The Korean Journal of Quaternary Research
Vol. 3, No. 1, 55~68 (1989)

Application of Numerical Methods in the Zonation and Correlation of Four Late Quaternary Pollen Data from Iowa

Hyung Keun Kim
Department of Earth Science, King Sejong University Seoul 133-747, Korea

수치분석의 도식화를 통한 第四紀 花分자료의 分帶 및 對比

金 亨 根
세종대학교 지구과학과

ABSTRACT

This paper presents examples of the computer-aided zonation and correlation of pollen data from the Late-glacial to Holocene stratigraphic sequences at four sites in central Iowa, U.S.A.

Spearman's rank correlation coefficient matrix and first four components of Principal components analysis plotted in a stratigraphic order are combined to provide an excellent zonation of the pollen data at each site.

Correlation of the four pollen sequences are conducted by Principal components analysis of the data sets combined in one. The first and second principal components successfully provide correlation lines that match fairly closely the zone boundaries of each pollen sequence. The third and fourth components, in contrast, are greatly different from site to site, representing the unique pollen assemblages at each site.

요 약

본 논문은 네지점에서의 제4기지층의 화분분석 자료들을 대상으로 각각의 자료를 보다 객관적으로 적절하게 分帶하고(zone) 對比하기(correlate) 위하여 컴퓨터를 이용한 수치분석과 도식화의 몇가지 방법을 제시한다. 그 하나는 두가지의 수치분석방법을 종합하여 分帶하는 것으로, 즉 스피어맨의 순위 상관계수의 도표와 主成分分析에서 얻어진 제1, 제2, 제3, 제4 주성분의 그래프를 비

교합으로서 서로 공통적인 그룹과 각각 특징적인 그룹을 종합하여 分帶한다.

둘째는 對比의 방법인데, 네지점의 자료들을 일단 하나로 합하여 이를 근거로 한 主成分分析을 실시하고, 그결과 얻어진 제1, 제2 주성분을 각지점에 해당하는 네개의 그룹으로 다시 나누어 도식화하면, 이들 네개의 그룹은 공통적인 변화추이를 보여주므로 서로간의 對比가 매우 용이해 진다. 그러나 제3, 제4 주성분들은 각 지점에 대해 고유의 양상과 변화를 보여 주는데 이는 각 지점이 가지는 특성을 나타내는 것으로, 이들 성분들은 對比에는 사용될 수 없다.

INTRODUCTION

Quaternary pollen analysis is quantitative in nature in that the data are the number of different types of pollen and spores counted in a sample. Pollen percentages of taxa are calculated from the pollen counts at each level of a stratigraphic section, which have always been the basis of pollen studies. Traditionally, pollen percentages have been presented in a pollen diagram. Pollen diagrams are often difficult to comprehend because the pollen data they represent are complex in the number of taxa and levels, and in their vertical (stratigraphic) variations.

The description and interpretation of pollen diagrams are generally facilitated by zonation, which groups the levels of similar pollen composition. A pollen zone, as defined by Gordon and Birk (1972), is 'a body of sediment with a consistent and homogeneous fossil pollen and spore content that is distinguished from adjacent sediment bodies by differences in the kind and frequencies of its contained fossil pollen and spores'. In this sense, a pollen zone is an assemblage zone which is a biostratigraphic unit (Cushing, 1967). A pollen zone describes and summarizes a particular division of a single stratigraphic sequence. It is thus a local assemblage zone which is unique to a certain site and time range (Cushing, 1967).

Local pollen zones are of limited use in the investigation of vegetational history within a particular area of study. Pollen records from several sites in the area should be compared and interpreted. If similar pollen assemblages are found to exist between sites, a regional pollen assemblage zone can be established by the comparison of the assemblages (Cushing, 1967). Thus the comparison is strickly a biostratigraphic correlation.

Conventionally, pollen zonation and correlation has been carried out visually from pollen diagrams, generally based on the previous experience of the investigators. This method has been increasingly opposed by those who criticized it for its intrinsic subjectivity, and many numerical methods has been devised for more objective zonation and correlation (e.g., Birks and Gordon, 1985).

Present paper provides examples of computer-aided zontion and correlation. Pollen data are from Late-glacial to Holocene stratigraphic sequences at four sites in central Iowa, U.S.A. (Kim, 1986). The purpose of this paper is:

1) to zone each of pollen data sets by utilizing and comparing two numerical methods, i.e., Spearman's rank correlation coefficient and Principal components analysis (PCA), and to examine the efficiency of the combined methods,
2) to correlate four pollen data sets by PCA and to examine the significance of similarities and differences in the pollen zones between

the sites.

Correlation coefficient matrices were constructed using the BMDP statistical package, and Principal components were obtained using the SAS package. To conduct the numerical analyses, twenty taxa that occurred consitently at four sites were selected (Table 1). Pollen percentages of each taxa were used as the raw data. Pollen percentages are based on the sum of AP(arboreal pollen) and NAP (nonarboreal pollen). Cyperaceae, aquatic pollen and spores are excluded from the sum (Kim, 1986).

Table 1. Principal components analysis of pollen data at four sites.

ZUEHL FARM

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	7.490204	2.566449	0.374510	0.374510
PRIN2	4.923755	2.862419	0.246188	0.620698
PRIN3	2.061335	0.623899	0.103067	0.723765
PRIN4	1.437436	0.520517	0.071872	0.795636

EIGENVECTORS

	PRIN1	PRIN2	PRIN3	PRIN4
PICEA	0.273551	-.201865	-.285590	0.175603
LARIX	0.284726	-.087347	0.096882	0.083192
ABIES	0.294458	-.002719	0.309386	0.010865
PINUS	-.250910	-.118731	0.172036	0.450179
FRAXINUS	0.219010	0.039411	0.295223	-.114710
ALNUS	0.234596	0.026057	0.355015	0.075370
BETULA	0.176954	0.113094	0.313423	0.113390
QUERCUS	0.019052	0.414123	-.116915	0.034824
ULMUS	-.225105	0.288234	0.070085	0.177531
OSTRYA/CARP.	0.149865	0.372121	0.061732	0.029459
CORYLUS	0.250166	0.172128	0.309383	-.047654
CARYA	-.084472	0.394876	-.178191	0.104600
JUGLANS	-.114249	0.367430	-.142927	0.154999
PLATANUS	-.261686	-.083562	0.238682	0.339681
ACER	0.023093	0.328077	-.109981	0.071838
CHENOPODS	-.256965	-.062044	0.070035	-.489168
GRAMINEAE	-.298414	-.090855	0.222980	0.077073
AMBROSIA	-.264409	0.032275	0.120898	-.427438
ARTEMISIA	0.167286	-.264613	-.341134	0.228287
TUBULIFLORAE	-.286126	-.105352	0.217582	0.237370

PILOT MOUND SITE

EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
8.423603	4.490160	0.421180	0.421180
3.933442	1.813548	0.196672	0.617852
2.119895	0.773527	0.105995	0.723847
1.346368	0.457801	0.067318	0.791165

EIGENVECTORS

PRIN1	PRIN2	PRIN3	PRIN4
0.292120	-.072234	0.311624	0.091588
0.240750	-.085341	0.362580	0.290094
0.271241	0.192412	-.067661	-.176261
-.241957	0.225570	0.092151	-.110956
0.301299	0.186899	-.042163	-.120616
0.252749	0.171233	0.038968	-.017601
0.219884	0.228584	-.224163	0.121560
-.204579	0.356247	0.149554	0.025720
0.102473	0.400814	-.241691	-.119354
0.256621	0.209473	-.141952	-.029782
0.225271	0.311666	-.199329	-.073786
-.231640	0.270103	0.265771	0.000471
-.221742	0.261609	0.239168	-.100859
-.203788	0.182676	0.110712	0.050071
0.086770	0.108425	0.052105	0.762912
-.228840	-.228495	-.364818	-.104293
-.184153	0.200608	0.242868	-.120932
-.180521	0.247943	-.212540	0.300705
0.215305	0.009777	0.420402	-.277123
-.188452	0.118185	-.085743	0.160524

COLO MARSH

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	6.971157	2.395898	0.348558	0.348558
PRIN2	4.575259	2.222363	0.228763	0.577321
PRIN3	2.352896	0.406298	0.117645	0.694966
PRIN4	1.946597	0.834544	0.097?30	0.792295

EIGENVECTORS

	PRIN1	PRIN2	PRIN3	PRIN4
PICEA	-.262635	-.118866	0.422226	-.034688
LARIX	-.244869	-.033621	0.388560	-.135199
PINUS	0.305346	0.112017	0.098060	0.106044
ABIES	-.256592	0.194276	-.166314	0.197463
FRAXINUS	-.215748	0.096635	-.142210	0.201090
ALNUS	-.242042	0.235440	-.039097	0.244452
BETULA	-.157826	0.323284	-.043103	0.095159
QUERCUS	0.217471	0.333705	0.065210	0.050277
ULMUS	-.050300	0.351928	-.075329	-.387095
OSTRYA/CARP.	-.233967	0.311417	-.169857	0.122568
CARYA	0.195006	0.297845	0.067160	-.207769
CORYLUS	-.256410	0.247721	-.150720	0.213045
JUGLANS	0.104990	0.328816	0.026506	-.376568
PLATANUS	0.277470	0.171770	0.162297	0.170637
ACER	-.061872	0.227653	0.089187	-.429845
CHENOPODS	0.202395	-.195332	-.414109	-.107905
GRAMINEAE	0.279594	0.172222	0.186290	0.187415
AMBROSIA	0.253905	0.125153	-.108768	0.248195
ARTEMISIA	-.115323	0.047926	0.508092	0.168526
TUBULIFLORAE	0.291315	0.087583	0.180773	0.269492

JEWELL SITE

EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
6.771626	1.962161	0.338581	0.338581
4.809465	2.568143	0.240473	0.579055
2.241322	0.975494	0.112066	0.691121
1.265827	0.216399	0.063291	0.754412

EIGENVECTORS

PRIN1	PRIN2	PRIN3	PRIN4
0.164922	-.221098	-.436223	0.152998
0.193312	-.193521	-.410047	0.173340
0.279172	-.167534	0.222703	-.247308
-.274544	-.096735	-.053944	0.268310
0.238013	-.102621	0.048454	-.520348
0.261022	-.163977	0.312135	0.237772
0.235270	-.120840	0.323299	0.278001
0.038643	0.423828	0.044587	0.170541
0.127347	0.396271	-.002551	0.084169
0.278065	0.222600	0.182408	-.028748
0.019726	0.426667	-.026535	0.130637
0.285493	-.036760	0.358255	0.130162
0.028753	0.384402	-.014289	0.068549
-.249462	0.119253	0.075511	-.108995
0.178390	0.135163	-.316553	0.131418
-.288832	-.078993	0.081102	-.224211
-.280607	-.031094	0.251424	0.136362
-.262844	0.015969	0.148449	0.032411
-.009927	-.242921	0.083541	0.471084
-.314371	-.088786	0.124613	0.110104

ZONATION OF POLLEN SEQUENCE

Many numerical methods have been proposed to zone pollen sequence by those who emphasized the importance of objectivity in the zonation (e.g., Gordon and Birks, 1972). To note, no two methods privide the zonation identical to each other. Further, some of the numerical zonations are often not very convincing. Blind utilization of them might result in the incorrect interpretation of pollen data. To avoid the pitfall that might be encountered, two numerical methods are adopted and the results are compared for the final decision.

Spearman's rank correlation coefficient was used by Ogden (1977) to compare surface pollen samples. Maher (1982) applied this measure to an 'agglomerative' method which creates a dendrogram for the zonation of a pollen diagram. He also constructed a correlation coefficient matrix that is similar to a 'divisive' procedure (see Gordon and Birks, 1972 for agglomerative and divisive methods in zonation). A 'modified divisive method' is adopted to create correlation coefficient matrices (e.g., Figure 1b).

Principal components analysis (PCA) is an ordination method which has been successfully used for the zonation of pollen diagrams (e.g., Birks, 1974; Pennington and Sackin, 1975). The results of PCA can either be plotted as a 2-dimensional scatter diagram (e.g., Figure la), or the component scores can be plotted stratigraphically (e.g., Figure 1c). This method is useful because the first few principal components usually convey most of the information on the variation of all the taxa included in the analysis. Percentages of twenty taxa are used as variables (attributes). The first four principal components include about 80% of total variance for each site (Table 1).

The zonation was conducted according to the following steps:

1) visually cluster the scatter plot of the first and second components,
2) visualy cluster the correlation coefficient matrix independently so that each cluster has low correlations with those above and below.
3) if the results of 1) and 2) do not agree on cluster boundaries, resolve the differences by inspecting the stratigraphic plots of the first four principal components, and
4) visually subdivide the clusters into zones, using the stratigraphic plots of the first four principal components.

Brief descriptions of the procedures and the results follow.

(1) Zuehl Farm Site: The first four principal components are represented on two separate ordination planes (Figure la). An arrow indicates the lowest level and a black dot the uppermost level for each plot. Coordinates of the first and second principal components can be grouped into four divisions. Coordinates of the third and fourth principal components represent more complicated distributions. A correlation coefficient matrix (Figure 1b) is separated into four divisions (solid horizontal lines) and two subdivisions (dashed lines). When the first four principal components are plotted stratigraphically to produce four different curves (Figure 1c), each curve shows its unique changing pattern. Combining the visually defined divisions of the four curves and and those of the correlation coefficient matrix, the pollen profile is is divided into eight zones.

(2) Pilot Mound Site: Both the ordination of the first and second principal components and the correlation coefficient matrix represent four distinct groups (Figure 2a,b). The zone boundries do not necessarily match between the two methods (Figure 2b,c). For instance, a boundary between zone 3 and zone 4 is established from the correlation coefficient matrix, while zone 4 and zone 5 are separated by Principal com-

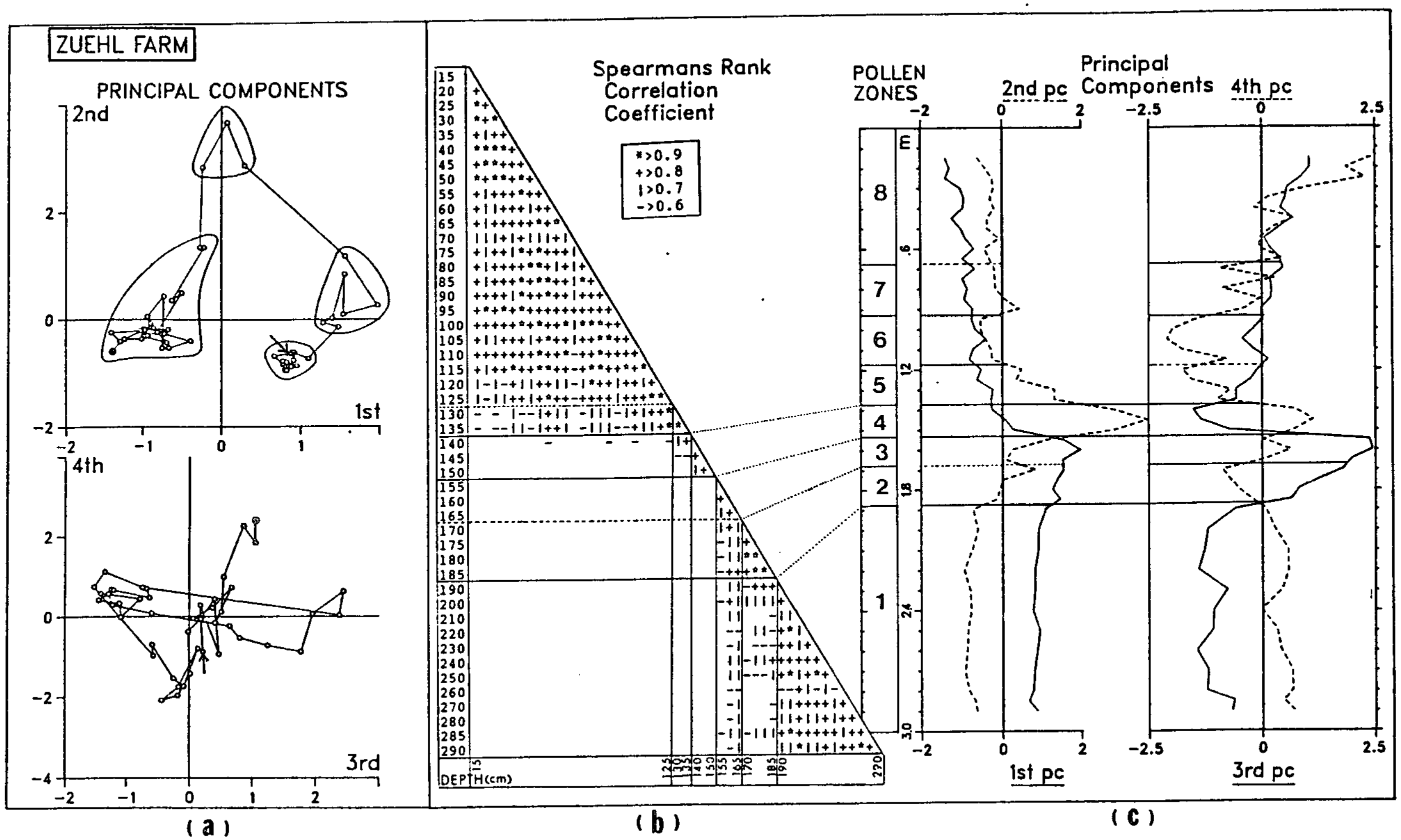

Fig. 1. 1. Zoning the Zuehl Farm Site data by Principal components analysis and Spearman's rank correlation coefficient.

a) First four principal components on two ordination planes, an arrow indicating the lowest level and a black dot the uppermost level for each plot

b) Spearman's rank correlation coefficient matrix: the values above 0.6 are expressed in four different symbols, and those below 0.6 in blanks.

c) Stratigraphic plots of the first four principal components, with pollen zones on the left.

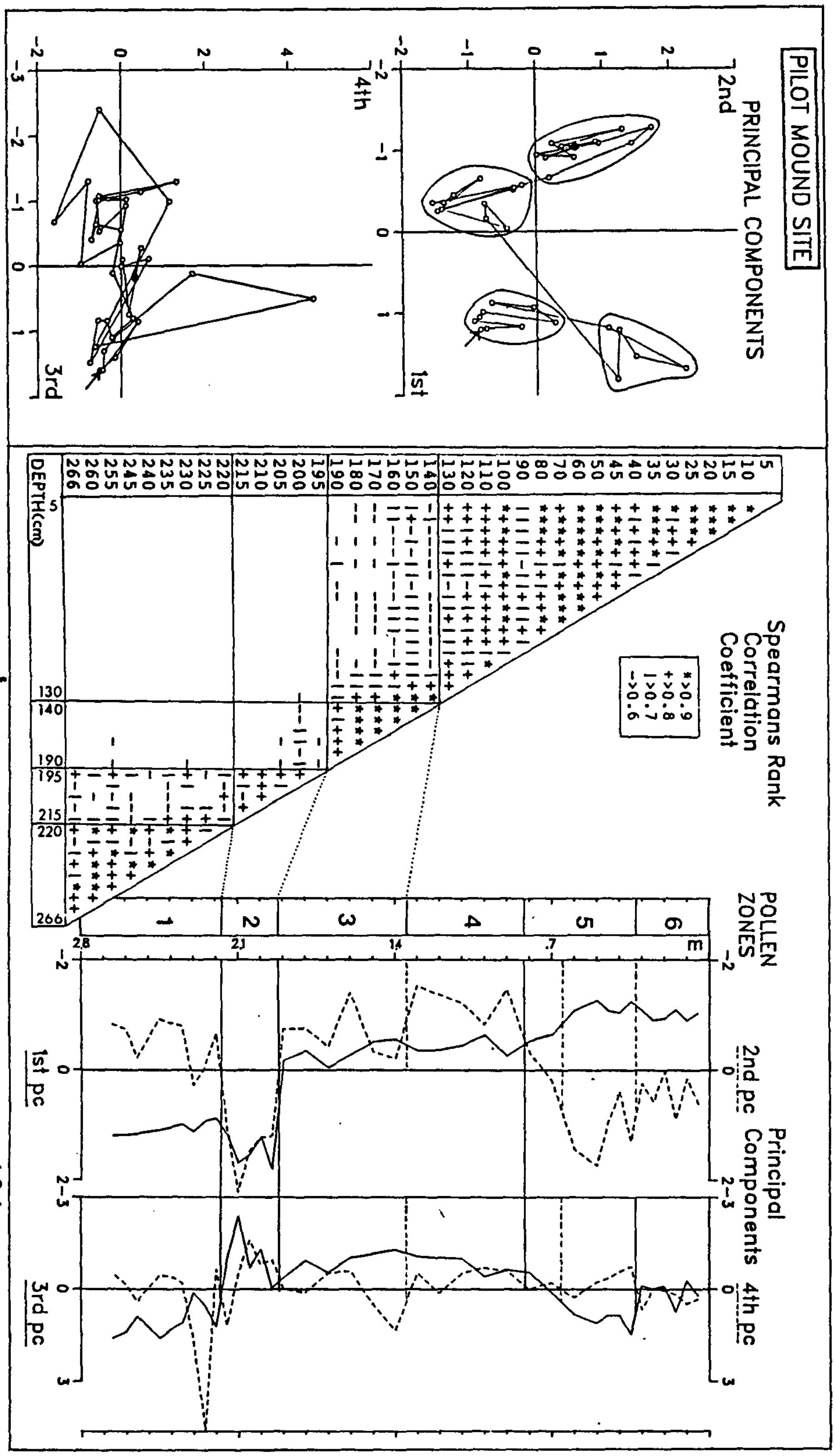

Fig. 2. Zoning the Pilot Mound Site data. Detailed explanations of the figure are same as in Figure 1.

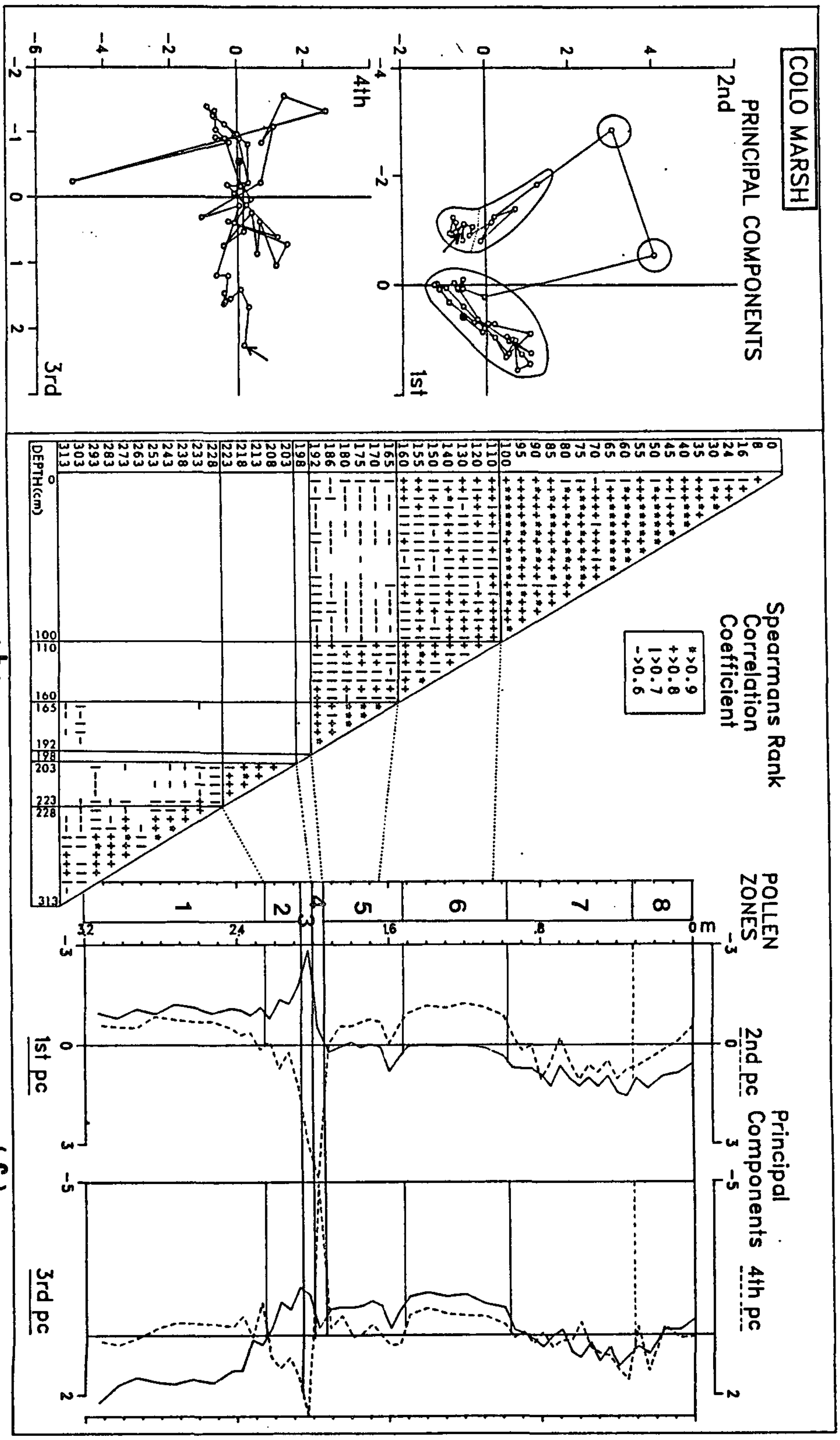

Fig. 3. Zoning the Colo Marsh data. Detailed explanations of the figure are same as in Figure 1.

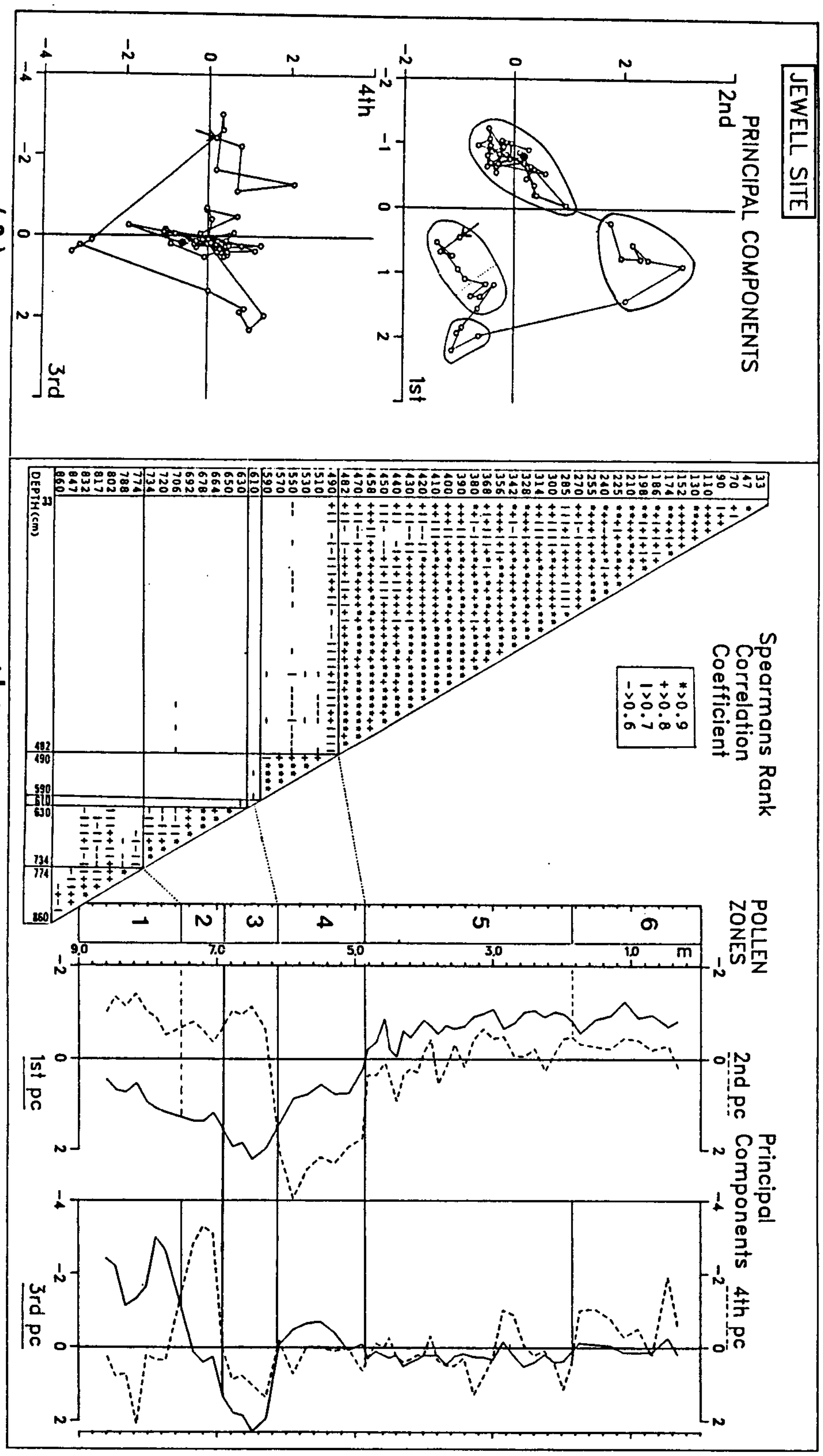

Fig. 4. Zoning the Jewell Site data. Detailed explanations of the figure are same as in Figure 1.

ponents analysis. The boundary of zone 5 and zone 6 is defined by the third and fourth principal components. Six pollen zones are thus created.

(3) Colo Marsh: The first and second principal components group the levels into four divisions, and the correlation coefficient matrix can be separated into six divisions (Figure 3a,b). The upper and lower boundries of zone 6 in the PCA deviate from those established in a correlation coefficient matrix (Figure 3b,c). The boundries in the PCA are adopted here because the curves of all four principal components consistently change at these boundries. The boundry between zone 7 and zone 8 is visually determined from the pollen diagrams; it is not conspicuous in either numerical method. Eight pollen zones are constructed.

(4) Jewell Site: The first and second principal components group the levels into four divisions, and the correlation coefficient matrix is grouped into five divisions (Figure 4a,b). A division at the level of 610cm in the correlation coefficient matrix is considered to be insignificant (Figure 4b,c). The third and fourth principal components divides the uppermost group into two (zone 5 and zone 6). The pollen profile is divided into six pollen zones.

CORRELATION OF POLLEN ZONES

Gordon and Birks (1974) devised several numerical methods of comparing pollen diagrams and constructed regional pollen zones that are defined solely on the basis of overall similarities in pollen composition. Birks and Berglund (1979) used a principal components ordination technique to compare two diagrams.

Table 2. Principal components analysis of pollen data at four sites, as combined in one.

	EIGENVALUE	DIFFERENCE	PROPORTION	CUMULATIVE
PRIN1	5.812355	1.705842	0.290618	0.290618
PRIN2	4.106513	2.014438	0.205326	0.495943
PRIN3	2.092075	0.250233	0.104604	0.600547
PRIN4	1.841841	0.784562	0.092092	0.692639

EIGENVECTORS

	PRIN1	PRIN2	PRIN3	PRIN4
PICEA	0.286443	-.213881	-.164104	0.350128
LARIX	0.275496	-.128380	-.137663	0.312991
ABIES	0.308402	0.028721	0.231219	-.062150
PINUS	-.236684	-.030034	0.313358	0.305350
FRAXINUS	0.237673	0.057684	0.109762	-.138752
ALNUS	0.273740	0.066950	0.374624	-.072479
BETULA	0.217336	0.120350	0.355655	-.072529
QUERCUS	-.156439	0.391444	0.011186	0.136062
ULMUS	0.039248	0.430882	-.172960	0.028164
OSTRYA/CARP.	0.194087	0.371495	0.078819	-.069303
CARYA	-.059993	0.369446	-.207588	0.168611
CORYLUS	0.226268	0.200377	0.363846	-.160932
JUGLANS	-.114108	0.371250	-.091329	0.134570
PLATANUS	-.269769	0.056017	0.216290	0.163300
ACER	0.101408	0.244240	-.205806	0.175525
CHENOPODS	-.244698	-.143529	-.103761	-.469970
GRAMINEAE	-.280543	0.025673	0.314960	0.175746
AMBROSIA	-.230471	0.044696	0.085165	-.078993
ARTEMISIA	0.157119	-.182878	0.100737	0.445338
TUBULIFLORAE	-.285167	-.053809	0.294970	0.205506

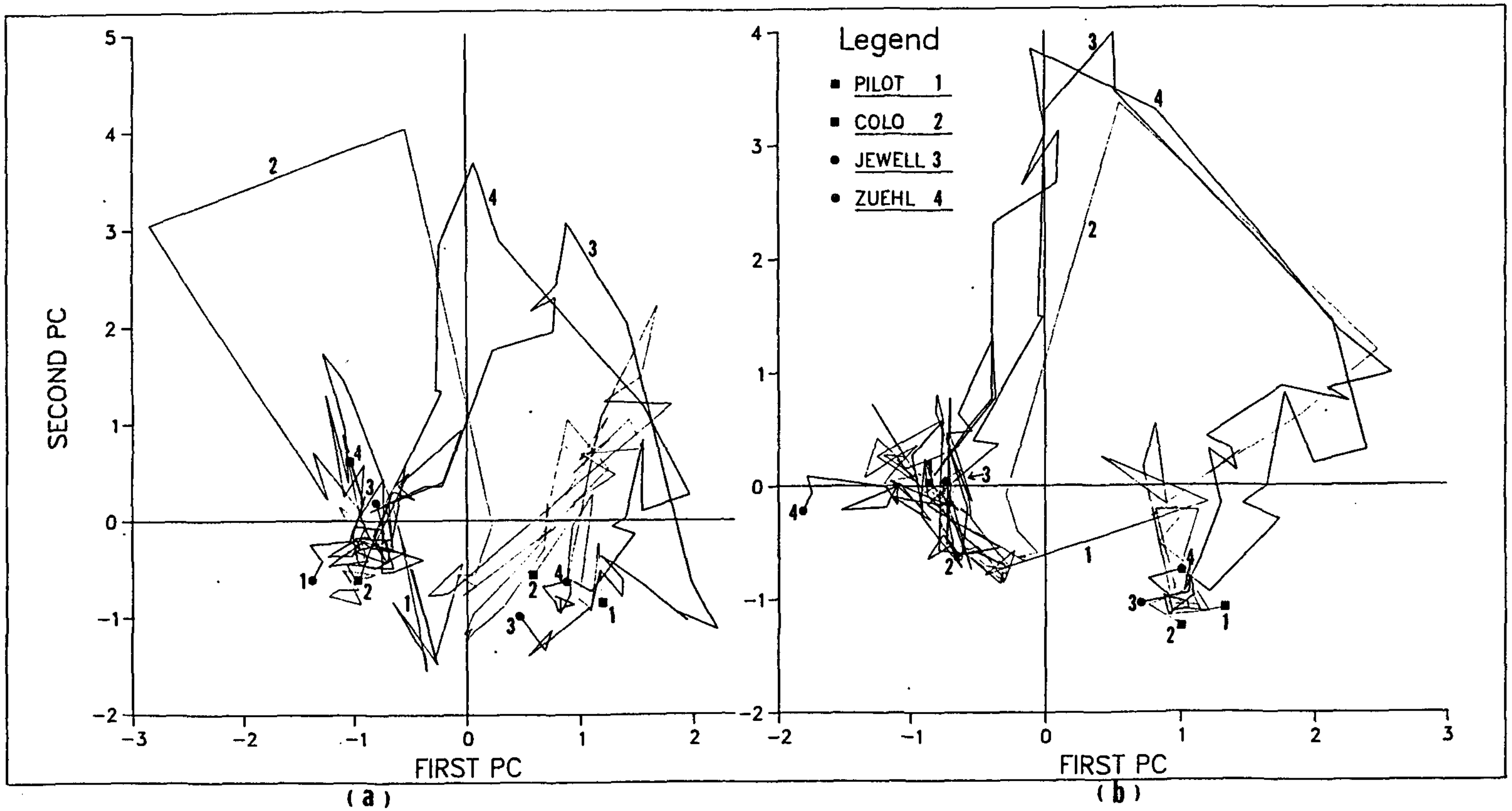

Fig. 5. The first and second principal components of the four data sets, configurated in one ordination plane.

a) Four paths are obtained from the separate numerical analyses of four data sets.

b) Four paths are obtained by a numerical analysis of four data sets combined in one and re-grouping into the original four sets.

Principal components analysis is used for the present purpose.

Figure 5a shows the configuration of the first and second principal components for the four data sets. Each set is obtained separately, and used for the zonation as described earlier. The figure indicates that the principal components at each site cannot be successfully combined for use in the correlation between the sites. The reason is apparent: each set of principal components is obtained on the basis of the variation of pollen data at each site, and the total variance would be different for the the different pollen data. Even if two levels at different sites contained exactly same pollen assemblages, they would be plotted at different positions in the ordination plane.

The correct this problem, four data sets are first combined and treated as if they were in one stratigraphic section. This process yields 188 levels from the four sites. Twenty taxa are used as variables (attributes), the same as for the Principal components analysis at each site. The first and second principal components account for abut 50% of the total variance, and the first four components account for 69% (Table 2). The pollen assemblage at each level, represented by the first and second principal components, is expressed as a point in the ordination plane. Thus, two identical pollen assemblages at different levels, whether or not they are from the same site, would share the same coordinates. After the calculations are completed, the levels are regrouped into the original four data sets, and plot-

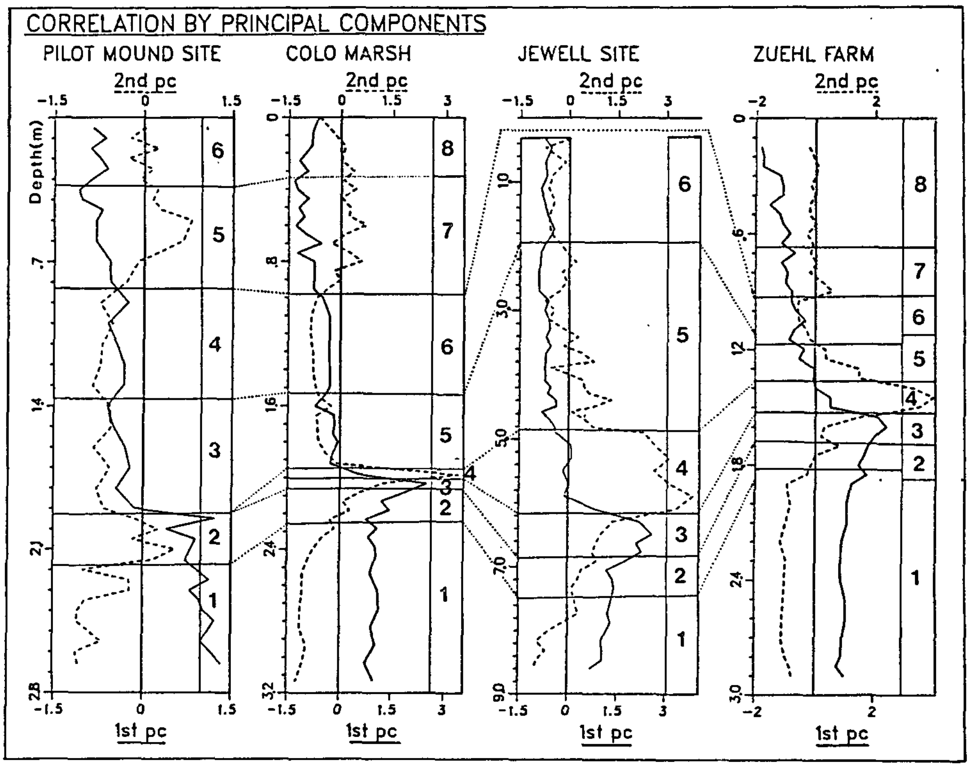

Fig. 6. Correlation of local pollen assemblage zones by the first and second components.
A solid curve represents the first principal component and a dashed curve is the second component for each site. Note that the correlation lines (dashed lines between the plot of each site) mostly match to the zone boundaries.

ted as four separate paths (Figure 5b). Each path nicely mathes the others, even though it is apparent that certain zones are absent from the Pilot Mound Site.

Stratigraphic plots of the first and second principal components from the four site are represented in Figure 6. The changing pattern of the curves is very similar between the sites, enabling a detailed correlation. Dashed lines in Figure 6 are the correlation lines which represent a biostratigraphic correlation.

DISCUSSION

It was shown that two different numerical methods are configurated and compared to sucessfully draw the zone boundaries. The correlation lines match the zone boundaries fairly closely, suggesting that local zones also have regional significance. With minor deviations in the zone boundaries, most local zones represent 'regional pollen assemblage zones' as defined by Cushing (1967). Seven regional pollen zones are established in central Iowa from local zones at the four sites.

It is noteworthy that correlation is conducted successfully with the first two principal components which include only 50% of total variance (Table 2). The efficiency of the first

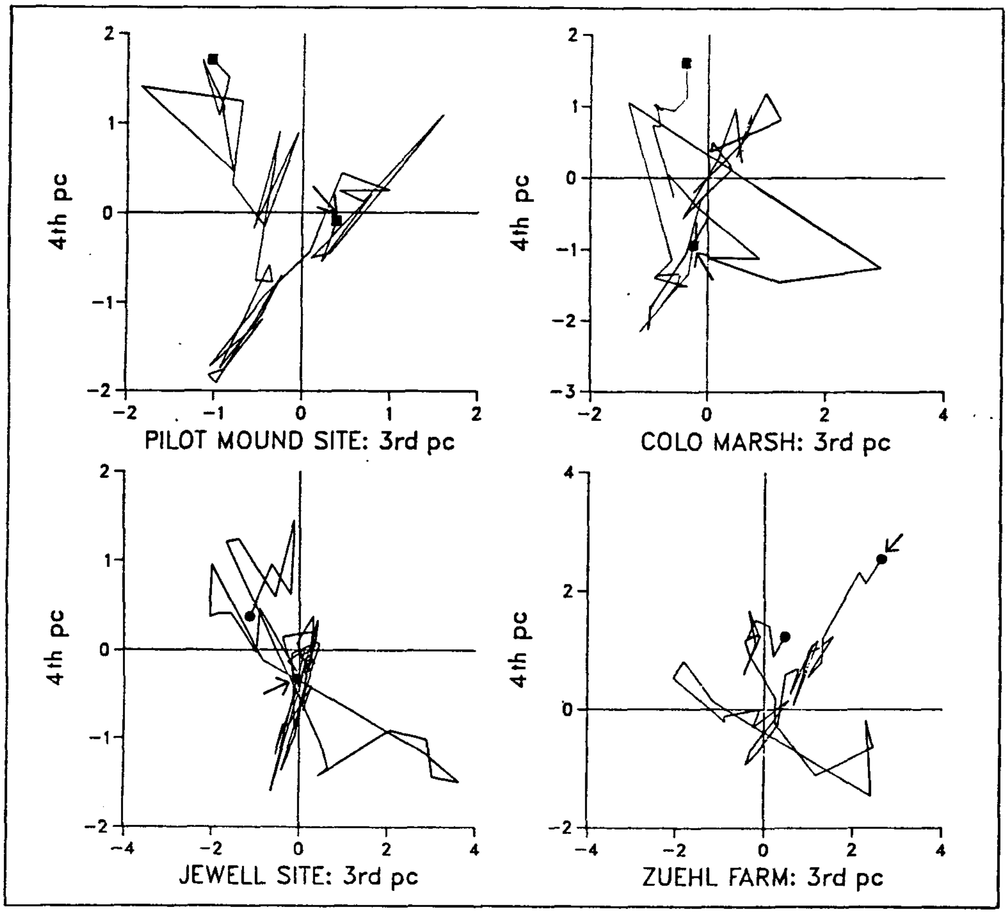

Fig. 7. The configuration of the third and fourth principal components of four data sets combined in one. Four separate plots are obtained by re-grouping the data into the original four data sets after the calculations are completed.

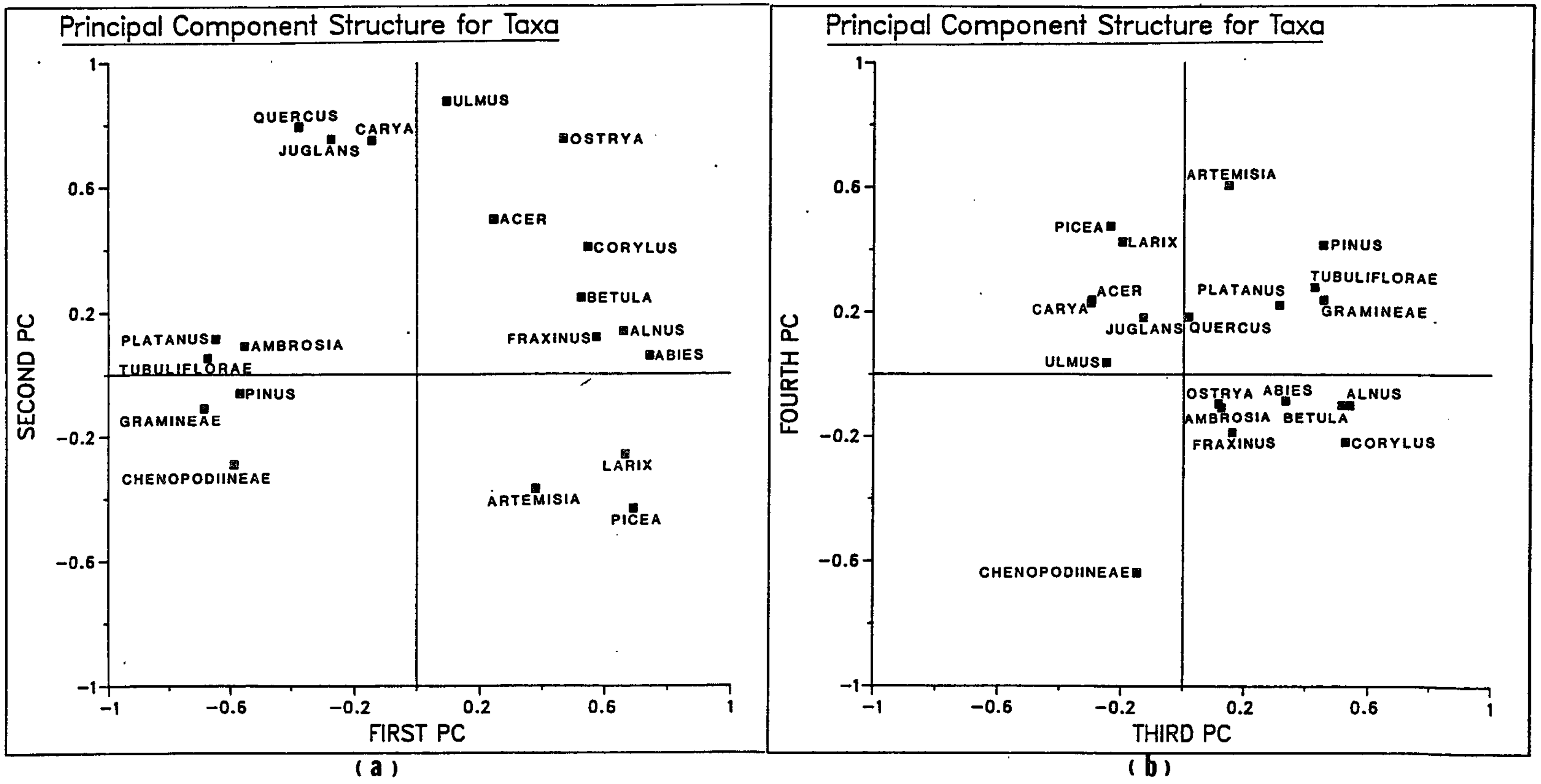

Fig. 8. The first four components of the principal component structure for taxa: a) first and second components and b) third and fourth components.

two components in correlation suggests that most of the variance that represents the similarity between sites in the pollen assemblages and their changing patterns is included in these components. In contrast, the distribution of the third and fourth principal components varies greatly between sites (Figure 7). The variability in the third and fourth components seems to be related to the unique features of each site.

The nature of the four principal components can be investigated by principal component structure, i.e., a measure of the contribution of the variables (taxa) to the principal components. It is defined as the statistical correlations between the taxa and the principal components. In Figure 8a and 8b, the distance of a taxon from the center of the axes is the proportion of its variance accounted for. Thus the tendency of taxa to be distributed far from the origin in the ordination plane of the first and second components indicates that all the taxa are largely accounted for by either or both of the first and second components of the PCA (Figure 8a).

The peripheral distribution of taxa is comparable to the distribution of the levels (and pollen zones as implied by groups of levels) in the ordination plane of the first and second principal components, although there is not a 1:1 mathematical correspondence between the two (Figure 5b). For instance, *Picea*, *Larix* and *Artemisia* in the lower right quadrant (Figure 8a) are comparable to the clustering of the lowermost levels (i.e., *Picea* zone) in the same quadrant (Figure 5b). The location of *Fraxinus*, *Abies*, *Alnus* and *Betula* in the upper right quadrant (Figure 8a) also matches the Picea-*Fraxinus* zone and *Alnus-Betula* zone in the same quadrant, and so on. This correspondance suggests that all the taxa contributed to some extent to create the similarity in the pollen assemblages and their changing patterns between sites.

The distribution of the third and fourth components of taxa are quite different in that most taxa are scattered around the center of the axes (Figure 8b). It indicates that each taxon contributed in quite different strength to the third and fourth principal components. It is apparent that the third and fourth principal components cannot be used for correlation, because the differential contribution of the taxa to these components reveals the dissimilarity in the pollen assemblage between sites.

REFERENCES

Birks, H.J.B., 1974. Numerical zonation of Flandrian pollen data. New Phytology, **73:** 351-358.

Birks, H.J.B., and Berglund, B.E., 1979. Holocene pollen stratigraphy of southern Sweden: a reapraisal using numerical methods. Boreas, **8:** 257-279.

Birks, H.J.B., and Gordon, A.D., 1985. Numerical Methods in Quaternary Pollen Analysis. Academic Press, 317pp.

Cushing, E.J., 1964. Redeposition in Late-Wisconsin pollen spectra from east-central Minnesota. Amer. Jour. Sci., **262:** 1075-1088.

Gordon, A.D., and Birks, H.J.B., 1972. Numerical methods in Quaternary paleoecology I. Zonation of pollen diagrams. New Phytology, **71:** 961-979.

Gordon, A.D., and Birks, H.J.B., 1974. Numerical methods in Quaternary paleoecology II. Comparison of pollen diagrams. New Phytology, **73:** 221-249.

Kim, H.K., 1986. Late-glacial and Holocene Environment in Central Iowa: A Comparative Study on pollen Data from Four Stes. Ph. D. Dissertation, Univ. of Iowa, U.S.A.

Maher, L.J., 1982. The palynology of Devils Lake, Sauk County, Wisconsin. In M.E. Ostrom (Ed.), Quaternary History of the Driftless Area. Univ. Wisconsin-extension, Geol. and Nat. Hist. Survey, Field Trip Guide Book, No. **5:** 119-125.

Ogden, J.G., III., 1977. Pollen analysis: state of art. Geographie Physique et Quaternaire, **31:** 151-159.

Pennington, W., and Sackin, M.J., 1975. An application of principal components analysis to the zonation of two late-Devensian profile. New Phytology, **75:** 419-453.

(Revised and Accepted: November 10, 1989)