THE PRINCIPAL COMPONENT ANALYSIS OF GENOTYPE-ENVIRONMENTAL INTERACTIONS AND PHYSICAL MEASURES OF THE ENVIRONMENT

JEAN M. PERKINS

Department of Genetics, University of Birmingham, Birmingham B15 2TT, England

1. Introduction

The linear regression analysis of the environmental plus genotype-environmental interaction components of a genotype's performance in each environment against a non-independent environmental component (derived from the average of all genotypes in each environment) as first proposed by Yates and Cochran (1938) has been criticised on statistical grounds by Freeman and Perkins (1971). However, Fripp (1972) has demonstrated that any bias introduced by the use of a non-independent environmental measure makes little difference either to the ranking of the genotypes according to the magnitude of their linear regression coefficients or to the proportion of the genotype × environmental variation accounted for by the heterogeneity of these regressions when compared with the results of analyses of regression against various independent but biological measures.

Where an independent assessment of the environment has been used and the linear regressions account for all the significant genotype-environmental interactions, the regression equations provide reliable predictions over both environments and generations (Bucio Alanis, Perkins and Jinks, 1969).

The analysis is found to have little predictive value when a large proportion of the genotype-environmental interactions cannot be explained by a linear regression. Under these circumstances, however, some insight into the nature of the residual variation can be gained by grouping the genotypes on the basis of the correlations between the deviations from linear regression of pairs of genotypes (Perkins and Jinks, 1968*b*).

In the present paper, the principal components analysis (see, for instance, Seal, 1964) of environmental and genotype-environmental interaction components of variation will be described and its relationship to the analyses of linear regression, against a non-independent environmental measure, and of deviations from linear regression explained. The results obtained from the multiple regression of genotypic performance in different seasons against physical measures of the environment will also be reported.

2. Material and data

Previous reports on the analysis of genotype-environmental interactions of a set of 29 inbred lines of *Nicotiana rustica*, from diverse origins, in each of ten environments showed that for the two characters, final height and flowering time, the linear regression analysis, against the non-independent environmental measure, accounted for only about half the mean square for genotype-environmental interactions (Perkins and Jinks, 1968*a*, *b*). Further analysis of the residual variation permitted the partitioning of the genotypes

into groups which showed significant positive correlations, over environments, between the deviations from linear regression of pairs of genotypes in the same group and significant negative correlations between pairs of genotypes in different groups (Perkins and Jinks, 1968b). For final height, it was found that the two major groupings obtained correspond with a single gene difference, non-mophead, M, and mophead, m. The non-mophead lines tend to grow steadily throughout the season with a straggly flowering head, while the mophead lines develop most rapidly later in the season and have a compact flowering head.

TABLE 1

The mean final heights of two sets of eight inbred lines of Nicotiana rustica (4 non-mophead, M, and 4 mophead, m) one grown in each of 10 years and the other in each of 9 years

							Year				
Variety	M/m	1949	1950	1957	1958	1959	1961	1967	1968	1969	1970
Set 1 (10 seasons)											
12	M	102·24	115·57	141·73	138·68	139·70	174·42	123·70	158·90	134·90	128·50
15	M	64·14	75·08	54·19	87·12	77·98	93·98	69·60	91·30	76·10	66·60
35	M	95·89	120·90	87·88	136·40	133·86	160·78	110·49	157·80	127·30	122·70
42	M	89·92	105·84	73·66	137·92	116·84	152·40	91·44	146·10	115·50	110·60
	$\bar{M}$	88·05	104·35	89·37	125·03	117·10	145·40	98·81	138·53	113·45	107·10
2	m	84·33	118·54	93·47	114·55	100·20	106·68	107·61	122·30	118·50	108·30
21	m	92·46	135·59	102·04	106·93	103·12	124·46	97·03	131·50	106·40	108·70
30	m	77·22	116·84	97·79	178·36	107·19	112·60	103·63	124·30	95·10	111·70
34	m	82·30	123·44	95·50	115·06	98·30	113·46	99·06	106·30	88·80	86·60
	$\bar{m}$	84·08	123·60	97·20	113·73	102·20	114·30	101·83	121·10	102·20	103·83
$\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$		86·06	113·98	93·28	119·38	109·65	129·85	100·32	129·81	107·83	105·46
$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$		1·99	−9·63	−3·92	5·65	7·45	15·55	−1·51	8·71	5·63	1·64
Set 2 (9 seasons)											
1	M	84·07	80·77	92·20	96·39	100·58	132·08		112·20	103·60	102·60
5	M	85·09	106·68	79·82	102·36	108·46	128·70		145·70	126·30	113·60
13	M	92·46	119·38	116·08	134·37	137·92	134·62		134·80	122·70	130·45
38	M	65·02	57·66	65·52	106·17	86·61	107·52		111·70	101·40	92·20
	$\bar{M}$	81·66	91·12	88·41	109·82	108·39	125·73		126·10	113·50	109·71
11	m	86·92	97·79	100·01	120·40	113·03	133·78		136·05	115·60	121·50
22	m	112·40	153·82	120·46	146·81	128·52	148·59		163·70	133·80	149·00
23	m	111·25	157·48	127·76	154·94	143·00	160·02		150·62	108·30	109·00
27	m	87·38	121·16	55·88	112·01	98·30	123·62		120·40	88·70	98·70
	$\bar{m}$	99·49	132·56	101·03	133·54	120·71	141·50		142·69	111·60	119·55
$\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$		90·57	111·84	94·72	121·68	114·55	133·62		134·40	112·55	114·63
$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$		−8·91	−20·72	−6·31	−11·86	−6·16	−7·89		−8·30	0·95	−4·92

The data in the present paper therefore consists of the mean final height (in centimetres) of two sets of eight inbred lines of $N.$ *rustica* from the larger set of 29 in each of 10 years and in each of 9 of these years respectively (see table 1). Each set of lines can be subdivided into two groups of four on the basis of the single gene difference, non-mophead, M, and mophead, m. For both sets the mean in each season of the non-mophead, $\bar{M}$, and of the mophead, $\bar{m}$, lines are also given in table 1 together with their average sum, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$, and average difference, $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$.

The sowing date and the seasonal average (over 4 months, May, June, July and August) of six environmental factors are given in table 2 for each

<div style="text-align:center">TABLE 2</div>

The sowing date and seasonal average (over May, June, July and August) of six environmental factors (see Section 2 for key to abbreviations)

Year	Sowing date	RH6	RH12	RAIN	SUN	MXT	MNT
1949	16/5	86·425	61·500	0·2371	6·6525	67·975	51·100
1950	4/5	88·850	67·275	0·3240	5·8425	65·600	51·325
1957	1/4	85·350	62·350	0·3613	5·5325	65·075	50·925
1958	10/4	89·625	67·600	0·4090	4·8550	64·550	51·325
1959	25/4	84·225	59·975	0·2162	6·7200	68·350	52·600
1961	6/6	86·000	62·050	0·2512	5·9150	64·450	50·500
1967*	9/4	88·325	65·800	0·3071	5·5150	64·900	50·925
1968	9/5	89·250	68·100	0·5148	4·4600	63·400	50·350
1969	23/4	88·625	68·175	0·4432	5·8450	65·625	51·550
1970	21/4	86·990	66·365	0·0890	5·9325	66·715	51·460

* 1967 is the year omitted for the second set of eight inbred lines.

year. The six environmental factors are percentage relative humidity at 6 a.m., RH6; percentage relative humidity at 12 p.m., RH12; daily rainfall in inches, RAIN; daily sunshine in hours, SUN; maximum air temperature in ° F, MXT; and minimum air temperature in ° F., MNT. These were computed from the records of the Edgbaston Observatory, Birmingham, which is situated 2 miles from the experimental field.

3. THE MODEL AND ITS RELATIONSHIP TO THE PRINCIPAL COMPONENTS ANALYSIS

In order to gain insight into the possible relationships between the regression of the mean performance against the non-independent environmental measure and the principal components analysis of the sum of squares —sum of products matrix of each set of non-mophead and mophead lines of *N. rustica* over environments, the following model may be considered.

<div style="text-align:center">Model</div>

<div style="text-align:center">Environment</div>

Line	1	2
M_1	$(1+\beta_1)\epsilon+(1+\beta_2)\delta$	$-(1+\beta_1)\epsilon-(1+\beta_2)\delta$
M_2	$(1-\beta_1)\epsilon+(1-\beta_2)\delta$	$-(1-\beta_1)\epsilon-(1-\beta_2)\delta$
m_1	$(1+\beta_1)\epsilon-(1+\beta_2)\delta$	$-(1+\beta_1)\epsilon+(1+\beta_2)\delta$
m_2	$(1-\beta_1)\epsilon-(1-\beta_2)\delta$	$-(1-\beta_1)\epsilon+(1-\beta_2)\delta$

The model gives the contrasting environmental and genotype-environmental interaction components of performance of two non-mophead, M_1 and M_2, and two mophead, m_1 and m_2, lines in each of two environments. The mean over all lines and environments and the genetic component of each line have been omitted from the model, since being constant for each line over environments, they will make no contribution to the sum of squares of the lines or to the sum of products between them over environments. The notation follows that of Perkins and Jinks (1968a) where:

$+\epsilon$ and $-\epsilon$ are the environmental components of the first and second environments respectively, derived in practice as the deviation of the average performance of all lines in an environment from the mean performance over

all lines and environments, *i.e.* the non-independent environmental component, ϵ_j, of Perkins and Jinks (1968a).

$1 + \beta_1$ and $1 - \beta_1$ are the two possible linear regression coefficients a line may have when, in practice, its mean performance in each environment is regressed against the corresponding value of ϵ, *i.e.* the $1 + \beta_i$ of Perkins and Jinks (1968a). In the model, $1 + \beta_1$ and $1 - \beta_1$ have each been allocated to both a non-mophead, M, and a mophead, m, line to ensure that which of the two contrasting regression coefficients a line may show with respect to ϵ is independent of whether it is a non-mophead or a mophead line.

$+\delta$ and $-\delta$ represent the opposing deviations from linear regression of the non-mophead and mophead lines respectively (see Perkins and Jinks, 1968b). In the model, *i.e.* after the overall mean and the genetic component of each line have been removed, the value of δ in each environment is equal to half the difference between the average of the non-mophead lines and the average of the mophead lines.

$1 + \beta_2$ and $1 - \beta_2$ represent the two contrasting coefficients a line may have with respect to $+\delta$ or $-\delta$.

A sum of squares-sum of products matrix can be derived algebraically for the four lines in the model over environments. The matrix so derived is given in the following table having extended the model to give the sum of squares or products of the lines over many environments ($\sum$ = the sum over environments) and having equated the expected value of $\sum \epsilon \delta$ to zero since ϵ and δ are obtained from the two independent orthogonal comparisons between the performances of the four lines, M_1, M_2, m_1 and m_2, in each environment,

$$\tfrac{1}{4}M_1 + \tfrac{1}{4}M_2 + \tfrac{1}{4}m_1 + \tfrac{1}{4}m_2 \text{ and } \tfrac{1}{4}M_1 + \tfrac{1}{4}M_2 - \tfrac{1}{4}m_1 - \tfrac{1}{4}m_2$$

respectively.

<div align="center">Sum of squares-sum of products matrix*</div>

Lines	M_1	M_2
M_1	$(1+\beta_1)^2\Sigma\epsilon^2 + (1+\beta_2)^2\Sigma\delta^2$	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 + (1+\beta_2)(1-\beta_2)\Sigma\delta^2$
M_2	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 + (1+\beta_2)(1-\beta_2)\Sigma\delta^2$	$(1-\beta_1)^2\Sigma\epsilon^2 + (1-\beta_2)^2\Sigma\delta^2$
m_1	$(1+\beta_1)^2\Sigma\epsilon^2 - (1+\beta_2)^2\Sigma\delta^2$	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 - (1+\beta_2)(1-\beta_2)\Sigma\delta^2$
m_2	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 - (1+\beta_2)(1-\beta_2)\Sigma\delta^2$	$(1-\beta_1)^2\Sigma\epsilon^2 - (1-\beta_2)^2\Sigma\delta^2$

Lines	m_1	m_2
M_1	$(1+\beta_1)^2\Sigma\epsilon^2 - (1+\beta_2)^2\Sigma\delta^2$	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 - (1+\beta_2)(1-\beta_2)\Sigma\delta^2$
M_2	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 - (1+\beta_2)(1-\beta_2)\Sigma\delta^2$	$(1-\beta_1)^2\Sigma\epsilon^2 - (1-\beta_2)^2\Sigma\delta^2$
m_1	$(1+\beta_1)^2\Sigma\epsilon^2 + (1+\beta_2)^2\Sigma\delta^2$	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 + (1+\beta_2)(1-\beta_2)\Sigma\delta^2$
m_2	$(1+\beta_1)(1-\beta_1)\Sigma\epsilon^2 + (1+\beta_2)(1-\beta_2)\Sigma\delta^2$	$(1-\beta_1)^2\Sigma\epsilon^2 + (1-\beta_2)^2\Sigma\delta^2$

The eigenvalues or latent roots of this matrix, when derived algebraically, are found to have the following values:

$$\lambda_1 = [2(1+\beta_1)^2 + 2(1-\beta_1)^2]\textstyle\sum\epsilon^2$$
$$\lambda_2 = [2(1+\beta_2)^2 + 2(1-\beta_2)^2]\textstyle\sum\delta^2$$
$$\lambda_3 = 0$$
$$\lambda_4 = 0$$

The terms $[2(1+\beta_1)^2 + 2(1-\beta_1)^2]$ and $[2(1+\beta_2)^2 + 2(1-\beta_2)^2]$ may each

* The third and fourth columns of this 4×4 matrix are tabulated under the first and second columns, respectively.

be generalised as the sum of regression coefficients squared over lines, $\sum(1+\beta_1)^2$ and $\sum(1+\beta_2)^2$ respectively. The eigenvalues then become:

$$\lambda_1 = \sum(1+\beta_1)^2\sum\epsilon^2$$
$$\lambda_2 = \sum(1+\beta_2)^2\sum\delta^2$$
$$\lambda_3 = 0$$
$$\lambda_4 = 0$$

and the matrix, $\mathbf{A}$, of the corresponding normalised eigenvectors for just the first two non-zero eigenvalues, λ_1 and λ_2, is found to be:

Matrix of normalised eigenvectors, $\mathbf{A}$

		Line		
Eigenvalue	M_1	M_2	m_1	m_2
λ_1	$\dfrac{(1+\beta_1)}{\sqrt{\Sigma(1+\beta_1)^2}}$	$\dfrac{(1-\beta_1)}{\sqrt{\Sigma(1+\beta_1)^2}}$	$\dfrac{(1+\beta_1)}{\sqrt{\Sigma(1+\beta_1)^2}}$	$\dfrac{(1-\beta_1)}{\sqrt{\Sigma(1+\beta_1)^2}}$
λ_2	$\dfrac{(1+\beta_2)}{\sqrt{\Sigma(1+\beta_2)^2}}$	$\dfrac{(1-\beta_2)}{\sqrt{\Sigma(1+\beta_2)^2}}$	$\dfrac{(1+\beta_2)}{\sqrt{\Sigma(1+\beta_2)^2}}$	$\dfrac{(1-\beta_2)}{\sqrt{\Sigma(1+\beta_2)^2}}$

It is possible to obtain a matrix, $\mathbf{Y}$, of scores in each environment for the two principal components, p_1 and p_2, whose sum of squares over environments are equal to the two non-zero eigenvalues, λ_1 and λ_2, respectively. If $\mathbf{X}$ is the matrix of original scores of the four lines, M_1, M_2, m_1, and m_2, in each environment, which for each line are in the form of deviations from the line average over environments as given in the model, and if $\mathbf{A}$ is the matrix of normalised eigenvectors of the non-zero eigenvalues then:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

For our simple model the 4×2 matrix of original scores, $\mathbf{X}$, is premultiplied by the 2×4 matrix of normalised eigenvectors of the two non-zero eigenvalues, $\mathbf{A}$, to give a 2×2 matrix of principal component scores in each environment, $\mathbf{Y}$, $i.e.$

$$\underset{2 \times 2}{\mathbf{Y}} = \underset{2 \times 4}{\mathbf{A}}\ \underset{4 \times 2}{\mathbf{X}},$$

whose elements are found to equal the following:

Matrix of principal component scores, $\mathbf{Y}$.

	Environment	
Component	1	2
p_1	$\sqrt{\Sigma(1+\beta_1)^2}\epsilon$	$-\sqrt{\Sigma(1+\beta_1)^2}\epsilon$
p_2	$\sqrt{\Sigma(1+\beta_2)^2}\epsilon$	$-\sqrt{\Sigma(1+\beta_2)^2}\delta$

It is clear that the first principal component is related to the general response of all lines to environmental differences as measured by the non-independent environmental component, ϵ, while the second component is related to the specific difference in response of the non-mophead and mophead lines as measured by δ.

If the model given accounts for most of the variation of the non-mophead and mophead lines over environments a number of relationships can be predicted between the principal components analysis and the analyses of linear regression, against the non-independent environmental measure, and of deviations from linear regression already referred to.

Provided that most of the variation of each line over environments is accounted for by the regression of its mean performance in each environment against the non-independent environmental measure, the sum of squares of the first principal component, λ_1, will be equal to the sum over lines of their individual regression sums of squares. Likewise, provided that most of the residual variation is accounted for by a further regression against the average difference in performance of the non-mophead and mophead lines in each environment, the sum of squares of the second principal component, λ_2, will be equal to the sum over lines of their individual remainder sums of squares when taken from the first regression analyses.

The weight of the ith line in the normalised eigenvector of the first eigenvalue, a_{1i}, is expected to equal the value of its regression coefficient, $1 + \beta_{1i}$, against ϵ divided by the square root of the sum over lines of the regression coefficients squared *i.e.* $(1 + \beta_{1i})/\sqrt{\sum_i (1 + \beta_{1i})^2}$.

The score of the first principal component in the jth environment, y_{1j}, is expected to equal the value of the corresponding non-independent environmental component, ϵ_j, multiplied by $\sqrt{\sum (1 + \beta_{1i})^2}$, *i.e.* $\sqrt{\sum (1 + \beta_{1i})^2}\epsilon_j$.

The sign of the weight of each line in the eigenvector of the second eigenvalue is expected to distinguish the non-mophead lines (positive sign) from the mophead lines (negative sign) as already achieved by the analysis of deviations from linear regression against ϵ_j (Perkins and Jinks, 1968b). The magnitude of the weight of the ith line, a_{2i}, is, according to its sign, a measure of the degree of non-mopheadedness or mopheadedness of the line and, as for the first eigenvector, is equal to the value of the line's regression coefficient, $1 + \beta_{2i}$, against δ divided by $\sqrt{\sum_i (1 + \beta_{2i})^2}$, *i.e.* $(1 + \beta_{2i})/\sqrt{\sum_i (1 + \beta_{2i})^2}$.

The score of the second principal component in the jth environment, y_{2i}, will equal the value of the corresponding average difference between the non-mophead and mophead lines, δ_j, multiplied by $\sqrt{\sum_i (1 + \beta_{2i})^2}$, *i.e.* $\sqrt{\sum_i (1 + \beta_{2i})^2}\delta_j$.

4. RESULTS

(a) *The principal components analysis*

As described for the model in the previous section, an 8×8 sum of squares-sum of products matrix, **S**, was derived for the mean final heights of the eight inbred lines (table 1) over the ten seasons for set 1 and over the nine seasons for set 2 and a principal components analysis was applied to each of these two matrices.

Because the scores of the eight inbred lines in each set are not interdependent over environments eight non-zero eigenvalues are obtained for each **S** matrix. These eigenvalues represent the sum of squares of the corresponding principal components over environments and each can be converted to a variance by dividing by the number of environments minus one

TABLE 3

The variance of each of the eight components from a principal components analysis of the variation and covariation of each set of eight inbred lines over seasons, their degrees of freedom and probability when tested against the error variance

	Set 1			Set 2		
Component	Variance	d.f.	P	Variance	d.f.	P
1	1916·1	9	***	1836·1	8	***
2	270·2	9	***	526·3	8	***
3	171·3	9	***	218·1	8	***
4	72·7	9	***	101·5	8	***
5	53·4	9	***	69·7	8	***
6	27·9	9	**	36·3	8	**
7	11·4	9	n.s.	32·4	8	**
8	1·3	9	n.s.	1·2	8	n.s.
Error	9·4	520		11·3	484	

n.s.; probability is non-significant
**; probability = 0·001 − 0·01
***; probability < 0·001

for each set, *i.e.* nine and eight respectively (see table 3). For each set, an error variance can be derived for the mean final height of each line in each environment from the mean variance of individuals within lines and environments. It can be seen in table 3 that when tested against this error the variances of the first six principal components in set 1 and of the first seven components in set 2 are significant. However, as indicated in table 4, the

TABLE 4

The correlations with the first and second principal components (p_1 and p_2) and the coefficients $(1 + \hat{\beta}_{1t}'s)$ of regression against $\hat{e}_j$ for the eight inbred lines, over ten seasons in set 1 and nine seasons in set 2

Set	Variety	M/m	Correlation with component		Coefficient of regression against $\hat{e}_j$
			p_1	p_2	$1 + \hat{\beta}_{1t}$
	12	*M*	0·79	0·37	1·09
	15	*M*	0·95	0·04	0·81
	35	*M*	0·98	0·09	1·57
	42	*M*	0·96	0·13	1·65
1 (10 seasons)	2	*m*	0·32	−0·23	0·62
	21	*m*	0·72	−0·54	0·80
	30	*m*	0·79	−0·43	0·81
	34	*m*	0·62	−0·60	0·66
Percentage of total variance			75·90	10·70	

86·60

Set	Variety	M/m	p_1	p_2	$1 + \hat{\beta}_{1t}$
	1	*M*	0·76	0·40	0·79
	2	*M*	0·89	0·24	1·22
	13	*M*	0·81	0·07	0·80
	38	*M*	0·83	0·47	1·13
2 (9 seasons)	11	*m*	0·92	0·32	1·02
	22	*m*	0·85	−0·23	0·96
	23	*m*	0·60	−0·70	0·91
	27	*m*	0·81	−0·43	1·17
Percentage of total variance			65·07	18·65	

83·72

variances of the first two principal components, p_1 and p_2, jointly account for
87 and 84 per cent. of the total variation in mean final height of the eight
inbred lines in sets 1 and 2 respectively, over environments.

In order to identify the first two components and compare them over sets,
the correlation of each line with each component rather than the weight of
each line in the normalised eigenvector of each component is given in table 4.
This shows that the first component represents a general response of all lines
to environmental differences since, for both sets, each line has a high
positive correlation with this component, with the exception of variety 2 in
set 1. The second component obviously represents a specific difference in
response to environmental differences of the non-mophead and mophead
lines since, in both sets, all the non-mophead lines are positively correlated
with it and all the mophead lines, with the exception of variety 11 in set 2,
are negatively correlated. The identity of these two components was exactly
predicted by the model.

To what extent are all the other predictions found to hold?

The mean final height of each inbred line in each environment was
regressed against the corresponding non-independent environmental
measure, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ in table 1, for both sets of lines. The regression coeffi-
cient, $1+\hat{\beta}_{1i}$, of each line and its standard deviation are given in table 4.
The weight of each line in the normalised eigenvector of the first component
was predicted from these coefficients using the formula, derived from the
principal components analysis of the model in the previous section,
$(1+\beta_{1i})/\sqrt{\sum(1+\beta_{1i})^2}$. The correlation over inbred lines between the
observed and expected weights was found to be 0·998 for both sets of lines.

The score of the first principal component in each environment was also
predicted for both sets by multiplying the estimate of the non-independent
environmental component, $\hat{\epsilon}_j$, in each environment by $\sqrt{\sum(1+\hat{\beta}_{1i})^2}$, i.e.
$\sqrt{(\sum 1+\hat{\beta}_{1i})^2}\hat{\epsilon}_j$, which corresponds with the formula derived from the com-
ponents analysis of the model in the previous section. The correlation over
environments between the observed and expected component scores is again
high for both sets, having a value of 0·991 in set 1 and a value of 0·999 in set 2.

The sum of squares of the first principal component, λ_1, is expected to
equal the sum over lines of their regression, against $\hat{\epsilon}_j$, sums of squares.
Their respective values are 17245·0994 and 16976·6338 in set 1 and
14688·5426 and 14639·0054 in set 2. Thus once more there is a good corres-
pondence between observed and expected, consistently in both sets.

The regression of the mean final height of each inbred line in each
environment against the corresponding average difference of the non-mop-
head and mophead lines, $\hat{\delta}_j$, has not been previously attempted. The
second principal component, however, has successfully distinguished
between the response to environmental differences of the non-mophead and
mophead lines in both sets. This was expected on both the basis of a
previous analysis of the deviations from linear regression, against $\hat{\epsilon}_j$, of these
lines as part of a larger set (Perkins and Jinks, 1968b) and on the basis of the
results of the principal components analysis of the model in the previous
section. The prediction of the weight of each line in the normalised eigen-
vector of the second component and of the score of this component in each
environment can be expected to be just as good as those for the first principal
component.

Provided that the residual variation after the regression of the score of each inbred line in each environment against the corresponding value of $\hat{\epsilon}_j$ is largely explained by a further regression against the corresponding value of $\hat{\delta}_j$, the sum of squares of the second principal component, λ_2, over environments is expected to equal the sum over lines of their remainder sums of squares when taken from the first regression analyses. The respective values of the sums of squares of the second component and of the total remainder are 2432·1337 and 5743·0442 in set 1 and 4210·1700 and 7901·3042 in set 2. The sum of squares of the first principal component over environments in both sets has been shown to be close in value to the total sum of squares of regression of the mean final heights of the inbred lines in each environment against the corresponding non-independent environmental component, $\hat{\epsilon}_j$. The sum of squares of the second component is expected to be smaller in magnitude than that of the total remainder after fitting the first regression, against $\hat{\epsilon}_j$, to the mean final heights of the inbred lines in each environment since four further components in set 1 and five further components in set 2 were shown to also make a significant contribution to this residual variation. Although specific combinations of the inbred lines are correlated with some of these further components it has not been possible to identify them from the known characteristics of the inbred lines concerned.

(b) *The multiple regression analyses*

The intention was to carry out a multiple regression analysis of the performance of each inbred line in each season against a corresponding set of physical measures of these seasons. The physical measures used were those described in section 2 and given in table 2 (omitting their scores in 1967 for the inbred lines in the second set). However, not all the variation over seasons in the mean final heights of most of the inbred lines could be explained in this way. Hence, other environmental factors, for which we have no information, or higher powered derivatives of the existing physical measures were required. In order to use higher powered derivatives it was first necessary to reduce the total number of physical measures since the sum of their linear and quadratic derivatives alone would exceed in number the total degrees of freedom available for fitting and testing the regression (number of environments minus one).

The six climatic factors, RH6, RH12, RAIN, SUN, MXT and MNT, were therefore summarised into three components obtained from a principal components analysis of their correlation matrix over seasons. Two such matrices were derived by correlating the scores of the climatic factors in all possible pairs over all ten seasons and then over nine of these seasons (omitting their scores in 1967) in order to correspond with the first and second sets of inbred lines respectively. Because most of the climatic factors have been measured on different, unrelatable, scales the correlation matrices rather than their variance-covariance counterparts were computed. Use of the correlation matrix simply means that all factors have been effectively rescaled to have the same unitary variance over environments (*c.f.* the diagonal of a correlation matrix). This rescaling is achieved by dividing the score of each climatic factor in each season by the standard deviation obtained from the square root of the variance of the factor over environments.

For both correlation matrices six non-zero eigenvalues were obtained, representing the variances of six corresponding principal components over

environments. However, for both sets, as indicated in table 5, the first three principal components jointly account for 95 per cent. of the total variance of all components over environments. The total variance of these three components was therefore accepted as summarising the total standardised variance of the six original climatic factors over seasons.

The correlation between the standardised scores of each climatic factor and the scores of each of the three components, p_1, p_2 and p_3, over seasons in the two sets are given in table 5. The correlations are consistent over both sets. RH6, RH12 and RAIN are associated with the first component in the opposite direction to SUN, MXT and MNT. This component reveals the common association of high temperature and low relative humidity and rainfall when the sun is shining and, conversely, of low temperature and high relative humidity and rainfall when the sun is not shining due to cloudiness. Since SUN is also most highly correlated with this component it will be

TABLE 5

The correlations of the six environmental factors with the first three principle components of each set over seasons

		Component		
Set	Environmental factor	p_1 " sun "	p_2 " humidity "	p_3 " rain "
	RH6	0·78	0·57	−0·23
	RH12	0·83	0·54	−0·03
1 (10 seasons)	RAIN	0·75	−0·17	0·61
	SUN	−0·94	0·09	−0·07
	MXT	−0·91	0·33	0·04
	MNT	−0·66	0·57	0·43
Percentage of total variance		66·82 +	18·09 + 10·23	= 95·14
	RH6	0·77	0·58	−0·23
	RH12	0·83	0·54	−0·03
2 (9 seasons)	RAIN	0·76	−0·17	0·59
	SUN	−0·94	0·09	−0·07
	MXT	−0·91	0·33	0·03
	MNT	−0·66	0·58	0·43
Percentage of total variance		66·70 +	18·44 + 9·96	= 95·10

referred to as the " sun " component. RH6, RH12 and MNT are the factors most highly correlated with the second component. For the purposes of description it will therefore be referred to as the " humidity " component. RAIN and to a lesser extent MNT are most highly correlated with the third component which, therefore, will be referred to as the " rain " component.

The scores on the three principal components of the ten and nine seasons, respectively, were obtained in much the same way as those for the two components of the model in section 3. That is, for each set, the matrix of the three normalised eigenvectors (corresponding with the first three eigenvalues) was post-multiplied by the matrix of the standardised scores of the six original climatic factors in each season in the form of deviations from the mean of each factor over environments. The component scores in each season are given in table 6 under the column headings, p_1, p_2 and p_3, respectively, for both sets. It appeared, from an examination of the original meteorological reports, that whether a season has a high or low score for each of these components depends upon the predominant disposition of cyclones and anticyclones for that season.

For both sets of seasons, the standardised deviation of each sowing date (when scored as the number of days after an arbitrary date in April) from the overall mean, s, is given in table 6. In both sets, the deviations have been divided by the standard deviation of sowing dates over seasons in order to make them comparable with the corresponding set of six standardised climatic factors, summarised by the three principal components. The quadratic derivatives of sowing date and the three climatic components, s_1^2, p_1^2, p_2^2 and p_3^2, are also given, for both sets, in table 6 as the deviations from their respective means over seasons. The linear and quadratic derivatives of sowing date and of the three principal components of the six climatic factors in each season given in table 5 constitute the two new sets of physical

TABLE 6

The standardised scores of sowing date(s) and of the first three principal components (p$_1$, p$_2$ and p$_3$) of the six environmental factors and their squared derivatives (s^2, p$_1^2$, p$_2^2$ and p$_3^2$) in each of ten seasons for set 1 and nine seasons for set 2

					Parameter			
Year	s	s^2	p_1	p_1^2	p_2	p_2^2	p_3	p_3^2
Set 1 (10 seasons)								
1949	0·9386	−0·0190	2·1562	1·0410	−0·3091	−0·8814	−0·4198	−0·3760
1950	0·3298	−0·7912	−0·4818	−3·3761	0·8559	−0·2443	−0·0835	−0·5452
1957	−1·3445	0·9077	0·2113	−3·5636	−1·5145	1·3168	0·4217	−0·3744
1958	−0·8879	−0·1116	−1·9061	0·0250	0·6858	−0·5066	0·4228	−0·3734
1959	−0·1268	−0·8839	3·8198	10·9827	0·2196	−0·9287	1·1241	0·7114
1961	2·0040	3·1160	0·3021	−3·5169	−1·6742	1·8260	−0·8058	0·0971
1967	−0·9386	−0·0190	−0·7590	−3·0321	−0·0464	−0·9747	−0·4191	−0·3766
1968	0·5835	−0·5595	−3·3249	7·4468	−0·6087	−0·6064	0·3002	−0·4621
1969	−0·2283	−0·8479	−0·7865	−2·9896	0·9808	−0·0149	0·8827	0·2270
1970	−0·3298	−0·7912	0·7691	−3·0167	1·4112	1·0146	−1·4229	1·4724
Set 2 (9 seasons)								
1949	0·8332	−0·1947	1·9712	0·3282	−0·2988	−0·8939	−0·4434	−0·3346
1950	0·2252	−0·8382	−0·5362	−3·2699	0·8080	−0·3303	−0·1311	−0·5140
1957	−1·4468	1·2043	0·1163	−3·5439	−1·4451	1·1051	0·3613	−0·4007
1958	−0·9908	0·0928	−1·8918	0·0215	0·6502	−0·5604	0·3445	−0·4125
1959	−0·2308	−0·8356	3·5469	9·0231	0·2110	−0·9387	1·0669	0·6071
1961	1·8972	2·7105	0·2049	−3·5154	−1·6032	1·5871	−0·7979	0·1054
1968	0·4785	−0·6599	−3·2416	6·9506	−0·5841	−0·6420	0·2025	−0·4902
1969	−0·3322	−0·7785	−0·8286	−2·8708	0·9303	−0·1177	0·7758	0·0707
1970	−0·4335	−0·7010	0·6590	−3·1231	1·3321	0·7913	−1·3784	1·3688

measures of the environment. Their scores in the two sets of seasons correspond with those of the two sets of inbred lines (table 1), respectively.

Twelve variates were derived for each set of data from the mean final heights of the eight inbred lines in each season. These were the four non-mophead and four mophead lines taken individually together with $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$, in the form of deviations in each season from their respective means over seasons. The score of each variate in each season was regressed, in a multiple regression analysis, against the corresponding set of environmental parameters, the physical measures of these seasons (table 6). The mean final heights of many of the inbred lines, particularly in the earlier seasons, are based on few individuals, since they were primarily grown for their maintenance by inbreeding. Their variances are therefore based on too few degrees of freedom to provide reliably estimated weights for a weighted

least squares analysis. An unweighted analysis has therefore been used. The results of these regression analyses are given in table 7. After deduction of the mean, only seven degrees of freedom to fit a seven parameter model and one degree of freedom to test it are available for the second set of data since observations are available for nine seasons only. The deviation of each line, $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ in the second set has, therefore, been regressed against the complete set of parameters several times, omitting one parameter at a time. The parameter which when omitted in general gave the best fit for the non-mophead lines and $\bar{M}$, the mophead lines and $\bar{m}$, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$, respectively, is indicated in table 7 and the corres-

TABLE 7

Multiple regression analysis for each inbred line, $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ in the two sets. The complete set of parameters was used for set 1 but one parameter had to be omitted for set 2. The one which when omitted gave the best fit for the non-mopheads and mopheads, in general, and for $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ is indicated

Set 1 (10 seasons)

Variety	M/m	Model	Reg M.S. (8 d.f.)	P	Rem M.S. (1 d.f.)	P	Error M.S.	Error d.f.	$\frac{\sigma^2_{reg} \times 100}{\sigma^2_{rem} + \sigma^2_{reg}}$ %
12	M	Complete	464·5	***	107·8	***	5·8	54	82
15	M	Complete	174·9	***	32·4	n.s.	10·6	67	100
35	M	Complete	599·0	***	238·8	***	10·8	66	72
42	M	Complete	735·4	***	4·8	n.s.	18·6	68	100
	$\bar{M}$	Complete	408·1	***	71·1	*	11·4	255	86
2	m	Complete	139·6	***	184·7	***	7·1	58	43
21	m	Complete	189·3	***	417·2	***	5·9	67	31
30	m	Complete	162·6	***	402·4	***	9·4	69	28
34	m	Complete	148·7	***	445·0	***	7·0	71	24
	$\bar{m}$	Complete	115·4	***	352·9	***	7·3	265	24
$\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$		Complete	208·4	***	185·4	***	9·4	520	53
$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$		Complete	53·4	***	26·9	n.s.	9·4	520	100

Set 2 (9 seasons)

Variety	M/m	Model	Reg M.S. (7 d.f.)	P	Rem M.S. (1 d.f.)	P	Error M.S.	Error d.f.	$\frac{\sigma^2_{reg} \times 100}{\sigma^2_{rem} + \sigma^2_{reg}}$ %
1	M	$-p_2^2$	269·8	***	4·2	n.s.	10·3	60	100
5	M	$-p_2^2$	494·9	***	36·9	n.s.	14·4	62	100
13	M	$-p_2^2$	197·0	***	269·4	***	20·9	58	41
38	M	$-p_2^2$	457·5	***	269·6	***	7·7	63	63
	$\bar{M}$	$-p_2^2$	286·6	***	1·1	n.s.	13·3	243	100
1i	m	$-p_2^2$	308·7	***	8·6	n.s.	13·6	61	100
22	m	$-p_1$	325·1	***	2·2	n.s.	10·6	59	100
23	m	$-p_1$	448·7	***	702·9	***	10·3	62	39
27	m	$-p_1$	537·9	***	40·8	***	2·9	59	93
	$\bar{m}$	$-p_1$	301·7	***	9·3	n.s.	9·3	241	100
$\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$		$-p_3$	254·9	***	7·1	n.s.	11·3	484	100
$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$		$-s$	39·0	***	0·2	n.s.	11·3	484	100

n.s.; probability is non-significant
*; probability = 0·01 − 0·05
***; probability < 0·001.

ponding regression analysis is given. Two other items have been included in table 7 for both sets. First, the significances of the regression and remainder items have been included when tested against the appropriate error mean square which is the mean variance of replicate individuals within environments. A fixed model has been assumed for the regression mean squares since the inbred lines are a selected sample. Secondly, the percentage of the sum of the regression plus remainder σ^2's, $\sigma^2_{reg} + \sigma^2_{rem}$, accounted for by the regression σ^2, σ^2_{reg}, has also been included. In table 8, the values of the

TABLE 8

The significant regression coefficients for the non-mophead and mophead lines, $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ of the two sets with non-significant remainder mean squares in table 7

	Set 1		Set 2			
M/m	M	M	M	M	$\bar{M}$	m
Variety	15	42	1	5		11
Parameter						
s	4·4	7·9	8·4	17·5	8·1	4·8
s^2	14·2	27·0	7·8	n.s.	8·4	10·0
p_1	−4·1	−8·7	−5·7	−10·6	−7·7	−8·2
p_1^2	1·8	4·4	n.s.	n.s.	1·2	1·7
p_2	13·2	27·8	n.s.	n.s.	n.s.	n.s.
p_2^2	n.s.	n.s.	—	—	—	—
p_3	4·1	5·6	10·2	16·7	11·4	6·5
p_3^2	n.s.	n.s.	24·1	31·1	23·8	23·0

	Set 2			Set 1	Set 2
M/m	m	$\bar{m}$	$\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$	$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$	$\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$
Variety	22				
Parameter					
s	6·4	7·8	7·4	n.s.	—
s^2	n.s.	10·8	9·6	6·4	n.s.
p_1	—	—	−2·7	−2·3	−4·3
p_1^2	7·6	7·5	5·3	n.s.	−3·4
p_2	37·3	37·8	24·8	n.s.	−18·1
p_2^2	49·2	38·9	25·0	−5·4	−21·1
p_3	−9·4	−5·7	—	4·4	8·8
p_3^2	−47·5	−41·9	−17·0	12·8	35·0

significant regression coefficients are listed for those variates which in table 7 had non-significant remainder mean squares.

5. DISCUSSION

In section 3, a model representing the contrasting environmental and genotype-environmental interaction components of a pair of non-mophead and a pair of mophead lines in each of two environments was constructed. A principal components analysis was applied to the sum of squares-sum of products matrix of these lines over environments. From this analysis it was possible to predict the variance over environments, the score in each environment and the weight of each line in the normalised eigenvector of the first

principal component in terms of estimates of parameters which can be obtained from certain regression analyses (Perkins and Jinks, 1968a, b).

In section 4a, the principal component and regression analyses were repeated on the mean final heights of two sets of four non-mophead and four mophead lines in each of ten and nine seasons respectively. The applicability of the model to the data was tested by comparing the observed and expected properties of the first two principal components.

These comparisons showed that the sum of squares over seasons and the score in each season of the first principal component are directly related to those of the non-independent environmental component, $\hat{\epsilon}_j$. This environmental component represents the general response of all lines to an environment since it is equal to the deviation of the average of all lines in an environment from the overall mean. The weight of each line in the normalised eigenvector of the first principal component is directly related to the linear regression coefficient, $1 + \hat{\beta}_{1i}$, obtained by regressing the performance of the line in each environment against the corresponding estimate of $\hat{\epsilon}_j$.

The score of the second component in each environment is directly related to the specific difference in response to an environment of the non-mophead and mophead lines, estimated from the average difference in their mean performances in each environment, $\hat{\delta}_j$. As predicted, the non-mophead and mophead lines can be distinguished by the sign of their weight (positive and negative, respectively) in the normalised eigenvector of this component. One mophead line in the second set, variety 11, was exceptional in showing a positive instead of negative sign.

The variance of the first two principal components over seasons jointly accounted for 87 and 84 per cent. of the total in sets 1 and 2 respectively. The variance over environments of the second component however did not adequately explain all that remaining after taking out the first component since in both sets several further but unidentifiable components were each shown to make a significant contribution.

In the present context, two identifiable characteristics of performance for the eight inbred lines of each set in the different seasons are related to the first two principal components, i.e. the environmental component, $\hat{\epsilon}_j$, which can be derived from $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ and the difference in response of the non-mophead and mophead lines, $\hat{\delta}_j$, equal to $\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$. Both of these characteristics can, therefore, be directly investigated further. This may take the form of an analysis of variance, where the different environments are derived from a deliberate hierarchical design, such as different sowing dates in each of several seasons, or a factorial design, such as all possible combinations of the presence or absence of a number of fertiliser treatments. Alternatively, the further analysis may take the form, as in the present context, of a multiple regression against physical measures of the environment. Even if a principal component cannot be identified, its scores in the different environments may still be subjected to either of the two kinds of further analysis described, whichever is appropriate, and this in itself may aid the identification of the component.

The multiple regression of the environmental scores against a set of physical environmental measures for each of the four non-mophead and four mophead lines and their derivatives, $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$, in each set are reported in section 4b. A number of conclusions may be drawn from the results of the regression analyses given in table 7. The remainder mean

squares of half of the non-mophead lines in both sets are non-significant. Of the remaining four non-mophead varieties the regression $\hat{\sigma}^2$ still accounts for over 60 per cent. of the total $\hat{\sigma}^2$, $(\hat{\sigma}^2_{\text{reg}} + \hat{\sigma}^2_{\text{rem}})$ with the exception of variety 13 in set 2 (41 per cent.). The regression $\hat{\sigma}^2$ for the mean of the four non-mophead lines in each environment, $\bar{M}$, accounts for 86 and 100 per cent. of the total in sets 1 and 2 respectively. Conversely, the mophead lines, with three exceptions, have large significant remainder mean squares so that their regression $\hat{\sigma}^2$'s account for 43 per cent. or less of the total. This is reflected by the regression $\hat{\sigma}^2$ for the mean of the four mophead lines in each environment, $\bar{m}$ in set 1 which accounts for only 24 per cent. of the total but not by the regression $\hat{\delta}^2$ of $\bar{m}$ in set 2, the set which contains the three exceptional varieties, 11, 22 and 27. That variety 11 resembles the non-mophead lines in having a non-significant remainder mean square is not surprising since like them it was shown to have a positive correlation with the second principal component.

Consistently, in both sets, the variation over environments of the specific difference between the non-mophead and mophead lines, as measured by $\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$ in each season, is completely accounted for by the multiple regression against the physical environmental measures. The multiple regression also completely accounts for the variation over environments of the general environmental component, as measured by $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ in each season, in the second set but accounts for just 53 per cent. in the first set.

In table 8 the regression values are given for those variates which showed a non-significant remainder mean square in table 7. The signs of the significant coefficients of the non-mophead lines are consistent across the two sets. The magnitudes are also consistent, with the exception of the linear derivative of the " humidity " component, p_2, and the quadratic derivative of the " rain " component, p_3^2, which show large, positive coefficients in sets 1 and 2 respectively but non-significant ones in the other set. The pattern of coefficients of variety 11, a mophead line, is characteristic of the non-mophead lines as expected from its correlation with the second principal component.

Comparisons between the non-mophead lines of sets 1 and 2 (including $\bar{M}$ of set 2) and the mophead variety 22 and the average of the mophead lines, $\bar{m}$, in set 2 reveal the following similarities and differences. The non-mophead and mophead lines show a similar positive relationship with both the linear and quadratic derivatives of sowing date, s and s^2. Therefore, over the range of sowing dates used the later that the lines are sown the taller they will become.

The major response of the non-mophead lines to the " sun " component is a negative relationship with its linear derivative, p_1, while the mophead lines show no response to this derivative but just a positive one to the quadratic derivative, p_1^2. The " sun " component was the first component extracted from the correlation matrix of the six climatic factors and therefore it accounts for most of their standardised variation over seasons (67 per cent. in both sets, table 5). This component represents the general relationship that when it is sunny it is hotter and drier but when it is sunless it is colder and wetter. The non-mophead lines tend to be shorter in hotter, drier seasons and taller in colder, wetter seasons. The mophead lines on the other hand tend to be shortest in the intermediate seasons and tallest in the two

E

kinds of extreme seasons, hot and dry or cold and wet (compare tables 1 and 6).

The second " humidity " component, p_2, accounts for 18 per cent. of the total variation of the six standardised climatic factors over seasons (table 5). The two humidity factors, RH6 and RH12, and minimum temperature, MNT, show the highest correlations, which are positive, with this component (table 5). A season, therefore, with an extreme score for this component is for a proportion of the time either humid and warm (especially at night) or arid and cold (especially at night). The non-mophead lines in general show no response to this component, but may show a positive response to its linear derivative, p_2, while the mophead lines display a large positive reaction to both the linear and quadratic derivatives, p_2 and p_2^2, of this component. When the major pattern of cyclones and anticyclones in a season is such as to give a high score for this component (humid and warm) the mophead lines and to a lesser extent the non-mophead lines tend to be taller while the converse is true for a season with a low score (arid and cold).

The third " rain " component, p_3, accounts for just 10 per cent. of the total variation of the six standardised climatic factors over seasons (table 5). The rain factor, RAIN, and, to a slightly lesser extent, minimum temperature, MNT, show the highest correlations, which are positive, with this component (table 5). It seems, therefore to be a precipitated form of the second component so that a season with an extreme score is either wet and warm (especially at night) or dry and cold (especially at night). It is with this component that the non-mophead and mophead lines show their greatest contrast in response. The non-mophead lines show a positive reaction to both the linear and, in general, quadratic derivatives, p_3 and p_3^2, of this component while the mophead lines show a negative response to both derivatives. In both cases the response to the quadratic derivative is by far the greater. This means that when the major pattern of cyclones and anti-cyclones in a season is such as to give a high score for both the linear and quadratic derivatives of this component (wet and warm) the non-mophead lines grow taller and the mophead lines shorter. When a season has a low score for both derivatives (dry and cold) the converse is true.

These similarities and differences are reflected by the general response of all lines to environmental differences in set 2 as registered by the environmental measure, $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$, and by the specific difference between the non-mophead and mophead lines in both sets 1 and 2, as measured by $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$. Thus in set 2, where the remainder mean square of the average of the non-mophead, $\bar{M}$, and mophead, $\bar{m}$, lines are non-significant, the coefficients of the variates $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ and $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ (table 8) are very close to those obtained by computing the average sum and average difference, respectively, of the corresponding coefficients of $\bar{M}$ and $\bar{m}$. The non-significance of the remainder mean square of $\bar{m}$ in this set cannot be accounted for by the inclusion of the non-mopheadlike variety 11 in the derivation of the average performance of the mophead lines in each environment alone. As noted in section 4b, the regression $\hat{\sigma}^2$ of each of two further mophead varieties, 22 and 27, accounts for 100 and 93 per cent., respectively, of their total $\hat{\sigma}^2$'s. The significant coefficients of $\frac{1}{2}\bar{M}-\frac{1}{2}\bar{m}$ in set 1 correspond with those in set 2 (table 8). The regression $\hat{\sigma}^2$ of $\frac{1}{2}\bar{M}+\frac{1}{2}\bar{m}$ in set 1 accounts for just 53 per cent. of the total

(table 7) though the signs of the coefficients are found, in general, to correspond with those in set 2.

In conclusion, therefore, the specific difference between the non-mophead and mophead lines, as predicted by the second principal component and measured by $\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$, can be completely explained by the multiple regression against the physical environmental measures in both sets. The difference appears to centre around their contrasting linear and quadratic responses to the presence or absence of rainfall. Thus the greatest average difference between the non-mophead and mophead lines $(\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m})$ is found, for both sets, to occur in 1950 (table 1). It is also this year that has the largest negative score for the quadratic derivative of the " rain " component, p_3^2, and it is reinforced by a negative sign for its linear derivative, p_3, (table 6). The largest positive score for the quadratic derivative, p_3^2, in 1970 is cancelled by the largest negative score of equal magnitude for the linear derivative, p_3 (table 6).

The general response of all lines to environmental differences, as predicted by the first principal component and measured by $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$, is only partially explained by the multiple regression in set 1 but completely explained in set 2. It is characterised by a positive response to both the linear and quadratic derivatives of sowing date, s and s_2 (table 6). The largest value of $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ in set 1 and the second largest in set 2 occur in 1961 (table 1). It is also this year which shows an extreme positive score for both the linear and quadratic derivatives of sowing date, s and s^2 (table 6). The extreme negative scores for these two derivatives, however, cancel one another out (table 6).

The general response of the lines as measured by $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ is also characterised by a negative relationship with the linear derivative, p_1, and a positive relationship with the quadratic derivative, p_1^2, of the " sun " component. The non-mophead and mophead lines will both be taller in a season with a low score for both derivatives, which however tend to be self-cancelling, or in a season with a low score for p_1 and a high score for p_1^2 (table 6). The year 1968 conforms to the latter requirement and from table 1 it can be seen that the average sum of the non-mophead and mophead lines, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$, has its second largest value in set 1 and its largest value in set 2 in this year. Both kinds of lines will be shorter in seasons with a high score for p_1 and an intermediate score for p_1^2. The year 1949 conforms with this requirement (table 6) and it can be seen from table 1 that $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ has its lowest value in this year for both sets.

The general environmental measure, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$, shows a high positive response to both derivatives, p_2 and p_2^2, of the " humidity " component. The mophead lines will make the greatest contribution to both responses though the non-mophead lines may reinforce the positive response to the linear derivative, p_2. The second lowest score of this linear derivative, p_2, together with the most negative score of the linear derivative of sowing date, s, in 1957 (table 6) are probably jointly responsible for the second lowest value of $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ for this year in both sets (table 1). The negative response of the average sum of the non-mophead and mophead lines, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$, to the quadratic derivative of the " rain " component, p_3^2, is due to the greater contribution of the mophead lines. Its effect is not detectable due to the positive response of the non-mophead lines to this derivative and due to the joint

response of the non-mophead and mophead lines to other physical para-
meters.

It is clear from this detailed analysis that the same phenotypic response
within a group of lines may be caused by different environmental para-
meters.   Conversely, different phenotypic responses between groups of lines
may be caused by the same environmental parameter.   We have shown that
a large proportion of the variation of the non-mophead and mophead lines
over environments can be explained in terms of physical measures of these
environments, especially if the specific responses of the individual lines are
averaged out by analysing the mean performances of $\bar{M}$, $\bar{m}$, $\frac{1}{2}\bar{M} + \frac{1}{2}\bar{m}$ and
$\frac{1}{2}\bar{M} - \frac{1}{2}\bar{m}$ in the different seasons.

The mophead lines of set 1 and the mophead line, variety 23, of set 2
are however notable exceptions in having large significant remainder mean
squares.   The seasons must, therefore, differ in some other environmental
parameter(s) which has not been included in the analysis and to which the
non-mophead lines are, in general, relatively insensitive and the mophead
lines sensitive.   The missing parameter may be an even higher powered
derivative of the existing physical measures.   The relatively higher coeffi-
cient values of the mophead lines in response to the quadratic derivative of
each physical parameter in table 8 compared with that of the corresponding
linear derivative suggests that this may well be so.   It may, alternatively, be
an entirely different environmental parameter which, in an experiment
deliberately designed to investigate the dependence of plant performance on
physical characteristics of the environment, would certainly have been either
measured or controlled, such as aspects of soil, micro-climatology and
husbandry.   Whatever the cause there would be too few degrees of freedom
available to include it in the present analyses.   One solution, as far as the
climatic-environmental measures are concerned, would be to derive a
correlation matrix for them and for as many of their higher powered deriva-
tives as is possible or desirable over all the seasons for which the nearest
meteorological station has records.   For example, an 18×18 matrix would
result from the correlation of the six climatic factors, chosen in the present
paper, and their quadratic and cubic derivatives over seasons.   A principal
components analysis would be applied to such a correlation matrix.   The
22 years which would be available in the present context, 1949-70, would be
sufficient to give a non-singular 18×18 matrix.   In this way, a few indepen-
dent principal components, which summarise both the linear and higher
powered derivatives of the climatic factors, would be obtained and the score
of each component, in those seasons for which the mean performances of the
lines in a set are available, could be extracted.

Two of the non-mophead lines in set 2 of table 1, varieties 1 and 5,
together with the generations that can be derived from a cross between them
have been extensively investigated because their mean performances and
sensitivities to environmental differences are controlled by relatively simple
genetical systems (Bucio Alanis, Perkins and Jinks, 1969; Jinks and Perkins,
1970; Perkins and Jinks, 1971).   Two further lines, the non-mophead line,
variety 12, and the mophead line, variety 2 (set 1, table 1), and the genera-
tions that can be derived from a cross between them have also been exten-
sively investigated, as a contrasting pair of lines, because they display two
extreme states of complexity in the genetical control of their mean perfor-
mances and environmental sensitivities (Jinks and Perkins, 1970; Perkins

and Jinks, 1971). It is of interest to note in table 7 that the variation of both varieties 1 and 5 over seasons has been completely explained by the multiple regression whereas that of both varieties 12 and 2 must be determined by a more complex reaction to environmental differences since the remainder mean squares of these lines are large and significant.

7. SUMMARY

1. The properties of the first two components obtained from a principal components analysis of the sum of squares-sum of products matrix for the character, mean final height, of two sets of eight inbred lines of *Nicotiana rustica* over ten and nine seasons, respectively, agree with those predicted by a model which was based upon the results of previous regression analyses.

2. The first principal component is directly related to the general response of lines to environmental differences as measured by the average of all the lines in each season.

3. The second principal component is directly related to the average difference between two sub-groups of four lines in each set of eight which differ primarily at a single locus for non-mophead, M, versus mophead, m.

4. The difference in response of the non-mophead and mophead lines to environmental differences is completely accounted for by a multiple regression against physical environmental measures, consistently across the two sets.

5. The general response of all lines to environmental differences is partially explained in set 1 and completely explained in set 2 by the multiple regression analysis.

6. A large part of the variation over seasons of each mophead line in set 1 and one such line in set 2 could not be explained in this way.

7. It is postulated that either further physical parameters or higher powered derivatives of the existing ones are required to account for this residual variation. A means of reducing the number of physical parameters, especially if climatic, is suggested using a principal components analysis.

9. REFERENCES

BUCIO ALANIS, L., PERKINS, JEAN M., AND JINKS, J. L. 1969. Environmental and genotype-environmental components of variability. V. Segregating generations. *Heredity*, 24, 115-127.

FREEMAN, G. H., AND PERKINS, JEAN M. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity*, 27, 15-23.

FRIPP, Y. J. 1972. Genotype-environmental interactions in *Schizophyllum commune*. II. Assessing the environment. *Heredity*, 28, 223-238.

JINKS, J. L., AND PERKINS, JEAN M. 1970. A general method for the detection of additive, dominance and epistatic components of variation. III. F_2 and backcross populations. *Heredity*, 25, 419-429.

PERKINS, JEAN M., AND JINKS, J. L. 1968a. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23, 339-356.

PERKINS, JEAN M., AND JINKS, J. L.   1968*b*.   Environmental and genotype-environmental
    components of variability.   IV. Non-linear interactions for multiple inbred lines.
    Heredity, 23, 525-535.
PERKINS, JEAN M., AND JINKS, J. L.   1971.   Analysis of genotype × environmental interaction
    in triple test cross data.   *Heredity, 26*, 203-209.
SEAL, HILARY L.   1964.   *Multivariate Statistical Analysis for Biologists*.   Methuen, London.
YATES, F., AND COCHRAN, W. G.   1938.   The analysis of groups of experiments.   *J. Agric. Sci.,
    28*, 556-580.