Noise elimination in gradient analysis

B. Tóthmérész

Ecological Institute, Lajos Kossuth University, Debrecen, H-4010 HUNGARY;
Fax. (36) 52 310 936; E-mail: TOTHMERB@HUKLTE51

Abstract: Robust nonlinear smoothing is used to improve the result of gradient analysis and to eliminate the distorting effect of rare species. The usefulness of the method is demonstrated by a case study. The result of ordinations of the plots of a forest-margin-glade transect of vegetation was improved by the method.

Introduction

Gradients in vegetation occur very frequently. In *direct gradient analysis* the vegetation samples are arranged using observed vegetational gradients. In *indirect gradient analysis* samples are arranged along axes generated from the vegetational data, frequently created by various types of ordination. There are many other situations when gradient-like data arise. The analysis of regeneration and degradation processes, time-series data, transect or transect-like data or simply a sequence of data also involves problems similar to gradient analysis. Sometimes it seems reasonable to consider succession as a 'community-gradient' (Whittaker 1970). When the x-values (space, time, succession series or any other real or abstract variable) are equally or nearly equally spaced then their structure is so simple and regular that often y receives most of the attention. Generally, we refer to such y-values as a *data sequence*. Time series serves as a typical example when the sequence comes about by recording a value for each successive time interval. When we are speaking about *gradient-like* data or *data sequence* we refer to such data where the x-values of the data are in order; usually, but not always, the x-values are equally spaced.

In ecology, sampling errors are very typical which are sometimes comparable in magnitude with the studied phenomenon or process. Studying the vegetation along an environmental gradient or studying the vegetation during a time period, there are many random factors that may bias the studied effect. Some of these arise by mistake during field work and could be eliminated by perfect or faultless sampling if it is possible at all. However, there are many other factors distorting the samples; e.g. unwarranted soil heterogeneity, microtopographic variability, yearly variability of the weather in a long-term study of succession or many other factors which cannot be controlled by the researcher.

It would be very useful to develop a method which eliminates sampling errors or is extremely tolerant to those effects which bias the original trends. The presented method fulfils some of these conditions. The method proved very effective in eliminating the bias due to the rare species in vegetational data as well.

Material

A semi-artificial data set was created from belt transect data obtained in an alkali-oak forest - glade mosaic complex. The forest part of the mosaic complex is a *Festuco pseudovinae-Quercetum roboris community* and the glades are covered by *Peucedano-Asteretum sedifolii* communities. The original data set included 24 plots; 4 plots were eliminated and the the remaining 20 plots were analyzed. The number of individuals of the species was recorded in each plot of 1 m by 1

Table 1. The number of individuals of *Galium aparine* in a clear-cut area for 10 m^2 plots.

Year	Actual counts	Smoothed data
1980	0	0
1981	0	0
1982	0	0
1983	0	0
1984	4	4
1985	5	4
1986	3	3
1987	0	3
1988	188	8
1989	8	9
1990	9	8
1991	5	9
1992	19	9
1993	9	9

m size; the survey was made by A. Molnár (Molnár 1988, 1989).

Method

The method is based on robust smoothing techniques proposed by Tukey (1977). I used the so-called tree-term running median smoothing procedure. This is similar to the traditional smoothers except that it uses medians rather than means. During the smoothing procedure the i-th value, y_i, is replaced by the median of y_{i-1}, y_i, and y_{i+1}. The problem of how to treat endpoints remains. The simplest solution is the so-called "copy-on" technique which means that there is no end-value smoothing just the first and last values are copied on. A more sophisticated procedure is to use the median of y_1, y_2, and $3y_2-2y_3$ as the first value. The last value can be calculated similarly. Usually the smoothing procedure is repeated until no further changes occur.

I illustrate the method by a field example. The number of *Galium aparine* individuals is displayed by column 2 of Table 1 during a 14 years-log period on a clear-cut area of a beech forest (Tóthmérész 1989). The beech-forest was clear-cut after the first year of the study. We can see that there are three different periods: at the beginning *Galium aparine* was not present (1980-

1983). During the second period the number of individuals was approximately 3-4 (1980-1987) and during the last period the number of individuals was higher; it may be 8-9 (1988-1993). The 188 individuals counted in the 9-th year may be real, indeed it was, but it has nothing to do with a smooth curve we would like to use to describe the change. The number of individuals were counted on 10 m^2 in a permanent plot. This year there may be an unusually dense patch and a more relevant result could have been produced by a sampling area of 100 m^2 or larger. The third column of Table 1 presents the result of smoothing with the copy-on technique; i.e. without end-point smoothing. We can see now that the three periods, as we have guessed earlier, can be separated very clearly with $x_1=0$, $x_2=3.5$, and $x_3=8.75$ average number of individuals. The separation of these stages is also supported by other characteristics of the vegetation and by multivariate statistical methods as well (Tóthmérész 1989).

Results

The species composition clearly indicates the glade-margin-forest zones (Figure 1). The result of PCoA, based on the number of individuals using Bray-Curtis dissimilarity, clearly demonstrates the difference of forest plots from the others but the margin and glade are not separated clearly (Figure 2). The first two dimensions explain 58.71% of the total sum of squares. Using the three-term running median smoothing with end-point smoothing the first two dimensions ex-
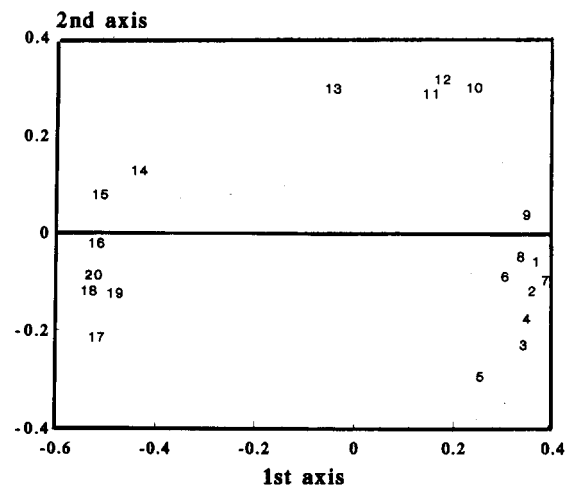

Figure 1. Ordination of plots from a glade-margin-forest transect based on the Jaccard index. Axes 1-2 explain 60.13% of the total sum of squares.

plain 69.10% of the total sum of squares. Thus, the improvement is 10.39%. The result of PCoA very clearly indicates the forest-margin-glade regimes (Figure 3). The number of species decreased from 56 to 37, i.e., 19 were eliminated.

There is no evident systematic structure in the ordination of the residuals (residual = original data - smoothed data, Figure 4). This suggests that the improvement of the ordination of the data set is not an artifact because the residuals show a "random" structure.
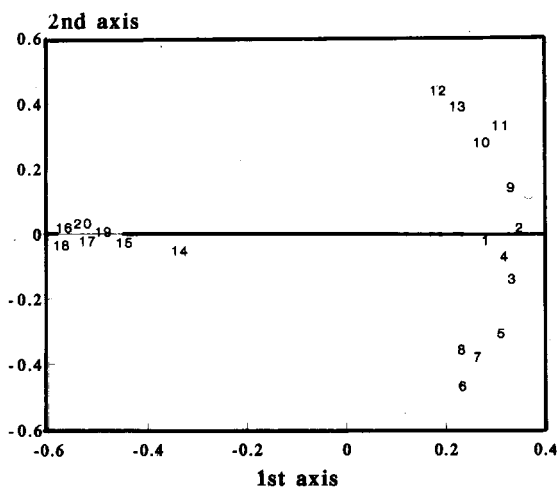

Figure 2. Ordination of plots based on the number of individuals using the Bray-Curtis measure. Axes 1-2 explain 58.71% of the total sum of squares.
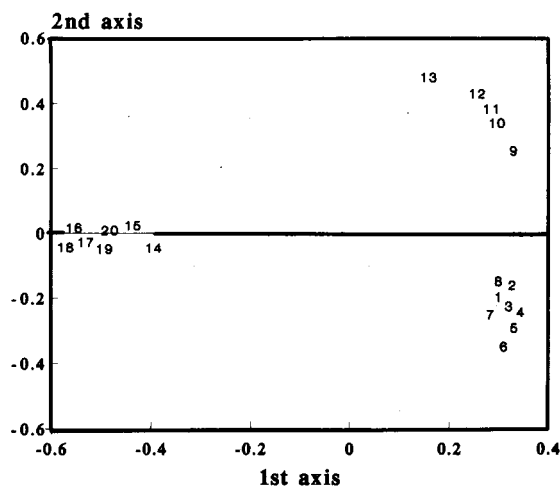

Figure 3. Ordination of plots after noise elimination using the Bray-Curtis measure. Axes 1-2 explain 69.1% of the total sum of squares.
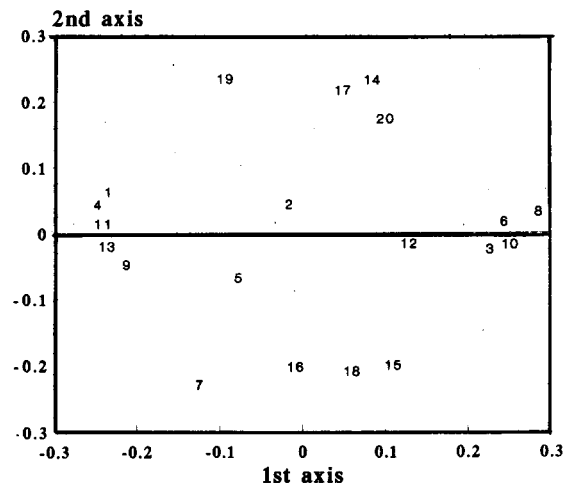

Figure 4. Ordination of the residuals. There is no systematic pattern in the result suggesting that just noise was eliminated from the data and no artefacts were produced.

Discussion

Noise elimination is a frequent problem in ecology. Gauch (1982) revealed that ordinations by themselves reduce noise becaue they selectively recover patterns affecting several species simultaneously in the first axes. The method presented in this paper is related to robust nonlinear smoothing; the "noise" is eliminated on the basis of temporal and/or spatial series structure of the data. Thus, it is important to stress that the presented meth-ods can be applied to "data sequences"; i.e. when the plots, communities or other objects to be studied can be arranged into a real or abstract series or sequence. I used the method in many different research problems successfully (in preparation). It was used to improve the result of a community classification in a field study of sand grassland communities when the high number of rare species heavily influenced the results and devaluated the real differences of the communities (Nagy et al. 1990). It was also used to identify stages of a successional series in a community developed on a clear-cut area of a beech forest (Tóthmérész 1991, 1992).

It is also important to stress that my experience with the method suggests that boundaries do not change using the procedures; and it is vital because a smoothing procedure which moves the border of the regimes would be unsatisfactory from ecological point of view.

The presented method looks very useful to improve and stress the structure of the data set. A careful study of the residuals can diminish the risk or producing artifacts. When there is no definite pattern of the residuals, we can use the improved results safely. For a definite pattern a detailed study is needed to examine the reasons producing the pattern. In this paper I just mentioned the simplest version of the procedure. There are more sophisticated smoothing methods not discussed here which deserve attention in the future.

References

Gauch, H. G. Jr. 1982. Noise reduction by eigenvector ordinations. Ecology 63:1643-1649.

Molnár, A. 1988. Sziki erdős-sztyep társulások klasszikus fitocönológiai és strukturális vizsgálata. Thesis, pp. 41.

Molnár, A. 1989. A bélmegyeri Fás-Puszta növényzete. Botanikai Közlemények 76: 65-82.

Nagy, M., Papp, M., Tóthmérész, B. and Précsényi, I. 1990. Classification and mapping of sand grassland communities on a protected area in Eastern Hungary. Acta Biologica Debrecina Supplement 1990: 25-34.

Tóthmérész, B. 1989. Szekunder szukcessziós folyamatok elemzése cluster-analizissel. Botanikai Közlemények 76: 35-49.

Tóthmérész B. 1991. Secondary succession after clear-cutting of a beech forest. Conference of the International Association for Vegetation Science, Eger, Hungary.

Tóthmérész B. 1991. A new, extremely powerful technique to analyze gradient-like data. Conference of the International Association for Vegetation Science, Eger, Hungary.

Tóthmérész, B. (ed.) 1992. Rejtek Project. Szekunder szukcesszió komplex és folyamatos vizsgálata mészköves, sekélytalajú bükkös tarvágása után. Research Report, pp. 53.

Tukey, J.W. 1977. Exploratory Data Analysis. Addison-Wesley.

Whittaker, R.H. 1970. Communities and Ecosystems. MacMillan, London.