ELSEVIER

Journal of Chromatography A, 686 (1994) 225–233

JOURNAL OF CHROMATOGRAPHY A

Statistical treatment of large digital chromatographic data sets

D.C. Freeman*, D.W. Byrd

Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA

First received 8 December 1993; revised manuscript received 8 June 1994

Abstract

Capturing vast amounts of digital chromatographic data is now routine, but it brings with it the onerous problems of data reduction, unification and analysis. The methods reported here provide a rapid means of storing, sorting, tabulating and analyzing large sets of chromatographic data.

1. Introduction

Static headspace gas chromatography (GC) is a sensitive method for analyzing the volatile and semivolatile hydrocarbons [1–4]. The method lends itself to automation, potentially allowing for the acquisition of large data sets with minimal labor. While headspace GC is a potent analytical tool, it has not been used extensively in large population studies. The correlation of retention time data across multiple samples is a significant problem that limits the ease whereby large data sets are accumulated and analyzed. Variability in retention time data among samples stems from human error in the manipulation of the samples as well as system related errors. Manual input of chromatographic data for statistical analysis is also a source of error. The automated static headspace sampling system and digital data storage system described in Byrd and Freeman [5] minimize both of these errors.

Here, we illustrate the assembly and analysis of large chromatographic data sets using the volatile and semivolatile terpene compounds produced by two subspecies of big sagebrush (*Artemisia tridentata* ssp. *tridentata*, i.e. basin big sagebrush, and *A. tridentata* ssp. *vaseyana*, i.e. mountain big sagebrush) and various hybrids between these two taxa [6]. Like most chromatographic studies we are interested in comparing mean concentrations of one or more compounds among various treatments, but in addition we are also interested in comparing the ability of these sagebrush taxa to regulate the production of volatile compounds among leaves of the same plant. The process of hybridization disrupts long established intergene coordination (co-adapted genes complexes), and thus the ability of a plant to regulate its body. We have investigated this disruption by comparing within plant similarity indices among plant populations.

The majority of volatile and semivolatile compounds produced by sagebrush are monoterpenes [7–14], i.e. stable volatile or semivolatile non-saponifiable lipids, linked to essential metabolic processes. We also examined the distributions and concentrations of sesquiterpenes, coumarins and auxins that, like terpenes, are derived from isoprene [11]. Terpenes and related compounds are believed to function as pollinator

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(94)00575-T

attractants, pesticides, phytotoxins, bacteriostats, predator defense compounds or insect larval abortives [7,8,10]. Several studies suggest that monoterpenes are taxa specific [6,8,10,12–15].

2. Methods

We examined a narrow hybrid zone in Salt Creek Canyon, UT, USA. There, basin big sagebrush occurs below 1740 m, while mountain big sagebrush occurs above 1860 m, with the hybrid zone found in between. Plants were collected from five sites, one from each of the two parental regions, and one from each of the three hybrid zones [6]. Twelve plants were sampled from each of the parental sites and from four plants in each of the hybrid sites. The plants were permanently marked and the same individuals were sampled in both the Spring and Fall (1991). For each season we collected two leaves per branch from two branches per plant. We also sampled two leaves from one flowering stalk (inflorescence) per branch in the Fall. The data set was derived from more than 400 samples. Since the concentrations of more than 100 compounds were determined using the protocols in Byrd and Freeman [5], the entire data set exceeded 40 000 data points. Manipulation of these data provided additional information (total hydrocarbons, percent area of a given component, mass of all components, mass of individual components and percent of area or mass of a component) and expanded the total data set to over 200 000 data points.

We used PeakSimple I software [16] to capture the chromatographic analog signals and converted them to an ASCII digital format. The digital storage consisted of a voltage file and a summary report file that included retention time, percent area, peak height, peak area and internal standard information. These summary files were imported directly to a spreadsheet (Lotus 2.1 [17]). These data were transposed for statistical procedures so that each sample represented a row of the matrix rather than a column, with retention time windows shown in columns. This step was repeated for each sample. Further reduction of data was made on this data set by selecting specific time windows from the data set.

3. Statistical analyses

Below, we use a subset of our data to illustrate several parametric and non-parametric statistical procedures that may be used to analyze chromatographic data, comment on the limitations of various tests and report results for the whole data set (for details see [18–25]. Parametric tests utilize the variance in testing for differences among the means of treatments. Two assumptions must be met: (1) the data must be normally distributed and (2) the variances must be the same for each treatment. If both these assumptions are valid, then these statistically powerful parametric procedures can be utilized. If these assumptions are not met and the data cannot be transformed so that the assumptions hold, then one should use the less powerful non-parametric procedures.

3.1. Comparing means among treatments

Parametric procedures

To illustrate the use of parametric procedures, we have compared the concentration of a compound with a retention time of 7.92 min among the sagebrush taxa using a simple oneway analysis of variance (Table 1) and a multiple range test (Table 2). The former test determines if there is a significant difference among the means, but does not determine which means are different from one another. A oneway analysis of variance (also known as a fixed effect or Model 1 analysis of variance) using all the data for peak 7.92 showed that the concentration differed significantly among the sites ($F_{4,125} = 13.86$, $P < 0.0001$). The a posteriori Student–Newman–Keuls multiple range test does determine which means differ significantly from each other. The Student–Newman–Keuls multiple range test (Table 2) showed that the concentrations at site 1 ($\bar{X} = 0.238 \pm 0.2053$) differed significantly from the concentrations at all other sites. However, sites 2, 3, 4 and 5 did not differ

Table 1
Data from one leaf per plant for four plants from each of the five sites used to illustrate the procedures of a oneway analysis of variance

Terpene concentration from sagebrush plants growing at five different sites, data are from the Fall of 1991					
Site	1.000	2.000	3.000	4.000	5.000
Data	0.764	0.021	0.114	0.000	0.138
	0.790	0.032	0.148	0.000	0.119
	0.812	0.030	0.130	0.000	0.113
	0.692	0.029	0.127	0.090	0.112
Oneway analysis of variance					
Total	Site 1	Site 2	Site 3	Site 4	Site 5
Mean $\bar{X}$	0.764	0.028	0.130	0.023	0.120
Sum	3.058	0.112	0.519	0.090	0.482
$(\sum\chi)^2/N$	2.338	0.003	0.067	0.002	0.058
$\sum(\chi^2)/N$	0.586	0.001	0.017	0.002	0.015
Total $\sum_i \sum_j \chi_{i,j}$	4.261				
Total sum of squares $\sum_i \sum_j \chi^2 - C$	2.346	0.003	0.068	0.008	0.058
$C[(\sum_i \sum_j \chi_{i,j})^2]/N$	0.908				

Total sum of squares = $T = \sum_i \sum_j \chi^2_{i,j} - C$	1.576	
Group sum of squares = $G = \sum_i [(\sum_j \chi_{i,j})^2]/N_i - c$	1.560	Error MS = E/DF = 0.015/15 = 0.001 Group MS = G/DF = 1.560/4 = 0.390
Error sum of squares = $E = T - G$	0.015	F = group MS/error MS = 390.000

The F value is compared against tabulated using the degrees of freedom among groups (one less than the number of groups) and the error degrees of freedom. In this case there are 20 data points. One degree of freedom is assigned to the mean, four to the groups and so there are 15 degrees of freedom for error.

significantly from one another. The means for these sites were 0.0472 ± 0.0535, 0.0266 ± 0.0531, 0.0106 ± 0.0299 and 0.0794 ± 0.1266, respectively. Despite the appealing result from the oneway analysis of variance its use is not appropriate for this data set. These chromatographic data have unequal variances among sites (as determined by Bartlett's test $F = 20.622$; see Table 3), and the data are not normally distributed (as determined by the Kolmogorov–Smirnov goodness of fit test; see [18]). Therefore one should not use the parametric procedure but should instead use a non-parametric Kruskal–Wallis procedure (Table 4).

Non-parametric procedures

The non-parametric Kruskal–Wallis test does not have the stringent assumptions of normality or homogeneity of variance. We have illustrated its use in Table 4 where we compare multiple means among treatments. Again, the results show a significant difference among the sites when all the data for the compound corresponding to peak 7.92 were used ($\chi^2 = 41.90$).

Table 2
Student–Newman–Keuls multiple range test

$SE = \sqrt{(S^2/2)(1/n_a - 1/n_b)}$; $q = (\bar{X}_b - \bar{X}_a)/SE$; S^2 is the error mean square from Table 1

Comparison	Difference $E\bar{X}_b - \bar{X}_a$	SE	q	Critical value of q	Conclusion
1 vs. 3	0.634	0.0158	40.126	4.367	$P < 0.05$
1 vs. 5	0.644	0.0158	40.759	4.367	$P < 0.05$
1 vs. 2	0.736	0.0158	46.582	4.367	$P < 0.05$
1 vs. 4	0.741	0.0158	46.898	4.367	$P < 0.05$
3 vs. 5	0.010	0.0158	0.633	4.367	$P > 0.05$
3 vs. 2	0.102	0.0158	6.456	4.367	$P < 0.05$
3 vs. 4	0.107	0.0158	6.772	4.367	$P < 0.05$
5 vs. 2	0.092	0.0158	5.822	4.367	$P < 0.05$
5 vs. 4	0.097	0.0158	6.139	4.367	$P < 0.05$
2 vs. 4	0.005	0.0158	0.316	4.367	$P > 0.05$

This test is used after an analysis of variance to determine which means differ from one another. See Table 1 for the analysis of variance. In performing this procedure the means must be ranked first. Critical values of q are obtained from the statistical tables.

Multivariate procedures

If the samples being analyzed contain more than one compound of interest, then the concentration of compound A in each sample is not independent of the concentration of compound B, and one must use a multivariate analysis of variance (MANOVA) as the parametric procedure. The MANOVA test carries with it all the assumptions of analysis of variance and the same tests are used to determine if the assumptions

Table 3
Bartlett's test, used to determine if the variances are homogenous among the various treatment groups

Sites	1.000	2.000	3.000	4.000	5.000
Data	0.764	0.021	0.114	0.000	0.138
	0.790	0.032	0.148	0.000	0.119
	0.812	0.030	0.130	0.000	0.113
	0.692	0.029	0.127	0.090	0.112
Mean	0.764	0.028	0.130	0.023	0.120
$SS = \sum(\bar{X}_1 - X_1)^2$	0.0082	0.000	0.0006	0.0061	0.0004
ν_i	The degree of freedom, i.e., $n_i - 1$				
ν_i	3.000	3.000	3.000	3.000	3.000
$s_i^2 = ss_i/\nu_i$	0.0027	0.0000	0.00012	0.002	0.0001
$\log s_i^2$	−2.5686	−4.6320	−3.6990	−2.6990	−4.0000
$\nu_i \log s_i^2$	−7.7059	−13.8961	−11.0969	−8.0969	−12.0000
$s_p^2 = \sum SS_i / \sum \nu_i$	0.0008		$\log s_p^2$	−3.1144	
B	$B = \ln(s_p^2) \cdot \left(\sum \nu_i\right) - \sum \nu_i \ln s_i^2$				
B	−17.9505				
C	$C = 1 + [1/3(k-1)]\left[\left(\sum_i 1/\nu_i\right) - 1/\sum \nu_i\right]$				
C	1.1333				
$B_c = B/C$	−15.8387		$P < 0.005$		

Data are the same as used in Table 1. B_c is compared to a χ^2 distribution. In this case the variances are not homogenous and one should use a non-parametric test as illustrated in Table 4.

Table 4
The Kruskal–Wallis test, a non-parametric procedure for comparing three or more means; it uses ranked data

Sites	1.000	2.000	3.000	4.000	5.000
Ranks (R_i)	18.000	4.000	11.000	2.000	15.000
	19.000	7.000	16.000	2.000	12.000
	20.000	6.000	14.000	2.000	10.000
	17.000	5.000	13.000	8.000	9.000
n_i	4.000	4.000	4.000	4.000	4.000
$\sum$rank (R_i)	74.000	22.000	54.000	14.000	46.000
R_i^2/n_i	1369.000	121.000	729.000	49.000	529.000
$\sum T = \sum(t_i^3 - t_i)$	$(3^3 - 3)$	t_i refers to the number of ties. In this case three numbers were tied			
$C = 1 - \left[\left(\sum T\right)/(N^3 - N)\right]$	0.9970				
$H = 12/[N(N+1)] \cdot \sum_i R^{2_i}/n_i - 3(N+1)$	16.914				
$H_c = H/C$	16.9650	$P < 0.005$			

The analysis is performed on the ranks and not the data itself. The data are given in Table 1. The distribution of H_c is approximated by that of a χ^2 with one degree of freedom less than the number of groups being compared.

are met. Some authors mistakenly use multiple "t" test (or other tests such as correlations —χ^2, etc.) for such comparisons, but this causes a build up of type 1 errors, i.e. the probability of rejecting a true hypothesis. The null hypothesis in most cases is that there is no difference among the treatments. So the use of multiple tests leads to the reporting of significant differences where none may exist. If one chooses to use multiple "t" tests one can correct for the build up of type 1 error by using the sequential Bonferroni test (see [19]).

To distinguish sources of variation in the data set, we recommend principal component analysis. The method can be used to obtain a visual display of multivariate data. Principal component analysis requires the determination of new axes in the multivariate space formed by the original variables. These new axes are usually generated using least squares regression techniques [20,23,26,27]. The first component (new axis) is that line which accounts for the greatest variation. The second axis is that line which is perpendicular to the first axis and accounts for the greatest amount of the remaining variation among all the possible lines that could be drawn perpendicular to the first line. The third component must be perpendicular to both the first and second components and similarly accounts for the greatest amount of the remaining variation (i.e. it is also chosen from among all the lines that are mutually perpendicular to the first and second components). The procedure can be carried on until all the variation is accounted for. Because each principal component is in fact a linear combination of the various chemical concentrations, one can compute the principle component score of each sample. These scores can then be plotted against each other to determine if the samples came from one, or more than one, population. We have done this for our terpene data. In Fig. 1 we have plotted the first versus the second principal component for terpenes sampled in the Fall for the two parental taxa. The results clearly show that the terpene profile of basin big sagebrush differs from that of mountain big sagebrush. Hybrids form intermediate distributions when compared to the parentals. While a number of pretreatments of the data are possible and influence the data, for this analysis we did not pretreat the data. Principal component analysis assumes linearity; if this assumption does not hold, there are non-linear ordinations that can be used, e.g. detrended principal components [20] or non-metric multidimensional scaling [21].

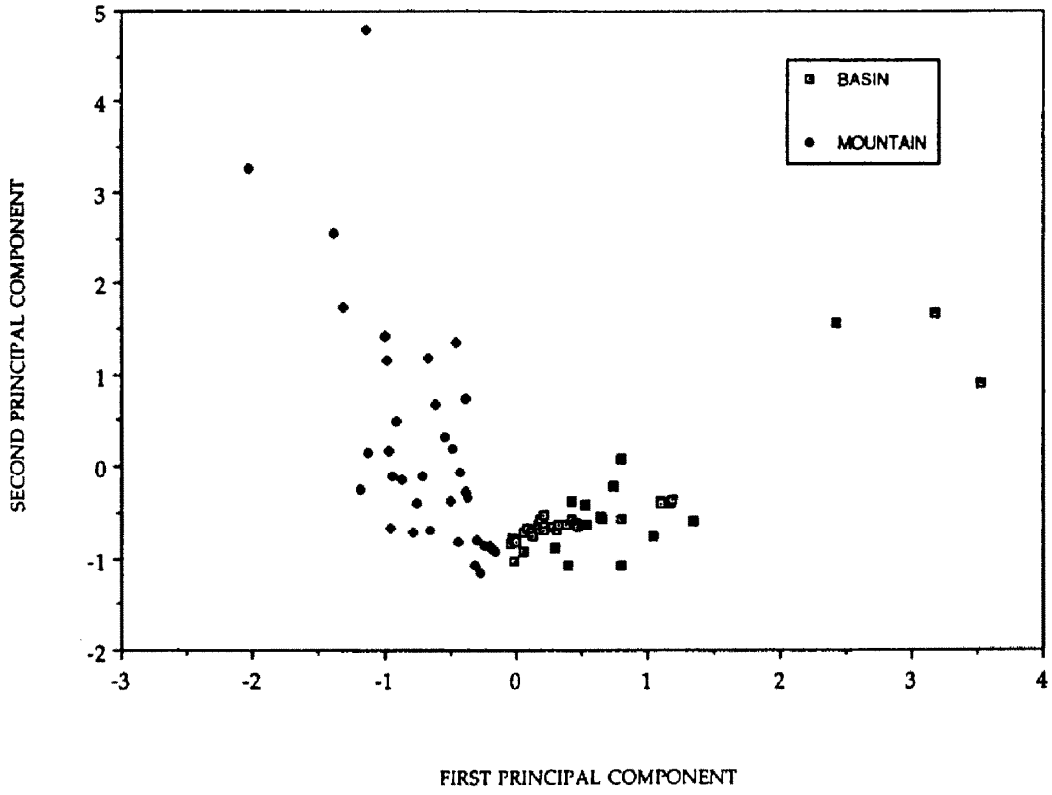

Fig. 1. Fall principal component analysis. First versus second component.

3.2. Within individual variation as a measure of developmental control, i.e. stress

Chromatographic data may also be used to determine if organisms are under stress. The premise for this is rather simple. Organisms that are not stressed tightly regulate the development of their own bodies, while stress interferes with this regulation. By examining this within individual variation in concentration or proportions of compounds, one can assess the degree of stress organisms are under [22,24].

One can use either a parametric approach, i.e. analysis of variance, or a non-parametric approach for this analysis. In our case, the parametric approach involves computing a within plant variance for the concentration or proportion of one or more compounds, and determining if the within plant variance differs among the taxa of sagebrush. Our full data set fails to meet the assumptions for the analysis of variance. We have opted instead to use various non-parametric Euclidean distance measures.

Distance measures

To compute the Euclidean distance, let A_1 be the concentration of one compound and B_1 be the concentration of a second compound taken from sample 1. Similarly, A_2 and B_2 are the concentrations found in sample 2 from the same individual (i.e. in our case multiple leaves on the same plant). Now, we simply plot the concentrations in Euclidean space where compound A represents the x axis and compound B represents the y axis. The formula for the Euclidean distance (ED) is given in Eq. 1.

$$\mathrm{ED}_{jk} = \sqrt{(X_{ij} - X_{ik})^2} \quad (1)$$

Each sample represents a point in this Euclidean space, and the distance between two points from the same individual represents a measure of

the dissimilarity, thus is an indicator of the degree of stress individuals are under [22–24]. These distances can then be analyzed using either analysis of variance or Kruskal–Wallis tests.

The results of our study show that mountain big sagebrush is less able to regulate the production of terpenes than either basin big sagebrush or the hybrids (Tables 5 and 6). The Euclidian distance described above will increase as the number of compounds is increased. Ludwig and Reynolds [23] have reviewed a large number of resemblance functions and advocated dividing the total Euclidian distance by the number of compounds. They refer to this as the mean absolute Euclidean distance (MAD, Eq. 2).

$$\mathrm{MAD}_{jk} = \sum |X_{ij} - X_{ik}| / S \qquad (2)$$

In our case the taxa also differ significantly for this distance measure as well (Tables 5 and 6).

Distance measures can also be used to compare the proportions of compounds among sam-

Table 5
Measure of within-plant similarity using resemblance functions computed with data from the Spring

	Basin	Hybrid	Mountain
Euclidean distance between leaves from the same branch	0.39 0.26	0.27 0.29	0.41 0.37
Euclidean distance among leaves from different branches	0.38 0.22	0.30 0.27	0.43 0.36
Chord distances between leaves from the same branch	1.30 0.15	1.26 0.11	1.22 0.19
Chord distances among leaves from different branches	1.29 0.15	1.27 0.10	1.22 0.18
Mean absolute distance between leaves from the same branch*	0.90 0.59	0.58 0.52	0.82 0.68
Mean absolute distance among leaves from different branches*	0.91 0.54	0.66 0.50	0.83 0.64
Jaccard index for leaves from the same branch	0.77 1.48	0.58 0.26	0.51 0.25
Jaccard index for leaves from different branches	0.49 0.26	0.55 0.26	0.48 0.22
Diversity index for leaves from the same branch	1.86_a 0.48	1.92_a 0.45	1.54_b 0.34
Diversity index for leaves from different branches	1.96_a 0.51	1.97_a 0.46	1.48_b 0.40

* Taxa differ significantly from one another at P 0.05.
[a,b] Means subscripted by the same letter do not differ significantly; means subscripted by different letters differ significantly at $P < 0.05$.

Table 6
Measure of within plant similarity using resemblance functions computed with data from the fall

	Basin	Hybrid	Mountain
Euclidean distance between leaves from the same branch	0.11_a 0.07	0.26_b 0.20	0.41_c 0.29
Euclidean distance among leaves from different branches	0.23_a 0.10	0.36_b 0.22	0.62_c 0.20
Chord distances between leaves from the same branch	1.21_a 0.13	1.29_b 0.08	1.30_b 0.09
Chord distances among leaves from different branches	1.23_a 0.13	1.31_b 0.07	1.36_b 0.12
Mean absolute distance between leaves from the same branch*	0.24_a 0.15	0.59_b 0.43	0.91_c 0.64
Mean absolute distance among leaves from different branches*	0.53_a 0.25	0.84_b 0.49	1.47_c 0.46
Jaccard index for leaves from the same branch	0.79_a 0.14	0.68_b 0.13	0.56_b 0.22
Jaccard index for leaves from different branches	0.70_a 0.16	0.53_b 0.22	0.53_b 0.16
Diversity index for leaves from the same branch	$1.85_{a,b}$ 0.40	2.06_b 0.39	1.74_a 0.36
Diversity index for leaves from different branches	1.70_a 0.40	2.02_b 0.37	1.70_a 0.34

Footnotes as in Table 5.

ples from the same individual. To do this one projects the data on to a unit circle. The length of the chord separating the data points (chord distance, CHD; Eq. 3) is the measure of dissimilarity [23,25].

$$\mathrm{CHD}_{jk} = \sqrt{2(1 - \mathrm{ccos}_{jk})} \tag{3a}$$

$$\mathrm{ccos}_{jk} = \sum(X_{ij}X_{ik})/\sqrt{\sum X_{ij}^2 \cdot \sum X_{ik}^2} \tag{3b}$$

Our data again showed that the mountain big sagebrush was less able to regulate the proportions of volatile and semivolatile compounds than the other two taxa (Table 5 or 6).

Much can be learned by examining the presence or absence of a compound. Jaccard's index (JI, see [23]) is based upon such data and simply examines the percent of compounds that are common to all samples. We have computed the percent of compounds in common among two leaves from the same branch.

$$\mathrm{JI} = a/(a + b + c) \tag{4}$$

where a = number of compounds made by both leaves, b = number of compounds made by leaf 1 that are not produced by leaf 2, c = number of compounds produced by leaf 2 that are not produced by leaf 1.

Surprisingly, not all leaves make all the compounds the plant is capable of making. Once again, mountain big sagebrush was less able to regulate its physiology than either of the other taxa (Tables 5 and 6).

Using the Shannon–Weaver information index (entropy H', Eq. 5) we can also examine the information (entropy) content of a given sample. H' increases as the number of compounds increase and as the concentrations of the compounds become more uniform.

$$-H' = -\sum p_i \ln p_i \tag{5}$$

where p is the frequency of the component measured.

This information provides an idea of the complexity of the chemical environment that must be dealt with by both predators and pathogens. Our results show that the information content varied significantly among the groups.

3.3. Sample size

The sample size required varies with the test being used. In general the greater the number of samples taken, the greater the confidence in the mean. For parametric procedures one can use the procedure given in Ref. [23] where $n = (s^2 t^2_{\alpha(2),n-1} F_{\beta(1),(n-1,\nu)})/d^2$, where s^2 is the sample variance from a pilot study, where $t^2_{\alpha(2),n-1}$ refers to the tabulated value of a two-tailed t distribution with $n-1$ degrees of freedom, $F_{\beta(1),(n-1,\nu)}$ is the tabulated value of an F distribution with $n-1$ and ν degrees of freedom, ν is the degrees of freedom from the pilot study and d is the half width of the desired confidence interval. The n in the above equation is the desired sample size and the equation must be solved by iteration. In our work, we have found greater variability among plants than within plants. Thus, one needs to sample many more plants per treatment than leaves within a plant. We have found that 10 plants per treatment and four leaves per plant (total of 40 samples per treatment) is adequate to discriminate among the sagebrush taxa.

4. Conclusions

We have provided a conceptual approach and background for statistical interpretations of large amounts of digitally stored chromatographic data in a routine and efficient manner. Examples have been provided to demonstrate the transformation of GC data into a format compatible for direct importation into statistics software programs and guidance on the proper use of statistical tools. We have demonstrated that the coupling of automated GC procedures and digital storage of data provide a viable method for the accumulation of large sets of data that can be subjected to statistical analysis. The approach used in this study is applicable to other chromatographic analytical techniques which provide for the digital storage of retention data.

References

[1] E. Drozf, *Anal. Chem.*, 16 (1973) 1132.
[2] R.G.S. Bidwell, *Plant Physiology*, MacMillan, New York, 1974.
[3] J.D. Cedarleaf, B.L. Welch and J.D. Brotherson, *J. Range Manage.*, 36 (1983) 492.
[4] J.A. Corkill, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 11 (1988) 211.
[5] D.W. Byrd and D.C. Freeman, *J. Chromatogr.*, 686 (1994) 235.
[6] D.C. Freeman, W.A. Turner, E.D. McArthur and J.H. Graham, *Am. J. Botany*, 78 (1991) 805.
[7] R. Croteau, *Chem. Rev.*, 67 (1987) 829.
[8] S.O. Duke, R.N. Paul, Jr. and S.M. Lee, *Biologically Active Natural Products*, (*ACS Symposium Series*, No. 380), American Chemical Society, Washington, DC, 1988, p. 318.
[9] W.W. Epstein and L.A. Gaudioso, *Phytochemistry*, 23 (1984) 2257.
[10] R.G. Kelsey, *Symposium on the Biology of Artemisia and Chrysothamus; General Technical Report INT-200*, Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, UT, 1984, pp. 375–388.
[11] A.T. Lihninger, *Biochemistry*, Worth Publishers, New York, 1979.
[12] R.G. Kelsey, J.R. Stephens and F. Sharizadeh, *J. Range Manage.*, 35 (1982) 617.
[13] E.D. McArthur, B.L. Welch and S.C. Sanderson, *J. Heredity*, 79 (1988) 268.
[14] B.L. Welch and E.D. McArthur, *J. Range Manage.*, 34 (1981) 380.
[15] E.D. McArthur, in K.L. Johnson (Editor), *Utah Shrub Ecology Workshop I*, Utah State Univ., Logan, UT, 1983, p. 3.
[16] *PeakSimple I*, SRI Co., Torrence, CA, 1988.
[17] *Lotus 123*, Lotus Development Corp., Cambridge, MA, 1985.
[18] J.H. Zar, *Biostatistical Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
[19] W.R. Rice, *Evolution*, 43 (1989) 223.
[20] H.G. Gauch, *Multivariate Analysis in Community Ecology*, Cambridge University Press, New York, 1982.
[21] M.J.R. Fasham, *Ecology*, 58 (1977) 551.
[22] D.W. Byrd, *Master's Thesis*, Department of Biology, Wayne State University, Detroit, MI, 1992.
[23] J.A. Ludwig and J.F. Reynolds, *Statistical Ecology*, Wiley, New York, 1988.
[24] D.C. Freeman, J.M. Emlen and J.H. Graham, *Genetica*, 89 (1993) 97.
[25] E.C. Pielou, *The Interpretation of Ecological Data*, Wiley, New York, 1984.
[26] M.H. Kaspar and W.H. Ray, *AIChE J.*, 38 (1992) 1593.
[27] S.T. Balke, *Quantitative Column Liquid Chromatography: A Survey of Chemometric Methods*, Elsevier, Amsterdam, 1984.