Growth-free canonical variates applied to fossil foraminifers

R. A. REYMENT and C. F. BANFIELD

The methods of the companion paper (Gower 1976), using two of the proposed methods for estimating growth vectors (principal components and factor analysis), are applied to two species of Paleocene planktic foraminifers. In most cases, the results for the principal components estimates differ but slightly from those yielded by a growth vector from maximum likelihood factor analysis. The growth-free canonical variate means can be applied to comparisons between species observed at different periods of time.

Richard A. Reyment, Paleontologiska institutionen, Uppsala universitet, Box 558, S-751 22 Uppsala; Colin F. Banfield, Rothamsted Experimental Station, Harpenden, Herts., United Kingdom, 1st March, 1976.

Introduction

T. P. Burnaby's interest in growth invariance grew out of his research on Carboniferous pelecypods, a group of invertebrates showing additive growth and lacking a terminal growth stage. He observed that samples of these pelecypods always consisted of a number of individuals at different stages of growth which made statistical comparisons of their means biologically misleading. He realized that unless it is possible to be precise about the growth stages of a species (this can be done for vertebrates with clearly identifiable terminal growth criteria, vertebrate teeth, crustaceans), many kinds of statistical analysis become pointless if applied to growth-confounded data. This is particularly true of multivariate statistical methods using generalized distances, linear discriminant functions or canonical variate analysis.

Burnaby (1966) found a logical solution to the problem of obtaining growth-invariant versions of the above-mentioned methods. He couched his development in terms of mixtures of growth stages, ecologically confounded size differences, and differences ascribable to sexual dimorphism. The theoretical concepts involved are treated in the companion paper by J. C. Gower (1976). Here, we shall be concerned with testing the theory on fossil data.

Method of analysis

Gower (1976) describes four methods for estimating a matrix, $\mathbf{K}$, the columns of which are k growth vectors. The coefficients of these growth vectors give linear combinations of the v variates that have been measured on each of the n fossil specimens. Two of the methods proposed by Gower for estimating these vectors, namely, the "external" estimation methods where concomitant variables are required, could not be applied to our study material. Such additional variables have to be highly correlated with age so as to express the growth of the specimens. As such variables cannot be found on foraminifers, nor, indeed, on almost all fossil material, only the two "internal" methods of estimation could be utilized.

The first of these methods of internal estimation depends on some results of Jolicoeur (1963), which suggest that the k growth vectors can be estimated by the first k latent vectors of the pooled within-

Table 1. Pooled within-groups covariance matrix and group means for *Subbotina pseudobulloides*
($N = 278$; groups $= 6$).

Pooled covariance matrix

0,0184					
0,0173	0,0203				
0,0170	0,0180	0,0226			
0,0165	0,0146	0,0159	0,0206		
0,0176	0,0198	0,0188	0,0157	0,0257	
0,0199	0,0210	0,0199	0,0155	0,0204	0,0298

Group means

						N
5,1391	4,9476	4,6280	4,6165	4,4182	4,2415	20
5,1501	4,9542	4,5415	4,6714	4,4530	4,2696	29
4,9748	4,7771	4,4850	4,4833	4,2821	4,0888	30
5,1837	5,0079	4,6270	4,6666	4,5271	4,3238	39
5,0404	4,8598	4,5178	4,5127	4,3508	4,1376	60
5,1845	5,0206	4,6998	4,6813	4,5481	4,3389	100

Table 2. Pooled within-groups covariance matrix and group means for *Globoconusa daubjergensis* (Brönniman) (N = 145; groups = 4).

Pooled covariance matrix

0,0240					
0,0207	0,0227				
0,0230	0,0215	0,0264			
0,0233	0,0194	0,0233	0,0285		
0,0210	0,0216	0,0216	0,0207	0,0255	
0,0216	0,0208	0,0216	0,0192	0,0202	0,0252

Group means

						N
5,0198	4,8471	4,6717	4,5225	4,4162	4,2545	26
5,0056	4,8723	4,6776	4,5338	4,4626	4,2588	21
5,0152	4,8263	4,6614	4,5466	4,4073	4,2385	38
5,0754	4,9154	4,7107	4,5974	4,5085	4,3407	60

groups covariance matrix, **W**, calculated from the logarithms of the variates. The growth effects are then considered to be the major source of variation within each group and can be represented by the first few principal components.

The second method of internal estimation obtains the growth effects by factor analysis. As many factors as seem consistent with the data are fitted and the first k of these are taken as the growth vectors. As in the first method, the k growth effects are considered to contribute the maximum variation in the fossil groups.

The estimation of the growth vectors is undoubtedly a troublesome part of the Burnaby approach and one to which he gave much thought. It seems to us, however, that in many cases, the internal methods of estimation here used may offer a reasonable, although necessarily inferior, alternative to the external methods.

Once **K** has been estimated, it is possible to obtain

$$\mathbf{Q} = \mathbf{I} - \mathbf{K}(\mathbf{K'K})^{-1}\mathbf{K'}$$

which projects the values of the specimens onto a space that is orthogonal to the space of **K**. These

values, so obtained, are free from growth effects.

As shown by Gower in the companion paper, a generalized inverse of the Rao type can be computed for **QWQ** and the inverse,

$$\mathbf{C} = \mathbf{Q}(\mathbf{QWQ})^{-}\mathbf{Q},$$

of Burnaby (1966) formed. Solution of the equation $(\mathbf{CG'G} - \lambda\mathbf{I})\mathbf{l}$, where **G** is the matrix of v variate means for each of the p groups, will give canonical variates **l** that are devoid of growth effects. Also, the Mahalanobis' D^2 distances between the means of groups i and j when projected onto the Q-space are given by

$$D_Q^2 = (\mathbf{g}_i - \mathbf{g}_j)\mathbf{C}(\mathbf{g}_i - \mathbf{g}_j).$$

Consequently, by using **GCG'**, it is possible to obtain the coordinates of the group means in the Q-space directly by the Q-mode technique of Gower (1966). The canonical variate loadings are then given by **CG'PR⁻**, where **P** is the matrix of coordinates obtained by the Q-mode technique and **R⁻** is the diagonal matrix of inverted non-zero latent roots. In the companion paper, Gower shows that the above algebra, and therefore the computational details, can be simplified for the principal components estimation of **K** and that it is not necessary to compute a generalized inverse of **QWQ** to obtain Burnaby's inverse, **C**.

To obtain the principal components estimate of **K**, the GENSTAT system (Nelder et al. 1973) was used and for the factor loadings estimates, the programs UFABY3 (Jöreskog & Van Tillo 1971) and REFADIV (Sörbom 1974). All the subsequent steps in the calculations, matrix manipulations and canonical variates, were obtained with the help of GENSTAT.

The study material

The methods were tested by means of data obtained from Malmgren (1974) on the Paleogene planktic foraminifers, *Globoconusa daubjergensis* (Brönniman) and *Subbotina pseudobulloides* (Plummer), from the type Danian in Southern Scandinavia. The six variables measured by Malmgren on both of these species are shown in Fig. 1.

The samples we selected for our analysis derive from a borehole put down at Limhamn, Skåne, Sweden. The specimens of *S. pseudobulloides* come from the levels: 1,00 m, 3,00 m, 9,30 m, 33,30 m, 40,50 m, and 67,20 m. The specimens of *G.*
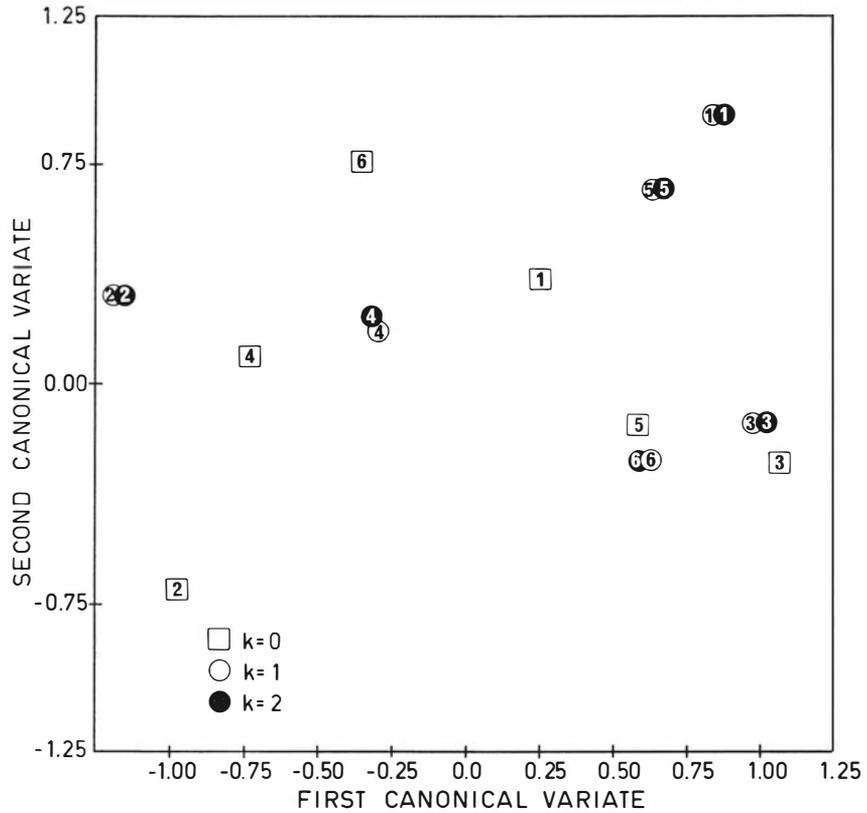
Table 3. Latent roots and vectors of the within-groups covariance matrix for *Subbotina pseudobulloides*.

1	2	3	4	5	6
Latent roots					
0,1129	0,0094	0,0070	0,0047	0,0022	0,0011
Latent vectors					
0,3860	−0,1291	0,2063	−0,1784	−0,3120	0,8140
0,4032	0,1582	−0,1402	−0,0539	−0,7730	−0,4386
0,4069	−0,1071	−0,0252	0,8992	0,1117	0,0363
0,3535	−0,7514	0,3035	−0,2505	0,1894	−0,3460
0,4288	0,0219	−0,7890	−0,2596	0,3463	0,0759
0,4626	0,6179	0,4717	−0,1630	0,3701	−0,1348

Fig. 2. Plot of the means for the first canonical variate against those for the second canonical variate for $k = 0$, 1, and 2 for *Subbotina pseudobulloides*.
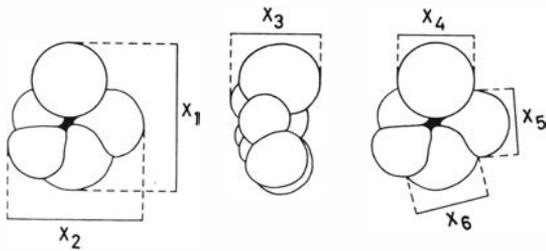

Fig. 1. Measurements made on the foraminiferal species analysed.

daubjergensis come from levels 0,00 m, 1,00 m, 3,00 m, and 9,30 m.

Results for principal components estimates

Jolicoeur (1963) pointed out that the logarithmic transformation of morphological variables tends

to make the results of statistical calculations independent of the order of magnitude of the variables. It has, as it were, a standardizing effect. The covariance matrix becomes almost independent of the order of magnitude and scale of the measurements on which it is based. He concluded that the first principal axis of such a covariance matrix corresponds to the general direction in which

Table 4. Latent roots and vectors of the within-groups covariance matrix for *Globoconusa daubjergensis*.

1	2	3	4	5	6
Latent roots					
0,1320	0,0087	0,0054	0,0031	0,0018	0,0013
Latent vectors					
0,4137	0,1267	−0,1883	0,1489	0,1624	0,8537
0,3919	−0,3396	0,1789	−0,1605	0,7899	−0,2223
0,4255	0,1076	−0,1313	−0,8293	−0,3166	−0,0462
0,4168	0,7659	0,0232	0,3028	0,0726	−0,3771
0,4034	−0,2629	0,7356	0,2150	−0,4224	0,0487
0,3972	−0,4488	−0,6112	0,3555	−0,2564	−0,2739

organisms have changed while growing and it may be looked upon as a path along which growth has progressed further for larger than for smaller individuals.

In our present study, six variates were measured on each of 278 specimens of *S. pseudobulloides* and each of 145 specimens of *G. daubjergensis* (see Fig. 1). The logarithms were taken of these two sets of data from which the pooled within-groups covariance matrices were obtained (representing 6 samples for the former species and 4 samples for the latter). This information is presented in Tables 1 and 2. From these matrices, the principal components were computed. These are given in Tables 3 and 4 for the two species.

Growth-free canonical variate analyses were then made for each of the species for $k = 0$, 1, and 2. The analysis corresponding for $k = 0$ is, of course, the standard one of canonical variates where no growth effects are extracted. The analyses with $k = 1$ and $k = 2$ correspond to the removal of one and two principal components as growth vectors, respectively.

Tables 5 through 12 contain the squared generalized distances, the canonical variate loadings and the canonical variate means resulting from the analyses of the two sets of specimens for the three levels of k. In Fig. 2, the means of the six samples are plotted relative to the first two canonical variate axes for $k = 0$, 1, and 2. The means for $k = 1$ and $k = 2$ are seen to be similar, which reflects the slight difference made to the

squared generalized distances when two principal components are removed instead of one (see Table 5). When, however, no principal components were taken out ($k = 0$), the configuration of the means was substantially different from those for $k = 1$, and 2. This illustrates the comparatively large change made to the distances by removing one or two principal components. It is this change that we are attributing to the variation due to growth.

For $k = 0$, the means of samples 3 and 6 are relatively far apart ($D^2 = 3,1345$) but once the presumed growth variation has been removed, these two samples are much closer together ($D^2 = 0,2947$). In fact, all the distances are reduced by

Table 6. Squared generalized distances for *Globoconusa daubjergensis* for $k = 0$, 1, 2.

$k = 0$				
1	0,0000			
2	0,6592	0,0000		
3	0,1904	0,7637	0,0000	
4	0,9841	0,6453	0,9417	.0,0000
$k = 1$				
1	0,0000			
2	0,6519	0,0000		
3	0,1890	0,7476	0,0000	
4	0,7671	0,5020	0,6865	0,0000
$k = 2$				
1	0,0000			
2	0,6252	0,0000		
3	0,0610	0,4768	0,0000	
4	0,7322	0,5015	0,3892	0,0000

Table 5. Squared generalized distances for *Subbotina pseudobulloides* for $k = 0$, 1, 2.

$k = 0$					
1	0,0000				
2	2,8677	0,0000			
3	1,9539	4,4016	0,0000		
4	1,5322	0,8623	3,6189	0,0000	
5	0,7524	2,9727	0,8738	1,9316	0,0000
6	1,3623	2,6748	3,1345	0,7847	2,2441	0,0000
$k = 1$					
1	0,0000				
2	2,8647	0,0000			
3	0,7654	3,0867	0,0000		
4	1,3496	0,7246	1,3157	0,0000	
5	0,2759	2,4150	0,7146	0,6821	0,0000
6	1,0088	2,3848	0,2947	0,7563	0,5923	0,0000
$k = 2$					
1	0,0000				
2	2,8430	0,0000			
3	0,7495	3,0120	0,0000		
4	1,3096	0,6042	1,3101	0,0000	
5	0,2202	2,2688	0,7021	0,6807	0,0000
6	0,9927	2,3103	0,2947	0,7515	0,5801	0,0000

Table 7. Canonical variate analysis for *Subbotina pseudobulloides* for $k = 0$

Latent roots

1	2	3	4	5
3,1008	1,3326	0,6374	0,2336	0,0235

Canonical variate loadings

1	2	3	4	5
$-1,6494$	$-2,0139$	$-17,4327$	$-2,5077$	15,2510
4,7737	0,3278	$-8,0457$	$-2,6102$	$-16,2890$
$-8,4805$	9,7136	0,9643	3,7573	$-0,0948$
5,6673	$-3,3072$	7,2132	8,2769	$-6,8715$
2,4804	0,9962	7,2153	$-7,5067$	2,5701
2,4981	$-1,1284$	7,9632	2,5453	5,4067

Coordinates of means

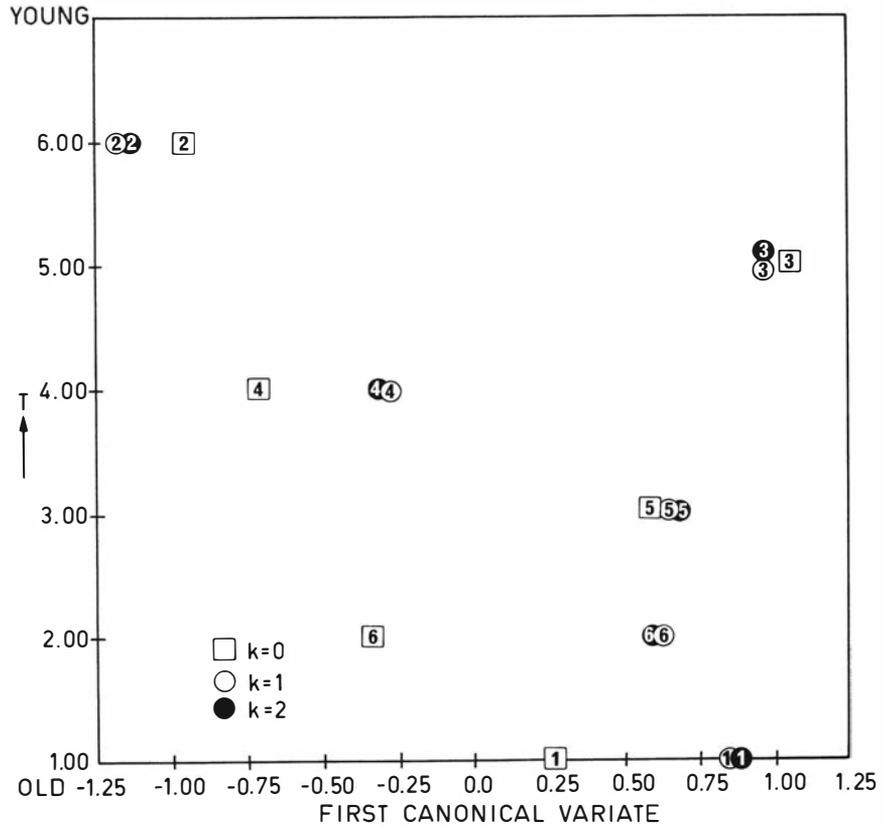| 1	2	3	4	5
$-0,2762$	0,3301	$-0,5245$	0,2498	0,0258
0,9390	$-0,7096$	0,0179	0,1451	$-0,0435$
$-1,0794$	$-0,3072$	0,4137	0,1008	0,0420
0,7064	0,1001	0,0143	$-0,2162$	0,1046
$-0,6122$	$-0,1777$	$-0,2669$	$-0,3043$	$-0,0652$
0,3224	0,7643	0,3455	0,0249	$-0,0636$

Fig. 3. Plot of the means of the first canonical variate against stratigraphical position for $k = 0$, 1, and 2 for *Subbotina pseudobulloides*.

the removal of principal components because the initial distances $(k = 0)$ are being partitioned into two components which express the distances projected onto the growth-space and those projected onto the space orthogonal to the growth-space. Even though these distances are smaller, the samples are more distinct because the minimum distances between sample means required for significance are based on reduced within-sample variation. So, for purposes of discrimination, removal of the growth effect is important.

The chronological order of the samples, from oldest to youngest, is 1—6—5—4—3—2. In Fig. 3, the sample means for the first canonical variate have been plotted against their chronological (stratigraphical) order. As might be suspected from the previous figure, it is samples 1, 4, and 6 that show the most divergence resulting from the removal of "growth" variation. What is perhaps more interesting when interpreting such a plot is the pattern of oscillation of the sample means with time

Table 8. Canonical variates analysis for *Subbotina pseudobulloides* for $k = 1$.

Latent roots

1	2	3	4	5
2,2298	0,6467	0,2667	0,0619	0,0001

Canonical variate loadings

1	2	3	4	5
1,5126	17,9109	0,4724	15,5306	−249,7511
−4,0535	7,2746	5,9664	−16,3598	−98,3832
12,9584	−1,6057	−2,6835	−1,4948	19,6511
−6,6390	−6,8857	−7,5410	−7,8315	52,2846
−1,6893	−7,5132	6,7275	4,1577	79,9098
−2,4882	−7,6466	−3,7076	4,7446	162,8214

Coordinates of means

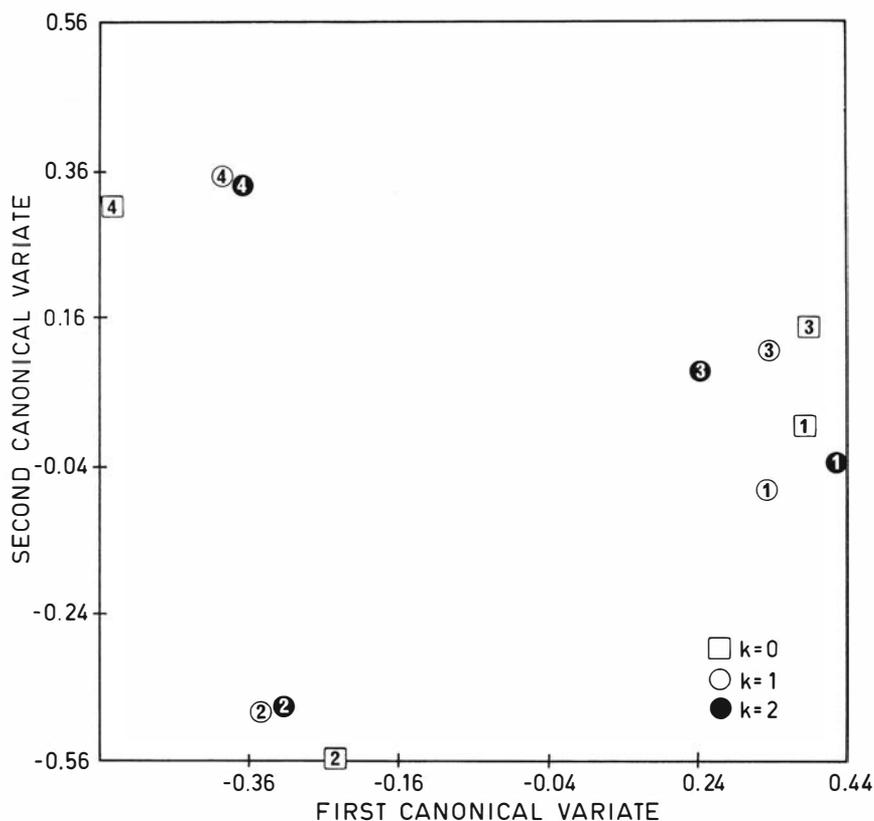| 1	2	3	4	5
0,4676	0,5034	−0,1826	−0,0667	0,0043
−1,1577	0,0329	−0,1904	−0,0278	−0,0033
0,5503	−0,3420	−0,2395	0,1348	−0,0010
−0,4440	−0,0502	0,2503	0,0903	0,0069
0,3066	0,2902	0,2569	0,0392	−0,0071
0,2772	−0,4342	0,1052	−0,1698	0,0001

Fig. 4. Plot of the means for the first canonical variate against those for the second canonical variate for $k = 0$, 1, and 2 for *Globoconusa daubjergensis*.

Table 9. Canonical variates analysis of *Subbotina pseudobulloides* for $k = 2$.

Latent roots

1	2	3	4
2,2010	0,6438	0,2140	0,0462

Canonical variate loadings

1	2	3	4
1,7033	17,8998	−0,9369	18,0750
−4,2569	7,4935	6,5773	−18,2704
13,1791	−1,8563	−1,4409	0,8396
−5,7706	−7,4621	−4,0342	−6,7154
−1,8026	−7,3548	7,2186	7,3689
−3,2226	−7,3141	−7,2916	−0,1189

Coordinates of means

1	2	3	4
0,4836	0,4905	−0,1487	−0,0714
−1,1369	0,0262	−0,1666	−0,0368
0,5500	−0,3500	−0,2420	0,1019
−0,4612	−0,0342	0,1913	0,0900
0,2913	0,3012	0,2148	0,0524
0,2731	−0,4338	0,1512	−0,1360

and how this has been altered by the removal of growth variation. Such a pattern through time indicates evolutionary changes and is useful for seeing whether similar changes occur for other species sampled at the same locations.

We now consider the results for the second species, *G. daubjergensis*. Table 6 shows the squared generalized distances between the means of the four samples. Samples 2, 3 and 4 were taken from the same stratigraphical levels as samples 2, 3 and 4 of the previous species. Here we see that the distances do not reduce so dramatically for $k = 0$, 1, and 2 as for *S. pseudobulloides* and that the removal of growth variation within samples, contributes less than in the other species. Fig. 4 shows the configuration of means to be similar for all three values of k. However, even though the distances have not been reduced much by the removal of growth variation, the minimum distance that separates the means of two distinct
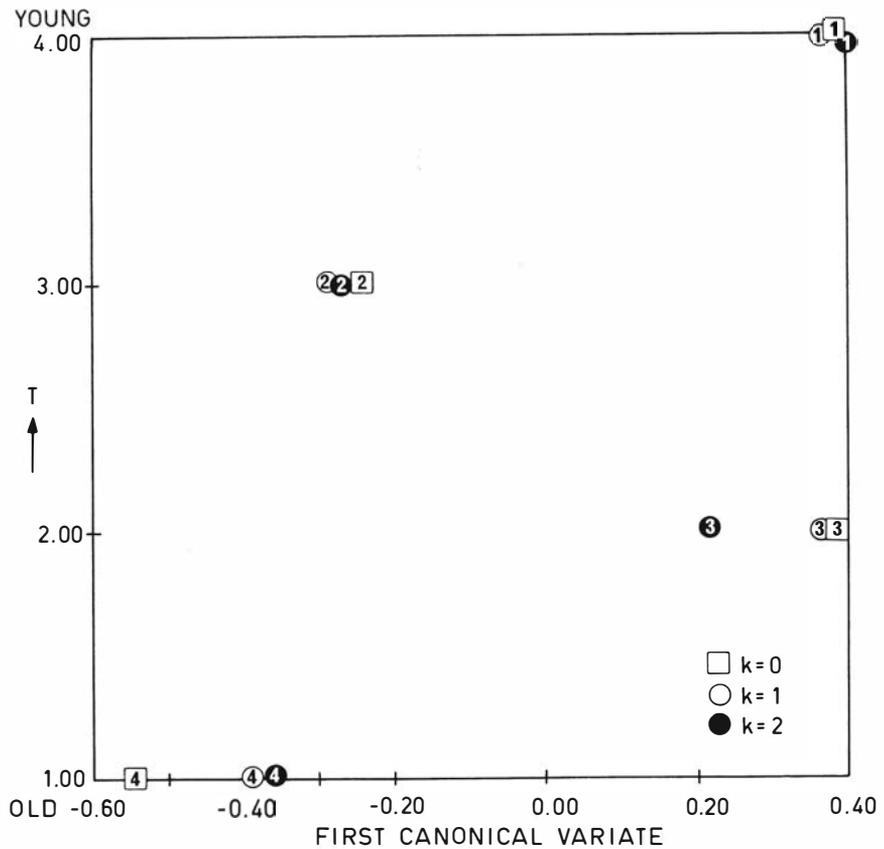
Fig. 5. Plot of the first canonical means against stratigraphical position for $k = 0$, 1, and 2 for *Globoconusa daubjergensis.*

samples may have been, and so the discrimination may have been improved.

Fig. 5 shows the plot of the means of the first canonical variate against stratigraphical order. The pattern is again a characteristic one of oscillation with time comparable with the pattern seen in Fig. 3 for *S. pseudobulloides*. However, these patterns could have been compared satisfactorily if the growth variation had not been extracted and the inclusion of more common samples would be required to show anything conclusive.

Results using factor loading estimates

As with principal components estimates of growth, the logarithms of the six variates measured on the specimens of the two foraminiferal species were used to obtain pooled within-sample co-

Table 10. Canonical variates analysis of *Globoconusa daubjergensis* for $k = 0$.

Latent roots		
1	2	3
0,6522	0,3056	0,0882

Canonical variate loadings		
1	2	3
−12,5609	11,6357	9,1919
3,4289	−6,1108	5,0991
−6,3700	−8,4808	1,2442
5,1141	2,0156	−11,3165
6,7507	−2,6028	−2,9314
7,3496	5,6981	0,3018

Coordinates of means		
1	2	3
−0,3888	0,0006	0,2138
0,2413	−0,4421	−0,0434
−0,3909	0,1407	−0,1994
0,5385	0,3007	0,0291

Tabzle 11. Canonical variates analysis of *Globoconusa daubjergensis* for $k = 1$.

Latent roots

1	2	3
0,5408	0,2623	0,0829

Canonical variate loadings

1	2	3
−16,0170	6,8459	8,6467
4,1708	−5,9667	4,6145
−5,8802	−11,7332	−1,2762
4,8337	3,8941	−10,8967
7,2445	−0,7606	−2,6349
6,4349	8,0106	1,9197

Coordinates of means

1	2	3
−0,3661	−0,0842	0,1984
0,3506	−0,3579	−0,0533
−0,3690	0,1000	−0,1953
0,3845	0,3421	0,0502

variance matrices, already listed in Tables 1 and 2. These were then subjected to factor analyses to fit as many factors as were consistent with the data. For both sets of data, either one or two factors were computed. As suggested by Gower (1976), if two factors were extracted, then only the first of these was attributed to growth effects. These growth vectors are given in Tables 13, 15, 17, and 20. The resulting squared generalized distances and the canonical variate loadings and means appear in Tables 13 through 20.

Kuhry et al. (1976) have considered in some detail the problem of estimating a multivariate

Table 12. Canonical variates analysis of *Globoconusa daubjergensis* for $k = 2$.

Latent roots

1	2	3
0,4350	0,2586	0,0029

Canonical variate loadings

1	2	3
−17,0445	9,0035	2,5782
2,0658	−5,7598	18,4632
−6,6665	−11,2458	−8,7230
10,1243	1,7221	2,7786
6,8059	−1,3918	−9,0252
5,3179	7,9576	−5,3043

Coordinates of means

1	2	3
−0,4226	−0,0308	0,0314
0,2856	−0,3810	−0,0017
−0,2191	0,0884	−0,0423
0,3561	0,3234	0,0126

Table 13. The growth vector obtained from a maximum likelihood factor analysis of the covariance matrix for *Subbotina pseudobulloides,* and the resulting squared generalized distances.

Growth vector

1	2	3	4	5	6
0,0650	0,0680	0,0660	0,0610	0,0730	0,0770

Squared generalized distances

1	0,0				
2	2,8295	0,0			
3	0,7655	2,7485	0,0		
4	1,2745	0,7646	1,0656	0,0	
5	0,3036	2,2235	0,6977	0,5443	0,0
6	0,9761	2,4933	0,2041	0,7716	0,5756	0,0

growth vector by a one factor solution. Sörbom's (1974) method of maximum likelihood factor analysis, in which no assumptions are made for uncorrelated errors, was thought to yield a satisfactory solution, as it takes biological constraints into account. Jöreskog & Van Thillo (1971) give a rapidly convergent algorithm for maximum likelihood factor analysis. In the following, we consider examples of the residual covariance approach.

We begin with our analyses of *S. pseudobulloides.* The growth vector estimated by extracting one factor using maximum likelihood factor analysis (Jöreskog & Van Thillo 1971) of the covariance matrix was used to compute squared generalized distances (Table 13), which were subsequently used in a canonical variates analysis (Table 14).

Table 14. Canonical variate analysis of *Subbotina pseudobulloides* for $k = 1$, resulting from maximum likelihood factor analysis of the covariance matrix.

Latent roots

1	2	3	4	5
2,0801	0,6447	0,2619	0,0523	0,0006

Canonical variate loadings

1	2	3	4	5
1,1807	17,8532	0,7124	16,3392	−8,1876
−3,6866	7,3876	5,5094	−17,1672	7,7698
13,0217	−1,4658	−2,9400	−1,0709	−1,0330
−6,4121	−6,9497	−7,6919	−7,6192	−1,2692
−1,4124	−7,4751	6,8346	4,3392	−6,2927
−2,4837	−7,7461	−3,3328	4,2078	7,9066

Coordinates of means

1	2	3	4	5
0,4680	0,5082	−0,1967	−0,0464	0,0094
−1,1417	0,0210	−0,1766	−0,0371	−0,0062
0,4650	−0,3482	−0,2298	0,1290	−0,0037
−0,3944	−0,0483	0,2547	0,0847	0,0154
0,2580	0,2897	0,2571	0,0258	−0,0161
0,3451	−0,4224	0,0912	−0,1560	0,0013

Table 15. The growth vector obtained from a one factor model using Sörbom's factor analysis of the covariance matrix for *Subbotina pseudobulloides,* and the resulting squared generalized distances.

Growth vector

1	2	3	4	5	6
1,0000	1,0410	1,1220	0,8240	1,1530	1,1850

Squared generalized distances

1	0,0				
2	2,8012	0,0			
3	0,8329	3,7600	0,0		
4	1,4497	0,5649	1,8069	0,0	
5	0,3288	2,8133	0,7078	1,0512	0,0
6	1,0833	2,0566	0,6150	0,7265	0,8531	0,0

Table 17. The growth vector obtained from the first factor of a two factor model using Sörbom's (1974) factor analysis of the covariance matrix for *Subbotina pseudobulloides,* and the resulting squared generalized distances.

Growth vector

1	2	3	4	5	6
0,6160	0,4399	0,5503	0,6200	0,4520	0,4868

Squared generalized distances

1	0,0				
2	2,8351	0,0			
3	0,8725	3,6630	0,0		
4	1,5275	0,8497	2,6769	0,0	
5	0,2230	2,6734	0,7766	1,4982	0,0
6	1,3598	2,6583	2,1584	0,7843	1,7876	0,0

The squared generalized distances are close to the values obtained for the principal components estimate of the growth vector (cf. Table 5, $k = 1$) and, consequently, the coordinates of the canonical variate means are found to be similar in all five significant vectors. Because of these similarities, we do not present plots for these results as the figures closely resemble Figs. 2 and 3.

Table 15 shows the growth vector ($k = 1$) when Sörbom's (1974) method is utilized to estimate one factor from the within-samples covariance matrix. His algorithm allows the residuals to be correlated as this may otherwise be considered an unrealistic constraint from the biological point of view. The squared generalized distances are again

similar to those obtained from principal components ($k = 1$), although some values are closer to the distances obtained by maximum likelihood factor analysis, discussed above. The canonical variate means and loadings obtained from these distances (Table 16) again only deviate slightly from those of the principal components approach. However, the distances did tend to be slightly larger and therefore the canonical variate means are further apart, showing that less variation has been removed by the elimination of growth by this particular estimate.

When a two-factor model is fitted to the co-

Table 16. Canonical variate analysis of *Subbotina pseudobulloides* for $k = 1$, resulting from Sörbom's (1974) factor analysis of the covariance matrix using a one factor model.

Latent roots

1	2	3	4	5
2,5655	0,6551	0,2348	0,1192	0,0014

Canonical variate loadings

1	2	3	4	5
2,6936	18,4595	1,6367	−15,6522	−6,5309
−4,9059	6,7436	3,9027	18,2701	6,1081
12,3047	−2,1909	−3,4935	1,8142	−1,0320
−7,3993	−7,0497	−8,0357	5,8695	−2,0153
−1,6604	−7,1700	7,1784	−4,5579	−5,8852
−2,8531	−7,5490	−2,8987	−4,2056	8,2503

Coordinates of means

1	2	3	4	5
0,4415	0,4763	−0,2325	0,1031	0,0112
−1,1708	0,0649	−0,1597	−0,0595	−0,0145
0,7220	−0,3246	−0,1318	−0,2171	0,0003
−0,5331	−0,0469	0,2197	−0,0429	0,0260
0,4258	0,3037	0,2994	−0,0196	−0,0180
0,1146	−0,4735	0,0049	0,2360	−0,0050

Table 18. Canonical variate analysis of *Subbotina pseudobulloides* for $k = 1$, resulting from Sörbom's (1974) factor analysis of the covariance matrix using the first factor of a two factor model.

Latent roots

1	2	3	4	5
2,6367	1,1181	0,4913	0,1303	0,0143

Canonical variate loadings

1	2	3	4	5
−7,3224	−12,0254	9,0853	12,4574	5,2225
6,6102	2,2470	13,7017	−12,4718	−6,0130
−11,1153	6,7834	−1,9761	−2,2157	−0,6151
6,7433	−0,5148	−8,7414	−7,8062	−5,3153
3,3333	4,1256	−3,9271	10,6473	−2,7078
4,1742	2,3431	−6,8647	−1,9328	8,8043

Coordinates of means

1	2	3	4	5
−0,4956	−0,0108	0,3189	−0,2357	0,0390
1,0333	−0,5345	−0,1292	−0,1024	−0,0289
−0,8184	−0,2394	−0,4794	0,0447	0,0254
0,6322	0,1979	0,1217	0,1790	0,0728
−0,4858	−0,2374	0,3229	0,1660	−0,0621
0,1342	0,8243	−0,1550	−0,0516	−0,0462

Table 19. The growth vector obtained from a maximum likelihood factor analysis of the covariance matrix for *Globoconusa daubjergensis* and the resulting canonical variate analysis.

Growth vector

1	2	3	4	5	6
0,0650	0,0620	0,0660	0,0650	0,0640	0,0640

Squared generalized distances

1	0,0		
2	0,6490	0,0	
3	0,1888	0,7433	0,0
4	0,7222	0,4771	0,6363	0,0

Latent roots

1	2	3
0,5214	0,2513	0,0815

Canonical variate loadings

1	2	3
−16,4909	5,7456	8,5771
4,4105	−5,8566	4,5265
−5,2693	−12,0752	−1,7009
4,7078	4,3670	−10,7580
7,2693	−0,3375	2,5649
5,8592	8,1931	2,1489

Coordinates of means

1	2	3
−0,3580	−0,1034	0,1940
0,3722	−0,3350	−0,0556
−0,3639	0,0920	−0,1941
0,3497	0,3464	0,0557

variance matrix, using the Sörbom (1974) method, and only the first factor is employed as an estimate of the growth vector ($k = 1$), we obtained the squared generalized distances of Table 17. They are much larger than the equivalent distances from the principal components, or the above factor analyses, and show that the removal of the growth effect is far less impressive. The resulting canonical variate analysis, given in Table 18, shows the means to be farther apart.

For *G. daubjergensis*, similar comparisons could be made between the results of the principal component and factor analyses as for *S. pseudobulloides*. For instance, Table 19 shows the squared generalized distances obtained when one factor is estimated using the method of Jöreskog & Van Thillo (1971) on the within-samples covariance matrix. These are similar to the principal component distances (Table 6, $k = 1$). The resulting canonical variate analysis loadings and means also compare favourably, so we have not included any plots as they correspond to Figs. 4 and 5 ($k = 1$).

Table 20 shows the growth vector obtained from the first of two factors estimated from the within-samples correlation matrix, using the program of Jöreskog & Van Thillo (1971). The distances are slightly smaller than those obtained by principal components (cf. Table 6, $k = 1$), but not those yielded by the previous factor analytical approach.

Conclusions

The results have shown us that using the principal components of the pooled within-samples covariance matrix, we could remove the major source of variation. This variation we have interpreted as being caused by the individuals of the foraminiferal species having been at different stages of growth when fossilized. As these growth differences interfere with discrimination between the samples on a purely evolutionary basis, it is important that they be eliminated. We found the patterns produced when the growth-free canonical variate sample means were plotted, especially against chronological order, permitted comparisons to be made be-

Table 20. The growth vector obtained from the first factor of a two factor model using maximum likelihood factor analysis of the correlation matrix for *Globoconusa daubjergensis*, and the resulting canonical variate analysis.

Growth vector

1	2	3	4	5	6
0,9220	0,9940	0,9010	0,8040	0,9110	0,8920

Squared generalized distances

1	0,0		
2	0,6554	0,0	
3	0,1624	0,7103	0,0
4	0,8433	0,5480	0,6469	0,0

Latent roots

1	2	3
0,5403	0,2894	0,0619

Canonical variate loadings

1	2	3
−16,8563	7,6439	5,7018
0,9558	−7,7764	0,4275
−5,4229	−9,6302	−1,4291
6,3592	4,1130	−10,0901
8,1180	−0,8067	−0,7999
7,8130	7,6087	4,9851

Coordinates of means

1	2	3
−0,4000	−0,0741	0,1640
0,3140	−0,3936	−0,0446
−0,3298	0,1348	−0,1733
0,4159	0,3329	0,0539

tween species so as to suggest common evolutionary changes.

The results of the factor analysis proved similar to those of the principal components, especially for $k = 1$. As the principal components approach simplifies the algebraic and computational aspects of the analysis, it is to be preferred.

Our work has been of a somewhat exploratory nature because of the absence in the literature of any previous applications of growth-free canonical variate analysis. We feel this technique can be of use in similar exercises where the major source of variation within populations obscures variation between populations. We are, however, not persuaded that any internal estimation of growth effects is superior to the use of external growth variables. It is to be hoped that future applications of growth-free canonical variates will endeavour to compare both approaches.

REFERENCES

Burnaby, T. P. 1966: Growth-invariant discriminant functions and generalized distances. *Biometrics 22*, 96—110.

Gower, J. C. 1966: A Q-technique for the calculation of canonical variates. *Biometrika 53*, 588—589.

Gower, J. C. 1976: Growth-free canonical variates and generalized inverses. *Bull. Geol. Inst. Univ. Uppsala, N.S. 7*, 1—10.

Jolicoeur, P. 1963: The degree of generality of robustness in *Martes americana. Growth 27*, 1—27.

Jöreskog, K. G. & Van Thillo, M. 1971: New rapid algorithms for factor analysis by unweighted least squares, generalized least squares and maximum likelihood. Research Memorandum 71—5, Princeton N.J. Educational Testing Service.

Kuhry, B., Marcus, L. F. & Sörbom, D. 1976: Factor solutions in multivariate allometry. Manuscript.

Malmgren, B. 1974: Morphometric studies of planktonic foraminifers from the type Danian of Southern Scandinavia. *Stockh. Contrib. Geol. 29*, 126 pp.

Nelder, J. A., and members of the Rothamsted Statistics Department, 1973: GENSTAT Reference Manual. Interuniversity Research Council Series. Scientific and Social Sciences Program Library, University of Edinburgh.

Sörbom, D. 1974: Detection of correlated errors in longitudinal data. *Research Report, Department of Statistics, University of Uppsala, No. 74—8*. 27 pp.