

The Analysis of Ten Years of Cotton Data

by E.R. Williams and P.H. van Ewijk*

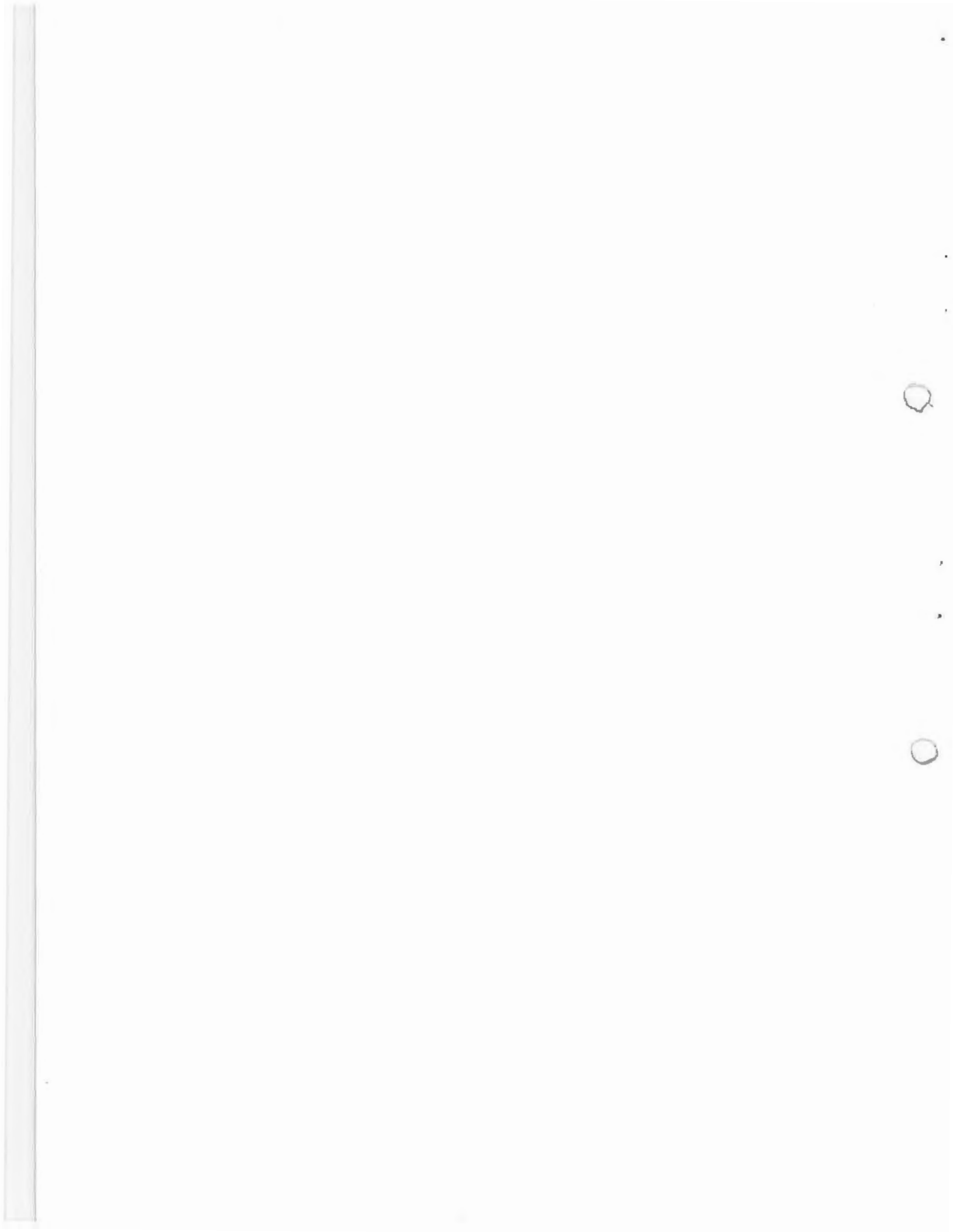
CSIRO, Division of Mathematics and Statistics

Canberra

* visiting from The Agricultural University

Wageningen

The Netherlands



1. Description of Project

1.1 Outline

The Australian Cotton Cultivar Trial (ACCT) has been conducted annually since 1974-75 at 8-12 sites per year in the major cotton growing districts in New South Wales and Queensland. The number of cultivars under test has varied from 18 to 25. The data collected in these trials have been standard at all sites and include yield, agronomic and quality data. Whilst these data have been analyzed on a year by year basis, and some groups of years have been subject to adaption analyses of various types, a comprehensive analysis of the data has not been done.

At present three breeding programs are operating, two in New South Wales and one in Queensland. This is based on previous experience and limited evidence that cultivars different from those in the Namoi (NSW) are required on the Darling Downs and Central Queensland. Likewise the present system of cultivar evaluation samples all cotton growing districts because the nature of cultivar variation from location to location (Genotype x Environment Interaction) is not well known.

The purpose of this project is to conduct a critical examination of the ACCT data from 1974 to 1984 and to try to learn more about the relative behaviour of cultivars over the testing sites and the effectiveness of locations to evaluate cultivar differences. We can then draw conclusions on the most efficient organization of the breeding and cultivar evaluation programs.

Over the ten years included in this study, 21 sites were used for cultivar trials. The abbreviations used for these sites are given in Table 1.2.1 and the actual distribution of the sites over years is presented in Table 1.2.2. It should be pointed out that "sites" 14 and 15 in Table 1.2.1 are separated out because the trials carried out there were unsprayed. For completeness we have included the results of these trials in subsequent tables but in general, performance was poor and these data were dropped from detailed statistical analyses.

Table 1.2.2 is quite sparse (i.e. only 87 out of the possible 210 site-year combinations eventuated) so for some of the later statistical analyses the number of sites was collapsed into 13 localities. The list and abbreviations are given in Table 1.2.3 and the distribution over years is presented in Table 1.2.4. The assumption made when combining data from more than one site at the same location is that cultivar behaviour is governed more by the properties of the various locations (for example, latitude, temperature, etc) rather than differences between sites at the same location. Collapsing Table 1.2.2 down to form Table 1.2.4 filled in the number of location-year combinations that actually eventuated (85 out of 130); note that for Moree and Myall Vale in 1981 two sites were used at the same locality. Also we repeat that "location" 11 in Table 1.2.3 represents two sites in NSW which were not sprayed and therefore have been kept separate, and in fact excluded from most statistical analyses.

Table 1.2.1

List of Sites

<u>No.</u>	<u>Abbreviation</u>	<u>Name</u>
1	bi	Biloela : Qld
2	bo	Boggabilla : Qld
3	dd	Darling Downs : Qld
4	em	Emerald : Qld
5	lo	Lockyer : Qld
6	mcp	Moomin Creek (Pechelba) : NSW
7	mcr	Moomin Creek (retr) : NSW
8	mog	Moree (Garah) : NSW
9	mon	Moree (Norwood) : NSW
10	mv	Myall Vale : NSW
11	mvi	Myall Vale (Insect resistant trial) : NSW
12	sg	St George : Qld
13	th	Theodore : Qld
14	unb	Unsprayed (Burrenda) : NSW
15	unw	Unsprayed (Waiwera) : NSW
16	wa	Warren : NSW
17	wnb	West Namoi (Beechworth) : NSW
18	wnk	West Namoi (Kerribee) : NSW
19	wnp	West Namoi (Pindara) : NSW
20	wnt	West Namoi (Titrees) : NSW
21	wnl	West Namoi (Lammermoor) : NSW

Table 1.2.2
Distribution of Sites over Years

bi	bo	dd	em	lo	mcp	mcr	mog	mon	Site		th	unb	unw	wa	wnb	wnk	wnp	wnt	wnl
									mv	mvi									
1		1							1		1		1	1				1	
1									1		1		1	1				1	
1		1							1		1	1		1				1	
1		1	1						1		1	1		1				1	
1		1	1						1		1			1				1	
1		1	1						1		1			1				1	
		1	1			1		1	1		1	1		1			1		
1		1	1	1			1	1	1	1	1			1					1
1	1	1	1	1	1			1	1		1	1		1		1			
1	1	1			1			1	1					1	1				

Table 1.2.3

List of Locations

<u>No.</u>	<u>Abbreviation</u>	<u>Name</u>
1	bi	Biloela : Qld
2	bo	Boggabilla : Qld
3	dd	Darling Downs : Qld
4	em	Emerald : Qld
5	lo	Lockyer : Qld
6	mc	Moomin Creek : NSW
7	mo	Moree : NSW
8	mv	Myall Vale : NSW
9	sq	St George : Qld
10	th	Theodore : Qld
11	un	Unsprayed : NSW
12	wa	Warren : NSW
13	wn	West Namoi : NSW

Table 1.2.4

Distribution of Locations over Years

Year	Location												
	bi	bo	dd	em	lo	mc	mo	mv	sq	th	un	wa	wn
74	1		1					1	1	1	1	1	1
75	1							1	1	1	1	1	1
76	1		1					1	1	1	1	1	1
77	1		1	1				1	1	1	1	1	1
78	1		1	1				1	1			1	1
79	1		1	1				1	1			1	1
80			1	1		1	1	1	1	1		1	1
81	1		1	1	1		2	2	1	1		1	1
82	1	1	1	1	1	1	1	1	1	1		1	1
83	1	1	1			1	1	1				1	1

1.3 Cultivars

The list of cultivars (with abbreviations) which have been used over the 10 years is given in Table 1.3.1. Unlike most genotype x environment x year studies (where the genotypes or cultivars remain much the same over the course of the study), here we are confronted with a large number of cultivars being replaced from year to year. This can be seen in Table 1.3.2 which records the presence of the 92 cultivars (numbered 1 through 92 as in Table 1.3.1) over the 10 years.

We note in Table 1.3.2 that only 2 cultivars have been present in all years, namely deltapine 16 and namcala. Other cultivars are closely linked genetically (for example there are several deltapine varieties). This fact was not used in the initial statistical analyses, but was looked at later on.

No.	Abbreviation	Name
1	dp16	12 deltapine 16
2	dpsm	deltapine smoothleaf
3	dps113	12 deltapine sl sel 13
4	kim 11	kindell 11
5	hop	hopicala
6	gl06	gl06
7	m8sup	m8 superokra
8	del277	delcott 277
9	st7A	70 stoneville 7A
10	st7Ane	stoneville 7A nectariles
11	basmlc	bayou sm 1
12	ac70	acala 70/185
13	nam	namcala
14	cr79	cr 79
15	cr74	cr 74
16	rexsm	rex smoothleaf
17	h62	h62-8
18	han1032	hancock
19	mcn1032	macnair 1032
20	st702	stoneville 702
21	ls/obG	ls/ob G090
22	ss/obG	ss/ob S081
23	ls/obF	ls/ob Field bulk
24	rivp	riverina poplar
25	rivg	riverina gold
26	dp61	4 deltapine 61
27	tampsp37	37 tamcott sp37
28	c4104	C coker 4104
29	m8	m8
30	g245	g245-2-11
31	c310	coker 310
32	mv81	mv081
33	70081	70-081-20-12-19-4
34	cr142	cr 142-45
35	stgln1	stoneville glabrous nect
36	dpglfr	deltapine glabrous frego
37	cr128	cr128-5
38	c312	C coker 312
39	des24	des 024
40	mcn220	macnair 220
41	sicot2	sicot 2
42	n40286	n40 286 hf
43	sicot1	12 sicot 1
44	n40286h	n40 286 h
45	n4039h	n40 39 h
46	n70281	n70 28/1
47	n70283	n70 28/3
48	dp55	55 deltapine 55
49	mo63br	mo63-277 br
50	cbr6	C coker br-600
51	dp80	80 deltapine 80
52	c315	C coker 315
53	csm511	C coker sm 511
54	n41572n	n41 57-2n
55	dp61ipr	deltapine 61-ipr
56	dp41	41 deltapine 41

Table 1.3.1. Cont.

No.	Abbreviation	Name
57	dp7146N	deltapine 714 6N
58	n40lh	n40 lh
59	sicotlf	sicot lf
60	sicot3	sicot 3
61	n40439g	n40 439 g
62	n40439h	n40 439 h
63	mo63j	mo63 277j
64	n39104	n39 10/4
65	n39428	n39 42/8
66	n70371	n70 37/10
67	n85338	n85 33/8
68	750073	75007-3
69	760231	76023-1
70	n74630	n74 630
71	n4039hf	n40 39 hf
72	n40439d	n40 439ds
73	n914109	n91 4/109
74	n917111	n91 7/111
75	n91830	n91 8/30
76	n919111	n91 9/111
77	siokra	siokra
78	n74480	n74 480
79	dp6177	deltapine 61-77
80	dp90	deltapine 90
81	n7469	n74 69
82	n74250	n74 250
83	n98283	n98 283
84	n98247	n98 247
85	n98249	n98 249
86	n98251	n98 251
87	n98308	n98 308
88	n98320	n98 320
89	n98329	n98 329
90	n98331	n98 331
91	n98335	n98 335
92	n98336	n98 336

Table 1.3.2

Distribution of Cultivars over Years

Year	1	2	3	4	5	6	7	8	9	12
74	111111111111111111									
75	11111111111111111111									
76	111111111111111111			111111						
77	111111111111111111			111 1 11						
78	1	11 11	1	1	11 1	111111				
79	1	11 11	1 1 1	1 1 1	11111					
80	1	11 11	1 1	1 1	1 1111111111111111					
81	1	1	1 1	1 1	1111	1 1	1111111111111111			
82	1	1	1	1	11 1 1	1 1	1 11 11	1 1111111111		
83	1	1	1	1	11 1	1 1	1 1	1 1	11111111111111	

1.4 Variates

Up to 10 variates were measured from each plot of each experiment. A summary of the variates measured, abbreviations and number of years measured is given in Table 1.4.1. When a variate was measured for less than 10 years it is the most recent year(s) that is (are) missing. Three of the variates (boll, ltpc and lyld) are quantitative, the others are fibre quality variates.

In more detail, a bulk sample is taken from each plot from which the following are determined:

- (i) boll = average seed cotton weight (in grams) from one boll.
- (ii) ltpc = weight of lint after ginning as a percentage of seed cotton weight before ginning.
- (iii) lyld = lint yield in kilograms per hectare.

The fibre quality variates are also determined from the bulk sample. The high-volume testing machine does one micronaire test and then three "comb" tests per sample. Each "comb" produces three pieces of information :

- (a) the length of fibres at the mean of the length distribution,
- (b) the length of fibres at the upper 2.5% point of the length distribution,
- (c) the breaking strength.

Micronaire (mike) is a complex character influenced mainly by fibre fineness and maturity. The units of mitre are relative rate of air flow and values are usually in the range 3.0 to 5.0. The length distribution is determined by light transmission through a stretched sample of fibres, from which the 50 and 2.5 percentiles (L50 and L2.5) are obtained. The leng = L2.5 (in inches), unir = $(L50/L2.5) \times 100$ and elongation (elon) is the amount of stretch (as a percentage of original length) of fibres before breaking point.

Strength (strg) is a measure of linear density and the units are grams per kilometer of length.

List of Variates

<u>Abbreviation</u>	<u>Name</u>	<u>No. of Years</u>
boll	boll weight	7
elon	elongation	9
leng	length	10
ltpc	lint percent	10
lyld	lint yield	10
mike	micronaire	10
strg	strength	10
unir	uniformity ratio	10
yain	yarn appearance index	8
yast	yarn strength	8

2. Preparation of Data Files

2.1 Collection of Raw Data

Although lyld had been analysed by others, year by year and for some groups of years, the other variates had not been processed. We decided to obtain the raw data from each separate experiment (87 of them) and for each variate and systematically re-analyse all data. The data came from two sources representing the NSW and QLD components of this study.

Unfortunately we had to contend with a considerable lack of uniformity in the way the data had been assembled between and within the two sources. Consequently the preparation of a uniformly structured raw data file took much more time than originally anticipated. We will not dwell on this aspect but rather list some of the problems that we confronted:

- (i) Some data recorded with variates in parallel, others serially (with different numbers of items per line).
- (ii) Instruction: Lack of adequate instructions on contents of data files as well as faulty and repeatedly changing instructions.
- (iii) Order: Different orders for plots, variates and locations even within the same data files.
- (iv) Identification: No information of locations and years in the data files. Cultivars were numbered 1,2,3... etc. Different codings were used for missing values (-1, -1.000, *, M, MP).
- (v) Format: Numbers in free format serial data often did not have a space separator. Within the same file formats were different for each location by year combination.
- (vi) Absence of records: Some of the parallel read data files had card images missing, necessitating a backtracking to original codings.
- (vii) Transformations: Different plot sizes meant separate scalings for total plot yields. Weights were recorded in any of grams, kilograms or pounds. In many cases it was unclear whether the weight of the sample pick had been added into the total plot weight. In a couple

Once the data had been collected and assembled into a consistent data file, the experiments from each site in each year were analysed according to their statistical structure. Most of the designs were lattice designs for 25 varieties and three replicates. A summary of the designs used is given in Table 2.2.1. A description of square lattice designs (one dimensional elimination of variation) and lattice square designs (two dimensional elimination of variation) is provided in Chapters 10 and 12 of Cochran and Cox (1957).

The field layout for the majority of the experiments was similar to that indicated in Figure 2.2.1, with either three or six replicates. There were some variations due for example to field size restrictions. In Figure 2.2.1 the furrows or field rows run away from the irrigation channel and the field columns parallel. The authors find this terminology confusing and so throughout this report, we will refer to "rows" and "columns" of a design where a row is in the same direction as a field column and a column is in the same direction as a field row.

The statistical package GENSTAT was used for the analyses. Each data set was screened for outliers by plotting residuals versus fitted values. We picked up a number of obvious "mispunchings" of data and corrected them. There were still a number of data values that we deemed as outliers but had no explanation for them. The list is given in Table 2.2.2; these values were set as "missing". There were also a number of missing values in the original data files. The package GENSTAT estimates missing values so that a designed experiment can still be analysed as such. However, for those site x year combinations where 50% or more of the data within one row or column (in the case of lattice squares) was missing, a randomized block analysis was carried out initially to obtain missing value estimates and then the lattice analysis was used. The reason for this two-stage procedure is that GENSTAT (not unexpectedly) can give spurious missing value estimates in situations where the blocking structure is complex and there are missing values concentrated within one or more blocking units.

There are some extra comments to be made about the details in Table 2.2.1.

(i) The designs used for 1978, all sites and 1979, were unusual in that they do not fit into the scheme of lattice square designs as specified by Cochran & Cox (1957). As a result the MACRO for lattice square designs in GENSTAT would not analyse these designs. This led to some statistical work on lattice square designs (Williams et al, 1986) and eventually a generalization of the GENSTAT MACRO. However for this study we simply used a square lattice analysis with rows as blocks.

(ii) In 1983 all designs except mv had six replicates. The lattice squares were obtained from the usual 25 x 3 designs by exchanging rows and columns for the second half. Only three replicates were used for the quality data so care must be taken to ensure that the three replicates are chosen as a set; this was not done for 1983 dd.

(iii) As mentioned the experimental layouts were in the main similar to Figure 2.2.1. An extreme departure was the layout used for 1978 em. The design was for three replicates of a 4 x 4 lattice square but the arrangement of the plots was for 16 rows and 2 columns.

Table 2.2.1

Experimental Designs used for separate experiments

Year	Sites	No. of Cultivars	No. of Replicates	Design
1974	bi,dd,mv,sg,th,unw,wa,wnt	18	3	randomised blocks
1945	bi,mv,sg,th,unw,wa,wnt	25	3	lattice square
1976	bi,dd,mv,sg,th,unb,wa,wnt	25	3	lattice square
1977	bi,dd,em,mv,sg,th,unb,wa,wnt	25	3	lattice square
1978	bi,dd,em,mv,sg,wa,wnt	16	3	square lattice
1979	bi,dd,em,mv,sg,wa,wnt	16	3	square lattice
45 1980	dd,em,sg,th	25	3	randomized blocks
24	mcr,mon,mv,wa,wnp	25	3	square lattice
3 1981	bi,dd,em,lo,mv,mvi,sg,th	25	3	square lattice
100	mog,mon,wa,wnl	25	6*	square lattice
1982	bi,dd,em,lo,mv,sg,th	25	3	lattice square
72	bo,mcp,mon,wa,wnk	25	6*	lattice square
26 1983	mv	25	3	lattice square
77	bi,bo,dd,mcp,mon,wa,wnb	25	6*	lattice square
32	dd (quality data)	25	3	square lattice

* data for 6 replicates available for 1yld and 1tpc ;
for other variates only 3 replicates available

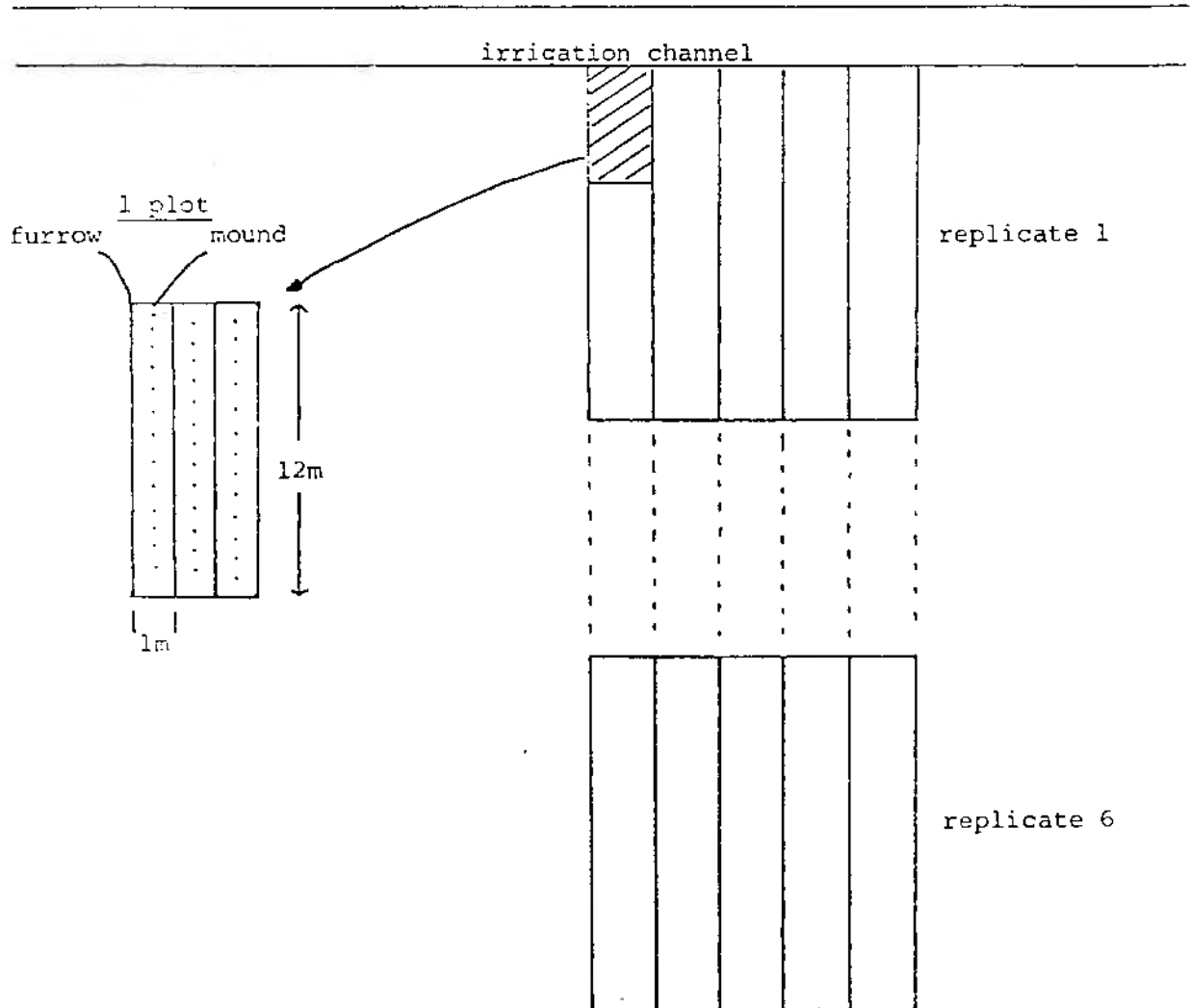
Table 2.2.2

List of Outliers

Variate	Year	Site	Replicate	Row	Col	Cultivar	Value
lyld	1975	mv	2	5	2	ls/obG	1020.2
lyld	1976	th	1	4	2	g245	877.7
lyld	1979	wa	2	2	2	dp61	2032.4
lyld	1981	lo	3	4	1	mcn220	592.8
unir	1976	wa	2	2	1	nam	4.6
unir	1978	bi	3	2	2	c310	21.6
unir	1982	wa	3	5	1	dp6177	4.5
strg	1978	bi	3	2	2	c310	42.0

Figure 2.2.1

Layout of cotton variety trial



2.3 Summary Data Files

From the analyses of variance of the individual experimental designs we accumulated the following information:

- (i) Adjusted cultivar means (with recovery of interblock information).
- (ii) The residual mean square and F-values for replicates, rows, columns (where applicable) and cultivars from each (intrablock) analyses of variance table.

In any year the cultivars tested were the same at all sites. Therefore the cultivar means can be conveniently summarized by a series of cultivar x location tables, one for each year and each variate. These are assembled in Appendix A and constitute the basic data file from which the subsequent cultivar x location x year study was conducted.

In addition, in Appendix B we give summary tables of the analysis of variance information listed in (ii) above. These tables will be analysed later. For now we simply observe that the location x year tables for the residual mean square has been scaled for convenience; details of the scaling factors are given in Table 2.3.1. The same scaling will be used in the subsequent Findlay-Wilkinson analyses of variance.

Scaling Factors for Variates when used in
Analysis of Variance Tables

<u>Variate</u>	<u>Scaling</u>
boll	-
elon	-
leng	$\times 10^3$
ltpc	-
lyld	$\times 10^{-4}$
mike	$\times 10$
strg	-
unir	-
yain	$\times 10^{-1}$
yast	$\times 10^{-4}$

3. Analysis of Cultivar x Location Tables

3.1 Summary so far

In the previous Chapters we have given details on how the tables in Appendices A and B were obtained; but for the remainder of this study these Appendices will represent the totality of the data to be considered. To reiterate some important (from the point of view of statistical approach) aspects of the contents of Appendix A:

(i) Over the 10 years of the study, on total of 21 sites and 92 cultivars were used. However within any one year the same group of cultivars were tested at the same collection of sites.

(ii) In some years there were two sites included from the same location; these data were combined over sites to reduce the total number of locations to 13.

(iii) One of the locations (un) was in fact only nominal and represented sites at which an unsprayed trial was conducted. This location was excluded from most statistical analyses.

Any analyses of the tables in Appendix B will be left to the next Chapter. For now we simply observe that the tables of residual mean squares from the individual trial analyses are fairly consistent and therefore it is reasonable to consider a constant variance over locations and over years. For example for the main variate of interest (lyld), if we exclude the value 4.93 for wa, 1981 which is high due to mite attack (P. Reid, pers. comm), the range of variation of residual mean squares is about the same as that mentioned by Patterson and Silvey (1980) in their comments on British variety trial data.

For each variate and for each year we have a complete cultivar (C) x location (L) table of means. Therefore it is sensible to take account of these complete tables and investigate the presence of CxL interaction within each table. Other options would be to look at subsets of the data or incomplete tables over years; however we feel that the advantage of having the complete CxL tables should not be ignored. With complete tables we can use the model and analysis outlined by Finlay and Wilkinson (1963), and before then by Yates and Cochran (1938).

The model is as follows:

$$y_{ij} = \mu + \tau_i + \beta_i \theta_j + \epsilon_{ij}, \quad (3.2.1)$$

where y_{ij} is the value of a variate for cultivar i at location j ; μ is a parameter for the general mean; τ_i is a parameter for cultivar i , $\sum_i \tau_i = 0$; θ_j is a parameter for location j , $\sum_j \theta_j = 0$, β_i is a regression coefficient for cultivar i ; and ϵ_{ij} is an error term, assumed to have zero mean and variance σ^2 .

The formulation of the model (3.2.1) allows for margins (or main effects) to be fitted in a CxL table and in addition there is a provision for cultivars to demonstrate different behaviour over locations (via the β_i). This decomposition of the CxL interaction into a component for linear regressions and a residual can be incorporated into an overall analysis of variance table, see for example Finlay and Wilkinson (1963, Table 2). We also refer to this paper for details on the interpretation of the β_i .

As a summary we note that:

- (i) When there is no interaction the β_i are all equal to 1.
- (ii) When $\beta_i > 1$, then cultivar i is said to be sensitive to locations, performing better at good locations than at poorer locations.
- (iii) When $\beta_i < 1$, then cultivar i is said to be stable, performing much the same at all locations.

(iv) Some β_1 can be negative; interpretation is more difficult but such cultivars would be major contributors to CxL interaction.

(v) The values of the β_1 must be viewed in conjunction with the estimated cultivar means ($\hat{\mu} + \hat{\tau}_i$) in order to judge the relative merits of cultivars.

The model (3.2.1) proposes that the interaction between cultivars and locations arises due to differences in the ranges of values from row to row (cultivar to cultivar) in the two-way CxL table, and furthermore the rows can be linearly regressed against the column averages. We could equally well think of switching this around and consider that the CxL interaction was deriving from differences in the ranges of values from column to column (location to location), or indeed the model could incorporate both concepts, i.e.

$$y_{ij} = \mu + \gamma_j \tau_i + \beta_i \theta_j + \varepsilon_{ij}, \quad (3.2.2)$$

where γ_j is a regression coefficient for location j . For a complete two-way table, we can very simply extend the analysis of variance table of Finlay and Wilkinson (1963) to give a partition of the CxL interaction into four orthogonal components. A sample analysis of variance table is given in Table 3.2.1. The framework of Table 3.2.1 has been used on the CxL table for each year and each variate. Results are given in Appendix C. To help relate the numbers in Appendix C to Table 3.2.1 we note that:

- (i) The numbers in column (2) of Appendix C give the number of locations (l)
- (ii) The numbers in column (3) of Appendix C give the number of cultivars (c)
- (iii) The entries (j) in the 'mean square' column of Table 3.2.1. refer to column (j) in Appendix C.

(iv) To obtain the Finlay-Wilkinson analysis of variance table for model (3.2.1), one can combine the lin.lin and dev.lin lines of Table 3.2.1 to get the regression line of Finlay-Wilkinson (1963, Table 2) on $l-1$ degrees of

deviations from regression line of Findlay-Wilkinson's Table 2 on $(l-2)(c-1)$ degrees of freedom.

(v) Likewise a different pairwise combination of the four components of the CxL interaction in Table 3.2.1. would correspond to a model where locations were assumed to have different ranges of values over cultivars.

The model (3.2.2) and corresponding analysis of variance table (Table 3.2.1) are very easily handled using the statistical package GENSTAT (Alvey et al, 1977). The results are summarized in Appendix C. We have already described columns (2) through (10) of Appendix C; we now explain the remaining columns, where we use the symbol (j) to refer to column j of Appendix C: Column (11) - the proportion of the CxL interactions sum of squares which is accounted for by linear regressions, i.e.

$$(11) = \frac{(7) + (l-2)(8) + (c-2)(9)}{(l-1)(c-1)(6)}$$

Column (15) - the mean square for regressions of locations on cultivar means, i.e.

$$(15) = \frac{(7) + (l-2)(8)}{l-1}$$

Column (16) - the deviations mean square from regression of locations on cultivar means, i.e.

$$(16) = \frac{(c-2)(9) + (l-2)(c-2)(10)}{(l-1)(c-2)}$$

Column (17) - the mean square for regressions of cultivars on location means, i.e.

$$(17) = \frac{(7) + (c-2)(9)}{c-1}$$

Column (18) - the deviations mean square from regressions of cultivars on location means, i.e.

$$(18) = \frac{(l-2)(8) + (l-2)(c-2)(10)}{(l-2)(c-1)}$$

Column (19) - the average (pooled) estimate of σ^2 for each year. The estimate is obtained from the tables of mean square errors (fmse) in Appendix B, by taking column averages and then dividing by three

(the number of replicates in individual trials). We note however that for the last three years of lyld and ltpc, some trials had six replicates so some adjustment is needed here.

Columns (20,21,22,23) - F ratios for testing the four components in the breakdown of the CxL interaction, i.e.

$$(20,21,22,23) = \frac{(15, 16, 17, 18)}{(19)} .$$

To further help in understanding the contents of Appendix C, we note that the mean squares in the first 5 lines in Table 2 of Findlay and Wilkinson (1963) correspond to columns (5), (4), (6), (17) and (18) respectively of Appendix C; and the residual in Findlay and Wilkinson's Table 2 corresponds to column (19) of Appendix C.

As well as obtaining the mean squares given in Appendix C, our GENSTAT analyses also give the estimates for the regression coefficients γ_j and β_i in (3.2.2). These values are given in Appendices D and E respectively. The estimated means $\hat{\mu} + \hat{\theta}_j$ and $\hat{\mu} + \hat{\tau}_i$ are given in Appendices F and G respectively.

Table 3.2.1

Sample Analysis of Variance Table

<u>Source</u>	<u>Degree of freedom</u>	<u>Mean square</u>
Locations (L)	$l-1$	(4)
Cultivars (C)	$c-1$	(5)
CxL interaction	$(l-1)(c-1)$	(6)
lin.lin	1	(7)
lin.dev	$l-2$	(8)
dev.lin	$c-2$	(9)
dev.dev	$(l-2)(c-2)$	(10)

3.3 Interpretation of Finlay-Wilkinson Analyses

The results in Appendices C, D and E represent an enormous reduction in the basic data file but of course are still too detailed for any clear picture. So let us reduce them further and look at some relevant summary statistics. In the first instance we want to know whether the CxL interaction is significant; the appropriate F test is the ratio of columns (6) and (19) in Appendix C. Average values (over years) for each variate are given in Table 3.3.1 along with the averages for columns (6) and (19). In forming Table 3.3.1 we note that it is reasonable to pool the quantities involved over years.

The degrees of freedom associated with the F statistics in Table 3.3.1 are large, in excess of 200 for both numerator and denominator so from a statistical point of view we would conclude that for all variates the CxL interaction is significant. However it does seem that lyld is presenting a much stronger result than the other variates.

As a first step in trying to interpret the CxL interaction we have used the Finlay-Wilkinson approach and Table 3.3.1 gives the proportion of the interaction accounted for by this approach (i.e. average over years of column (11) in Appendix C). The proportions are much the same over all variates. It has been argued that by only accounting for about 30% of the CxL interaction makes the Finlay-Wilkinson model open to criticism. However we could equally well look at the F ratios in, say columns (20) - (22) of Appendix C for lyld, and argue that relatively speaking the variation accounted for by regressions is high.

The model (3.2.1) is an attempt to describe the significant CxL interaction. Other methods were investigated to try and achieve this aim, for example:

- (1) Looking for pattern in tables of residuals (i.e. after removing margins from a CxL table). Here we were looking for consistent unusual behaviour of locations and/or cultivars over CxL tables for several years.

interaction other than the conclusions which can be drawn from the Findlay-Wilkinson regression coefficients based on about 30% of the interaction information. We suspect that similar conclusions could be obtained using other data exploratory methods (see for example Byth and Mungomery, 1981). We now look in more detail at the contents of Appendix C in relation to the Findlay-Wilkinson breakdown of the CxL interaction:

boll : The F value (Table 3.3.1) is quite low so we would not expect too many dominant features. We notice that the residual mean square from the individual analyses (column (19)) decreases over the years, presumably due to better experimentation. There is a high F value for locations in 1977 (5.37) referring to Appendix D we see that the extreme slopes are dd (1.52) and em (0.42). Another feature is the high lin.lin value for 1977 (1.12). What the lin.lin component tells us is whether there is any relation between the means and the regression slopes (i.e. in (3.2.2)) relation between the γ_j and θ_j or the τ_1 and β_1). A discussion of this point together with earlier references is given by Shorter in Byth and Mungomery (1981, page 266). When the lin.lin component is high then we have a desirable situation in that the locations (cultivars) with the highest mean values are also the most sensitive (have the highest regression coefficients) and so summarization is simplified.

elon : As with several variates the 1974 values seem very erratic; here a noticeably low value of the average error causes some larger F values in columns (20) - (23). In general for later years the residual decreases (as happened with boll and also with some other variates). Other features include a high lin.lin value for 1981 (1.25) and high location F values for 1980 and 1981 (5.82 and 6.36), although there are no particularly high values for regression coefficients in Appendix D.

- leng: The 1974 (in particular) and 1975 data show signs of being very noisy (see columns (20)-(23)). A check of residuals for the two-way table in 1974 reveal some unusual values for dd so we suspect the original data from this site. Other features are high lin.lin for 1976 and 1977 (2.11 and 5.25) and several high location F values.
- ltpc: The residual is highish for 1974 but stable for remaining years and lin.lin is high for 1974 - 1978 inclusive and also 1983. Location F values are high for 1975 (5.85) and 1978 (8.08), the latter value due to em(1.46), mv(1.40) and dd(0.53).
- lyld: From Table 3.3.1 we see that this is the variate that demonstrates the highest CxL interaction and is therefore the one that most effort should be placed in interpretation. We have already mentioned however that no consistent patterns had been picked up in the CxL tables over years apart from what we can get out of the Findlay-Wilkinson analyses (although we are aware that similar information could be obtained from other techniques). For lyld the residual mean squares are fairly constant over years (unlike with most other variates). The lin.lin value are high for all years with the exception of 1983, where instead the location deviations mean square is high (10.64). The locations F value is high for all years and although the cultivar F values are also high (column (22)), it is clear that locations are the major source of the variation in the CxL tables and so the lin.lin components should in the main be interpreted in terms of locations rather than cultivars (we note that lin.lin is a symmetric component in that if there is a linear relationship between location means and slopes then there will also be one between cultivar means and slopes). From Appendix D we can extract the major contributors to the high locations F values; without listing them all it is clear that wn and mv feature strongly with bi and sg to a lesser extent.

- mike: Residual mean squares are reasonably consistent over years. The lin.lin component is high for 1974, 1978, 1981 and 1982. Location F values are high for several years, but high regression coefficients are spread across a number of locations.
- strg: High F values for columns (21) and (23) in 1974 due to large residuals for dd in the CxL table (we have already mentioned that the dd data is to be questioned for this year). The lin.lin component is high for 1976 and there are a number of high location F values.
- unir: From Table 3.3.1 we note that the average F value and in particular the average proportion of interaction for this variate are not high. There are no particularly interesting features in the tables apart from a high lin.lin value for 1975.
- yain: There are high residual means squares for 1974 and 1975 and these are accompanied by high deviations mean squares (column (10)) indicating a general level of noisiness in the data. Other features are high lin.lin for 1975 and 1978.
- yast: This variate had a high average F value (Table 3.3.1). There were high F values in columns (21)-(23) for 1974 and 1975 indicating some noisiness in the data. Otherwise the lin.lin component was not particularly high, although the locations F value was high for several years, indicating that high locations regression coefficients may not be tied to high location means.

Table 3.3.1

Summary values from Appendix C

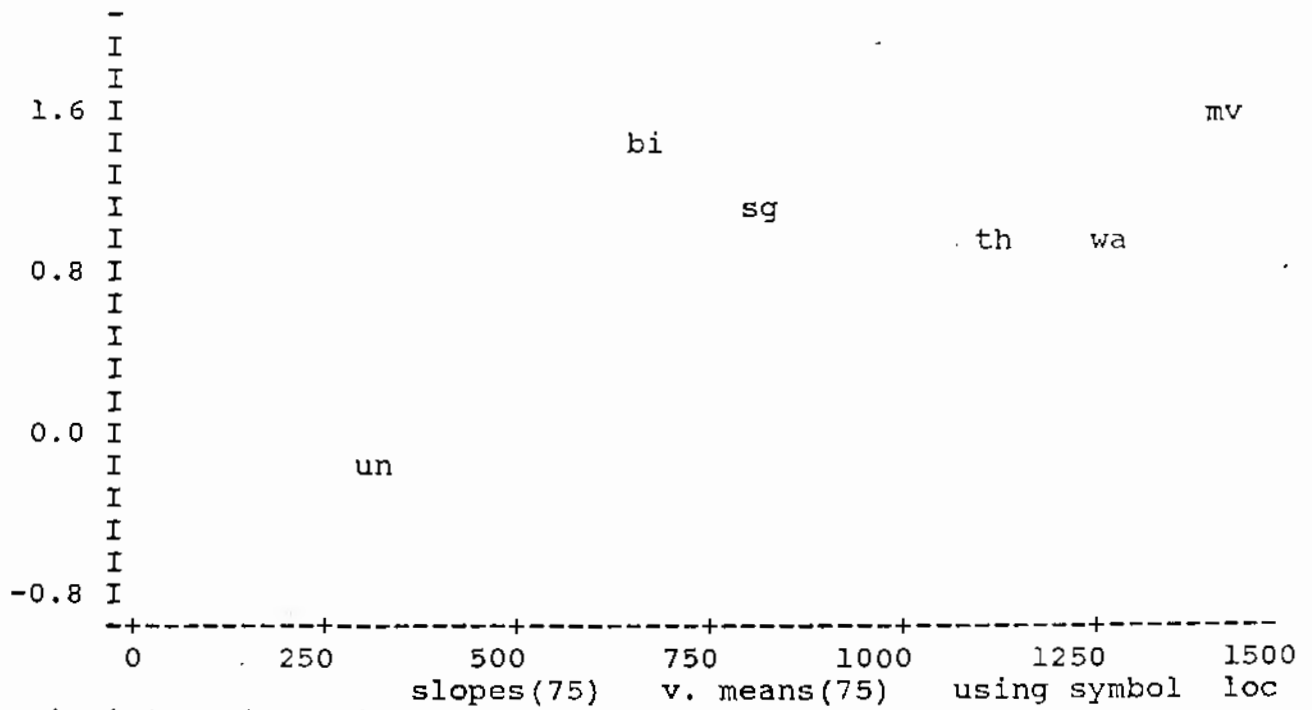
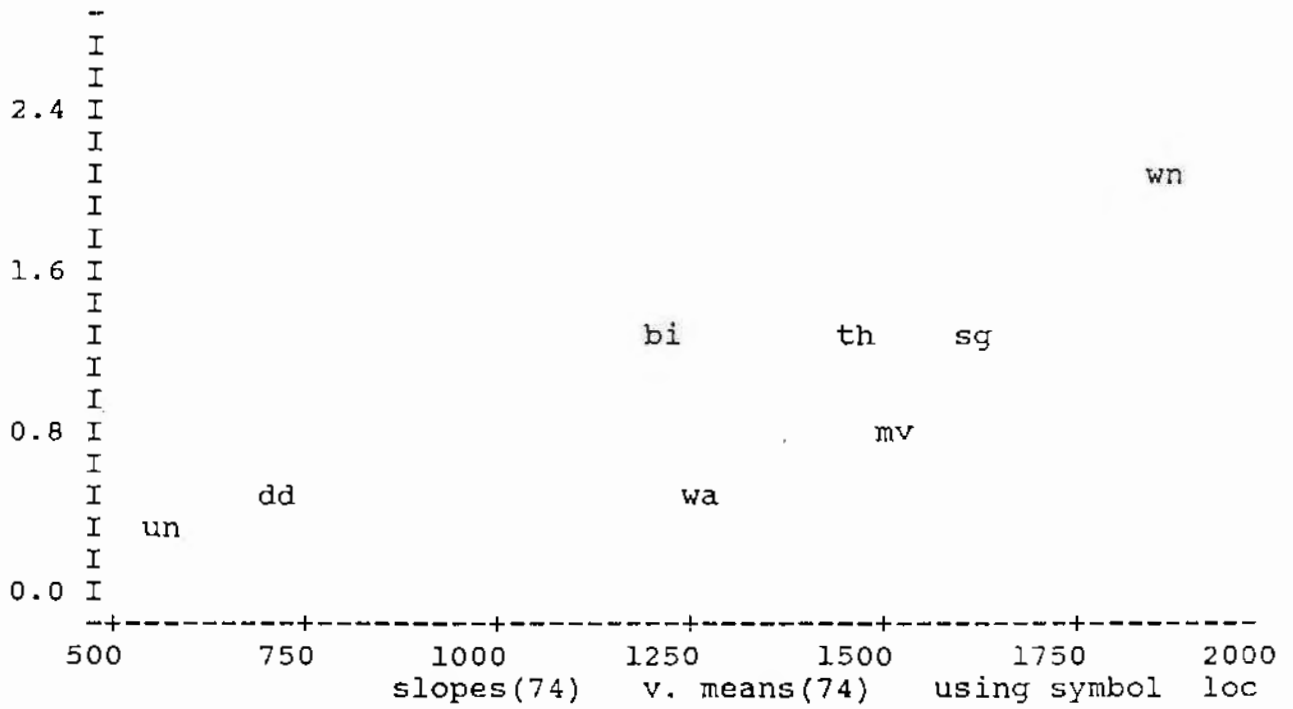
	CxL	Av. residual	Av. F	Av. propn.
	mean square	mean square	value	variation
boll	0.10	0.07	1.38	0.20
elon	0.13	0.09	1.54	0.24
leng	0.32	0.16	1.98	0.26
ltpc	0.71	0.48*	1.47*	0.28
lyld	1.28	0.41*	3.27*	0.31
mike	0.45	0.22	2.06	0.33
strg	0.54	0.29	1.85	0.32
unir	0.61	0.37	1.65	0.19
yain	1.49	1.04	1.44	0.27
yast	0.85	0.40	2.13	0.29

* based on 3 replicates for individual trials

In the last Section we discussed the interpretation of the Findlay-Wilkinson analysis results in Appendix C. In particular we highlighted the importance of the lin.lin term (column (7)) in pointing towards a relationship between means and slopes for either cultivars or locations, although it is fairly evident from the results that location differences are the primary source of the significant CxL interaction and therefore we should concentrate on these. The variate lyld, as well as being the most important variate also exhibited the biggest CxL interaction so in this Section we elaborate on the interaction of the lin.lin component. In the Figure 3.4.1 we plot the slopes (from Appendix D) against the means from (Appendix F) for the variate lyld and separately for each of the 10 years. The lin.lin component (column (7) of Appendix C) is in fact the regression mean square from a linear regression of slopes on means and we see from Figure 3.4.1 that the best relationships are obtained for 1977, 1978 and 1980. This Figure also helps to identify desirable locations, eg. mv and wn. Further insight could be gained by relating these graphs to climatic conditions in various years. Finally, we emphasise that when lin.lin is low but other components of Appendix C are high such as us the case in lyld: 1983, we have to go further than Figure 3.4.1 to get an interpretation.

Figure 3.4.1

lyld:slopes:means 1974 - 1983



points coinciding with point wa

wn

