Iranian Journal of Forest
Vol. 16, No. 5, Winter 2025
pp. 45-55

DOI: 10.22034/ijf.2025.337849.1867

Research Article

The Effect of Data Transformation on Detrended Correspondence Analysis in Vegetation Studies (Case study: Kermanshah oak forests)

Naghmeh Pakgohar[1], Javad Eshaghi Rad[2*], Abbas Banj Shafiei[2], and Jalil Alavi[3]

[1]Ph.D in Forestry, Faculty of Natural Resources, Urmia University, Urmia, I.R. Iran
[2]Prof., of Forestry, Faculty of Natural Resources, Urmia University, Urmia, I.R. Iran
[3]Associate Prof., of Forestry, Faculty of Natural Resources, Tarbiat Modarres University, Noor, I.R. Iran

(Received: 10 May 2022; Accepted: 14 October 2024)

Abstract
Vegetation sampling is a fundamental practice in ecological research, often quantified by estimating species cover. This process typically involves transforming the data to account for the presence of common and rare species, as well as zeros in the dataset, which can significantly influence the analysis results. The subsequent analysis of these datasets aims to elucidate community patterns and dynamics within various ecosystems. However, the optimal transformation method for multivariate analysis remains a subject of ongoing debate among ecologists, as different transformation techniques can yield varying interpretations of ecological data. In this study, we aimed to evaluate the effects of different data transformations on the results of detrended correspondence analysis (DCA), specifically within oak forests (*Quercus brantii* Lindl.) located in the Zagros region of Iran. This region is characterized by its unique biodiversity and ecological significance, making it an ideal setting for such investigations. To achieve our objectives, we selected three distinct forest patches characterized by similar slopes and altitudes, ensuring that environmental variables were controlled. Vegetation sampling was conducted at five specific distances—0, 25, 50, 100, and 150 meters—along three transects that were spaced 200 meters apart from each other. This systematic approach allowed us to obtain a comprehensive representation of species distribution across the forest patches. We utilized the Braun-Blanquet cover percentage and the van der Maarel scale to prepare our datasets, ensuring consistency and reliability in our measurements. Each dataset underwent various transformations, including log, square root, and general relativization transformations. Subsequently, we applied DCA to each transformed dataset and compared the resulting ordination outcomes through Procrustes analysis, a method that quantifies the similarity between two datasets. The findings revealed that both log-transformed and square-root transformed datasets significantly enhanced the DCA results by effectively decreasing the variation present in the dataset. Procrustes analysis demonstrated that the concordance between the log-transformed and square-root transformed datasets and the raw data was significantly higher than that of the other transformations evaluated. Importantly, our results indicated that general relativization transformations were unsuitable for DCA analysis, as they did not adequately represent the underlying ecological relationships. Consequently, we recommend performing data transformations, particularly log or square root transformations, prior to conducting ordination analyses. This approach will not only enhance the reliability of the results but also facilitate more accurate ecological interpretations, ultimately contributing to a deeper understanding of community dynamics in forest ecosystems.

Keywords: General relativization, log transformation, Procrustes analysis, Square root, Zagros forests.

1. Introduction
Understanding vegetation structure and composition, through identification and description of the plant communities and their relation to the environment is a crucial step (Al Harthy & Grenyer, 2019) that affects conservation and management planning (Kent et al., 2012). Community ecologists seek to assess the complex abiotic and biotic relationships within ecosystems (Chahouki, 2013). They have applied multivariate approaches for understanding community data (McCune & Grace, 2002).

The structure of data has a considerable

*Corresponding Author Tel: +984432770489 Email: j.eshagh@urmia.ac.ir

effect on the output of multivariate analysis (McCune & Grace, 2002; Roberts, 2008; Lengyel & Podani, 2015). Vegetation community data sets consist of many zeros, which mostly affects dominant species (Legendre & Bocard, 2018). Thus, preparation of data before multivariate analysis is important (Jackson, 1993; Tukey, 1977; Zuur et al., 2010; Legendre & Bocard, 2018). Data transformation is the application of a mathematical treatment to convert data from one format to another (McCune & Grace, 2002); you get the data in exactly the form you need (Wickham & Grolemud, 2016). A vegetation ecologist may apply data transformation before performing multivariate analysis. Reasons for data transformation include improving the normality of the data (reducing the skewness of species data and decreasing zeros), increasing variation, and changing the weight of species or samples (McCune & Grace, 2002; Legendre & Bocard, 2018).

Data transformation means applying a linear or non-linear function to the data. It commonly includes logarithmic, square-root, and relativization (McCune & Grace, 2002). The selection of a suitable data transformation is essential (Tukey, 1977; Legendre & Bocard, 2018). However, most scientific articles do not discuss it (Roberts, 2008; Legendre & Borcard 2018). Lengyel et al. (2018) evaluated the effects of data transformation on the stability of clustering, and their results showed that, with changing the data transformation, classification stability was considerably changed. Legendre & Bocard (2018) evaluate the box-cox transformation for community composition data. Lengyel et al. (2018) studied optimal cluster number and varying scale of transformation power to reach stable classification. Tichy et al. (2020) assessed the optimal transformation of species cover using various transformations (power transformation, logarithmic transformation, and pseudo-species cut level) and applied a flexible-beta clustering algorithm. They recommended power transformation with exponent. 0.5 for unsupervised classification. Probably the most commonly used transformations for percentage cover are logarithmic (Dooley & Collinus, 1984; Gartner & Reif, 2004; Hardtle et al., 2006) and square-root (Molder et al., 2008), but other transformations are also applied (Tichy et al., 2020). Since then, it is recommended that the percentage cover values be transformed prior to calculation (Legendre & Legendre, 2012; milauer & Lep, 2014; Wildi, 2017).

Ordination is a multivariate technique (McCune & Grace, 2002) which is most often used in ecology to explore and explain the hidden patterns of structure and to visualize the gradient of vascular plant species composi-tion (Legendre & Legendre, 1998; McCune & Grace, 2002; García-Mijangos et al., 2021). Ordination methods have been widely used in plant community analysis (e.g., Xiaping et al., 2006; Ahmad & Yasmain, 2011; Khan & Hussain, 2013; Garca-Mijangos et al., 2021; Golizadeh et al., 2020). The ordination techniques are divided between indirect and direct gradient analysis (Gauch, 1982): Indirect gradient analysis utilizes only the species by sample matrix, essentially asking the species what the most important gradients are (Gauch, 1982). DCA and NMDS are the two most popular methods for indirect gradient analysis. DCA is based on an underlying model of species distributions, the unimodal model, while NMDS is not (McCune & Grace, 2002). Also Detrended correspondence analysis (DCA) is one of the most commonly used ordination methods which is a development of reciprocal averaging (= correspondence analysis) that avoids its two major disadvantages. Several studies have applied DCA (e.g., Diekmann et al., 1999; Schuman et al., 2003; Uotila & Kouki, 2005; Hardtle et al., 2005; Husain & Malk, 2006; Ruokolainen & Salo, 2006; Hardtle et al., 2006; Xianping et al., 2006; Molder et al., 2008; Super et al., 2013; Gulshan et al., 2014; Golizadeh et al., 2020; For example, Diekmann et al. (1999) assessed beech (*Fagus sylvatica*) forest communities in the Nordic countries by multivariate analysis. In order to standardize the data so that all data from the sample plots became comparable, all abundance values were transformed to a presence/absence scale.

Despite multiple attempts to transform vegetation datasets; the optimal data transformations for ordination methods are unclear. In this paper, we aim to assess the effect of commonly used data transformations in a vegetation dataset on the DCA method, one of the superior indirect ordination methods, and compare DCA results based on log transformations, square root transformations, and general relativization transformations on raw data and rescaled datasets to increase our robust understanding of plant communities.

1.1. Materials and methodsStudy Area

The research area is located in the Zagros forests in Kermanshah province (Iran) between 34° 09′ and 34° 14′ N and 46° 39′ and 46° 49′ E. Three oak (*Quercus brantii* Lindl.) patches, located on southern aspect, with similar slope and elevation conditions were selected. The mean annual temperature is 15°C, and mean annual rainfall is 500 mm (Eshaghi Rad et al., 2017).

1.2. Floristic sampling

This study was conducted in Cheharzebar Forests, Kermanshah. Vegetation sampling in each of the three stands was performed at distances of 0, 25, 50, 100, and 150 m along three transects placed 200 m apart (45 sampling points in total). For vegetation data collection, at each sampling point, two 400 m^2 (20 × 20 m) plots were designed for tree and shrub layers and five 0.25 m^2 (0.5 × 0.5 m) plots were implemented for herbaceous vegetation which were located orthogonally to the left and right of the transect. In each plot, all species were identified, and their percent cover and abundance of all species were estimated separately using the Braun-Blanquet scale (Ellenberg, 1974).

1.3. Data analysis

We tested the influence of cover transformation on vegetation ordination. The dataset consisted of 45 plots and 130 species. Two data matrices were derived from the same dataset for preparation. The first matrix included the Braun-Blanquet cover percentage, and the second matrix was rescaled with the van der Maarel scaling method (van der Maarel, 1979). There are multiple approaches to transforming a dataset. Transformed with log, square root and general relativization transformations are the most commonly used. We applied the data transformations on the Braun-Blanquet cover percentage data and van der Maarel scaling. Each data type was subject to three data transforms. In total, eight data matrices were prepared for this study that each one referred to as a dataset.

Before processing began, we assessed the gradient length of the raw dataset. The results indicated that dataset exhibited a compositional turnover > 1.5 standard deviations (Ter Braak & Prentice, 1988). Therefore, detrended correspondence analysis (DCA) was more suitable method to employ for the dataset. Based on the Kaiser-Guttman and Broken Stick criteria, principal component analysis and correspondence analysis were not used for the dataset (Froniter, 1976; Guttman, 1954). Then DCA was applied to all data sets (Hill & Gauch, 1980). The effect of data transformation is assessed using total variance (Legendre and Gallagher, 2001), eigenvalue of ordination axes (Diekmann et al., 1999) and length of gradient (Ter Braak & Prentice, 1988) to determine whether data transformation reduced variation within the dataset. Finally, the results of DCA analysis on different datasets were compared by Procrustes analysis. The details of all methods used were explained further below.

1.4. Data transformation

- Logarithmic transformation: This compresses high values and spreads low values by expressing the values as orders of magnitude. Log transformation is often useful when there is a high degree of variation within variables or samples, and generally, it is useful for many types of environmental and habitat datasets (McCune & Grace, 2002). However, it is not recommended for count data (O'Hara & Kotze, 2010),

Square-root transformation: converts data with a Poisson distribution into a normal distribution (Tabachnick, 2007). This transformation is similar to the log transformation. However, it is less drastic than a log transformation and it does not require special treatment with zeros.

- General relativization: rescales individual rows (or columns) in relation to some criterion based on the other rows (or columns) Relativization is an extremely important tool that all ecologist should understand multivariate statistics (McCune & Grace, 2002).

1.5. DCA

DCA is a commonly used method in vegetation analysis that is used to find patterns within vegetation community datasets. This ordination method belongs to the unimodal ordination methods based on correspondence analysis. DCA is an eigenvector ordination technique based on reciprocal averaging, but its performance has improved since the arch and compress effects are eliminated in this ordination method (McCune & Grace, 2002). Ordination analysis was conducted using PC-ORD software Version 4 (McCune & Mefford, 1999).

1.6. Procrustes analysis

This technique involves the comparison of two or more ordinations (Jackson, 1995). The degree of concordance between ordinations is estimated using a rotational-fit algorithm (Minchin, 1987; Legendre & Legendre, 1998). This widely used technique calculates the correlation between two ordinations, assesses statistical significance (Gower, 1971; Jackson, 1991), and minimizes the sum of squares between the two ordinations (Peres-Neto & Jackson, 2001). Low Procrustes residuals (m12) indicate that the two ordinations have strong congruence (Levis et al., 2014). Statistical analyses were carried out using R software and the vegan package version 3.1.0. (Oksanen et al., 2013).

2. Results

2.1. Ordination results

Table 1 presents the eigenvalues, total variation, and gradient lengths on axes one and two of DCA based on the different datasets in terms of cover percentage value. In the raw dataset, the eigenvalues of the first axis were 0.25, and the eigenvalue of the second axis was 0.12. However, the values of the first axis of the square-root and log transformations were lower than those of the raw dataset, whereas the values for general relativization were higher than the raw dataset eigenvalue. The total variation explained by the raw dataset was 2.54; the value of total variation in square-root and log transformations were less than in the raw data set but increased in the general relativization transformation. The gradient lengths in the square-root and log transformed data were shorter than those of the raw dataset. However, all three DCA statistics increased with general relativization. The eigenvalues for DCA axis two were identical for square-root and log transformed datasets.

Results of the DCA ordination of the data sets that were rescaled using the van der Maarel scaling method are presented in Table 2. The results indicate that the eigenvalues for the first axis of the raw dataset were 2.75. The eigenvalues for the first axis in the square-root and and log-transformed datasets were 0.23 and 0.24 (i.e., less than the raw data set eigenvalue). The eigenvalue of the general relativization transformation data set was 0.42. The total variation in DCA decreased with log and square-root transformations but increased with the general relativization transformation. The length of the gradient in the general relativization transformation dataset was the highest among other datasets.

2.2. Procrustes analysis

A permutation test (PROTEST) indicated a relatively strong correlation between the log-transformed data and the raw data with a p-value of 0.001 (Tables 3 and 4). The residuals (arrows in Figures 1 and 2) represent the degree of congruence between the ordination results of the two data sets based on the sample scores. Low Procrustean residuals were observed in several datasets (e.g., log and square-root transformed). The poorest correspondence (Procrustes residuals >0.6) was observed between the raw data sets and general relativization in van der Maarel rescaled data. Additionally, the log-transformed data and the general relativisation had the lowest correlation. M_{12} indicates the similarity between two datasets in the protest analysis. If M_{12} is less than 0.6, the similarity between two datasets is considered high. Furthermore, the results of comparisons between the original dataset and other data transformations, as well as those the van der Maarel rescaled dataset and other transformations, were similar (Figures 1 and 2)

Table 1. Results of DCA ordination with different data types (original data)

Data type	Total explained variation	Length of gradient (standard deviation units)	Eigenvalue	
			Axis 1	Axis 2
Raw data	2.54	2.33	0.25	0.12
Square-root	2.41	2.04	0.24	0.13
Log(x+1)	2.48	2.15	0.23	0.13
General relativization	4.47	2.97	0.38	0.25

Table 2. Results of DCA ordination with different data types (using ordinal scale of van der Maarel)

Data type	Total variation	Length of gradient (standard deviation units)	Eigenvalue	
			Axis 1	Axis 2
Raw data	2.90	2.75	0.26	0.18
Square root	2.49	2.18	0.23	0.13
Log(x+1)	2.60	2.29	0.24	0.13
General relativization	6.51	3.84	0.42	0.30

Table 3. Correlation and M_{12} between different data types used in DCA (original data)

Data type	p-value	M_{12}	r
Raw data * square root data	0.001*	0.28	0.85
Raw data * log data	0.001*	0.15	0.92
Raw data * general relativization	0.001*	0.17	0.91

* Significant at the 0.05 level.

Table 4. Correlation and M_{12} between different data types used in DCA (ordinal scale of van der Maarel)

Data type	p-value	M_{12}	r
Raw data * square root data	0.001*	0.20	0.89
Raw data * log data	0.001*	0.20	0.89
Raw data * general relativization	0.001*	0.69	0.56

* Significant at the 0.05 level.

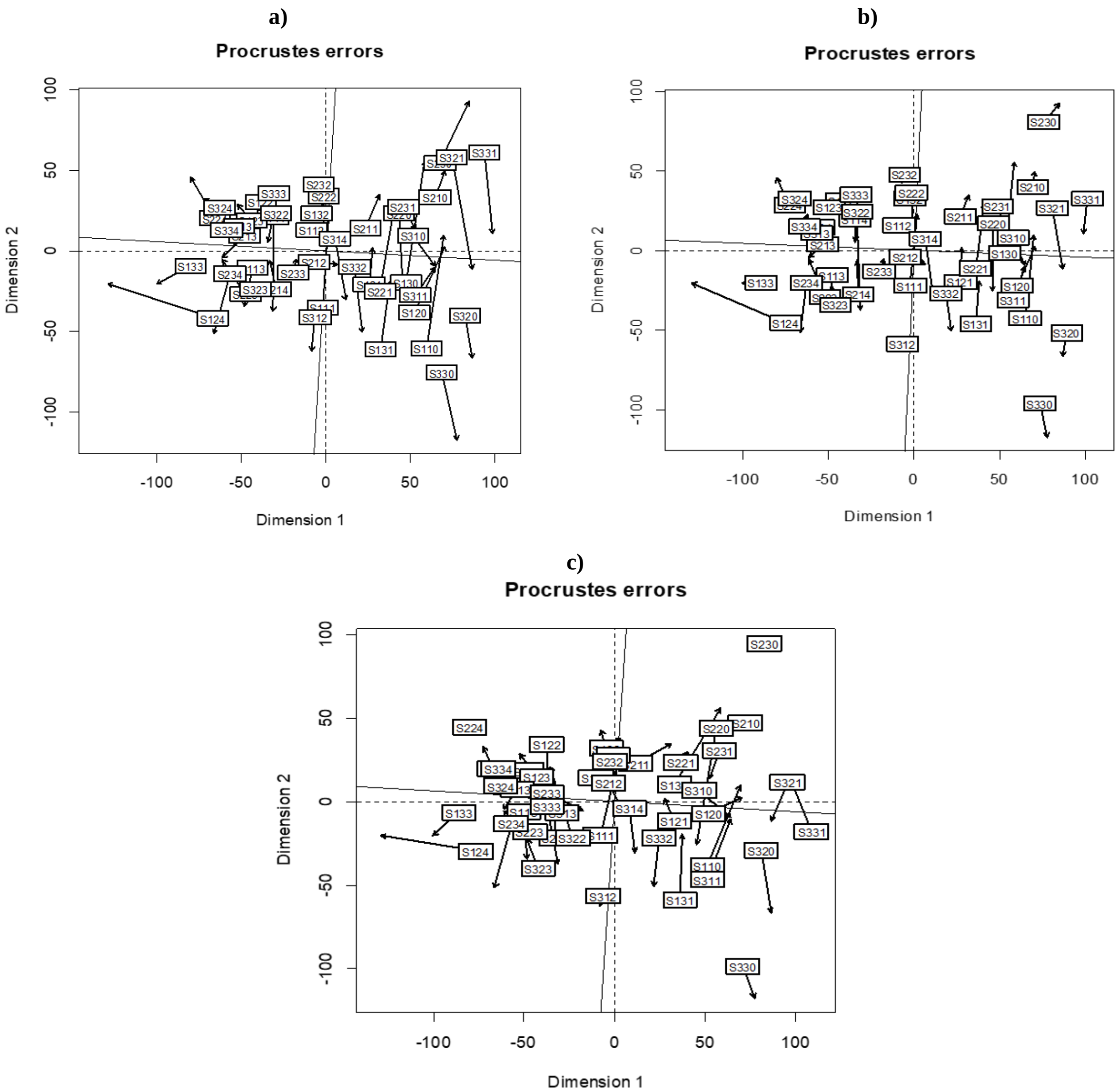

Figure 1. Configuration obtained from pairwise Procrustean rotation between four two-dimensional ordination solutions. (a) raw dataset and square-root dataset, (b) raw dataset and log dataset, (c) raw dataset and general relativization datasets

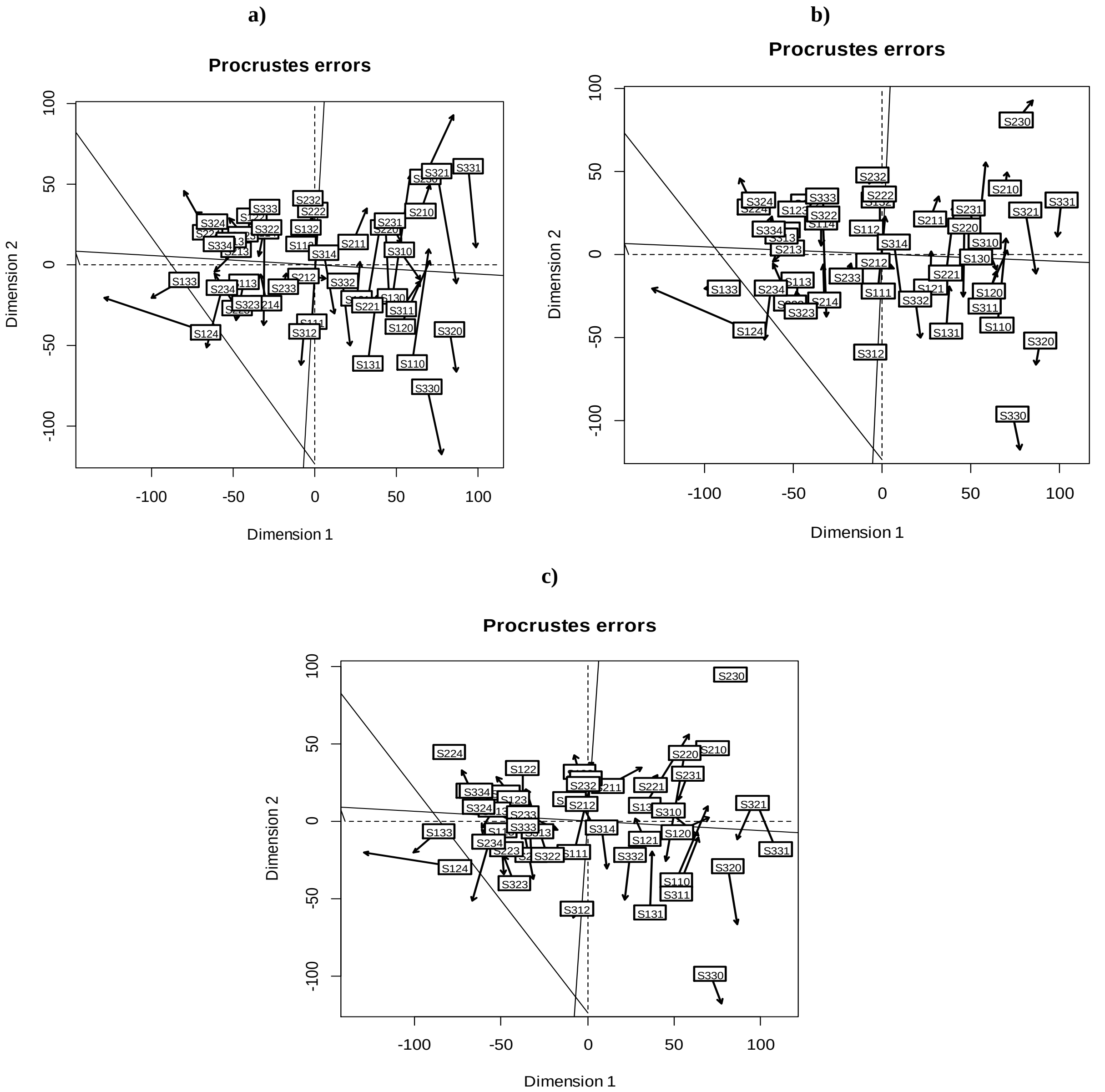

Figure 2. Configuration obtained from pairwise Procrustes rotations between four two-dimensional ordination solutions in which the data have been scaled using the van der Maarel scaling method. (a) raw data-set and square-root dataset, (b) raw dataset and log dataset, (c) raw dataset and general relativization data

3. Discussion

As we know, cover values are very important for vegetation analyses (see also Lengyel & Podani, 2015; Lengyel et al., 2018). Data transformation is one of the most crucial parameters while applying multivariate techniques (Lengyel & Podani, 2016) and can reduce noise and the zero effect in a vegetation data set (McCune & Grace, 2002). Our results indicated that data transformations significantly improved the ordination results and were appropriate for community compositional data. This finding was in line with Legendre & Gallagher (2001). The present study illustrated that the result of the DCA was different among different data sets. According to the results, the log transformation and square root transformation increased DCA effectiveness while decreasing total variance and length of gradient in data set A. Also, Procrustes analysis indicated that the concordance between the log-transformed dataset and the square-root-transformed dataset as well as with the raw data was high. The results also confirm previous studies that have indicated the utility of log transformation is useful for community analysis; for example, Hill & Gauch (1980) applied log transformations in DCA to reduce the effect of the dominant species, or Roberts (2015) used log transformation to achieve convex transformation before classification of the vegetation dataset. In addition, Legendre & Bocard (2018) suggested using log transformations for ecological datasets.

The effectiveness of square-root transformation was similar to that of log transformation, but it was less drastic in normalizing data (McCune & Grace, 2002). As our results showed, square-root and log transformations yielded similar outcomes. Moreover, Maindonald & Braun (2007) recommended using square root transformation to normalize count data. Lengyel et al. (2018) stated that weak power transformations (with the exponent close to 1) that preserved the differences in original abundance patterns, would yield a higher cluster number, while strong transformations (the exponent approaching 0) that significantly reduce abundance differences and identify the optimal number of clusters in the lower half of the cluster range. Legendre & Bocard (2018) pointed out that the square root reduces the asymmetry of modestly asymmetric data distributions before subjecting data to linear methods of analysis. Tichy et al. (2020) recommended using the square-root transformation for percentage cover before unsupervised classification. Also, Lengyel et al. (2018) found that, with power transformation, low exponents (near-zero) classifications best explained patterns of vegetation communities. Pakgohar et al. (2021) applied the Hellinger transformation to raw abundance data obtained by taking the square root and concluded that the Hellinger transformation can improve clustering results.

Experience shows that data transformations can help achieve symmetry in ecological data, making there a choice between square-root and log transformations (Roberts, 2008). General relativization is useful when a Euclidean distance measure is used. Detrended correspondence analysis uses a chi-square distance measure, so general relativization is inappropriate for DCA (Roberts, 2008). The results indicated that using the original data or the van der Maarel rescaled data made no difference in the efficiency of DCA.

4. Conclusions

In the present study, after analyzing the results of DCA with one dataset, two scales of datasets and three data transformations, it was found that data transformations can be a useful tool to improve the output of analysis and provide reliable and repeatable results; they decreased the heterogeneity of the data set and enhanced the performance of DCA. Specially, square-root and log transformations were good choices among the assessed data transformations. Before using statistical models, researchers should be aware of the benefits of data transformation. We believe that our study shows the importance of selecting an appropriate data transformation before multivariate analyses and using the DCA technique.

References

Ahmad, S., & Yasmin, T. (2011). Vegetation classification along Hanna Lake, Baluchistan using ordination techniques. *Pakistan Journal of Botany*, *43*(2), 863-872.

Chahouki, M.A.Z. (2013). Classification and ordination methods as a tool for analyzing of plant communities. Multivariate analysis in management, engineering and the sciences, 221p.

Diekmann, M., Eilertsen, O., Fremstad, E., Lawessonal, J., & Aude, E. (1999). Beech forest community in Nordic countries – a multivariate analysis. *Plant Ecology*, *140*(2), 203-220. https://doi.org/10.1023/A:1009768225798.

Dooley, K., & Collins, S. (1984). Ordination and classification of Western Oak Forests in Oklahoma. *American Journal of Botany*, 71, 1221-122. https://doi.org/10.1002/j.1537-2197.1984.tb11977.

Eshaghi Rad, J., Soleimani, F., & Khodakarami, Y. (2017). Comparison of flora at the edge and within Oak forests in southern slopes of Kermanshah forests. *Applied Biology*, *30*(1), 19-36. https://doi.org/10.22051/jab.2017.2986. (In Persian)

Froniter, S. (1976). Study of the decrosisaca of valcure pro pers in principal components analysis. *Biology and Ecology*, *25*, 67-75.

García-Mijangos, I., Berastegi, A., Biurrun, I., Dembicz, I., JaniŠová, M., Kuzemko, A., ... & Dengler, J. (2021). Grasslands of Navarre (Spain), focusing on the Festuco-Brometea: classification, hierarchical expert system and characterisation. *Vegetation Classification and Survey*, *2,* 195. https://doi.org/10.1002/10.3897/VCS/2021/69614

Gartner, S., & Reif, A. (2004). The impact of forest transformation on stand structure and ground vegetation in the southern Black Forest, Germany. *Plant and Soil*, *264*, 35-51. https://doi.org/ 10.1023/B:PLSO.0000047751.25915.77

Gartner, S., & Reif, A. (2004). The response of ground vegetation to structural change during forest conversion in the Southern Black Forest. *European Journal of Forest Research*, *124*, 221-231. https://doi.org/ 10.1007/s10342-005-0065-7

Gauch, H.G. (1982). Multivariate Analysis in Community Ecology. Cambridge: Cambridge University Press https://doi.org/10.1017/CBO9780511623332

Gulshan, A., Salma, S., Malik, M., & Saeed, S. (2014). Probe the community structure of vegetation from Piedmont downhill to alluvial plains of Dera Ghazi Khan. *International Journal of Agriculture and Crop Sciences*, *7*, 91-99.

Gower, J.C. (1971). Statistical methods for comparing different multivariate analyses of the same data. In: Mathematics in the Archaeological & Historical Sciences (Eds F.R. Hodson, D.G., Kendall & P., Tautu). NewYork: University Press

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19*,149-161. https://doi.org/10.1007/BF02289162

Al Harthy, L., & Grenyer, R. (2019). Classification and ordination of the main plant communities of the Eastern Hajar Mountains, Oman. *Journal of Arid Environments*, *169*, 1-18. https://doi.org/10.1016/j.jaridenv.2019.05.017

Hill, M., & Gauch, O. (1980). Detrended Correspondence Analysis: an improved ordination technique. *Vegetati*o, *42*, 1-3,47-58. https://doi.org/10.1007/BF00048870

Hardtle, W., Redecker, B., Assmann, T.h., & Meyer, H. (2006). Vegetation responses to environmental conditions in floodplain grasslands: Prerequisites for preserving plant species diversity. *Basic and Applied Ecology*, *7*, 280-288. https://doi.org/10.1016/J.BAAE.2005.09.003

Hardtle, W., VON OHIMB, G., & Westphal, C. (2005). Relationships between the vegetation and soil conditions in beech and beech-oak forests of northern Germany. *Plant Ecology*, *177*(1),113–124. https://doi.org/10.1007/s11258-005-2187-x

Husain, Z., & Malk R. N. (2006).Classification and ordination of vegetation communities of the Lohibehr reserve forest and its surrounding areas, Rawlpindi, Pakistan. *Pakistan Journal of Botany*, *38*, 3: 543-558.

Jackson, D.A., & Somers, K.M. (1991). Putting things in order: the ups and downs of detrended correspondence analysis. *American Naturalist*, *137*, 74-712. https://doi.org/10.2307/2462603

Jackson, D.A. (1993). Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardization, measures of association, and ordination methods. *Hydrobiologia*, *268*, (1), 9-26. https://doi.org/10.1007/BF00005737

Jackson, D.A. (1995). PROTEST: a PROcrustean randomization TEST of community environment concordance. *Ecoscience*, *2*, 297–303. https://doi.org/10.1080/11956860.1995.11682297

Kent, M., (2012). In: Vegetation Description and Data Analysis: A Practical Approach, 2 ed.Wiley-Blackwell, Chichester, UK, and Hoboken, NJ.

Khan, M., & Hussain, F. (2013). Classification and ordination of vegetation in Tehsil Takht-e-Nasrati District Karak, Khyber Pakhtunkhawa, Pakistan. *Journal of Ecology and the Natural Environment*, *5*, 3, 30-39.

Lengyel, A., Roberts, D.W., & Botta-Dukát, Z. (2019). Comparison of silhouette-based reallocation methods for vegetation classification. *Journal of Applied Statistics, 32*(1), 1-10. https://doi.org/10.1101/630384

Lengyel, A., Landucci, F., Mucina, L., Tsakalos, J.L., & Botta-Dukát,Z.(2018) Joint optimization of cluster number and abundance transformation for obtaining effective vegetation classifications. *Journalo f Vegetation Science*, *29*(2), 336–347.https://doi.org/10.1111/jvs.12604

Lengyel, A., & Podani, J. (2015). Assessing the relative importance of methodological decisions in classifications of vegetation data, *Journal of Vegetation Science*, *26*(4), 804-15. https://doi.org/10.1111/jvs.12268

Legendre, P., & Borcard, D. (2018). Box–Cox-chord transformations for community composition data prior to beta diversity analysis. *Ecography*, *41*(11), 1820-1824. https://doi.org/ 10.1111/ecog.03498.

Legendre, P., & Legendre, L. (1998). Numerical Ecology (2nd English ed.) Amsterdam: Elsevier.

Legendre, P., & Gallagher, E. (2001). Ecologically meaningful transformation for ordination of species data. *Oecologia*, *129*(2), 271-280. https://doi.org/10.1007/s004420100716

Leps, J., & Smilauer, P. (2003). Multivariate Analysis of Ecological Data using Canoco. Cambridge: University Press, https://doi.org/10.1017/CBO9780511615146

Levis, E., Cakiroglu, A., Bucak, T., Odgaard, B., & Davidson, T. (2014). Similarity between contemporary vegetation and plant remains in the surface sediments in Mediterranean lakes. *Freshwater Biology*, *59*(4), 724-736. https://doi.org/10.1111/fwb.12299

McCune, B., & Grace, G. (2002). Analysis of Ecological Communities. Oregon: MjM Software Design.

McCune, B., & Mefford, M.J. (1999). PCORD. Multivariate Analysis of Ecological Data, Version 4. Oregon:MjM Software.

Minchin, P.R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, 69(1-3), 89-107. https://doi.org/10.1007/BF00038690

Maindonald, J., & Braun, W.J. (2007). Data analysis and graphics using R: An example-based approach (2nd Eds.). Cambridge University Press.

Molder, A., Bernhardt, M., & Schmidt, R. (2008). Herb-layer diversity in deciduous forests: Raised by tree richness or beaten by beech*?*. *Forest Ecology and Management*, *256*, 272–281. https://doi.org/10.1016/j.foreco.2008.04.012

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., & Minchin, P.R. (2013). vegan: Community Ecology Package. *R package version*, *2*(0-2), 1-295. (http://vegan.r-forge.r-project.org/)

Peres-Neto, P.R., & Jackson, D.A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*, *2*, 169–178. https://doi.org/10.1007/s004420100720

Pakgohar, N., Eshaghi Rad, J., Gholami, G.H., Alijanpour, A., & Roberts, D.W.A. (2021). Comparative Study of Hard Clustering Algorithms for Vegetation Data. *Journal of Vegetation Science*, e13042. https://doi.org/10.1111/jvs.13042

Ter Braak, C.J., & Prentice, I. C. (1988). A theory of gradient analysis. In Advances in Ecological Research Academic Press.

Roberts, D.W. (2015) Vegetation classification by two new iterative reallocation optimization algorithms. *Plant Ecology*, *216*(5), 741-58. https://doi.org/10.1007/s11258-014-0403-2

Ruokolainen, L., & Salo, K. (2006). Differences in performance of four ordination methods on a complex vegetation dataset. *Annales Botanici Fennici*, *43*(4), 269-275.

Roberts, S. (2008). Transform your data. *Nutrition*, *24*(5), 492-494. https://doi.org/10.1016/j.nut2008.001.004

Schuman, M., White, A., & Withman, W. (2003). The effect of harvest – Created gaps on plant species diversity composition and abundance in a Maine oak-pine forest. *Forest Ecology and Management*, *176*(1-3), 543-561. https://doi.org/10.1016/S0378-1127(02)00233-5

Super, L., Vellend, M., & Bradfield, G. (2013). Urban ecology in action: vegetation changes in Pacific Spirit Regional Park, Vancouer, BC Canada. *Davidsonia*, *23*(1), 21-31.

O'Hara, R., & Kotze, J. (2010). Do not log-transform count data. Nature Precedings, 1-1. https://doi.org/10.1038/npre.2010.4136.1

Tukey, J.W. (1977). Exploratory Data Analysis. London, Amsterdam, Don Mills, Ontario, Sydney:Addison-Wesley https://doi.org/10.1002/bimj.4710230408

Tabachnick, B.G., Linda, S., & Fidell. (2007). Using Multivariate Statistics (5th ed.). Boston: Pearson.

Tichý, L., Hennekens, S.M., Novák, P., Rodwell, J.S., Schaminée, J.H., & Chytrý, M. (2020). Optimal transformation of species cover for vegetation classification. *Applied Vegetation Science*, *23*(4), 710-717. https://doi.org/10.1111/avsc.12510

Uotila, A., & Kouki, J. (2005). Understorey vegetation in spruce-dominated forests in eastern Finland and Russian Karelia: Successional patterns after anthropogenic and natural disturbances. *Forest Ecology and Management*, *215*, 113–137. https://doi.org/10.1016/j.foreco.2005.05.008

Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc.

Xianping, Z., Mengben, W., Bo, S.h., & Yang, X. (2006). Quantitative classification and ordination of forest communities in Pangquangou National Nature Reserve. *Acta Ecologica Sinica, 26*(3), 754−761. https://doi.org/10.1016/S1872-2032(06)60013-9

Van der Maarel, E. (1979). Transformation of cover-abundance values in phytosociology and its effect on community similarity. *Vegetatio*, 39(2), 97–114. https://doi.org/10.1007/BF00052021

Zuur, A.F., Ieno, E., & Meesters, E. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*, 3-14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

مجلۀ جنگل ایران، انجمن جنگلبانی ایران
سال شانزدهم، شمارۀ ۵، زمستان ۱۴۰۳
ص ۵۵-۴۵
مقاله پژوهشی

شناسه دیجیتال (DOI): 10.22034/ijf.2025.337849.1867

اثر تبدیل داده‌ها بر تحلیل تطبیقی قوس‌گیری‌شده(DCA) در مطالعات پوشش گیاهی (مطالعه موردی: جنگل‌های بلوط کرمانشاه)

نغمه پاک‌گهر[۱]، جواد اسحاقی راد[۲]*، عباس بانج شفیعی[۲] و سید جلیل علوی[۳]

[۱] دانش‌آموخته دکتری، گروه جنگداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران
[۲] استاد، گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه، ایران
[۳] دانشیار، گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه تربیت مدرس، نور، ایران

(تاریخ دریافت: ۱۴۰۱/۰۲/۲۰؛ تاریخ پذیرش:۱۴۰۳/۰۷/۲۲)

چکیده

نمونه‌برداری از پوشش گیاهی یک مرحله اساسی در پژوهشهای بوم‌شناختی است که بیشتر با برآورد پوشش گونه‌ها انجام می‌شود. این فرآیند بیشتر شامل تبدیل داده‌ها برای درنظر گرفتن حضور گونه‌های عمومی و نادر و همچنین صفرها در مجموعه داده‌هـا اسـت کـه می‌توانـد تـأثیر زیادی بر نتایج تحلیل‌ها داشته باشد. تحلیل‌های بعدی این مجموعه داده‌ها با هدف یافتن الگوها و پویـایی جامعـه در اکوسیسـتم‌های مختلـف انجام می‌شود. با این حال، روش بهینه تبدیل داده‌ها برای تحلیل های چندمتغیره همچنان موضوعی مورد بحث در میان بـوم شناسـان اسـت، زیرا روش‌های مختلف تبدیل می‌توانند تفسیرهای متفاوتی از داده‌های بوم شناختی به همراه داشته باشند. در این مطالعه، به ارزیـابی تـأثیرات تبدیل‌های مختلف داده‌ها بر نتایج تحلیل تطبیقـی قوس‌گیری‌شـده (DCA)، بـه‌ویژه در جنگل‌هـای بلـوط (*Quercus branti* L.) واقـع در منطقه زاگرس ایران، پرداختیم. این منطقه به خاطر تنـوع زیسـتی منحصـر بـه فـرد و اهمیـت اکولـوژیکی آن، مکـان ایـده‌آلی بـرای چنـین بررسی‌هایی است. برای دستیابی به اهداف پژوهش، سه منطقه جنگلی مجزا با شیب و ارتفاع از سطح دریا مشابه انتخاب کردیم تـا متغیرهـای محیطی مشابه باشند. نمونه‌برداری از پوشش گیاهی در پنج فاصله مشخص ۰، ۲۵، ۵۰، ۱۰۰ و ۱۵۰ متر، در طول سـه خـط نمونـه کـه ۲۰۰ متر از یکدیگر فاصله داشتند، انجام شد. این رویکرد سیستماتیک به ما این امکان را داد تا نمـای کـاملی از توزیـع گونـه‌ها در سراسـر منـاطق جنگلی به دست آوریم. از درصد پوشش براون-بلانکه و مقیاس ون در مارل برای آماده‌سازی مجموعه داده‌ها استفاده کردیم و اطمینان حاصل کردیم که اندازه‌گیری‌ها قابل اعتماد هستند. هر مجموعه داده تحت تبدیل‌های مختلفی قرار گرفت، از جمله تبـدیل لگـاریتمی، ریشـه مربـع و تبدیل‌های نسبی عمومی. پس از آن، روش DCA برای هر مجموعه داده تبدیل شده، اجرا شد و نتایج به‌دست‌آمده با روش پروکراستز، روشـی که شباهت بین دو مجموعه داده را کمی‌سازی می‌کند، مورد مقایسه قرار گرفت. یافته‌ها نشـان داد کـه هـر دو مجموعـه داده تبدیل‌شـده بـه لگاریتم و ریشه مربع به‌طور قابل توجهی نتایج DCA را، با کـاهش مـؤثر تغییـرات موجـود در مجموعـه داده، بهبـود بخشـیدند. نتـایج روش پروکراستز نشان داد که هم‌خوانی بین مجموعه داده‌های تبدیل‌شده به لگـاریتم و ریشـه مربـع و داده‌هـای خـام بـه‌طور چشـمگیری بـالاتر از روشهای دیگر تبدیل‌ها ارزیابی‌شد. نتایج ما نشان داد که تبدیل‌های نسبی عمومی برای تحلیل DCA نامناسب بودند، زیرا نتوانستند روابط بوم شناختی موجود در داده ها را به‌خوبی نمایان کنند. بنابراین، توصیه می‌کنیم که قبل از انجام تحلیل‌های رسته بندی، از تبدیل داده ها، به‌ویژه تبدیل لگاریتمی یا ریشه مربع استفاده شود. این رویکرد نه تنها قابلیت اعتماد به نتایج را بهبود می‌بخشـد بلکـه تفسـیر اکولـوژیکی دقیـق‌تر را تسهیل می‌کند و در آخر به درک عمیق‌تری از پویایی جامعه در اکوسیستم‌های جنگلی کمک می‌کند.

واژه‌های کلیدی: تبدیل نسبی عمومی، تبدیل لگاریتمی، تحلیل پروکراستز، جنگل‌های زاگرس، ریشه مربع.

*نویسنده مسئول شماره تماس: ۰۴۴۳۲۷۷۰۴۸۹ ایمیل: j.eshagh@urmia.ac.ir