Assessment of the Applicability of Combining the General Standardization Procedure with Principal Component Analysis in Classifying the Structure of Species Composition in Forest Stand

Han, H., Song, J.-E., Seol, A. & Chung, J.-S.

Keywords: Arched distortion, contraction effect, detrended correspondence analysis, general standardization procedure, principal component analysis

Abstract: This study assesses the applicability of combining the general standardization procedure (GSP) with principal component analysis (PCA) in ecological data analysis. For this purpose, we conducted a comparison of the performance of combined GSP and PCA with detrended correspondence analysis (DCA) in classifying the structure of species composition in natural forest stands at the Kwangneung Experimental Forest. Prior to analyzing the data using PCA, it was standardized using GSP to avoid the tendency of involutions accompanying PCA. Results showed that both approaches were successful in classifying stand species composition and provided ordination surfaces showing three separate groups-coniferous, deciduous and mixed forests. In terms of total explained variance, GSP-PCA seemed to demonstrate better performance than DCA. However, in classifying species composition structures by visual inspection of the ordination graphs, the ordination plots obtained using DCA were more effective than those generated using GSP-PCA.

Received August 31, 2009; Accepted November 17, 2009

1. Introduction

Ordination is the ordering of objects along axes according to their resemblances. The major objective of ordination is to achieve an effective data reduction, expressing high-dimensional relationships in a low-dimensional space (McCune and Mefford, 1999).

In forest community ecology, ordination techniques have been widely used for arranging or ordering of species and/or communities (sample units) along environmental gradients. Typically, the data from vegetation surveys are high dimensional arrays, where the dimensionality equals the overall number of recorded species. However, the high number of data on species occurrence makes it difficult to derive ecological information directly from the raw data (Mahecha *et al.*, 2007). Thus, techniques of dimensionality reduction for extracting and visualizing the essential properties of vegetation data are required to interpret inherent ecological information within the data. An effective dimensionality reduction of data can be achieved by ordination techniques.

Traditionally, principal component analysis (PCA) has been used as a primary ordination approach in ecological data analysis. However, especially in high-dimensional data arrays, PCA produces serious distortions of the expected positions of communities due to a nonlinear structure of ordinary ecological data like coenoclines or coenoplanes (Branko and Ranka, 1994).

To address the limitation of PCA in ecological data analysis, Branko and Ranka (1994) suggested a generalized standardization procedure (GSP) which allows a wide choice of data transformations simply by varying values of a single standardization parameter. Experiments by these workers using simulated as well as field data sets revealed that arched distortions can be corrected when GSP prior to PCA. Compared to PCA, GSP-PCA had better performance in expressing the structure of analyzed coenoclines without the serious distortions of data.

Table 1. Area distribution of the each stand in Kwangneung Experimental Forest

Category	Total	Natural forest				Artificial forest
		Sub-total	Coniferous f.	Deciduous f.	Mixed f.	
Area (ha)	1,122.2	527.4	115.6	214.3	197.5	594.8
Share (%)	(100.0)	(47.0)	(10.3)	(19.1)	(17.6)	(53.0)

On the other hand, among the existing ordination techniques, detrended correspondence analysis (DCA) has the ability to overcome the major problems of other existing ordination techniques. It flattens out the misleading arch and corrects the contraction in scale of an ordinate data swarm (Pielou, 1984). In general, DCA turned out to have better performance than PCA when used in ecological analysis. However, only a few studies examined the performance of DCA against that of GSP-PCA. The latter could be an alternative of PCA in ecological data analysis.

The objective of this study is to assess the applicability of GSP-PCA to ecological data analysis by comparing its performance with that of DCA in classifying the structures of species composition in natural forest stands.

2. Materials and Methods

2.1. Study area

The study area, which is the Kwangneung Experimental Forest (KEF) of the Korea Forest Research Institute (KFRI), is located at the west-central portion of the Korean peninsula, and covers about 1,122 ha. The forest (37°45'N, 127°10'E), which is located in the suburban area of Seoul, is mainly composed of unique old-growth forests consisting of broad-leaved trees in the middle zone of a temperate forest (FRI Korea, 1994).

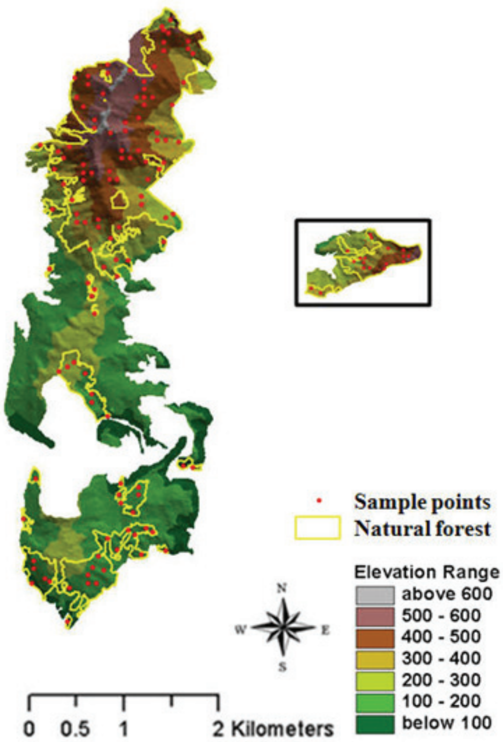

Figure 1. The distribution of sample plots in the study area

The KEF has one of the oldest natural forests in South Korea. The old-growth natural forest of the KEF has been protected from forest management activities, human disturbances and civilian access except for research (Lim *et al.*, 2003). The area of natural forest is about 527 ha consisting of coniferous, deciduous and mixed forests (Tab. 1).

2.2. Field surveys

The data were collected from the natural forest of the KEF in 2008. One hundred five (105) circular sample plots covering 0.05 ha were installed in coniferous, deciduous and mixed stands which had remained untouched for at least 10 years preceding the measurements (Figure 1). The diameters at breast height (DBH $\geq$ 6, cm) and the heights of all trees were measured in each sample plot for ecological data analysis.

2.3. Data analysis

To compare the performance of ecological ordination techniques in classifying forest stand species composition, PCA and DCA ordination techniques were used based on the abundances of the tree species in each sample plot. GSP was also applied prior to PCA to compare the suitability of combining the two techniques against using just one of these techniques. A GSP is described as:

[1] $$X_{(i,j)} = \frac{x_{(i,j)}}{\left(\sum_{i=1}^{n} x_{(i,j)}^{q}\right)^{1/q}}$$

where $X_{(i,j)}$ is the transformed value by sample plots, $X_{(i,j)}$ is the original value, and q is a constant that determines the effect of the transformation.

The value of the standardization coefficient q determines the efficiency of GSP in eliminating curvilinear distortions. The standardization by sample plots is especially sensitive to variation in q-values. Branko and Ranka (1994) reported that the optimal results for GSP in case of standardization by sample plots were obtained with $q = 0.5$. Thus, in the present study using GSP the value of 0.5 for q was used. The procedure was performed using PC-ORD package 4.25 for Windows (McCune and Mefford, 1999).

3. Results and Discussion

3.1. PCA ordination

Table 2 shows the cumulative values of explained total variance for the different ordination approaches. As expected, PCA with untrans-

Table 2. Cumulative values of explained total variance by the ordination approach

Axis	PCA	GSP-PCA	DCA
1	0.402	0.577	0.467
2	0.666	0.776	0.686

formed data gave the lowest values of explained variance due to the arch effect (Gauch, 1977). The arch effect arises because the second axis of ordination can have a strong relation to the first axis while at the same time be uncorrelated with it (Hill and Gauch, 1980). In the case of the arch effect, an axis is incurved resulting in greater difficulties of interpretation although the explained total variance of ordination axes is relatively high. Figure 2 shows the results of PCA wherein the arch effect was observed indicating an arch-shaped to the first axis. The dashed curve in Figure 2 gives a misleading idea of the axis.

3.2. GSP-PCA ordination

The quality of ordination obtained by combining the use of PCA and GSP (standardization by sample plots) is better than that obtained using PCA only. In particular, the use of GSP followed by PCA was able to address the limitation of PCA through the correction of the arch effect (Fig. 3). Branko and Ranka (1994) reported that the combination of GSP and PCA improved the PCA results. They also suggested that the combination of GSP and PCA would produce acceptable results with various data sets. Similarly, the arch effect in the previous PCA ordination graph was not observed using GSP. In terms of value of explained total variance, GSP-PCA exhibited higher values than PCA.

The other major fault of other ordinations is the distortion of relative distances between sample plots on its axes. In this study, GSP-PCA

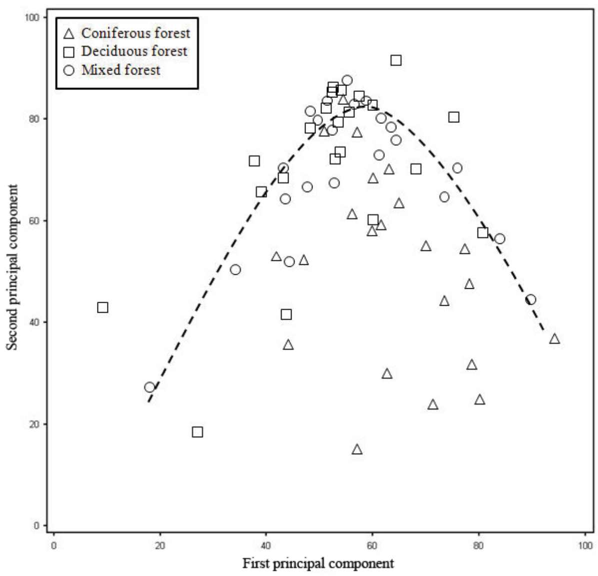

Figure 2. PCA ordination of sample plots based on forest types

was able to correct the arch effect but a compression of the distances between sample plots was observed (Fig. 3).

The distance between samples in the ordination graph is very important in the analysis of the results, thus, compression of the distance between sample plots makes it difficult to interpret the results of ordination.

3.3. DCA ordination

The usefulness of DCA was tested on the same data that were extracted from the KEF using other techniques. DCA eliminates the arch effect by detrending (Gauch, 1982). It means that centerization has a zero mean for the axis by segments.

In this study, DCA no longer showed the arch effect (Fig. 4). Detrending of the first axis using the DCA resulted in an easier interpreta-

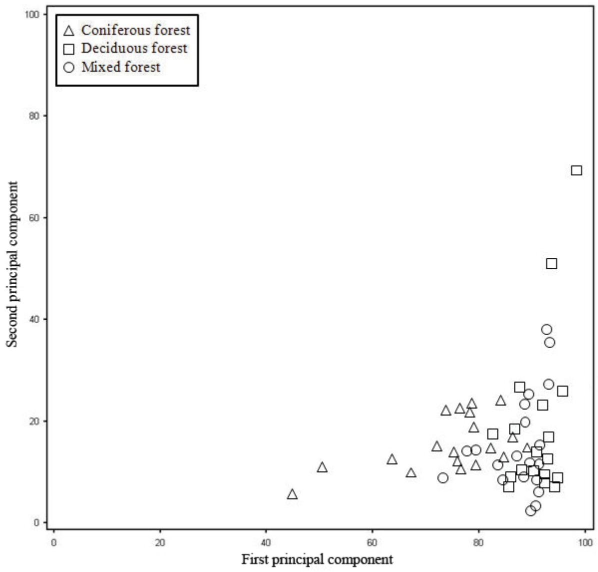

Figure 3. GSP-PCA ordination of sample plots based on forest types

tion of the ordination results compared to when other techniques were used. In DCA, the sample plots were ordered clearly based on the tree species composition of forest stands. The arch effect was eliminated by detrending the data swarm into several segments with dividing lines perpendicular to the first axis.

DCA can also eliminate the distortion of relative distance between sample plots by rescaling. Rescaling shifts sample scores along each axis such that the average width is equal to constant. Figure 4 shows that all the sample plots were well positioned with no compression of distance between the sample plots.

3.4. Comparison of the performance of the different ordination techniques

Gauch (1982) suggested that DCA gives useful results and permits

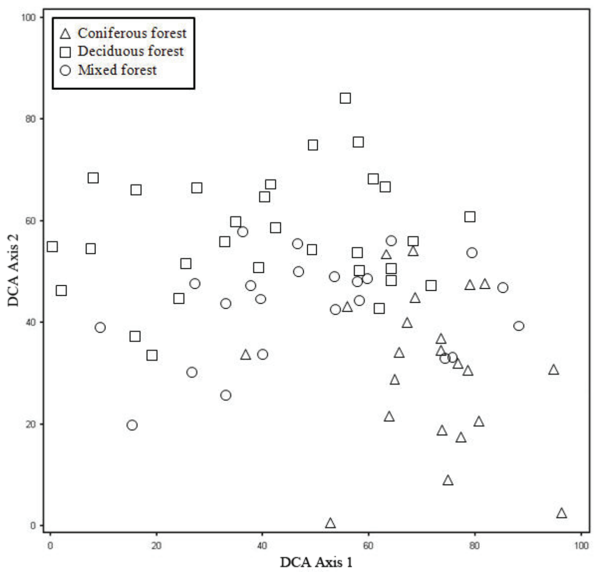

Figure 4. DCA ordination of sample plots based on forest types

ecologically correct interpretations to be derived from confusing multivariate data. Although PCA has the merit that it is the simplest and most straightforward approach, PCA involutes opposite ends of a coenospace, producing results that may be difficult to interpret (Kenkle and Orlóci, 1987). As seen in Figure 2, the arch effect was observed in PCA.

The combination of GSP and PCA eliminated the arch effect and resulted in the highest values of explained total variance. However, the limitation of the GSP-PCA approach is that there is the compression of distance between sample plots that makes the classification of forest stands based on tree species composition difficult.

On the other hand, the classification of sample plots based on tree species composition was easily achieved using DCA. This technique corrected the arch effect as well as the scale contraction effect through detrending and rescaling procedures. In the visual inspection of the ordination graphs to classify species composition structures, the ordination plot obtained using DCA was more effective than those obtained using GSP-PCA. DCA proved to be the best among the other techniques of ordination that were used in this study.

4. Conclusion

This study provides an estimate of the applicability of the GSP-PCA ordination in ecological data analysis. The use of GSP-PCA prevented the occurrence of the arch effect that results from the use of only PCA as ordination technique. However, the contraction effect was observed when GSP-PCA ordination was used. The distortion of relative distances between samples hindered the interpretation of ordination results even though the explained total variance of GSP-PCA was relatively successful.

DCA corrected the problems encountered using GSP-PCA and PCA only in classifying the sample plots based on tree species composition. Also, DCA had a better visual performance over GSP-PCA in ordering the sample plots. The results obtained indicate that GSP-PCA may have performed better than PCA but still its performance in classifying sample plots by forest types is not satisfactory.

However, it should be pointed out that the complexity of the ecological data with respect to species characteristics and/or life environment makes it difficult to interpret the data. Thus, in order to have a good assessment of the applicability of the ordination techniques that can be used to analyze ecological data more precisely, cautious attention and additional consideration on the various environmental gradients are required.

Acknowledgement

The authors would like to thank the Korea Forest Research Institute for their cooperation and research support.

References

Branko, K. and Ranka, P. (1994) A generalized standardization procedure in ecological ordination: Test with principal components analysis, *J. Veg. Sci.* 5: 259–262.

Forestry Research Institute of Korea (1994) *Kwangneung Experimental Forest*, Utkomunwhasa Press, 910p. (in Korean).

Gauch, H. G. (1977) *ORDIFLEX – A Flexible Computer Program for Four Ordination Techniques*: *Weighted Averages*, *Polar Ordination, Principal Components Analysis*, *and Reciprocal Averaging*, Cornell University, 185p.

Gauch, H. G. (1982) *Multivariate Analysis and Community Ecology*, Cambridge University Press, 312p.

Hill, M. O. and Gauch, H. G. (1980) Detrended correspondence analysis: An improved ordination technique, *Vegetatio* 42: 47–58.

Kenkle, N. C. and Orlóci, L. (1987) Applying metric and nonmetric multidimensional scaling to ecological studies: some new results, *Ecology* 67: 919–928.

Lim, J., Shin, J. H., Jin, G. Z., Chun, J. H. and Oh, J. S. (2003) Forest stand structure, site characteristics and carbon budget of the Kwangneung Natural Forest in Korea, *Korean J. Agri. Forest Meteorol.* 5: 101–109.

Mahecha, M. D., Martinez, A., Lischeid, G. and Beck, E. (2007) Nonlinear dimensionality reduction: Alternative ordination approaches for extracting and visualizing biodiversity patterns in tropical montane forest vegetation data, *Ecol. Info.* 2: 138–149.

McCune, B. and Mefford, M. J. (1999) *Multivariate Analysis of Eco-*

logical Data Version 4, MjM Software.

Pielou, E. C. (1984) *The Interpretation of Ecological Data*, John Wiley & Sons, 167p.