Non-additive combining abilities

By NEIL GILBERT*

Biophysics Department, University of Edinburgh,
West Mains Road, Edinburgh 9

Statistical methods are commonly based on additive models. In particular, the analysis of the yields of a set of crosses by additive 'general combining abilities' makes no genetic assumption, apart from the equal importance of the two parents. Such analyses in terms of g.c.a.'s are always successful to some extent, but rarely account for all the observed variation. This paper considers the simultaneous estimation of *two* constants—additive and multiplicative—for each parent. The data can then decide for themselves how far they are additive, and how far multiplicative. In those cases where the additive method accounts for nearly all the 'genetic' variance, there is naturally little prospect of significant improvement by a non-additive hypothesis.

1. ALGEBRA

$k \times l$ *table.* We consider all possible crosses between k male parents and l female parents. The model used is

$$y_{ij} = m + a_i + b_j + c_i d_j + \text{remainder}$$

where y_{ij} is the yield of the cross between the ith male and the jth female;

m is the general mean;

a_i, b_j are the male and female additive combining abilities;

c_i, d_j are the multiplicative combining abilities;

$\sum a_i = \sum b_j = \sum c_i = \sum d_j = 0$ and one arbitrary restriction must be placed on $\sum c_i^2$ or $\sum d_j^2$.

The model is evidently not restricted to the genetical case, but can be used generally on any two-way table. It no longer predicts (as does the purely additive model) that one parent will always outshine another, whatever the third parent to which both are mated. The least-squares estimates of m, a_i and b_j are the same as in the purely additive case. The analysis therefore consists of finding the interactions z_{ij} by correcting y_{ij} for main effects of the two-way table, and applying the model

$$z_{ij} = c_i d_j + \text{remainder}$$

to the $k \times l$ matrix $\{\mathbf{z}\}$ composed of the elements z_{ij}. From the least-squares equations it is easily shown that $\sum c_i^2 \sum d_j^2$ is the (largest) latent root of the matrix $\{\mathbf{z}\mathbf{z}'\}$ (or of the matrix $\{\mathbf{z}'\mathbf{z}\}$); and that the vectors c_i, d_j are the corresponding latent vectors of $\{\mathbf{z}\mathbf{z}'\}$ and $\{\mathbf{z}'\mathbf{z}\}$ respectively. Extraction of latent roots is, of course,

* Present address: John Innes Institute, Bayfordbury, Hertford, Herts.

F

child's play to a computer. The 'interaction' sum of squares $\sum z_{ij}^2$ is reduced by an amount $\sum c_i^2 \sum d_j^2$, and its degrees of freedom from $(k-1)(l-1)$ to $(k-2)(l-2)$. Similarly, by taking the first and second latent roots and vectors, the model

$$z_{ij} = c_i d_j + e_i f_j + \text{remainder}$$

reduces the 'remainder' d.f. to $(k-3)(l-3)$; and so on until all the variation is exhausted. From this point of view, therefore, the fitting of additive main effects may be regarded as an arbitrary reduction of the d.f. to $(k-1)(l-1)$, which would more naturally be achieved by means of latent roots and vectors. This explains the apparent paradox that, whereas additive main effects need not (necessarily) reduce the remainder mean square, multiplicative main effects must do so (since one latent root is always larger than the others). In other words, in the absence of true multiplicative effects, the m.s. for 'multiplicative main effects' is expected to exceed the 'error' m.s. The point must be borne in mind when judging the statistical significance of the multiplicative terms. If we impose the further arbitrary constraints $c_i = a_i$, $d_j = \lambda b_j$, estimation of λ absorbs Tukey's (1949) 'one d.f. for non-additivity'. The $k \times l$ table will not be discussed further, but it has to be considered in order to understand the $k \times k$ *diallel cross*, consisting of the $\frac{1}{2}k(k-1)$ crosses between k parents (omitting selfs and reciprocals). The model becomes

$$y_{ij} = m + b_i + b_j + c_i c_j + \text{remainder} \tag{1}$$

or

$$y_{ij} = m + b_i + b_j - c_i c_j + \text{remainder} \tag{2}$$

as opposed to the orthodox

$$y_{ij} = m + t_i + t_j + z_{ij} \tag{3}$$

where $\sum t_i = 0$ and z_{ij} is the interaction, or special combining ability.

Models (1) and (2) are identical in theory—apart from a factor $\sqrt{(-1)}$ in the c's—but must in practice be considered separately if we want real values of c.

Since

$$m + b_i + b_j + c_i c_j \equiv (m - \theta^2) + (b_i - \theta c_i) + (b_j - \theta c_j) + (c_i + \theta)(c_j + \theta)$$

the restrictions $\sum b_i = \sum c_i = 0$ are truly arbitrary. Since

$$m + b_i + b_j + c_i c_j \equiv m + (b_i - \tfrac{1}{2}c_i^2) + (b_j - \tfrac{1}{2}c_j^2) + \tfrac{1}{2}(c_i + c_j)^2$$

the model can equally be regarded as quadratic. The general combining abilities t_i of model (3) are easily estimated (Yates, 1947). Model (1) is not so simple; in particular, the estimate of b_i is not the same as that of t_i. From the least-squares equations it can be shown that model (1) is equivalent to fitting

$$z_{ij} = c_i c_j + \frac{c_i^2 + c_j^2}{k-2} - \frac{\sum c^2}{(k-1)(k-2)} + \text{remainder} \tag{4}$$

Model (2) fits the same expression to $-z_{ij}$. This is the method used in the numerical analysis. However, it is clearly unsatisfactory to regard the model purely as a further analysis of the special combining ability z_{ij}, since (4) involves k, the number

of parents, which can, of course, take any value we please. Similarly, the simple model

$$z_{ij} = c_i c_j + \text{remainder}$$

would be unsatisfactory when translated into a model for y_{ij}. Although the analysis is made in practice by fitting (4) to z_{ij}, that is only a convenient way of fitting (1) to y_{ij}. The reduction in $\sum z_{ij}^2$ due to fitting (4) is

$$\sum c_i z_{ij} c_j = \frac{1}{2(k-2)}\left[\left(k-2+\frac{1}{k-1}\right)\left(\sum c^2\right)^2 - k\sum c^4\right]$$

and the variance of the difference between two estimated c's is approximately

$$\frac{2k}{k-2}\frac{s^2}{\sum c^2}$$

where s^2 is the error variance of y_{ij}. This approximation is obtained from the least-squares matrix by assuming all values of c_i^2 equal. If the original errors are Normally distributed, the errors of the c_i's will not be Normal.

The values of c_i were found iteratively by a special computer programme. The equation for c_i is

$$(k-2)\sum_{j\neq i} z_{ij} c_j = \sum c^3 + \left(k-2+\frac{1}{k-1}\right)c_i \sum c^2 - kc_i^3$$

Convergence is slow, and oscillations tend to develop unless damped. From the discussion of the $k \times l$ table above, we can expect several sets of values of c_i to satisfy these equations. We are interested in one particular set. Consequently, the initial values supplied to the computer, if wildly wrong, may affect the final result.

The values of c_i were therefore obtained as follows. The sums of squares $Z_i = \sum_j z_{ij}^2$ were found, and $Z_i \div \sqrt{\left(\sum_i Z_i\right)}$ was taken as a preliminary estimate of c_i^2. These estimates of c_i^2 were inserted into the matrix $\{\mathbf{z}_{ij}\}$ as diagonal terms, and the first latent root was then extracted (corresponding to model (1)). The same estimates of c_i^2 were also inserted into the matrix $\{-\mathbf{z}_{ij}\}$, and the first latent root again extracted (corresponding to model (2)). One of these latent roots was always much larger than the other, so indicating whether model (1) or (2) was chosen by the data. As mentioned above, (1) and (2) are two forms of the same thing; to allow the data to decide between them is a necessary part of the least-squares estimation. The corresponding latent vector provided the starting values of c_i, which were fed into the programme for fitting (4). The analysis is thus rather complicated in practice, but that is merely a consequence of using diallel cross data; in principle it is very simple. Using the present Rothamsted computer, the method takes a lot of time. Only the examples reported here have been analysed, and I do not expect to do any more until a faster machine becomes available. Anyone who has access to a fast computer, and wishes to programme the present analysis, would do well to write to me first.

2. GENETIC MODEL

The analysis can be illustrated by a crude genetic model, very similar to that of Hayman (1954a). The analysis in no way depends on the model—it would be unacceptable if it did—and the reader is warned against identifying the two. Thus, apparent 'dominance' may easily be due to multiplicative gene action or similar curvature of the genotype–phenotype relation—with very different consequences under selection (Gilbert, 1961a). But the model, although strictly incredible, does illuminate certain aspects of the analysis. In particular, it *suggests* that the multiplicative statistical model will serve genetical data better than non-genetical.

In developing the genetic model, I make the same assumptions as Hayman, namely

(1) additive independent gene effects $\pm d$ (homozygotes) or $+h$ (heterozygote);
(2) two alleles at each locus;
(3) homozygous diploid parents;
(4) alleles independently distributed between parents,

together with the further simplification that the values of the genetic effects d and h are the same for all loci. We then have

(a) phenotype of ith parent $= p_i = M + 2(l_i - N)d + \text{remainder}$
(b) phenotype of F_1 cross $i \times j$

$$= M + [(l_i - N)d + l_i h] + [(l_j - N)d + l_j h] - l_i l_j h/N + \text{remainder}$$

where M is a general mean;

d, h are additive and dominance effects;
l_i is the number of 'good' loci in the ith parent;
$2N$ is the total number of segregating loci.

(b) is clearly of the same form as models (1) or (2), depending on whether h is negative or positive. In other words, if (b) were true, (1) or (2) would absorb the whole of the variation between the crosses, apart from error. We might be tempted to write $i = j$ in model (1), so expecting that, if the parents are inbred, $p_i = m + 2b_i + c_i^2 + \text{remainder}$. This expectation must be false, for we do not get (a) by substituting $i = j$ in (b). From (b) it is easily shown that—if the genetic model were true—c_i and $(b_i - \frac{1}{2}p_i)$ would both be concerned exclusively (apart from certain constants) with dominance and would therefore be completely correlated, since each would be a linear function of l_i. The further expectations, that the correlations $c_i \times b_i$ and $c_i \times p_i$ would each be unity, depend rather more heavily on the assumption of equal genetic effects at different loci. All these correlations will, of course, be diluted by 'environmental' error.

3. F_1 DATA OF HAYMAN (1954b)

Although, as pointed out above, model (1) should properly be regarded as fitting constants b and c simultaneously, the analyses of variance will be presented in terms of the reduction in $\sum z_{ij}^2$ that is achieved by fitting (4). In this way we can see

if the additive-multiplicative model performs better than the conventional additive one.

Table 1. *Analysis of variance of plant height of* Nicotiana rustica

	d.f.	m.s.
General c.a.	7	24,132
Special c.a.	20	1864
Reciprocal diffs.	28	683
Error	64	413
On fitting model (1) this becomes:		
Additive c.a.	7	24,132
Multiplicative c.a.	7	3825
Remainder	13	809
Reciprocal diffs.	28	683
Error	64	413

As pointed out in Section 1, we are in effect selecting the largest root of a matrix; and so, no valid significance test can be made on the m.s. for 'multiplicative c.a.'. We might adopt one of Bartlett's (1951) tests for the largest root of a matrix as an approximation, or evaluate the null distribution by Monte Carlo. But at present we can only remark that the c's absorb an impressive amount of the 'special c.a.' variance.

The correlations discussed in Section 2 are (6 d.f.):

$$\begin{array}{ll} c_i \times b_i & 0 \cdot 828 \\ c_i \times p_i & 0.427 \\ c_i \times (b_i - \tfrac{1}{2} p_i) & 0 \cdot 866 \end{array}$$

The signs of these, and similar, correlations can be reversed by reversing the signs of the c_i's (which leaves the rest of the analysis unaltered). Consequently, a two-tailed significance test should be used.

4. F_1 AND F_2 DATA OF KINMAN AND SPRAGUE (1945)

Table 2. *Analysis of variance of maize yield by model* (1)

	d.f.	F_1 m.s.	F_2 m.s.
Additive c.a.	9	424·2	363·2
Multiplicative c.a.	9	218·1	59·3
Remainder	26	83·3	12·7

There is no estimate of error here, but the multiplicative constants have considerably reduced the remainder m.s. Furthermore, the correlation between F_1 and F_2 values of c_i is 0·725 (8 d.f.), whereas the correlation between F_1 and F_2 values of z_{ij}—from which the c_i's were estimated—is only 0·257 (34 d.f.). The multiplicative constants can thus account entirely for the observed positive correlation

between 'special combining abilities' z_{ij} in successive generations; the correlation between F_1 and F_2 'remainders' is -0.237 (25 d.f.). The F_1 correlations are:

$$c_i \times b_i \qquad\qquad 0.113$$
$$c_i \times p_i \qquad\qquad 0.144$$
$$c_i \times (b_i - \tfrac{1}{2}p_i) \qquad 0.300$$

(The values of p_i suffer from severe inbreeding depression.)

There remains a possibility that a purely multiplicative model would serve as well as the additive-multiplicative one. The F_1 data were therefore analysed according to the hypothesis

$$y_{ij} = d_i d_j + \text{remainder}$$

The values of d were estimated by least-squares, giving:

Table 3. *Analysis of variance*

	d.f.	F_1 m.s.
Multiplicative c.a.	9	418·6
Remainder	35	119·4

The purely multiplicative model therefore absorbed no more variation than the purely additive one. This one case suffices to remove the possibility that the additive-multiplicative model is merely superfluous.

5. TOMATO YIELDS OF GILBERT (1961b)

The data comprise F_1 and F_2 generations in 1958 and 1960, and selected F_3 and F_4 in 1960. It was shown in the original paper that different F_3 and F_4 families had responded differently to selection. There is therefore no reason to expect the F_3 and F_4 analyses to show much similarity to the F_1 and F_2. In fact, the F_1 and F_2 in both years opted for model (2), but the F_3 and F_4 for model (1).

Table 4. *Analyses of variance*

					m.s.		
		1958		1960			
	d.f.	F_1	F_2	F_1	F_2	F_3	F_4
Additive c.a.:							
Regr. on parents	1	1827	4882	13,393	21,400	21,409	19,262
Remainder	16	225	230	703	382	643	624
Interactions:							
Multiplicative c.a.	17	230	93	804	393	812	1173
Remainder	118	57	43	198	184	289	389
Error (120 d.f. in 1958,							
608 d.f. in 1960)	—	48	28	169	169	169	169

The values of c_i, with parents arranged in groups according to Williams and Gilbert (1960), are shown in Table 5. Correlations (16 d.f.) between c_i's are:

$$1958\ F_1 \times 1960\ F_1 \quad 0.668 \qquad 1958\ F_1 \times 1958\ F_2 \quad 0.621$$
$$1958\ F_2 \times 1960\ F_2 \quad 0.881 \qquad 1960\ F_1 \times 1960\ F_2 \quad 0.596$$
$$1960\ F_3 \times 1960\ F_4 \quad 0.606$$

(The 5% two-tail confidence value is 0.53.) These figures are all very much higher than the corresponding correlations between the interactions z_{ij}. Those correlations which are expected to be large, if the c_i's are to be genetically meaningful, are so; the correlations of F_1 or F_2 with F_3 or F_4 are smaller. It might be thought that, if model (2) were fitted to F_3 and F_4 (as it was to F_1 and F_2), the resultant c_i's would be correlated with the F_1 and F_2 c_i's, although they are no longer true least-squares

Table 5. *Values of* c_i ($\sqrt{(oz.)}$ *per plant*)

| | 1958 | | 1960 | | | |
	F_1 (± 0.469)	F_2 (± 0.249)	F_1 (± 0.559)	F_2 (± 0.526)	F_3 (± 0.532)	F_4 (± 0.504)
Group A:						
Ailsa Craig	−3.96	−1.07	−2.16	−1.53	−0.40	−2.05
Harbinger	−1.61	−0.26	−1.11	−0.32	1.72	3.89
Plumpton King	−0.20	−0.56	−1.43	0.36	−1.59	−1.00
Crackerjack	−0.71	−0.61	−1.00	−0.15	0.17	1.09
Group B:						
Vetomold	1.15	0.16	1.64	0.33	−0.36	−1.70
Potentate	1.09	0.15	1.79	0.17	−2.06	1.18
Potential	1.09	−0.62	1.49	0.24	1.59	0.88
Vagabond	0.92	0.27	3.46	0.77	1.17	0.05
Comet	−1.50	−0.89	0.59	0.29	1.41	−0.70
Baby Lea	2.62	4.56	2.28	6.59	0.55	−0.72
Other varieties:						
Kondine Red	−0.14	−0.62	−1.09	−0.38	3.03	1.88
LMR 1	−1.28	−0.10	−1.92	−2.13	1.93	1.51
ES 1	−0.87	−0.62	−1.51	−1.91	−0.21	0.02
Delicious	1.08	−0.43	−1.61	−0.44	−4.95	−2.94
Radio	0.02	−0.19	−1.92	−0.98	−1.25	−2.30
Downes' Seedling	−0.47	0.03	0.27	−0.89	−0.60	−0.45
Exhibition	2.04	0.60	1.33	0.10	0.31	3.32
Moneymaker	0.72	0.19	0.91	−0.14	−0.44	−1.95

The figures in brackets are standard errors.

The F_1 correlations are:

	1958	1960
$c_i \times b_i$	−0.550	−0.207
$c_i \times p_i$	−0.750	−0.425
$c_i \times (b_i - \frac{1}{2}p_i)$	0.387	0.348

(The signs of these correlations can be reversed by reversing the signs of the c_i's.)

estimates. This expectation is not fulfilled. In fact the F_3 and F_4, having responded differentially to selection, give patterns of c_i quite different from F_1 and F_2. In Table 5, the varieties within groups A and B—thought on other grounds to be similar —show similar values of c_i in F_1 and F_2, but not in F_3 or F_4. It is possibly of interest that Baby Lea, as a variety, has previously performed extremely erratically.

<center>6. DISCUSSION</center>

To those accustomed to an additive way of thinking, it may appear that, since c_i is estimated from interactions z_{ij}, the value obtained will depend entirely on which other parents are included. In other words, the value of c_i might not truly characterize the ith parent—in the way that an additive combining ability does—but only the set of crosses as a whole. Such an argument vanishes as soon as the possibility of (partially) multiplicative genetic effects is admitted. Consideration of model (4) shows that elimination of one parent from the diallel cross would not seriously alter the *relative* values of the remaining c_i's. The large correlations between F_1 and F_2 values of c_i, expected on almost any genetic theory, indicate that the c_i's are genuinely describing some characteristic of the parents. The poor correlations between c_i and b_i show that c_i is not merely measuring the same thing as b_i. The poor correlations between c_i and $(b_i - \frac{1}{2}p_i)$ show that the genetic model of Section 2 is possibly valuable but certainly untrustworthy.

The present analysis may usefully be compared with that of Hayman (1954a). Hayman's very ingenious method depends rather crucially on unverifiable genetic assumptions, and involves a searching analysis of second-order statistics. I am afraid that the results can rarely mean what they purport to mean. Students of quantitative inheritance perhaps pay too much attention to second-order statistics altogether. The present method is free of both these troubles. It is concerned with the individual parents, whereas Hayman's analysis is more interested in the set of parents as a whole. The present method certainly does not try to sort out additive gene effects, dominance, etc. The use of such terms—except in a purely nominal way—is, I submit, bound to be misleading in these situations where the true genetic system is unimaginably complicated.

It may well be asked, what is the use of the method advanced in this paper? Regarding the study of quantitative inheritance as an end in itself, the method can and does improve our understanding. It is true that any behaviour can be explained in terms of multiple interactions between additive genetic effects. But unless some kind of pattern can be discerned, such an 'explanation' does nothing to improve our comprehension of the phenomena. The additive-multiplicative model has proved capable of describing remarkably well—although still not perfectly—the rather large bodies of data analysed here. It will be interesting to compare the magnitudes of c_i in F_1 and F_2, and to see how well the analysis works for backcrosses and for parents that are not inbred. I am convinced that genuine progress in the subject must come from cautious statistical description of this kind, rather than from analyses dependent on unverified genetical hypotheses. The results presented

in this paper suggest very strongly that a non-additive way of thinking is more helpful than the additive, however easy the latter may be statistically.

Various analyses, notably selection indices and polygenes, have been advocated as aids to practical breeding. None of the claims has so far been substantiated. In plant breeding, it is relatively easy to achieve any (reasonable) specification; the major difficulty is rather to recognize what kind of plant is really wanted, as the history of Proctor barley shows. The main need is not for elegant statistical frills, but for effective ways of reducing the numbers of plants that have to be grown. There is more scope for refined analysis in animal breeding, where individuals are expensive and no vegetative multiplication is possible. Bearing in mind its limitations, the concept of heritability is certainly useful, and perhaps no single parameter could be more useful. But it is often asserted that 'genetic correlations' enable us to predict changes in one character consequent on selection for another character. This assertion, which depends on assumptions of additivity and of linear relations between the characters, has not been adequately tested by experiment. Until such elementary information has been obtained, refinements such as that proposed in this paper seem superfluous. Neither the present method, nor any similarly sophisticated one, offers practical help to the breeder in dealing with specific problems. But any knowledge of the general pattern of quantitative inheritance will be worth having.

7. SUMMARY

An additive-multiplicative model is found to describe diallel cross data rather well. Estimates of the model's parameters are highly correlated from generation to generation.

REFERENCES

BARTLETT, M. S. (1951). The effect of standardization on a χ^2 approximation in factor analysis. *Biometrika*, **38**, 337–344.

GILBERT, N. (1961a). Polygene analysis. 2. Selection. *Genet. Res.* **2**, 456–460.

GILBERT, N. (1961b). A tomato selection experiment. *Genet. Res.* **2**, 361–372.

HAYMAN, B. I. (1954a). The theory and analysis of diallel crosses. *Genetics*, **39**, 789–809.

HAYMAN, B. I. (1954b). The analysis of variance of diallel tables. *Biometrics*, **10**, 235–244.

KINMAN, M. L. & SPRAGUE, G. F. (1945). Relation between number of parental lines and theoretical performance of synthetic varieties of corn. *J. Amer. Soc. Agron.* **37**, 341–351.

TUKEY, J. W. (1949). One degree of freedom for non-additivity. *Biometrics*, **5**, 232–242.

WILLIAMS, W. & GILBERT, N. (1960). Heterosis and the inheritance of yield in the tomato. *Heredity*, **14**, 133–149.

YATES, F. (1947). The analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity*, **1**, 287–301.