Applications of Canonical Discriminant Analysis in Horticultural Research

J.G. Cruz-Castillo¹, S. Ganeshanandam², B.R. MacKay³, G.S. Lawes⁴, C.R.O. Lawoko², and D.J. Woolley⁴
Department of Plant Science, Massey University, Palmerston North, New Zealand

Horticultural research often focuses on narrowly defined sectors of highly integrated systems that involve many intercorrelated variables. Despite tacit acknowledgment of such systems, scientists often present results of each variable separately. This approach ignores relationships between variables and their multivariate structure, and erroneously promotes a univariate view of the system under examination. While there is an obvious practical limit to the number of variables that can be examined, important information regarding the holistic interaction between the dependent and independent variables is overlooked. This not only represents lost opportunities, but also an inefficient use of resources during execution of the research.

Multivariate statistical techniques allow simultaneous analysis of multiple measurements of the individuals being investigated (Hair et al., 1987). While the original derivations of most of these techniques occurred about 50 years ago, the comparatively recent advances in computer power and capacity have revolutionized multivariate statistical analysis and it is now applied in many disciplines. In the early 1980s, there was little enthusiasm for using these techniques to analyze horticultural experiments (Swallow, 1981), but recently they have been used to interpret responses to treatments (Lawes et al., 1990; Perez-Gonzales, 1992; Yourstone and Wallace, 1990).

Canonical discriminant analysis (CDA) is a multivariate statistical technique that can identify differences among groups of individuals (or treatments) and improve understanding the relationships among the variables measured within those groups. Fisher (1936) was the first to use the technique, also known as Fisher's discriminant analysis (Mardia et al., 1979) or canonical variate analysis (Everitt,

1978). Groups are usually defined *a priori* by some criterion external to the set of variables measured, e.g., rootstocks (Cruz-Castillo et al., 1992), carbohydrate status (Prins et al., 1990), supermarkets selling flowers (Behe et al., 1992), tree species (Majer et al., 1992), or swede (*Brassica napus* L., Napobrassica Group) cultivars (Cole and Phelps, 1979). The objective of this paper is to present a description of CDA and demonstrate its applicability and potential in horticultural research.

UTILITY OF CDA

CDA determines how best to separate or discriminate two or more groups of individuals, given quantitative measurements of several variables on these individuals. For example, in studying storage disorders of fruit, those variables most strongly associated with a particular disorder, and their relationships with each other, can be assessed. CDA finds linear functions of quantitative variables that maximally separate two or more groups of individuals while keeping variation within groups as small as possible. This approach distinguishes several uncorrelated canonical discriminant functions (CDF) or canonical variables. These are linear combinations of the original variables that best separate the means of groups of observations relative to within-group variation (Rencher, 1992). The maximum number of CDFs is equal to the number of variables or one less than the number of groups, whichever is smaller. The first CDF, CDF₁, yields the maximum possible variation between groups with respect to within-group variation, reflecting group differences to the greatest degree possible. CDF₂ reflects group differences not displayed by CDF₁, with the condition of no correlation between CDF₁ and CDF₂. Similarly, CDF₃ is not correlated with CDF₁ and CDF₂; it reflects group differences not displayed by CDF₁ and CDF₂, and so on (Manly, 1986). Absence of correlation means that each CDF extracts a unique dimension of information from the data set.

In performing CDA, the hope is that the first few CDFs will account for almost all differences between groups. This approach enables a simple graphical representation of the relationships between groups by plotting the values of these CDFs (canonical scores) for sample observations. Besides identifying outliers in the data (Afifi and Clark, 1990), these plots may also be used to assign a new observation with unknown "grouping" to an existing group. The canonical scores of the new individual are calculated for the first two (or few) CDFs and its position plotted. The new individual is assigned to the group whose mean is closest to its position. This is an

empirical approach to allocating individuals with unknown grouping; discriminant analysis provides a more formal procedure [see Krzanowski (1988) for details].

Principal component analysis (PCA) is another multivariate statistical technique occasionally used in horticultural research (Broschat, 1979; Cruz-Castillo et al., 1991; Iezzoni and Pritts, 1991). It is important to recognize the practical differences between PCA and CDA to avoid confusion when using either technique. PCA is similar to CDA, but it ignores the group structure of the data (or individuals). Like CDA, PCA aims to find fewer dimensions (preferably two or three) than the original number of variables to explain the total variation in the data. PCA calculates linear combinations of the original variables, termed "principal components," that are orthogonal (noncorrelated) to each other, which maximizes variation among the original variables. In contrast, CDA maximizes variation between the groups of individuals while minimizing within-group variation of the original variables. In practical terms, PCA should be used to investigate overall variation in data, without concern for a specific grouping of observations, by means of linear relationships among measured variables under study. It is more appropriate to use CDA when it is important to separate known groups or *a priori* groupings, and to identify major sources of difference between groups.

Between-group differences of multivariate data from designed experiments are frequently analyzed by multivariate analysis of variance (MANOVA). In this context, the "union" of these variables, rather than individual response, is of intrinsic interest. Although CDA shares the basic aim of MANOVA, it provides canonical coefficients that yield relative information on each variable in distinguishing between groups. CDA provides visual descriptions of differences between the groups; MANOVA does not.

ASSUMPTIONS FOR CDA

Data for performing CDA should be replicated within and between the groups. Data need not be standardized before analysis, because unlike PCA, the outcome of a CDA is not affected by scaling individual variables (Manly, 1986). CDA operates under the assumption that the within-group covariance structure is homogeneous for all groups. Failure of this assumption reduces the reliability of any significance test used to determine the number of CDFs adequate to represent the data (Kshirsagar, 1972). These tests also require that data within groups have multivariate normal distributions. However, normality

Received for publication 12 July 1993. Accepted for publication 12 Jan. 1994. Acknowledgments to the Consejo Nacional de Ciencia y Tecnología, Mexico; Banco de México, Mexico; Massey Univ. Agricultural Research Foundation; and New Zealand Ministry of External Relations and Trade for financial assistance to J.G. Cruz-Castillo in his graduate studies; E.W. Hewett and C.B. Christie, Dept. of Plant Science, and R.J. Brook, Dept. of Statistics, for their constructive comments; and J.D. Majer, School of Environmental Biology, Curtin Univ. of Technology, Australia, for providing the eucalyptus data set. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

¹Graduate Student.

²Lecturer, Dept. of Statistics.

³Lecturer.

⁴Senior Lecturer.

and homogeneity assumptions are not always considered absolute prerequisites for using this technique (Cliff, 1987; Kendall, 1975; Krzanowski, 1988).

Sample size plays an important role in many multivariate techniques, including CDA. Estimates of parameters, particularly of variances and covariances, are more reliable with large samples. The combination of too many variables and few observations (i.e., sample sizes) usually causes unreliable estimates. Even with large samples, CDA may not provide meaningful results where there are too many variables. Currently, there is no optimum sample size. However, when CDA is effected on data with few variables, sample sizes ≈ 10 times the number of variables in each group would provide reliable estimates. Samples with fewer than the corresponding number of variables should be avoided, as computational problems will arise.

INTERPRETING CDF

CDFs are the canonical weights of the original variables. They provide information about the discriminatory power of each variable and are usually standardized to aid interpretation, e.g., by plotting the standardized canonical scores of individuals in each group for each CDF. When many groups are under study and visual discrimination is difficult, scores of dominant CDFs may undergo analysis of variance (ANOVA) followed by an independent multiple comparison test.

A distinction is made between interpreting each CDF and evaluating the contribution of each original variable to that CDF. Absolute values and signs of standardized canonical coefficients are used to rank variables in order of their contribution and to characterize the function. Cruz-Castillo et al. (1992) illustrates both situations using kiwifruit [*Actinidia deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson] vines. Five response variables were measured on 'Hayward' vines topworked onto eight different root systems. In this example, x_1 denotes the number of fruit removed by thinning, x_2 the percentage of flowers occurring in a compound dichasium at full bloom, x_3 the number of flowers per shoot, x_4 vine yield (kilograms per vine), and x_5 the percentage of large fruit (≥ 98 g fresh weight).

CDA yielded a dominant $CDF_1 = -0.47x_1 - 0.34x_2 + 0.25x_3 + 0.84x_4 + 0.70x_5$, in which the absolute values of the coefficients rank the variables. The higher values of the canonical coefficients x_4 and x_5 indicate that separation

between the rootstocks was most strongly influenced by total yield and fruit size. A high CDF_1 value signifies a rootstock characterized by high yields of large fruit. If these coefficients had been negative, then a high value for CDF_1 would identify a low-yielding rootstock with small fruit. In both cases, the relative contribution of each variable is the same; the interpretation, however, is different.

Using correlations between CDFs and original variables enhances interpretation of CDFs (Afifi and Clark, 1990), indicates how individual variables contribute to the CDF (Rencher, 1992), and highlights the relationship between original variables and CDFs. These correlations are generally called the "total canonical structure" (SAS, 1989).

When some of the original variables are highly intercorrelated, interpretations of the total canonical structure and standardized canonical coefficients of the CDF can be different. In these instances, it is simpler to make inferences using the correlations rather than CDF coefficients (Afifi and Clark, 1990). It is important to realize, however, that correlations of the total canonical structure merely show the importance of each variable independent of other variables, without providing information about the multivariate contribution of that variable. In standardized form, the canonical coefficients of each CDF provide information about the joint contribution of the variables to that CDF (Rencher, 1992).

The CANDISC procedure of SAS (SAS, 1989) produces relevant information such as the standardized canonical coefficients and the percentage of group variation explained by CDFs, correlation coefficients between CDFs and original variables, and standardized canonical scores for plotting observations. A likelihood ratio test signals the number of significant CDFs.

APPLYING CDF TO EMPIRICAL EXAMPLES

Three further examples, considered below, illustrate applying and interpreting CDA in horticultural research. Examples from studies with kiwifruit rootstocks and *Zantedeschia* cultivars demonstrate using CDA to identify the most important variables contributing to differences between groups. The *Zantedeschia* study also shows how CDFs provide insight into important associations between variables. CDA of a study investigating *Eucalyptus* leaf nutrient levels shows how complex differences within, and between, groups are characterized.

Table 2. Standardized canonical coefficients (SCC) and correlation coefficients between canonical discriminant functions (CDF_1 , CDF_2) and variables of four 'Hayward' kiwifruit selections topworked on five rootstocks.

Variable	CDF_1		CDF_2	
	SCC	r^2	SCC	r^2
Bud burst (%)	0.25	0.30	-0.16	-0.02
Floral bud burst (%)	0.00	-0.02	-0.15	0.01
Number of flowers/shoot	-0.27	0.12	0.33	0.32
Trunk diameter (mm)	0.13	0.49	-0.46	-0.03
Yield (kg/vine)	0.85	0.81	1.00	0.52
Mean fruit weight (g)	0.87	0.75	-0.59	-0.29
Fruit °Brix	-0.23	-0.30	0.49	0.36
Canonical correlation	0.76		0.32	
Eigenvalue	1.39		0.11	
Variance explained (%)	85.0		7.0	

²Simple correlation between each response variable and its respective CDF.

Table 3. Means of standardized canonical scores of the first canonical discriminant function (CDF_1) for four 'Hayward' selections topworked on five kiwifruit rootstocks.

Rootstock	CDF_1^z
1	-0.07 b
2	1.43 a
3	-1.58 d
4	1.12 a
5	-0.90 c

^zMean separation within each column by Fisher's protected LSD test ($P \leq 0.05$).

Multivariate comparisons of kiwifruit rootstocks. Lawes et al. (1990) hypothesized that the cropping performance of 'Hayward' kiwifruit vines varied with rootstock. Their experiment investigated the presence and nature of this variation. Each of four 'Hayward' selections was topworked on five clonal rootstocks. Each rootstock-scion combination was replicated six times. For each vine, data were collected for percent bud burst, percent floral bud burst, number of flowers per shoot, trunk diameter, fruit yield, average fruit size, and fruit °Brix at harvest. Data of the four 'Hayward' selections were combined for each of the five rootstocks. ANOVA was used to compare the five rootstocks for each variable. Significant differences ($P \leq 0.05$) existed between the vines on different rootstocks with respect to trunk diameter, yield, and mean fruit size (Table 1).

ANOVA does not show how different rootstocks compare with respect to all variables considered together, or how those variables may be interrelated. CDA simultaneously examines differences in the seven variables between vines, and indicates the relative contribution of each variable to rootstock discrimination. With data for seven variables and five groups, the maximum number of CDFs is four. Using nonstandardized data, CDF_1 accounted for 85% of between-rootstock variation (Table 2). The other three CDFs explained the remaining 15% between-rootstock varia-

Table 1. Measured variables of four 'Hayward' selections topworked on five different kiwifruit rootstocks.^z

Rootstocks	Bud burst ^y (%)	Floral bud burst ^y (%)	No. flowers/shoot ^y	Trunk diam (mm) ^x	Yield (kg/vine) ^x	Mean fruit size (g) ^x	Fruit SSC °Brix ^y
1	57.4	92.4	3.6	33.7	47.9	93	7.2
2	61.1	93.3	3.7	35.2	53.6	102	6.8
3	56.4	93.6	3.6	30.0	29.4	90	7.4
4	58.3	91.6	3.7	33.8	51.8	102	7.0
5	55.4	91.7	3.5	32.1	33.5	92	7.1

^zMeans of 24 replications/rootstock treatment.

^ySSC = soluble solids concentration. No significant differences among rootstocks ($P \leq 0.05$).

^xSignificant differences among rootstocks ($P \leq 0.05$).

Table 4. Standardized canonical coefficients (SCC) and correlation coefficients between the first three canonical discriminant functions (CDFs) and variables on four calla cultivars.

Variable	CDF ₁		CDF ₂		CDF ₃	
	SCC	r	SCC	r	SCC	r
Bud number	-0.12	0.44	0.11	-0.11	-0.038	-0.29
Tuber circumference	-0.01	-0.08	-0.56	-0.49	0.59	0.43
Tuber thickness	0.68	0.57	0.33	0.25	0.56	0.52
Shoot number	0.45	0.67	0.02	0.02	-0.14	-0.29
Plant height	0.05	0.30	0.50	0.41	-0.46	-0.49
Flower life	-0.10	0.19	-0.07	0.21	-0.12	-0.22
Days to flower	0.97	0.78	-0.17	0.06	0.36	-0.17
Duration of pollen shed	-0.31	-0.26	0.37	0.40	0.19	0.20
Peduncle length	0.14	0.17	-0.47	0.24	-0.23	-0.29
Spathe length	-0.32	-0.14	0.87	0.74	0.27	0.05
Primary flower number	-0.15	0.01	-0.27	-0.49	-0.16	-0.20
Percent variance explained	59.1		34.4		6.5	

Table 5. Summary statistics of scores between cultivars of the first two canonical discriminant functions (CDFs).

Cultivar	n	CDF ₁		CDF ₂	
		Mean	SE	Mean	SE
Black-eyed Beauty	13	2.34	0.29	0.42	0.36
Harvest Moon	25	-0.89	0.19	0.96	0.17
Pink Opal	15	-0.18	0.31	-0.88	0.29
Ruby	15	-0.37	0.17	-1.08	0.19

tion. As CDF₂, CDF₃, and CDF₄ account for negligible proportions of between-rootstock variation, only CDF₁ required further examination. This was confirmed by the likelihood ratio criterion test. This test showed that only CDF₁ was significant at the 5% level. Standardized canonical coefficients of CDF₁ (Table 2) are weighted toward fruit yield and average fruit weight. Discrimination between rootstocks may be based on the values of CDF₁, with high values signifying rootstocks with high yields of large fruit. Mean separation of canonical scores for CDF₁ revealed vines on rootstocks 2 and 4 were significantly different from vines on rootstocks 1, 3, and 5. Vines on rootstock 3 had the lowest values (negative in this case), indicating that they were vines with low yields of small fruit (Table 3). The predominance of CDF₁ in accounting for variation among rootstocks suggests that future comparisons need only focus on measurements of fruit yield and size.

Aggregating and interpreting information from several variables. Cultivar evaluation is an important component in introducing a new crop. Armitage (1986) presents a systematic approach for initial screening of species or cultivars with potential as pot plants or cut flowers. Subsequent development, however, is hindered by problems associated with subjectively assessing available selections. MacKay and Lawoko (unpublished) used CDA to provide objective discrimination between calla (*Zantedeschia* hybrid) cultivars, an important cut-flower crop in New Zealand (Funnell, 1992). Data consisted of vegetative and flowering characteristics of four cultivars ('Black-eyed Beauty', 'Harvest Moon', 'Pink Opal', and 'Ruby'). Tuber circumference and thickness and the number of dominant buds on the tuber were measured before planting. Plant height, number of shoots and primary flowers, flower peduncle and spathe length, days from

planting to flower, duration of pollen shed, and flower longevity were recorded as the first flower on each plant reached commercial harvest. Canonical scores of major CDFs were plotted and subjected to ANOVA.

Cultivars were maximally discriminated by two CDFs. CDF₁ discriminated on tuber thickness and flowering precocity (Table 4). High values of CDF₁ corresponded to plants with thick tubers, possibly a reflection of relative storage reserves, and delayed flowering characteristics. Cultivars suitable for cut-flower production would require low scores for this function, since flowering precocity is a favored characteristic. ANOVA of canonical scores revealed clear discrimination of 'Black-eyed Beauty' from the other cultivars on CDF₁ (Fig. 1, Table 5). 'Harvest Moon' scored significantly lower than 'Pink Opal' (SE of difference = 0.041), but was not distinguishable from 'Ruby'. Analysis of these scores also suggests that 'Black-eyed Beauty' and 'Pink Opal' are more variable in the characteristics encompassed by CDF₁ than the other cultivars. CDF₂ discriminated on tuber circumfer-

ence, plant height, and flower size (Table 4), with high scores associated with tall cultivars bearing large flowers from small-circumference tubers. Plant height and tuber size may have little relevance to a cut-flower producer, while the importance of flower size will depend on the anticipated market. 'Black-eyed Beauty' and 'Harvest Moon' were discriminated from 'Pink Opal' and 'Ruby'. As with CDF₁, 'Black-eyed Beauty' and 'Pink Opal' were more variable cultivars.

Both functions were consistent in their description of an inverse relationship between productivity and tuber size. Thick tubers were associated with delayed flowering (CDF₁), whereas small-circumference tubers were associated with large flowers (CDF₂). This result suggests that smaller-grade tubers may be more productive than large ones, with increased profitability through decreased input costs. This raises an important issue worthy of further investigation.

The discriminant functions are also useful for what they do not emphasize. CDA focuses on those attribute combinations that maximally distinguish one group, in this case one cultivar, from another. In this example, neither flower productivity, peduncle length, nor keeping quality was dominant in the discriminant functions, indicating that the cultivars did not markedly differ in these variables.

Characterizing changes in response between and within groups. Nitrogen, P, and K concentrations were analyzed in the foliage of two *Eucalyptus* species growing in Western Australia. Foliage was sampled in the upper and lower canopy of jarrah (*E. marginata* Donn. ex. Sm.) and marri (*E. calophylla* R. Br. ex. Lindl.). Ten trees of each species and canopy level combination were selected. Young and mature leaves were collected from each canopy level of each species. All material

Fig. 1. Canonical scores of the first two canonical discriminant functions (CDFs) of calla cultivars. (Δ) 'Black-eyed Beauty', (○) 'Harvest Moon', (□) 'Pink Opal', (◇) 'Ruby'.

was oven-dried, ground, and mineral content determined by standard methods (see Majer et al., 1992).

Only tree species and canopy levels were considered in the study. Hence, information from mineral data ignoring leaf age was grouped within the levels of the two experimental factors. The aim was to discriminate between the groups of leaf samples and relate this to the possibility that leaf nutrients might influence the groupings obtained. Majer et al. (1992) related nutrient differences to differences between species because the species differ in root physiology and therefore may also differ in their ability to extract nutrients from the rhizosphere. The species also may possess different internal storage strategies; for example, species with low foliar nutrient levels may sequester nutrients in other parts of the plant.

For illustration purposes, two CDAs were performed using all three nutrient response variables (Table 6). The first CDA investigated the difference between jarrah and marri independent of canopy level, while the second CDA differentiated the combined effect of species and canopy level.

As the maximum number of CDFs that could be obtained in the first case is one, a dummy second CDF was created with random observations from $N(0,1)$ for plotting purposes. Clearly, the samples from jarrah and marri trees were well separated by CDF_1 (Fig. 2). Potassium had the largest (and positive) standardized canonical coefficient and the largest correlation coefficient with CDF_1 (Table 6). However, canonical and correlation coefficients associated with N and P concentration were also high. This behavior indicates that K, and to a lesser extent N and P, strongly influenced discrimination of foliage from the two species. Signs of the standardized canonical coefficients indicate that K and N are opposite in effect to P in separating jarrah and marri trees.

The second CDA clearly discriminated all four groups (Fig. 3). Samples from upper and lower canopies for marri are well separated along the direction of CDF_2 , while virtually no difference is shown between them along CDF_1 . Canopy level behavior of jarrah samples is almost opposite to marri in that separation along CDF_1 is more prominent than that along CDF_2 . This behavior is consistent with the fact that two CDFs were required to explain most of the variation in the data (Table 6). More attention, however, should be given to the interpretation from CDF_1 , as it accounted for $\approx 77\%$ of the variation. Again, CDF_1 was highly influenced by K and could account for the separation between species and, to a lesser extent, between the canopy level of jarrah trees. However, in separating the canopy levels of marri trees along CDF_2 , K behaves opposite to both N and P.

SUMMARY

These case studies, taken from divergent areas of horticulture, demonstrate how CDA improves data interpretation. By reducing di-

Table 6. Summary of two canonical discriminant analyses performed on *Eucalyptus* foliar nutrient data. Total canonical structure and standardized canonical coefficients are presented in that order below the nutrients.

Groups discriminated	CDF	Percent variation explained	N	P	K
Jarrah and marri	1	100	0.66	0.68	0.98
			0.81	-0.70	2.27
Lower and upper regions of jarrah and marri	1	76.8	0.71	0.73	0.97
			0.86	-0.44	2.26
	2	22.7	-0.68	-0.53	0.15
			-1.10	-1.11	1.63

Fig. 2. Canonical scores of the first canonical discriminant function for leaf mineral element concentration in two *Eucalyptus* species. (Δ) Jarrah, ($\circ$) marri.

Fig. 3. Sample scores for leaf mineral element concentration in upper and lower canopy strata from jarrah and marri trees, plotted on first and second canonical discriminant axes. (Δ) Jarrah lower canopy, ($\circ$) jarrah upper canopy, ($\square$) marri lower canopy, ($\diamond$) marri upper canopy.

mensionality of the data sets, CDA summarized important differences between predetermined groups while recognizing the complex relationships between many characteristics; it is not possible to gain such insight using univariate statistical techniques. The assumptions required for valid use of CDA are no more rigorous than those of commonly used univariate techniques. Interpretation of canonical discriminant functions is straightforward and objective; analyzing the significant difference between groups is aided by plots of the canonical scores and by standard mean separation techniques. Procedures for performing CDA are available in many statistical computing packages. The examples presented demonstrate the scope and potential for CDA in interpreting results for the complex and highly integrated systems typical of horticultural research.

Literature Cited

- Afifi, A.A. and V. Clark. 1990. Computer-aided multivariate analysis. Van Nostrand, New York.
- Armitage, A.M. 1986. Evaluation of new floricultural crops: A systems approach. *HortScience* 21:9–11.
- Behe, B.K., T.A. Prince, and H.K. Tayama. 1992. Market segmentations of supermarket floral customers. *HortScience* 27:459–462.
- Broschat, T.K. 1979. Principal component analysis in horticultural research. *HortScience* 14:114–117.
- Cliff, N. 1987. Analyzing multivariate data. Harcourt, San Diego.
- Cole, R.A. and K. Phelps. 1979. Use of canonical variate analysis in the differentiation of swede cultivars by gas–liquid chromatography of volatile hydrolysis products. *J. Sci. Food Agr.* 30:669–676.
- Cruz-Castillo, J.G., G.S. Lawes, D.J. Woolley, and H. Varela-Alvarez. 1991. Rootstock influence on kiwifruit vine performance. *New Zealand J. Crop Hort. Sci.* 19:361–364.
- Cruz-Castillo, J.G., B.R. MacKay, G.S. Lawes, and D.J. Woolley. 1992. Canonical discriminant analysis in kiwifruit rootstock research. *Acta Hort.* 313:143–148.
- Everitt, B. 1978. Graphical techniques for multivariate data. Heinemann, London.
- Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Ann. Eugenics* 7:179–188.
- Funnell, K.A. 1992. *Zantedeschia*, p. 683–704. In: A. de Hertogh and A. Le Nard (eds.). The physiology of flower bulbs. Elsevier, Amsterdam.
- Hair, J.F., R.E. Anderson, and R.L. Tatham. 1987. Multivariate data analysis with readings. Macmillan, New York.
- Iezzoni, A.F. and M.P. Pritts. 1991. Applications of principal components to horticultural research. *HortScience* 26:334–338.
- Kendall, M. 1975. Multivariate analysis. Charles Griffin, London.
- Krzanowski, W.J. 1988. Principles of multivariate analysis. A user's perspective. Clarendon Press, Oxford.
- Kshirsagar, A.M. 1972. Multivariate analysis. Marcel Dekker, New York.
- Lawes, G.S., D.J. Woolley, D. Zhu, and G. Cruz. 1990. An early evaluation of rootstock effect in kiwifruit. XXIII Intl. Hort. Congr. Florence, Italy. Abstr. 2:4188.
- Manly, B.F.J. 1986. Multivariate statistical methods: A primer. Chapman, London.
- Majer, J.D., H.F. Recher, and S. Ganesanandam. 1992. Variation in foliar nutrients in Eucalyptus trees in eastern and western Australia. *Austral. J. Ecol.* 17:383–393.
- Mardia, K.V., J.T. Kent, and J.M. Bibby. 1979. Multivariate analysis. Academic, London.
- Perez-Gonzales, S. 1992. Associations among morphological and phenological characters representing apricot germplasm in central Mexico. *J. Amer. Soc. Hort. Sci.* 117:486–490.
- Prins, M.C., G. Jacobs, and K.I. Theron. 1990. Evaluation of carbohydrate status of apple nursery trees as a prediction of the subsequent growth in the orchard. XXIII Intl. Hort. Congr. Florence, Italy. 2:4163.
- Rencher, A.C. 1992. Interpretation of canonical discriminant functions, canonical variates, and principal components. *Amer. Statistician* 46(3):217–225.
- SAS Institute. 1989. SAS user's guide: Statistics. version 6, 4th ed. vol. 1. SAS Institute, Cary, N.C.
- Swallow, W.H. 1981. Statistical approaches to studies involving perennial crops. *HortScience* 16:634–636.
- Yourstone, K.S. and D.H. Wallace. 1990. Effects of photoperiod and temperature on rate of node development in indeterminate bean. *J. Amer. Soc. Hort. Sci.* 115:824–828.