

# THE CASE FOR INDIRECT GENETIC CONTROL OF SEQUENTIAL TRAITS AND THE STRATEGY OF DEPLOYMENT OF ENVIRONMENTAL RESOURCES BY THE PLANT

JOHN E. GRAFIUS AND ROGER L. THOMAS  
*Michigan State University*

Received 24.vii.70

## 1. INTRODUCTION

THE continuum of environmental resources may be thought of as a finite sump which can be taxed by the plant to supply the necessary factors for growth and development. As a general thesis herein, resources used at one stage of development are not available for use later on, nor are the resources assumed to be equally available throughout the season. The form of the argument hinges on a recurrence formula which when applied to traits in a developmental sequence can produce an oscillatory effect. An attempt will be made to show that when an oscillatory or wave form of curve obtains, direct genetic control of the more remote traits in the sequence is replaced by indirect control.

The successful variety is one which has adopted a strategy which will maximise the final returns (yield) from the use of available resources as it elaborates the developmental sequence. While the outcome is recorded in terms of environmental input associated with a sequence of traits, such as heads per unit area, seeds per head and kernel weight, no attempt will be made to disguise the fact that these morphological characteristics are outward manifestations of the whole metabolic process of growth and development.

The concept of an oscillatory or wave form of curve for developmental sequences is not new as an imaginative paper by Adams (1967) will attest. Nickell *et al.* (1969) touched on this in a paper on negative progress in selection for yield and Thomas *et al.* (1970a) present what they hope are convincing arguments for a sequential effect. Grafius (1969) developed the use of correlation between sequential traits as a measure of stress and presented such measures graphically as contour lines. The mathematical approaches used here may be found in a delightful little book by Maynard Smith (1968) and in a book on differential equations by Powell and Wells (1950).

This paper attempts to develop a general statement with regard to the oscillatory nature of the stress induced by competition for environmental resources by the traits in a developmental sequence. It also purports to show that the outcome depends upon the strategy of deployment of environmental resources by the plant. For present purposes it is assumed that strategy originates in the genetic code which determines the first terms in the oscillatory sequence. In the example given, a successful variety (high yield) has a markedly different strategy than an unsuccessful one.

## 2. THEORETICAL CONSIDERATIONS

Consider a series of sequential traits which are independent genetically but which interact physiologically. Suppose that the environmental resources available throughout the growing season are limited in total amount so that quantities used early influence subsequent events. In other words, we cease to think of traits as such but instead consider them only as sumps which trap environmental resources.

At the moment, we define environmental resources on a very broad basis to include such things as water, nutrients, carbohydrates, proteins, temperature, light and  $\text{CO}_2$ . No attempt will be made to distinguish between stress caused by limitation of external environmental resources from that caused by, say, the inability of the plant to fill as many ovaries as it lays down (internal stress).

Each trait in a developmental sequence will be transformed by subtracting the mean and dividing by the standard deviation so that the arrays for the various traits all have a mean of zero and a variance of 1. We take the plus and minus deviations from the mean as a measure of the amount of environmental resources used. Admittedly, this simplification does not seem to recognise differences in input required for a difference of one standard deviation in trait  $x_1$  and one in trait  $x_2$ . The argument is at least partly satisfied by the fact that each array is in units of standard deviations and that given normal distribution, each array is quite similar to its neighbouring arrays and has the same mean. A standard deviation gain for trait  $x_1$  is assumed to exhaust the environmental resources available for trait  $x_1$  just as fully as a standard deviation gain for trait  $x_2$  exhausts the resources available for  $x_2$ . The fact that a standard deviation for  $x_1$  may use more or less of the total environment than a standard deviation for  $x_2$  is taken care of by the constants  $b$  and  $c$  in the recurrence relationship which follows.

Let  $x_1, x_2, x_3, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_p$  be a series of sequential traits in standard measure, where each trait is a link in the chain of events but where each trait differs from the preceding only in the amount of environmental resources it uses to express itself.

We write

$$x_{n+2} + bx_{n+1} + cx_n = 0. \quad [1]$$

The constant coefficients are to be determined by least squares regression analysis by writing

$$\hat{x}_{n+2} = \beta_1 x_{n+1} + \beta_2 x_n \quad [1a]$$

from which it is found that  $\beta_1 = -b$  and  $\beta_2 = -c$ , where  $\beta_i$  is the standard partial regression coefficient. Since the array mean is zero there is no constant term in [1a].

To solve [1] we guess the solution  $x_n = A\lambda^n$  which upon substituting in [1] results in

$$A\lambda^{n+2} + Ab\lambda^{n+1} + Ac\lambda^n = 0,$$

from which it is seen that  $\lambda^2 + b\lambda + c = 0$ . The roots are:

$$\lambda_1 = -\frac{b}{2} + \frac{\sqrt{b^2 - 4c}}{2}, \quad \lambda_2 = -\frac{b}{2} - \frac{\sqrt{b^2 - 4c}}{2}.$$

Formal mathematics gives the complete solution as

$$x_n = A\lambda_1^n + B\lambda_2^n, \tag{2}$$

where  $A$  and  $B$  are constants chosen to fit initial conditions. We will not use [2] here except to discuss what happens. If  $b^2 - 4c$  is positive, this form

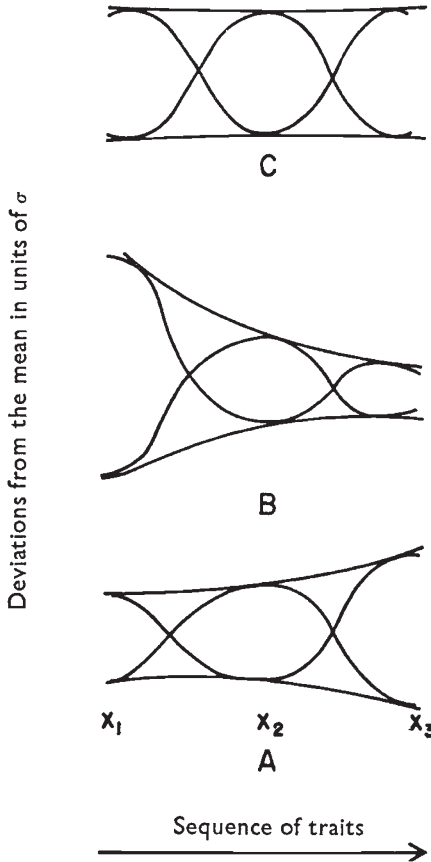


FIG. 1.—Examples of oscillatory curves. *A*: divergent, *B*: convergent, *C*: continuous oscillation. One intuitively expects type *B* to fit most biological situations although type *C* could be imagined for some situations where there was no carry-over of effects beyond the trait immediately preceding the one in question. Type *A* could exist only for a small number of traits.

is satisfactory. If, however,  $b^2 - 4c$  is negative,  $\lambda_1$  and  $\lambda_2$  are complex and a trigonometric form is needed. It turns out that the solution is similar in form to the solution of the differential equation

$$\frac{d^2x}{dn^2} + b \frac{dx}{dn} + cx = 0,$$

an equation which is common in wave mechanics (5). The solution is

$$x_n = c^{n/2}(k_1 \cos n\theta + k_2 \sin n\theta) \tag{3}$$

where  $\theta = \cos^{-1}\left(\frac{b}{-2\sqrt{c}}\right)$  and  $k_1$  and  $k_2$  are chosen to satisfy initial conditions. Again we will not use this equation except for general discussion. If, in equation [3],  $c < 1$  then the oscillations converge (*i.e.* the height of the wave gets smaller as  $n$  increases) when  $n$  gets large because  $c^{n/2}$  goes to zero. If  $c > 1$ , then the oscillations diverge as  $n$  increases and  $c^{n/2}$  gets very large.

Several examples of oscillatory curves are given in fig. 1. One would expect type *B* to be the most common in nature although type *C* could exist in situations where there was no carry-over beyond the immediately preceding trait. Situation *A* could exist only for a short series of traits.

To summarise:

1. If  $b^2 > 4c$ , there are no oscillations and [2] describes the behaviour of  $x$ , which expands without limit if  $\lambda_1 > 1$ .

2. If  $b^2 < 4c$ ,  $x$  oscillates and the amplitude decreases or increases depending on whether  $1 > c > 1$ . If  $c < 1$ , the oscillations converge.

### 3. EXAMPLE

#### (i) *Materials and methods*

Thirty-six experimental lines of oats representing six crosses were grown in replicated plots at two locations in Michigan. The locations represent two contrasting environmental areas. Location 1 is typical of the Thumb area of Michigan and the soil is a highly productive clay loam. Rainfall and growing conditions are generally optimal. Location 2 is characteristic of the Kalamazoo prairie soil which is typically a clay loam underlain by sand and gravel, and is only moderately productive, even in years when rainfall is adequate.

The individual plots were four rows wide and 4 metres long. The rows were 0.3 metres apart. There were four replications at each location.

Data were collected for yield ( $W$ ), number of heads per plot ( $X$ ), and average kernel weight ( $Z$ ). The number of seeds per head ( $Y$ ) was inferred from the equation  $Y = W/XZ$ .

Let the components of yield of oats be such a series as described in the previous section where the number heads per unit area ( $X$ ) =  $X_1$ ; the number of seeds per head ( $Y$ ) =  $X_2$ ; and the average kernel weight ( $Z$ ) =  $X_3$ .

Since the original traits are in different units we place them on the same scale by allowing

$$x_i = \left( \log \frac{X_i}{\sigma_i} \right) - \log \frac{\bar{X}_i}{\sigma_i}.$$

With this transformation, the various  $x_i$  traits are now on the same scale with a mean of zero and a variance of 1. Logs are used because the yield component traits affect each other multiplicatively, *i.e.*  $Z$  is affected by total seed number  $XY$ . It is emphasised that the various  $x_i$ 's are now + and - deviations from a mean of zero in units of  $\sigma$ .

#### (ii) *Results*

The patterns of  $x_1$ ,  $x_2$  and  $x_3$  for the data from the 288 plots (36 varieties,

four reps, two locations) are given in table 1. Convincing evidence of non-random distribution of plus and minus values can be seen. The run of zero (no change of sign) category is much less than expected, the deficiency being distributed at random among the remaining categories.

TABLE 1

*Pattern of relationship between  $x_1$ ,  $x_2$  and  $x_3$ . A plus indicates a value greater than the mean. There are 288 observations (36 varieties, 4 replications, 2 locations). A "run" equals the number of times the sign changes in the pattern*

Class in terms of runs	Pattern			Observed frequency	Expected frequency
	$x_1$	$x_2$	$x_3$		
Run of 0	{ - + }	{ - + }	{ - + }	24	72
Run of 1	{ + - + - }	{ - - + + }	{ - + - + }	165	144
Run of 2	{ - + }	{ + - }	{ - + }	99	72
				288	288

The analysis of variance for  $x_3$  is given in table 2. The variety effect is highly significant but is determined largely ( $R^2 = 66$  per cent.) by the regression of  $x_3$  on  $x_1$  and  $x_2$ . There is very little direct effect of variety on  $x_3$ . The variety exerts its control through  $x_1$  and  $x_2$ .

TABLE 2

*Analysis of variance of  $x_3$  for variety effect and subdivision of degrees of freedom. The variety effect is highly significant but 66 per cent. ( $R^2 = 0.66$ ) of the variation of  $x_3$  is determined by the regression of variety effect on  $x_1$  and  $x_2$ . The direct control of  $x_3$  by variety plays a very minor role*

Source of variation	d.f.	S.S.	M.S.	$R^2$
Total	287	0.3238	—	—
Varieties	35	0.1837	0.0052**	—
Regression $x_3 = f(x_1, x_2)$	2	0.1220	0.0610**	0.66
Error <i>a</i>	33	0.0617	0.0019	—
Within varieties	252	0.1401	0.0006	—
Regression $x_3 = f(x_1, x_2)$	72	0.0506	0.0007*	0.36
Error <i>b</i>	180	0.0895	0.0005	—

The regression within varieties over locations is significant ( $P < 0.05$ ) and explains 36 per cent. of the variation within varieties, ignoring location. Apparently, this is the percentage of determination of  $x_3$  by  $x_1$  and  $x_2$  within varieties which is independent of the variety mean and is therefore common to all varieties in the set. This fraction is associated with the developmental sequence, independent of the genetic effect due to varieties.

There is no direct test for location effect in this case since the degrees of freedom are too few. Location does exert a powerful effect, however, as

indicated in table 3. The  $b$  and  $c$  coefficients are significantly different between locations when the regression about the variety means was calculated within each location. It will be recalled that the relationship  $b^2 - 4c$  and the size of  $c$  determine the nature of the oscillatory relationship between  $x_1$ ,  $x_2$  and  $x_3$ . Consequently, differences in these coefficients between locations indicate significant differences in developmental stress patterns.

TABLE 3

*Least squares estimates of the constant coefficients  $b$  and  $c$  in the recursion formula*  
 $x_{n+2} + bx_{n+1} + cx_n = 0$  *for locations 1 and 2*

Location *	$b$	$c$
1	0.6	0.7
2	0.4	0.2

\* Location 1 coefficients significantly ( $P_b < 0.05$ ,  $P_c < 0.01$ ) greater than those of location 2.

The pattern of behaviour of  $x_1$ ,  $x_2$  and  $x_3$  can be predicted by the relationship  $b^2 - 4c$ , as indicated earlier. In the present example the regression for variety means had beta values  $\beta_{31.2} = -0.67$  and  $\beta_{32.1} = -0.61$  and since  $\beta_{31.2} = -b$ ,  $b = 0.67$  and since  $\beta_{32.1} = -c$ ,  $c = 0.61$ . From this we find  $b^2 - 4c = -2.0$ , which on the basis of our earlier conclusion means that  $x$  in [1] will oscillate and that as  $n$  increases it will converge ( $c < 1$ ).

TABLE 4

*Plus and minus deviations from the mean for the five highest and five lowest lines out of a total of 36 a location 1 and location 2 and their average. The top and bottom lines for location 1 and 2 are not necessarily the same but the patterns of deviations show remarkable constancy with respect to  $x_2$ . Evidence of oscillatory nature of the patterns is striking*

Relative yield	Location 1			Location 2			Average *		
	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$
Five top lines	+	+	-	-	+	+	+	+	0
	+	+	-	-	+	+	+	+	-
	-	+	+	-	+	+	0	+	+
	+	+	-	-	+	+	+	+	-
Five bottom lines	-	+	-	+	+	-	-	+	-
	-	+	-	+	+	-	-	+	-
	+	-	+	-	-	-	+	-	+
	-	-	+	-	-	-	-	-	+
	-	-	+	+	-	+	+	-	+

\* Based on the five best and five poorest lines averaged over both locations.

Table 4 represents non-random groupings of the five lines with the highest and five with the lowest yields. The top lines show a surplus of plus values for  $x_1$ ,  $x_2$  and  $x_3$  and the poor lines show a surplus of minus values, as would be expected. But the striking feature is the consistency of  $x_2$ . Additional evidence of this relationship is presented in table 5. A variety must have a high value for  $x_2$  to have high yield for this environment.

Inasmuch as selections from 14 crosses are represented in tables 4 and 5 this observation relative to  $x_2$  appears to be meaningful for these environments. Other environments may have different requirements.

Observed and calculated values for  $x_3$  are given in table 6 for the five

TABLE 5

*Additional evidence of the oscillatory nature of plus and minus deviations from the mean for 36 additional oat lines from eight different crosses grown in replicated yield trails at three locations in Michigan in 1968. The patterns are given for five top yielding and the five lowest yielding lines at each location and the same varieties are not necessarily represented here. Notice the consistency for  $x_2$*

Relative yield	Location 1			Location 2			Location 3		
	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$
Five top lines	+	+	-	+	+	-	-	+	+
	+	+	-	-	+	+	-	+	0
	+	+	0	0	+	0	-	+	-
	+	+	-	-	+	-	-	+	-
	+	+	-	+	+	-	+	+	-
Five bottom lines	-	-	+	-	-	+	-	-	-
	+	-	+	+	-	+	+	-	0
	+	-	-	-	-	+	-	-	+
	-	-	+	-	-	-	+	-	+
	+	-	-	-	-	+	-	-	+

highest yielding lines and the five lowest. In general, the expected values for  $x_3$  are negative for the better lines and positive for the poor lines. It will be recalled that (table 2) the degree of determination of the variation in  $x_3$  by  $x_1$  and  $x_2$  is 66 per cent. for all 36 varieties and hence the error

TABLE 6

*Observed and calculated values for  $x_3$ . The calculated value of  $x_3$  is based on the equation  $x_{n+2} + bx_{n+1} + cx_n = 0$ , where, in this case,  $b = 0.67$ ,  $c = 0.61$ . The initial values of the sequence  $x_n$  and  $x_{n+1}$ , are the observed values for the five highest and the five lowest yielding lines for the average of both environments. The error of estimate for  $x_3$  for all varieties is  $(1 - R^2) = 34$  per cent.*

Varieties	Number	Sequential trait number			
		$x_1$	$x_2$	$x_3$ (calc.)	$x_3$ (observed)
Five highest	1	+0.024	+0.030	-0.035	0.000
	2	+0.017	+0.074	-0.060	-0.041
	3	0.000	+0.031	-0.021	+0.010
	4	+0.005	+0.072	-0.048	-0.055
	5	+0.003	+0.032	-0.023	-0.002
Five lowest	6	-0.024	-0.010	+0.021	+0.013
	7	+0.018	-0.057	+0.027	+0.017
	8	-0.033	-0.014	+0.030	+0.012
	9	-0.010	-0.051	+0.040	+0.014
	10	+0.005	-0.057	+0.035	+0.033

of estimate for all varieties is  $(1 - R^2) = 34$  per cent. As an example of how the recurrence formula is used:

$$x_3 + 0.67x_2 + 0.61x_1 = 0$$

$$x_3 = -0.67x_2 - 0.61x_1.$$

Substituting the values from the top line of table 6:

$$x_3 = -0.67(0.030) - 0.61(0.024) = -0.035.$$

While the above is the identical answer one would get with standard regression techniques, more is involved. The knowledge that the data fit an oscillatory pattern which converges as  $n$  increases encourages us to draw and examine the graphs in fig. 2.

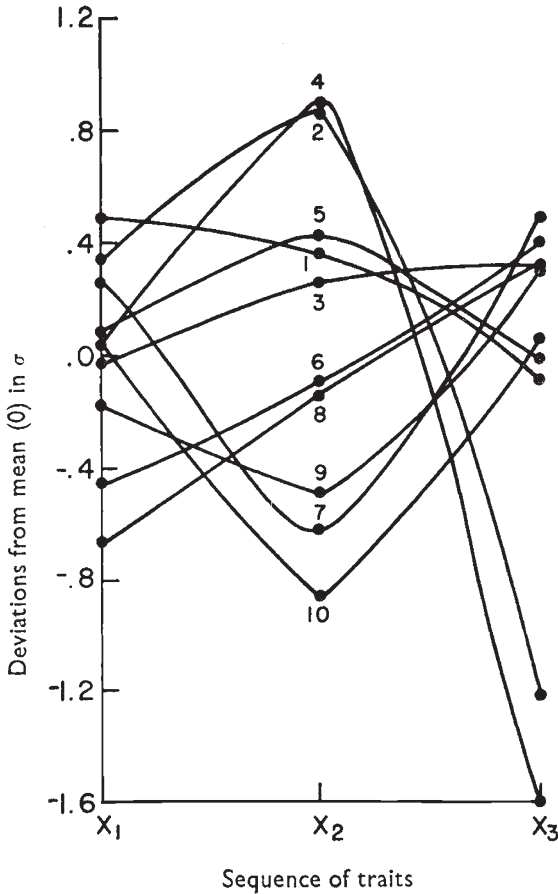


FIG. 2.—Graph of data from table 4. The five highest yielding lines (numbers 1-5) all had a high  $x_2$  value. The five lowest yielding lines (numbers 6-10) all had a low  $x_2$  value. The graphs illustrate the oscillating nature of the patterns and indicate the strategy needed for success. For example, the low yielding line 7 is an inverse image of the high yielding line 2. The numbers on the central axis ( $x_2$ ) refer to the variety numbers in table 4.

The graphs in fig. 2 illustrate the strategies leading to success or failure. A successful variety in this set and in this environment must have a plus value for  $x_2$ . It is interesting to note that curves 2 and 4 are almost exact transposes of curves 7, 9 and 10. Apparently, starting at the same  $x_1$ , success or failure depends on whether the variety elects high or low  $x_2$  as the next move. This choice must be largely under genetic control. Once



$x_2$  is fixed,  $x_3$  is largely determined by the rules of the game, the oscillatory wave form. Line 8 is an apparent exception in that the curve is continuously upward, but it still fits the thesis. Given minus values for  $x_1$  and  $x_2$ , then  $x_3$  should be plus.

#### 4. DISCUSSION

The use of input values as a measure of the trait permits a general treatment of the wave nature of reaction of traits in a sequential series. Conversion of the data to standard measure makes the array for each trait similar to the arrays for neighbouring traits so that a gain of one standard deviation in  $x_1$  exhausts the environmental resources available for  $x_1$  just as fully as a unit gain for trait  $x_2$  exhausts the resources available for  $x_2$ . The observation that a unit gain for  $x_1$  may use more or less of the total resources than a unit gain for  $x_2$  is taken care of by  $b$  and  $c$  in the recursion formula

$$x_{n+2} + bx_{n+1} + cx_n = 0.$$

The solution to this equation permits some general statements to be made about the relationship of traits in a developmental sequence:

If  $b^2 > 4c$ , there are no oscillations and  $x$  expands without limit if the roots of the equation exceed 1.

If  $b^2 < 4c$ ,  $x$  oscillates and the amplitude decreases or increases depending on whether  $1 > c > 1$ . If  $c < 1$ , the oscillations converge.

The subscripts  $n$  in the recursion formula indicate that once  $b$  and  $c$  are found the reaction of a series of  $n$  traits can be described. This would not happen in nature for a very long series of traits since other factors would invariably come into play. But the formula does encourage speculation about series of four or even five traits. For example, number of ears per plant, number of rows per ear, number of kernels per row and average kernel weight form such a series in maize.

The intuitive acceptance of the oscillatory convergent curve ( $B$  in fig. 1) as the most plausible biologic form means that we must also accept the fact that the remote traits in the sequence are not (or only weakly) under direct control of the genetic code for that trait. Control is largely indirect through the determination of the initial traits in the series. This seems at variance with the observed fact that the end traits in a series (seed weight, for example) frequently have the highest heritabilities. Perhaps this is just an illusion. Thomas *et al.* (1970*b*) show that removal of the correlated effects of  $x_1$ ,  $x_2$  and  $x_3$  from  $x_4$  (wheat has four components of yield) in wheat drives the heritability of  $x_4$  (kernel weight) to zero. Such cul-de-sacs are not uncommon, as witness the attempt by breeders to get large seed size incorporated into the phenotype of the new short-strawed wheat varieties. Note, we are not saying it cannot be done, but the chance of success will be enhanced if cognisance is given to the probable paucity of direct genetic control of kernel weight. It may also be possible to raise the entire curve through improved cultural practices so that the mean for seed weight is satisfactory even though  $b$  and  $c$  remain unchanged.

The oscillatory nature of the developmental pattern for traits  $x_1$ ,  $x_2$  and  $x_3$  in oats indicates that a definite strategy is necessary for success (high yield) in the present experiment. Given a common value for  $x_1$ , whether or not a variety succeeded depended on whether it elected high

$x_2$  or low  $x_2$ . Other environments may require different strategies, which will be indicated by the size of  $c$  and by the nature of the difference  $b^2 - 4c$ .

Implicit in the concept of oscillatory convergence is the fact that the more remote a trait from the origin of the sequence, the less the direct genetic control. In the case of  $x_3$ , we have a trait which when analysed separately exhibits a high heritability. In reality, we have just shown that  $x_3$  is only weakly controlled by the genetic code for  $x_3$ . Control is mainly indirect in what happens to  $x_1$  and  $x_2$ . Once  $x_1$  and  $x_2$  are determined, the oscillatory nature of the sequence takes control. Sixty-six per cent. of the variation in  $x_3$  was explained by  $x_1$  and  $x_2$ , leaving a remainder of 34 per cent. for direct control and experimental error.

### 5. SUMMARY

1. We have attempted to give a general description of the nature of the interrelationship of a series of traits in any developmental sequence.

2. Solution of the recursion formula  $x_{n+2} + bx_{n+1} + cx_n = 0$  (where the various  $x_n =$  different traits in the sequence) permits a general statement regarding the nature of any developmental sequence, providing  $n$  is small.

3. If the solution indicates an oscillatory form, and  $b$  and  $c$  are significant, direct genetic control of the more remote traits in the sequence can become very weak. In some cases the variation in the last trait is largely controlled by the magnitude of the first traits in the series.

4. In an example, it was found that direct genetic control of  $x_3$  was very weak, the size of  $x_3$  being largely under control of prior events  $x_1$  and  $x_2$ . Apparently starting at the same  $x_1$  value, success (high yield) or failure of a variety depended on the election of high or low  $x_2$  as the next move. This choice was largely under genetic control. But once  $x_2$  was determined,  $x_3$  was largely determined by the rules of the game, the oscillatory wave pattern.

5. The strategy of a successful variety in this experiment called for election of high  $x_2$ .

*Acknowledgments.*—It is a pleasure to acknowledge the help given by Dr M. W. Adams and C. M. Harrison in their review of this manuscript. Financial support by the Malting Barley Improvement Association is gratefully acknowledged.

Contribution from the Michigan Agricultural Experiment Station, East Lansing, as Paper No. 5012.

### 6. REFERENCES

- ADAMS, M. W. 1967. Basis of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*. *Crop Science*, 7, 505-510.
- GRAFIUS, J. E. 1969. Stress: A necessary ingredient of genotype  $\times$  environment interaction. 2nd International Barley Genetics Conference. Pullman, Washington.
- NICKELL, C. D., AND GRAFIUS, J. E. 1969. Analysis of a negative response to selection for high yield in winter barley, *Hordeum vulgare* L. *Crop Sci.*, 9, 447-451.
- POWELL, J. E., AND WELLS, C. P. 1950. *Differential Equations*. 205 pp. Ginn and Company, N.Y.
- SMITH, J. MAYNARD. 1968. *Mathematical Ideas in Biology*. 152 pp. Cambridge University Press.
- THOMAS, R. L., GRAFIUS, J. E., AND HAHN, SANG KI. 1970a. Stress: an analysis of its source and influence. *Heredity* 26, 423-432.
- THOMAS, R. L., GRAFIUS, J. E., AND HAHN, SANG KI. 1970b. Genetic analysis of correlated sequential characters. *Heredity* 26, 177-188.