FACTOR ANALYSIS AS A METHOD IN SYNECOLOGICAL RESEARCH

L. F. M. FRESCO

Laboratorium voor Plantenoecologie, Groningen.

SUMMARY

Factor analysis of floristic data is employed to solve relations between vegetations and environmental characteristics. This paper pictures the inverse approach. By means of peat-bog vegetations the validity of the technique has been demonstrated.

1. INTRODUCTION

As proved in psychological research factor analysis offers means for explaining a number of observed variables by a smaller number of underlying factors. GOODALL (1953), DAGNELIE (1960), MOORE (1968) and others have used this method to analyse vegetation data. In general laborious statistical techniques are not readily accepted. Notably Moore considers such methods too expensive in particular when resulting information additional to that gleaned from "classical" vegetation tables is slight. Investigations carried out at the Plant Ecology Laboratory, Groningen have once more led to the opinion that from an ecological point of view the information yielded by this method is exceedingly great.

Besides the "indirect" or R – approach whereby after an ordination of species and environmental characteristics the scores of the stands on the factors are calculated the inverse "direct" or Q – approach has yielded most valuable information concerning relations within plant communities as those referring to interplays with environment. In brief the paper discusses statistical aspects of this method, illustrated by sample data collected during a students' course.

2. FACTOR ANALYSIS

Test or attribute data can be assembled in a matrix; assuming m tests or attributes and n individuals as m points in a V_n (vectorspace) or as n points in a V_m. Factor analysis offers a possibility of reducing dimensions to any number desired. The method employed rests on the presumption that a linear correlation coefficient between tests can be regarded a linear combination of contributing independant factors affecting the test scores. By means of iteration, continued estimation converging to the desired value, the loading of each test on a factor can be found resulting in calculated "eigenvectors" of the correlation matrix. Each eigenvector has an "eigenvalue" composed of the sum of squares of the loadings and in the case of m tests representing m times that part of the original variance, explained by the factor. Rotation of factors around the origin in some cases simplifies the interpretation (IVIMEY COOK & PROCTOR 1967).

Synecological research practice generally dictates quadrats to be subjected to analyses leading to data referring to the presence of species as well as environmental characteristics. With the inverse approach or Q-type analysis attributes of species are measured and the data collected are treated theoretically. This procedure may appear to be statistically less well founded but under certain conditions as discussed in paragraphs to follow has proved to yield interesting results. ORLOCI (1966) has elucidated the Q-type analysis. He concludes that the use of the correlation coefficient is inadequate and that a "weighted similarity coefficient" might be better employed. Nevertheless use has been made of the correlation coefficient instead of Orloci's coefficient in this investigation. This decision is based on the following considerations. "Individuals" to be tested are defined as "all plants belonging to species X within area Y" and the universe to be investigated as "all individuals within area Y". Presuming all individuals of the universe are tested, the Q-type analysis is transposed to an R-type analysis and thus the correlation coefficient can be used.

A computer program in Algol 60 (Alcor) as described by Burema and modified by van der Weele was used.

3. THE VEGETATION

During June 1968 vegetational data and soil samples were collected by third year students in the "Harense Wildernis" a peat-bog area of some 3 hectares, about 10 kilometers south of Groningen. The reserve is owned and managed by the Netherlands Society for the Promotion of Nature Reserve. The vegetation consists for the greater part of marshy forest, for a smaller part of *Festuca rubra* grasslands where in former days blocks of peat were stored, furthermore of reed and sedge swamps (old peat-holes).

Individual species lists of twenty quadrats were made. Forested areas were not included. By subdivision of each quadrat into 64 subquadrats of each species frequencies were scored. Quadrats measured 2 by 2 meters. The total number of species was 101, including mosses and liverworts. The maximum number of species found in a quadrat was 35, viz. in quadrat nr. 5.

Brief description of the 20 plots:

quadrats 1, 2, 3 and 4: Moderately moist grasslands. *Festuca rubra, Holcus lanatus, Calamagrostis canescens, Galium palustre;*

quadrats 5, 6, 7 and 8: Marshy grounds, outside the boundaries of the reserve proper. Borders on a meadow. Burned infrequently.
Sedge swamp with much *Carex vesicaria* and *Carex acuta.*
March plants such as *Lathyrus palustris* and *Dryopteris cristata;*

quadrats 9 and 10: Sedge swamps, rich in species. Additional to the species mentioned above: *Carex paniculata* and *Carex lasiocarpa;*

quadrats 11 and 12: A low situated grassy path. Some grass species and much *Juncus effusus* and *Potentilla erecta;*

quadrats 13, 14, 15 and 16: Various aspects of the reed and sedge swamps within the reserve proper. *Carex riparia, Mentha aquatica, Viola palustris, Peucedanum palustre;*

quadrats 17 and 18: A low lying hayfield with marsh plants such as *Dryopteris thelypteris* and *Galium palustre* and species such as *Poa trivialis* and *Ranunculus repens;*

quadrats 19 and 20: A very low lying marsh, situated north of the reserve. *Carex rostrata, Lemna minor, Glyceria fluitans.*

4. ANALYTICAL RESULTS

Fig. 1 presents the position of the 20 quadrats on the prime two factors. The eigenvalues of the five most important factors are: 4.98, 3.49, 2.02, 1.75 and 0.97.

Hence the prime two factors explain $(4.98 + 3.49)/20 \times 100\% = 42.35\%$ of the variance and the prime three 52.45%.

Samples were taken of the top soil to a depth of 20 centimeters.

The following laboratory analyses were carried out:

1) moisture content expressed in vol. % H_2O ($_{H_2O}$)
2) N_{total} by Kjeldahl (N)
3) % organic matter (org)
4) pH_{H_2O} (pH)
5) C/N ratio (C/N)

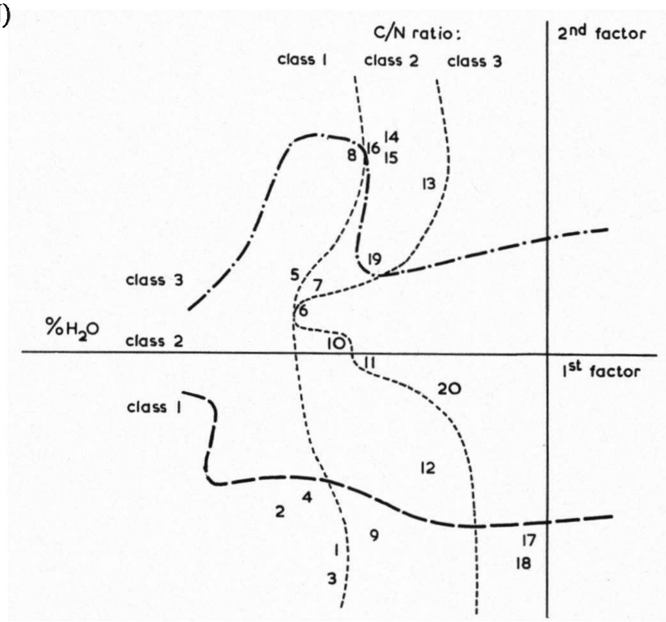

Fig. 1. The position of the stands on the plane defined by factors 1 and 2. Partition of the cluster on the ground of soil C/N ratio and soil moisture condition. H20: class 1 represents lowest analytical results; C/N: class 1 represents highest analytical results.

Spearman rank correlation coefficients between these characteristics and the loadings of the stands on the factors were calculated:

	H20	N	org	pH	C/N
factor I	–0.03	0.18	0.23	0.12	0.49*
factor II	0.86**	–0.91**	0.77**	0.19	–0.10
factor III	–0.21	0.17	–0.03	0.04	0.08

+ = reliability ≧ 95% ++ = reliability ≧ 99%

The conclusion may be drawn that the main environmental characteristic affecting the floristic composition of the quadrats is related to C/N ratio. Since a number of quadrats has been chosen on old paths and nearby meadows the thesis should be posed that the first factor represents human influence.

Striking is the fact that no correlation could be observed between either the total nitrogen or the organic matter content and the loadings on factor 1. This might point to the importance of C/N ratio as an ecological indicator. Since a statistically warranted location of the quadrats was not performed conclusions regarding nature and importance of the relation between C/N ratio and the vegetation of the area as a whole must remain provisional.

The second factor seems to represent soil moisture conditions. Soil analytical results do not permit deductions regarding the meaning of the third factor. At this stage due consideration should be paid to *fig. 2*, representing the plane, defined by second and third factors, composed of parts indicating presence and absence of a number of species in the quadrats. It may be observed that parti-

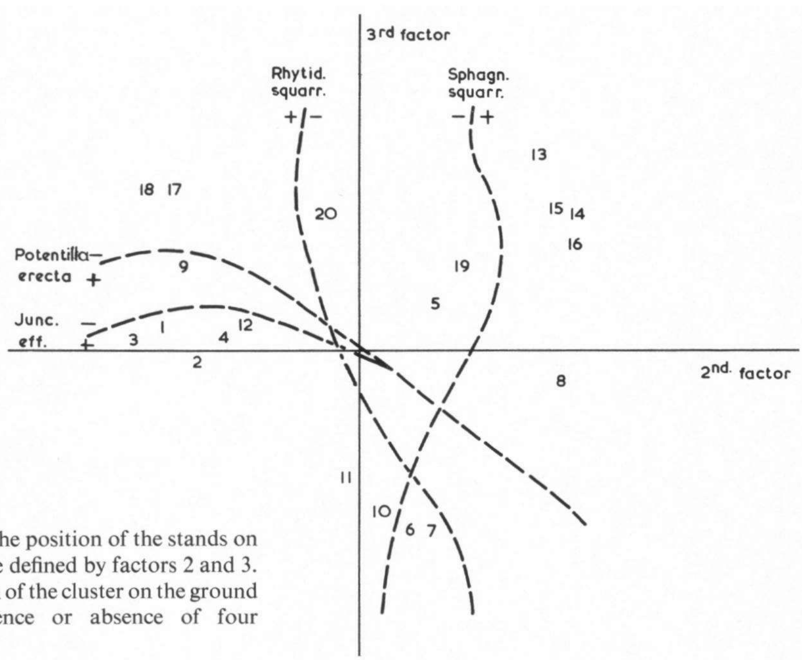

Fig. 2. The position of the stands on the plane defined by factors 2 and 3. Partition of the cluster on the ground of presence or absence of four species.

tion significant for presence of the mosses *Sphagnum squarrosum* and *Rhytidiadelphis squarrosus*, known to be indicative of respectively relatively moist and arid soil conditions projected unto the second factor confirms the expectation the second factor should be identified with soil moist conditions. Species aiding differentiation into positive and negative parts of the third factor are *Juncus effusus* and *Potentilla erecta*. Both species denote an invironment that has reached an advanced stage of eutrofiation, alternatively a transitional zone between oligotrophic and eutrophic environments. This third factor might be called "degree of eutrofiation".

A remarkable feature is the irrelevance of soil pH, varying between 3.6 and 5.8, in explaining vegetational differences of the 20 quadrats.

5. DISCUSSION

The importance of factor analysis for synecological research seems to be underlined by the results of this study. The fruits of the method are not limited to quantified environmental components important to explain vegetational variation; they also point to environmental elements overlooked and to be included in future research. Furthermore the technique offers the possibility to rank in order of importance environmental characteristics as means of explaining differences in vegetational composition.

Factor analysis offering an approach that is multidimensional as well as comparing between dimensions has qualities superior to the onedimensional vegetation table method. The critisism that extensive computer time makes factor analysis too expensive does not hold against the 12 minutes that were needed by the Telefunken TR4 computer to complete calculations made use of in this study.

More thorough investigations, applying factor analysis, will be reported upon in the near future.

AKNOWLEDGEMENT

The author wishes to thank prof. dr. D. Bakker, director of the Plant Ecology Laboratory, for his interest, dr. E. van der Maarel of the Nijmegen Geobotanical Institute for the fruitfull discussion and ir. F. Hagenzieker for correcting the text.

REFERENCES

BRAY, J. R. & J. T. CURTIS (1957): An ordination of the upland Forest communities of southern Wisconsin. *Ecol. Monogr.* **27**: 325–349.

DAGNELIE, P. (1960): Contribution à l'étude des communautés végétales par l'analyse factorielle. *Bull. Serv. d. l. Carte Phytogeogr.* V: 7–70 & 93–191.

GEER, J. P. VAN DE (1967): *Inleiding in de multivariate analyse*. Arnhem.

GOODALL, D. W. (1953): Factor analysis in plant sociology. 3d *Intern. Biometric Conf. Bellagio.*

— (1954): Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Austr. J. Bot.* **2**: 304–324.

IVIMEY COOK, R. B. & M. C. F. PROCTOR (1967): Factor analysis of data from an east Devon heath: a comparison of principal component and rotated solutions. *J. Ecol.* **55**: 405–414.

MAAREL, E. VAN DER (1966): *Over vegetatiestructuren, -relaties en -systemen, in het bijzonder in de duingraslanden van Voorne.* Diss., Utrecht.

MOORE, J. J. & A. O'SULLIVAN (1968): Zur mathematischen Bestätigung der tabellarischen Abgrenzung der Pflanzengesellschaften. *Ref. Symp. Intern. Ver. f. Veg. forsch.*, Rinteln (in press).

ORLOCI, L. (1966): Geometric models in ecology. I. The theory and application of some ordination methods. *J. Ecol.* **54**: 193–215.

SCHAAFSMA, W. (1961): Faktoranalyse. *Rapport TW8, Math. Inst. Rijksuniversiteit Groningen.*