhang et al, 2006)

DCA and CA analyses should be run with the 'downweight rare species' option selected. We generally do not recommend NMS with the Euclidean distance measure; it performed the worst empirically, and has no advantages over the other methods (Culman et al, 2008)

Among the widely used ordination techniques for the plant community analysis Canonical Correspondence (CA) has shown to be superior to others such as PCA (Gauch, 1982). Most community data sets are heterogeneous and contain one or more gradients with lengths of at least two or three half-changes, which makes CA results ordinarily superior to PCA results. However, with relatively homogenous data sets with short gradients, PCA maybe better (Palmer, 1993). Despite the considerable superiority of the CA over PCA, CA is not superior to DCA, which corrects its two major faults such as "arch effect" and "compression of end of first axis" (Gauch, 1982; Kent & Coker, 1992).

For complex and heterogeneous data sets, DCA is distinctive in its effectiveness androbustness (Gauch, 1982). Comparative tests of different indirect ordination techniques have shown that DCA provides a good result (Cazzier & Penny, 2002). This study found that DCA provides better results than CA results (Malik & Husein, 2006).

For example all ordination techniques, used in North East rangeland of Semnan, clearly indicated that gypsum, EC, slope are the most important factors for the distribution of the vegetation pattern.

In the present study, combination of CCA, DCA and RA results showed that *Ar.aucheri-As.spp-Br.to, Artemisia sieberi-Erotia ceratoides, Ar.sieberi-Zy. eurypterum and Zy. eurypterum -Ar. sieberi* types correlated with A.W2, gr2, O.M2 and clay1 factors and clay in 0-20 depth indicates *Ar.aucheri-As.spp-Br.to* type. *H.strobilaceum* type has strong relationship with soil salinity and heavy texture. This species showed a trend to high soluble rate, salinity and clay percent. S. rosmarinus types indicate soils with light texture and this type directly related to pH and lime percentage while St.barbata-A.aucheri type shows an inverse relation with these factors.

I fact, analysis with DCA gave results similar to CCA, suggesting that there is a relatively strong correspondence between vegetation and environmental factors; with the difference that the DCA is less isolated the site. CCA better shows differences between types. RA shows relationship between sites and factors, like the CCA analysis. RA axis 1 has an eigenvalue of 0.86. RA axis 2 with an eigenvalue of 0.017 is less important. Total variance (inertia) in the species data is 0.8887.In this method eigenvalue of RA axis1 was higher than CCA and DCA axis1. This study reflects that a spatial approach dealing with the most distinctive species of vegetation communities can yield similar results to those obtained with costly physico-chemical analysis and based on complex matrices of plant communities.

Similarity as this study, also Jafari et al (2003) in their study in Hoz-e-Soltan Reigion of Qom Province, showed that PCA analysis indicates that Halocnemum strobilaceum type has direct relationship with Salinity, Lime, pH and Loam.

May this series of papers serve to enhance the understanding and the proper and creative use of ordination methods in community ecology. Finally, understanding relationships between environmental variables and vegetation distribution in each area helps us to apply these findings in management, reclamation, and development of arid and semi-arid grassland ecosystems (Alisauskas, 1998). The ability to factor out covariables and to test for statistical significance further extends the utility of CCA.

Understanding the relationships between ecological variables and distribution of plant communities can provide guidance to sustainable management, reclamation and development of this and similar regions. In this sense, these results increase our understanding of distribution patterns of desert vegetation and related major environmental factors in the North East of Semnan. The results will also provide a theoretical base for the restoration of degenerated vegetation in this area. Understanding the indicator of environmental factors of a given site leads us to recommend adaptable species for reclamation and improvement of that site and similar sites (Zhang et al, 2005)

Appendix

Artemisia sieberi-Erotia ceratoides.   *A.sieberi-E.ceratoides*
Halocnemum strobilaceum   *H. strobilaceum*
Artemisia sieberi–Zygophylom eurypoides.   *A.sieberi-Z.eurypterum*

Zygophylom eurypterum- Artemisia sieberi.	*Z.eurypterum –A.sieberi*
Artemisia aucheri-Astragalus spp.-Bromus tomentellus	*A.aucheri-As.sp.-Br.tomentellus*
Seidlitzia rosmarinus.	*S.rosmarinus*
Slope (%)	slope
Gravel (%)	gr
Clay (%)	clay
Silt (%)	silt
Sand (%)	sand
Available moisture (%)	A.W
Gypsum (%)	gyp
Lime (%)	Lim
PH(acidity)	pH
Electrical conductivity (ds/m)	EC
Organic matter (%)	O.M
Elevation (meter)	elevate

Code 1 is related to the soil characteristics were measured in the first layer (0–20 cm) Code 2 is related to the soil characteristics were measured in the second layer (20–80 cm)

Author details

Mohammad Ali Zare Chahouki

Associate Professor, Department of Rehabilitation of Arid and Mountainous Regions,
Natural Resources Faculty, University of Tehran, Iran

6. References

[1] Alisauskas, R. T. 1998. Winter range expansion and relationships between landscape and morphometrics of midcontinent Lesser Snow Geese. Auk 115(4):851-862.

[2] Anderson, M.J. & Ter Braak, C.J.F. (2002): Permutation tests for multi-factorial analysis of variance. Journal of Statistical Computation and Simulation (in press)

[3] Bakus Gerald J, 2007. Quantitative Analysis of Marine Biological Communities Field Biology and Environment. WILEY-INTERSCIENCE, A John Wiley & Sons, Inc., Publication,453p

[4] Beals, E. W. 1984. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. Adv. Ecol. Res. 14:1-55

[5] Bradfield, G.E. and Kenkel, N.C. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. Ecology 68: 750–753.

[6] Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Mon. 27:325-49

[7] Causton, D. R. 1988. An introduction to vegetation analysis. Unwin Hyman, London.

[8] Curtis, J. T., and R. P. McIntosh. 1951. An upland forest continuum in the prairie-forest border region of Wisconsin. Ecology 32:476-96

[9] Culman, S.W. H.G. Gauch, C.B., Blackwood & J.E, Thies, 2008. Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. Journal of Microbiological Methods 75 (2008) 55–63

[10] Daviesa, K.W., J.D, Batesa, R.F, Millerb,2007. Environmental and vegetation relationships of the Artemisia tridentata spp. wyomingensis alliance. Journal of Arid Environments 70 (2007) 478–494

[11] De'ath, G. 1999. Principal curves: a new technique for indirect and direct gradient analysis. Ecology 80:2237-53

[12] Dı́ez, I., A. Santolaria, J.M. Gorostiaga, 2003. The relationship of environmental factors to the structure and distribution of subtidal seaweed vegetation of the western Basque coast (N Spain). Estuarine, Coastal and Shelf Science 56 (2003) 1041–1054

[13] Dunn, C. P., and F. Stearns. 1987. Relationship of vegetation layers to soils in southeastern Wisconsin forested wetlands. Am. Midl. Nat. 118:366-74.

[14] Feoli, E. and Orl'oci, L. 1979. Analysis of concentration and detection of underying factors in structured tables. Vegetatio 40: 49–54.

[15] Gauch, H. G., Jr. 1982a. Multivariate Analysis and Community Structure. Cambridge University Press, Cambridge.

[16] Gauch, H. G., Jr. 1982b. Noise reduction by eigenvalue ordinations. Ecology 63:1643-9

[17] Gittins, R. (1985). Canonical analysis. A review with applications in ecology. Berlin: Springer-Verlag.

[18] Guiller, A., A, Bellido & L, Madec,1998. Genet ic Distances and Ordinat ion : The Land Sna il He lix aspe rsa in Nor th Afr ica as a Test Ca se. Syst . Biol . 47(2) : 208- 227

[19] Goodall, D. W. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Austral. J. Bot. 1:39-63

[20] Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley-Interscience, New York, Chichester, Brisbane, Toronto.

[21] Greig-Smith P (1983) Quantitative plant ecology, 3rd edn. Blackwell, London

[22] Hill, M. O. 1979. TWINSPAN - A FORTRAN programme for arranging multivariate data in an ordered two-way table by classification of individuals and attributes. Cornell University, Ithaca, New York.

[23] Hill, M.O., Bunce, R.G.H. & Shaw, M.V. (1975): Indicator species analysis, a divisive polythetic method of classification, and its application to survey of native pinewoods in Scotland. Journal of Ecology, 63: 597–613

[24] Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. J. Ecol. 61:237-49

[25] Hill, M. O. and Gauch, H. G. 1980. Deterended correspondence analysis, an improved ordination technique. Vegetatio 42:47-58.

[26] Fu, B.J., S.L, Liu, K.M,Ma1 & Y.G, Zhu,2004. Relationships between soil characteristics, topography and plant diversity in a heterogeneous deciduous broad-leaved forest near Beijing, China. Plant and Soil 261: 47–54,

[27] Jafari, M., M.A, Zare Chahouki., A, Tavili & H, Azarnivand, 2003.Soil-Vegetation Rellationships in Hoz-e-Soltan Region of Qom Province, Iran. Pkistan Journal of Nutrition 2(6):329-334

[28] Jongman, R. H. G., ter Braak, C. J. F. & van Tongeren, O. F. R. (1987). Data analysis in community and landscape ecology. Wageningen: Pudoc [new edition: 1994, Cambridge: Cambridge University Press].

[29] Kendall, M.A., Widdicombe, S., 1999. Small scale patterns in the structure of macrofaunal assemblages of shallow soft sediments. J. Exp. Mar. Biol. Ecol. 237, 127–140.

[30] Kenkel, N. C., and L. Orloci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67:919-928.

[31] Kent, M., and P. Coker. 1992. Vegetation description and analysis: a practical approach. Belhaven Press, London.

[32] Kent, M. (2006) Numerical classification and ordination methods in biogeography. Progress in Physical Geography 30, 399-408

[33] Krebs, Ch.J. 1999. Ecological methodology. Addison-Welsey educational publishers. 620pp

[34] Legendre, P., and L. Legendre. 1998. Numerical Ecology, 2nd English Edition. Elsevier, Amsterdam.

[35] Legendre, P & E.D, Gallagher, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia :129:271

[36] Lepˇs, J. and ˇ Smilauer, P. 2003. Multivariate Analysis of Ecological Data using CANOCO. Cambridge University Press, Cambridge.

[37] Malik, R.N., & S.Z, Husain, 2006. Classification and Ordination of vegetation communities of the lohibehr reserve forest and its surrounding areas. Rawalpini. Pak. J. Bot., 38(3): 543-558.

[38] Mucina, L. 1997. Classification of vegetation: past, present and future. J. Veg. Sci. 8: 751–760.

[39] McCune, B. and M.J. Mefford. 1999. PCORD. Multivariate Analysis of Ecological Data, Version 4. MjM Software Design, Gleneden Beach, Oregon, USA.

[40] McCune, B., Grace J.B. and D.L. Urban. 2002. Analysis of Ecological Communities. MjM Software Design, Gleneden Beach, Oregon.

[41] McIntosh, R. P. 1985. The Background of Ecology. Cambridge University Press, Cambridge, Great Britain.

[42] Orl'oci, L. 1978. Multivariate Analysis in Vegetation Research. 2nd ed. Junk, The Hague.

[43] Palmer, M. W. 1986. Pattern in corticolous bryophyte communities of the North Carolina piedmont: Do mosses see the forest or the trees? Bryologist 89:59-65

[44] Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74:2215-30

[45] Peet, R. K. 1980. Ordination as a tool for analyzing complex data sets. Vegetatio 42:171-4

[46] Pielou, E. C. 1984. The Interpretation of Ecological Data: A Primer on Classification and Ordination. Wiley, New York.

[47] Prentice, I. C. 1977. Non-metric ordination methods in ecology. J. Ecol. 65:85-94

[48] Pritchard, N.M & J.B, Anderson, 1971. Observations on the use of Cluster analysis in botany with an ecological example. Journal of Ecology,Vol. 59,No.3, pp.727-747

[49] Shepard, R.N. (1962): The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika, 27: 125–139

[50] Swan, J.M.A. 1970. An examination of some ordination problems by use of simulated vegetational data. Ecology 51: 89–102.

[51] Tavili, A. & Jafari, M., 2009. Interrelations Between Plants and Environmental Variables. Int. J. Environ. Res., 3(2):239-246

[52] ter Braak, C. J. F. 1985. CANOCO - A FORTRAN program for canonical correspondence analysis and detrended correspondence analysis. IWIS-TNO, Wageningen, The Netherlands.

[53] ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167-79

[54] ter Braak, C. J. F., and C. W. N. Looman. 1987. Regression. Pages 29-77 in R. H. G. Jongman, C. J. F. ter Braak and O. F. R. van Tongeren, editors. Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen, The Netherlands.

[55] ter Braak, C. J. F., and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio 65:3-11

[56] ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Adv. Ecol. Res. 18:271-313

[57] ter Braak, C. J. F., and P. Šmilauer. 1998. CANOCO reference manual and User's guide to Canoco for Windows: Software for Canonical Community Ordination (version 4). Microcomputer Power, Ithaca.

[58] Van der Maarel E.,1979. Transformation of cover-abundance values in phytosociology and its effect on community similarity. - Vegetatio, 38: 97 – 114

[59] Van der Meer, J. (1991). Exploring macrobenthos-environment relationship by canonical correlation analysis. Journal of Experimental Marine Biology and Ecology, 148, 105-120.

[60] Varis, O. (1991). Associations between lake phytoplankton community and growth factors – a canonical correlation analysis. Hydrobiologia, 210, 209-216.

[61] Whittaker, R. H. 1967. Gradient analysis of vegetation. Biol. Rev. 42:207-64

[62] Whittaker, R. H. 1969. Evolution of diversity in plant communities. Brookhaven Symp. Biol. 22:178-95

[63] Wildi, O., 2010. Data analysis in vegetation ecology. A John Wiley & Sons, Ltd., Publication, 211pp

[64] Williams, WT. & Lambert, J.M. (1959). Multivariate methods in plant ecology.I, Association analysis in plant communities .J. Ecol.47,83-101.

[65] Wolda H (1981) Similarity indices, sample size and diversity. Oecologia 50:296–302

[66] Zare Chahouki, M.A., Khalasi Ahvazi, L. & Azarnivand, H.,2010. Environmental factors affecting distribution of vegetation communities in Iranian Rangelands. VEGETOS Journal, 23(2): 1-15.

[67] Zhang W H, Lu T, Ma KM, et al. Analysis on the environmental and spatial factors for plant community distribution in the arid valley in the upper reach of Minjiang River. Acta Ecologica Sinica, 2004, 24(3): 532–559

[68] Zhang_, Y.M., Y.N. Chen, B.R. Pan, 2005. Distribution and floristics of desert plant communities in the lower reaches of Tarim River, southern Xinjiang, People's Republic of China. Journal of Arid Environments 63 (2005) 772–784

[69] Zhang ,B, Valentine,I, Kemp,P, Lambert,G, 2006. Predictive modelling of hill-pasture productivity: integration of a decision tree and a geographical information system. Agricultural Systems 87 (2006) 1–17.