Acta Bot. Neerl. 22(6), December 1973, p. 599–607.

APPLICABILITY OF FACTOR ANALYSIS IN AUTECOLOGICAL RESEARCH, WITH PARTICULAR REFERENCE TO CHAMAENERION ANGUSTIFOLIUM (L.) SCOP.

J. VAN ANDEL and H. J. M. NELISSEN

Biologisch Laboratorium, Vakgroep Oecologie, Vrije Universiteit, Amsterdam

SUMMARY

Factor analysis is introduced into autecological research as a helpful method in order to interpret the results of investigations in which a lot of variables per individual, population, or species must be determined one or several times during the period of growth. Growth patterns of different individuals, populations, or species can simply be compared with each other. This is illustrated in two examples with reference to *Chamaenerion angustifolium*, and related to the distinction of ecotypes.

1. INTRODUCTION

When investigating the plant's behaviour, ecologists always have to choose one or more parameters in order to determine the plant's response to influences from the environment. For certain purposes it may be justified to choose only a few parameters, for instance number and weight of grains (corn), weight and area of leaves (lettuce). With investigations concerning the plant's growth behaviour sensu lato, a lot of parameters must be determined. HARPER (1967) stated that the strategy of the life cycle is an ecologically fascinating but neglected subject of study. He suggested that the ways in which different species of plants allocate their limited sources represent ways of describing the behaviour of a plant, which will be of great ecological interest when sufficient examples have been studied for generalizations to be made. The (proportional) distribution of the dry weight of plant parts through the life cycle ("life cycle strategy") can be shown in a diagram.

The problem arises, however, in what way such sorts of diagrams can be compared mathematically without loss of information. Some kinds of multivariate analysis are available, which can be applied to solve this problem.

The present paper deals with factor analysis, a statistical method which reduces a lot of variables to a few factors (which find expression in the variables). Two ways of analysis are possible: a) comparison of variables of one individual at several stages, b) comparison of variables of several individuals at one point of time. The calculating methods for both a) and b) are identical. Three-dimensional methods of analysis to combine these two have as yet been developed insufficiently.

The present paper introduces factor analysis into autecological research. The second method of analysis (b) is applied. After the account for the method used, two examples will be given with relation to *Chamaenerion angustifolium*.

2. METHOD

With each of the plant individuals or population samples in experimental use, six growth characteristics (parameters) were determined, henceforth indicated as "variables": 1) Shoot length, 2) Shoot fresh weight, 3) Root fresh weight, 4) Total dry weight (shoot plus root), 5) Leaf area, 6) Number of leaves.

The leaf area was determined with the aid of an apparatus designed by the Technical and Physical Engineering Research Service (T.F.D.L.) at Wageningen, the Netherlands (SCHURER 1971).

Correlation coefficients between each pair of variables were calculated from the values of each individual or sample. They were arranged as a correlation matrix. After this factor analysis was performed according to ADAM et al. (1971, Chapter 4), using ÜBERLA (1968) as background information.

Some additional comment must be given:

1. Orthogonal rotation of the co-ordinate system occurred in a visual way.
2. To decide whether a variable correlated with a certain common factor (in order to determine the factor-structure), the factor-loading of that variable (= the correlation coefficient between variable and factor) had to have a value of 0.350 or higher.
3. ÜBERLA (1968) gave the following formula for each variable i:

$$u_i^2 = 1-h_i^2 = b_i^2+e_i^2$$

The overall variance is 1, and h_i^2 is the communality. The unique variance (uniqueness u_i^2) is subdivided into a specific variance (specificity b_i^2) and a residual variance e_i^2.

In ADAM et al. (1971) the residual variance is neglected. Also in our calculations the following formula was applied:

$$b_i^2 = 1-h_i^2$$

4. The factor-scores were calculated with the aid of Gauss multiplicators α_{ik} (i equation systems, k factors), as can – with some effort – be deduced from ADAM et al. (1971).

When one factor was extracted from the variables, α_{11} was calculated as follows:

	1	2
1	A_{11}	1
2	1	0
1	$t_{11} = A_{11}$	1
2	$u_{21} = \frac{1}{t_{11}}$	$t_{22} = -u_{21}$

$$\alpha_{11} = -u_{21} = \frac{1}{A_{11}}$$

When two factors were extracted from the variables, α_{11}, $\alpha_{12} = \alpha_{21}$ and α_{22} were calculated as follows:

	1	2	3	4
1	A_{11}	A_{12}	1	0
2	$A_{21} = A_{12}$	A_{22}	0	1
3	1	0	0	0
4	0	1	0	0
1	$t_{11} = A_{11}$	$t_{12} = A_{12}$	1	0
2	$u_{21} = \frac{t_{12}}{t_{11}}$	$t_{22} = A_{22} - u_{21}t_{12}$	$t_{23} = -u_{21}$	1
3	$u_{31} = \frac{1}{t_{11}}$	$u_{32} = \frac{t_{23}}{t_{22}}$	$t_{33} = -u_{31} - u_{32}t_{23}$	$t_{34} = -u_{32}$
4	0	$u_{42} = \frac{1}{t_{22}}$	$t_{43} = t_{34}$	$t_{44} = -u_{42}$

$\alpha_{11} = -t_{33}$ $\qquad \alpha_{12} = \alpha_{21} = -t_{34}$ $\qquad \alpha_{22} = -t_{44}$

The factor analysis according to ADAM et al. (1971) with the additions mentioned above could easily be performed with the aid of an IME calculating machine.

5. RESULTS

Example 1. Four populations compared one with another

From 4 habitats each, located in the Netherlands, 30 samples of a population of *Chamaenerion angustifolium* were taken in June, July, and August, 1971. A sample consisted of a shoot with the roots, taken from a soil sample of 20 × 20 centimeters. This method of population sampling was applied because the species develops vegetatively by horizontal roots from which new shoots can sprout.

Habitat 1 is located in the C.R.M. Reserve "De Bruuk" near Groesbeek. The soil is loess-loamy. *C. angustifolium* grows under a plantation of *Populus* and *Alnus*.

Habitat 2 is located in the "Veluwehul" near Putten. The soil is sandy. *C. angustifolium* occurs among several types of shrub, heather, and grass.

Habitat 3 is situated on one of the sand dunes on Schiermonnikoog. *C. angustifolium* is a dominating species.

Habitat 4 is located at the "Bloemberg" in the North Holland Dune Reserve near Heemskerk, on a sand dune with *C. angustifolium* growing among some types of shrub and grass.

Table 1. Data related to Example 1: Samples of populations of *Chamaenerion angustifolium* in 4 habitats (1971).

Habitat		1	2	3	4
Sampling area (m^2)		17 × 40	20 × 26	10 × 20	15 × 15
Minimum population age (Moss 1936)		6	12	16	21
	June	98	36	30	58
Mean shoot length (cm)	July	155	41	93	101
	Aug.	157	32	70	91
	June	35	5	8	14
Mean shoot fresh weight (g)	July	45	4	19	23
	Aug.	41	3	19	17
	June	13	11	15	22
Mean root fresh weight (g)	July	15	6	15	17
	Aug.	15	9	17	24
	June	8	4	6	11
Mean total dry weight (g)	July	13	4	11	13
	Aug.	14	4	12	18
	June	589	129	129	266
Mean leaf area (cm^2)	July	863	113	348	366
	Aug.	500	53	322	209
	June	62	25	24	36
Mean number of leaves	July	59	25	46	44
	Aug.	60	19	42	26

Some data about the populations involved are given in *table 1*. Factor analysis led to the extraction of two factors. The co-ordinate system with the factor loadings on the axes was rotated 57 degrees in clockwise direction, for June and July as well as for August. The new factor-loadings and their derivatives are shown in *table 2*. From this table it can be deduced that in June and July, as well as August, factor 1 influenced the root weight in particular, and as a consequence the total dry weight, whereas factor 2 influenced the shoot parameters (length, weight, leaf area, and number of leaves).

What the two factors are in concreto can only be hypothesized. They are not wholly independent of each other, for orthogonal rotation did not seem most suitable.

Concerning the strategy of plant reactions two slightly interacting reaction systems must be distinguished in this example. It must be kept in mind that this result may have been caused by the sampling method. At all events, it is justified to compare the reaction patterns of the four populations. In order to do this, the factor-scores were calculated for each sample (the average population scores as well as the areas including about 90 per cent of the sample values

Table 2. Factor-loadings and factor-structure (Example 1). Between brackets: percentage of overall variance.

	Variable	Factor-loading after rotation		Communality	Specificity	Factor structure	
		1	2	h_i^2	b_i^2	1	2
June	1	0.255	0.822	0.741	0.259		×
	2	0.233	0.862	0.797	0.203		×
	3	0.934	0.108	0.884	0.116	×	
	4	0.933	0.358	0.997	0.003	×	×
	5	0.160	0.962	0.951	0.049		×
	6	0.152	0.932	0.892	0.108		×
				(87.7%)	(12.3%)		
July	1	0.156	0.642	0.436	0.564		×
	2	0.335	0.908	0.936	0.064		×
	3	0.852	0.296	0.814	0.186	×	
	4	0.772	0.606	0.963	0.037	×	×
	5	0.244	0.949	0.961	0.039		×
	6	0.320	0.729	0.629	0.371		×
				(79.0%)	(21.0%)		
August	1	0.078	0.259	0.073	0.927		
	2	0.290	0.931	0.951	0.049		×
	3	0.898	0.125	0.822	0.178	×	
	4	0.860	0.496	0.984	0.016	×	×
	5	0.201	0.966	0.973	0.027		×
	6	0.042	0.888	0.791	0.209		×
				(76.6%)	(23.4%)		

per population are shown in *fig. 1*). So the six variables involved have been reduced to two factors, without loss of information.

As can be deduced from *fig. 1*, shoot development (factor 1) dominated in June, partly at the expense of the reserve food in the roots (negative tendency of factor 2). In July both shoot and root growth occurred. In August root growth and/or reserve food supply (factor 2) dominated, partly at the expense of shoot growth (negative tendency of factor 1). Some differences concerning the time of development appear between the four populations, but this cannot sufficiently explain the general differences in factor-scores between the population samples.

It is of great ecological interest to examine whether the population differences as shown in *fig. 1* were due to habitat differences or to ecotypic population differences as well.

Concerning the habitats, we repeat that the habitats 3 and 4 are situated in the dunes (NE exposition without shading). Habitat 2 has an arboreal environment on sandy soil, habitat 1 is dominated by trees on a loess-loamy soil.

Concerning the populations it should be realized that the populations in

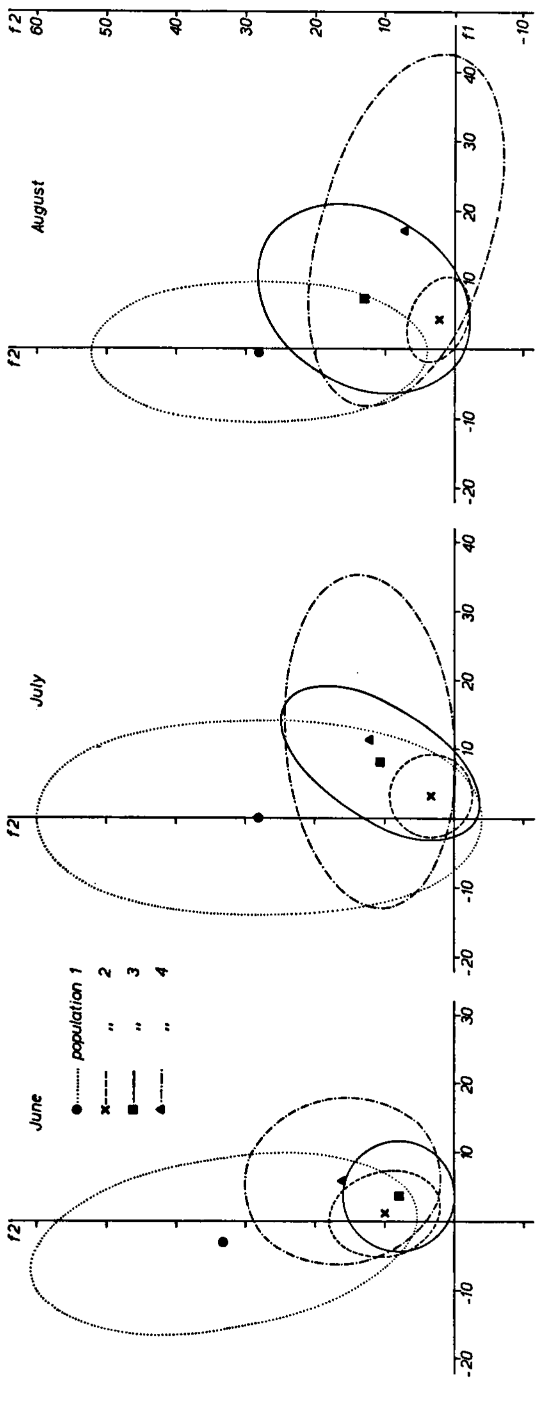

Fig. 1. Example 1. Mean factor-scores (f_1 and f_2) of each of the four populations, in June, July, and August 1971. The areas encircled around the mean values include about 90 per cent of the individual sample scores.

the habitats 2, 3, and 4 are much older than the population in habitat 1 (see *table 1*). The growth of the population in habitat 2 was strongly kept down by re-establishing wood vegetation; the plants did not flower.

The problem whether or not the population differences were due to genetic differences will be further discussed in Example 2.

Example 2. Plants grown from seed of three populations

Seed of *Chamaenerion angustifolium* was gathered from the populations in habitat 1, 3, and 4 (see Example 1) in August, 1971. The population in habitat 2 did not flower.

In late autumn, seedlings grown in a greenhouse to the rosette stage were planted on plots in an experimental field, nine seedlings per square meter. During the next spring the plots were fertilized with ranges of nitrogen, phosphorus, and potassium. The effects of the fertilizers will be reported in another paper. In this scope we only want to compare the three groups of plants grown with an equal NPK-regime.

In August, 1972, from each plot with nine plants, five were taken at random. Note that these plants are really individuals, whereas in Example 1 they had been sampled in another way. Of each group of 80 plants the six variables concerned were determined. The groups are indicated as 1, 3, and 4, according to the original populations.

Some data about the three plant groups are given in *table 3*. Factor analysis led to the extraction of only one common factor. The factor-loadings and their derivatives are shown in *table 4*. The factor-loadings in this table agree to a great extent with the most relevant loadings in Example 1 (*table 2*). Furthermore, the shoot length (variable 1) shows a relatively high "specificity" (b_i^2) in both examples. So this variable is less influenced by the common factor, as compared to the other variables. The only obvious difference between both examples is the number of common factors. This may be due to sampling differences or to the differences in age between the plants in this example and the population samples in Example 1.

It is interesting, now, to compare the factor-scores of the plants in the three groups. As can be deduced from *fig. 2*, the differences between the groups are

Table 3. Data related to Example 2: *Chamaenerion angustifolium* plants grown on an experimental field from seeds of the populations in 3 of the 4 habitats mentioned in Example 1 (August 1972).

Habitat from which the seeds had been gathered	1	3	4
Mean shoot length (cm)	76	78	63
Mean shoot fresh weight (g)	154	185	168
Mean root fresh weight (g)	37	31	48
Mean total dry weight (g)	59	52	49
Mean leaf area (cm^2)	2899	3223	3423
Mean number of leaves	874	1423	775

Table 4. Factor-loadings and factor-structure (Example 2).

Variable	Factor-loading	Communality h_i^2	Specificity b_i^2	Factor structure
1	0.593	0.352	0.648	×
2	0.961	0.924	0.076	×
3	0.871	0.759	0.241	×
4	0.982	0.964	0.036	×
5	0.941	0.885	0.115	×
6	0.865	0.748	0.252	×
Percentage of overall variance:		77.2	22.8	

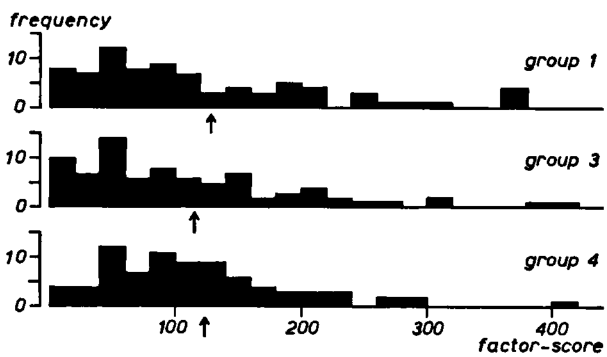

Fig. 2. Example 2. Factor-score frequency of three groups of plants grown from seed of population 1, 3, and 4, respectively (see Example 1), harvested in August 1972. Arrows indicate the mean scores.

negligible. The chances are that the population differences shown in Example 1 are not due to ecotypic variation on a genetical basis.

4. DISCUSSION

Factor analysis is a helpful method in order to interpret the results of investigations in which a lot of variables per individual (population, species, etc.) must be determined. Individual growth patterns, determined by measurement of various variables (at several stages), can be described and compared by reducing all measurements to only a few factor-scores.

In general factor analysis produces the following results:

1. The number of factors that influence the variables (by extraction of factors);
2. The extent to which a factor finds expression in each of the variables (the factor-loading);
3. The localization of the individuals in the co-ordinate system for the factor-scores.

It can only be hypothesized whether a certain factor indicates environmental circumstances or genetic circumstances, or both. The problem which factor in concreto manifests itself in the variables concerned, must be investigated experimentally.

The factor-scores are estimated values, so they cannot be used for further statistical analysis.

In the examples reported in this paper, the conclusions concerning the growth patterns are, of course, only related to the 6 variables measured. The greater the number of ecological parameters, the more valuable the interpretation of growth patterns can be.

In our examples we used 6 variables in order to introduce the factor analysis into autecological research. The choice of the variables was based on the various ways in which they indicate growth responses. WHITEHEAD & SINHA (1966) applied first-component analysis – a factor analytical method in which the "specificity" (b_i^2) of the variables is neglected – to taximetric results with the *Stellaria media* complex. They chose 12 variables, all related to the flowers. In their study, as in ours, the choice of the variables is a rather subjective one. It may be commendable to start with as many types of variables as possible and select by factor analysis those which prove to be important in relation to the problem concerned.

In this paper the applicability of factor analysis in autecological research has been illustrated in two examples related to *Chamaenerion angustifolium* and showing the applicability of the method in order to distinguish ecotypes, if present.

ACKNOWLEDGEMENTS

The authors wish to thank the Managerships of the Reserves mentioned for consenting us to work in the areas concerned, Prof. Dr. W. Ernst (Münster, Westfalen) for commenting upon the manuscript, and Mr. W. Vlijm for correcting the English text.

REFERENCES

ADAM, J., J.-H. SCHARF & H. ENKE (1971): *Methoden der statistischen Analyse in Medizin und Biologie*. VEB Verlag Volk und Gesundheit, Berlin.

HARPER, J. L. (1967): A Darwinian approach to plant ecology. *J. Ecol.* **55**: 247–270.

MOSS, E. (1936): The ecology of Epilobium angustifolium with particular reference to rings of periderm in the wood. *Amer. J. Bot.* **23**: 114–120.

SCHURER, K. (1971): Direct reading optical leaf area planimeter. *Acta Bot. Neerl.* **20**: 132–140.

ÜBERLA, K. (1968): *Faktorenanalyse*. Springer-Verlag, Berlin/Heidelberg/New York.

WHITEHEAD, F. H. & R. P. SINHA (1966): Taxonomy and taximetrics of Stellaria media (L.) Vill, S. neglecta Weihe and S. pallida (Dumort) Pire. *New Phytol.* **66**: 769–784.