

# THE USE OF ENVIRONMENTAL VARIABLES IN THE INTERPRETATION OF GENOTYPE-ENVIRONMENT INTERACTION

J. T. WOOD\*

*National Vegetable Research Station, Wellesbourne, Warwick, England*

Received 12.xii.74

## SUMMARY

The method proposed by Hardwick and Wood (1972) for relating genotype-environment interactions to measures of environmental variables is extended and two examples are discussed.

## 1. INTRODUCTION

THE interpretation of observed interactions between genotypes and the environment has been discussed on numerous occasions. Several authors, including Yates and Cochran (1938), Finlay and Wilkinson (1963), and Perkins and Jinks (1968), have proposed that where a number of genotypes have been tested in several environments, the yield of a genotype should be regressed on the mean yield of all genotypes in each environment. The magnitude of an individual regression coefficient can be interpreted as a measure of the stability of a genotype, a low value indicating stability and a high value instability.

More recently the use of independent measures of the environment has been considered by Fripp (1972), Hardwick and Wood (1972), Perkins (1974), and Jinks and Connolly (1973). Fripp discussed the regression of the yield of genotypes on the yield of control genotypes and also the use of direct measures of environmental variables. Hardwick and Wood showed how to find the linear function of a set of environmental variables which would best explain the observed genotype  $\times$  environment interaction. Perkins (1972) also considered the problem of finding functions of the environmental variables which would best explain the interaction. She calculated principal components of weather variables and then used functions of the first few components as predictors. A disadvantage of this is that an environmental variable which is of little importance in determining the response of the genotypes may make a big contribution to one or more of the first principal components. It would thus carry undue weight in interpretation of the interaction. In a situation where it is not known which environmental variables are important it is convenient to use the genotype yield data directly to determine the function of the environmental variables to be used.

The effect of discarding particular environmental variables can be readily examined. In this paper the method Hardwick and Wood (1972) is extended to the situation where more than one function is necessary to explain the interaction, and two examples are considered. In the second example, data from Perkins (1972) are examined so that the proposed method

\* Dr Wood is currently employed by the CSIRO Division of Mathematics and Statistics, P.O. Box 1965, Canberra City, A.C.T. 2601, Australia.

can be compared with the application of principal components analysis to the environmental data as a preliminary operation.

## 2. THEORY

Suppose the performance of  $m$  genotypes is measured in  $n$  environments,  $p$  environmental variables are measured in each environment, and  $y_{ij}$  is the performance of the  $i$ th genotype and  $x_{hj}$  is a measure of the  $h$ th environmental variable in the  $j$ th environment. Let  $y_{i.}$  be the mean performance of the  $i$ th genotype, and  $\mathbf{Y}$  be the  $m \times n$  matrix with elements,  $(y_{ij} - y_{i.})$ . Similarly let  $\mathbf{X}$  be the  $p \times n$  matrix with elements  $(x_{hj} - x_{h.})$ . The problem is to find functions of the environmental variables which can be used to explain the variance in performance of the genotypes from environment to environment.

First consider the situation with only one genotype. A common approach is to regress the yield of the genotype on the environmental variables using the model

$$y_{ij} - \mu_i = \alpha_1 x_{1j} + \dots + \alpha_p x_{pj} + e_{ij}, \quad (1)$$

where  $\mu_i$  is the expected value of  $y_{i.}$ , the  $\alpha$ 's are regression coefficients, and  $e_{ij}$  is a random error.

Here  $\alpha_1 x_{1j} + \dots + \alpha_p x_{pj}$  can be considered as a measure of the  $j$ th environment. However, if the process is repeated for a number of genotypes a different measure of the environment is likely to be obtained in each case, and this will be of little value in interpreting genotype-environment interactions. It would be more satisfactory to choose an environmental measure which is a good predictor of the performance of all the genotypes, if this is possible.

Consider the predictor

$$E_j = \alpha_1 x_{1j} + \dots + \alpha_p x_{pj}$$

and suppose that the performance of the  $i$ th variety is

$$y_{ij} = \mu_i + (1 + \beta_i)E_j + e_{ij}, \quad (2)$$

where  $\beta_i$  is a parameter which measures how much the response of the  $i$ th genotype to environmental variation varies from the average response. Thus  $y_{ij}$  is partitioned into four components, a genotypic component,  $\mu_i$ , an environmental component,  $E_j$ , an interaction component,  $\beta_i E_j$ , and a random error term.

This is very similar to the model used by Perkins and Jinks (1968) and numerous other authors, the only difference being the way in which the  $E_j$ 's are arrived at. It is convenient to write  $\Theta\phi_i$  for  $(1 + \beta_i)$ , and to impose the constraints,  $\Sigma\phi_i^2 = 1$ ,  $\Sigma E_j^2 = 1$ ,  $\Sigma E_j = 0$ . Hardwick and Wood (1972) described the estimation of the parameters of equation (2) by least squares. For least squares estimates of the parameters

$$\mathbf{Y}\mathbf{E} = \Theta\phi, \quad (3)$$

and

$$\phi' \mathbf{Y}\mathbf{X}' = \Theta \mathbf{E}'\mathbf{X}', \quad (4)$$

where  $\mathbf{E}$  is the vector whose elements are  $E_j, j = 1, \dots, n$ ;  $\boldsymbol{\alpha}$  is the vector whose elements are  $\alpha_i, i = 1, \dots, p$ ; and  $\boldsymbol{\phi}$  is the vector whose elements are  $\phi_i, i = 1, \dots, m$ . Since  $\mathbf{E} = \mathbf{X}'\boldsymbol{\alpha}$ , equations (3) and (4) can be combined to give

$$(\mathbf{XY}'\mathbf{YX}' - \Theta^2\mathbf{XX}')\boldsymbol{\alpha} = 0. \tag{5}$$

Then  $\Theta^2$  is the largest eigenvalue of this equation and  $\boldsymbol{\alpha}$  is the corresponding eigenvector. Equation (3) is used to derive  $\boldsymbol{\phi}$ .

Instead of the single environmental measure,  $E_j$ , several environmental measures,  $E_{kj}, k = 1, \dots, s$  may be needed, where

$$E_{kj} = \alpha_{1k}x_{1k} + \dots + \alpha_{pk}x_{pj}.$$

In this case  $\mathbf{E}_k$  is the vector whose elements are

$$E_{kj}, j = 1, \dots, m,$$

and if the  $\mathbf{E}$ 's are constrained to be orthogonal, it can be shown (Rao, 1964) that the  $\alpha$ 's are the eigenvectors corresponding to the  $s$ th largest eigenvalues in equation (5) and corresponding vectors  $\boldsymbol{\phi}_k$  can be derived as before.

TABLE 1  
 $\alpha$ 's, coefficients of variables in first environmental component

Environmental variables	All variables	Plant data only	Soil data only
N% in plant	0.0238	-0.0559	—
P% in plant	-0.0262	0.2100	—
K% in plant	-0.0008	-0.0468	—
Ca% in plant	0.0825	-0.0653	—
Mg% ( $\times 10$ ) in plant	0.0309	0.0772	—
Fe% ( $\times 10^2$ ) in plant	-0.0217	-0.0444	—
Soil pH	-0.0229	—	-0.0081
P (mg/100 g) in soil, acetate extract	0.0831	—	-0.0138
P (mg/100 g) in soil, bicarb. extract	0.2230	—	0.2300
K (mg/100 g) in soil, acetate extract	0.0090	—	0.0309
Ca (mg/100 g) in soil, acetate extract	-0.0997	—	-0.0133
Mg (mg/100 g) in soil, acetate extract	-0.0547	—	0.0124

### 3. EXAMPLES

#### (i) Data of Davies and Snaydon

Eight populations of the grass species *Anthoxanthum odoratum* and one population of each of four other grass species were grown in 16 environments. The environments were eight contrasting soils, ranging in pH from 4.5 to 8.0, and two harvests; different nitrogen fertiliser applications were confounded with the two harvests. There were three replicates. Similar experiments using fewer soils are discussed in Snaydon (1970).

Twelve environmental variables were measured for each soil (table 1). The six plant variables were measured at each harvest; the six soil variables were measured only once. It was convenient to standardise the environmental variables so that each had a variance of one. The estimates of the coefficients,  $\alpha_{ik}$ , for different variables can then be compared directly.

The statistical analysis described in section 2 was run first on all environmental variables, then on plant and soil variables separately. The  $\alpha$  values

(table 1), so obtained, suggest that phosphate is the major factor determining the relative performance of the populations. The same conclusion is reached whether plant or soil variables are considered. The proportion of the relevant sum of squares accounted for by this model was 93 per cent when all environmental variables are used, 76 per cent when only plant analysis data were used and 90 per cent when only soil analysis data were used.

The populations and species differed in response to environmental variables, as measured by the  $\phi$  values (table 2). In some cases the response of populations, measured by this technique, could be compared with previous

TABLE 2  
 *$\phi$ 's, responses of populations to first environmental component*

Genotypes	All environmental variables	Plant data only	Soil data only
<i>Anthoxanthum odoratum</i> , Park Grass 4 <sub>2</sub> U	0.3174	0.3155	0.3196
<i>Anthoxanthum odoratum</i> , Park Grass 4 <sub>2</sub> L	0.3059	0.3052	0.3068
<i>Anthoxanthum odoratum</i> , Park Grass 3U	0.2780	0.2888	0.2795
<i>Anthoxanthum odoratum</i> , Park Grass 3L	0.2709	0.2688	0.2724
<i>Anthoxanthum odoratum</i> , Roundoak Common	0.2754	0.2948	0.2768
<i>Anthoxanthum odoratum</i> , Hazely Heath	0.2311	0.2300	0.2342
<i>Anthoxanthum odoratum</i> , Watlington	0.2237	0.2245	0.2246
<i>Anthoxanthum odoratum</i> , Whiteknights	0.2848	0.2854	0.2891
<i>Agrostis tenuis</i>	0.2589	0.2741	0.2604
<i>Festuca rubra</i>	0.3910	0.3826	0.3834
<i>Helictotrichon pubescens</i>	0.1719	0.1592	0.1696
<i>Holcus lanatus</i>	0.3809	0.3648	0.3772

measures of response to phosphate, which was apparently the most important environmental variable. For example, the two populations (4<sub>2</sub>U and 4<sub>2</sub>L) from plots of the Park Grass Experiment at Rothamsted that had received phosphate fertiliser continually since 1956, had larger  $\phi$  values than those from unfertilised plots (3U and 3L). Equivalent differences in response to phosphate have already been found in sand culture experiments (Davies and Snaydon, 1974). Similarly the  $\phi$  values of the three grass species:

$$F. rubra = H. lanatus > H. pubescens$$

are equivalent to the order of response to phosphate in sand culture determined by Gittens (1963); *A. tenuis* was not investigated in Gittens' experiments. The differences in response to environmental variables, determined by this technique, therefore appear to match previous results obtained under closely defined conditions, and appear to have a relatively simple ecological explanation.

(ii) *Data from Perkins (1972)*

The second example uses data from Perkins (1972). The data consist of the mean final heights of two sets of eight inbred lines of *Nicotiana rustica*, one grown in each of 10 years and the other in each of 9 years. The environmental variables are the sowing date and six climatic variables. The total of the environment and genotype  $\times$  environment sums of squares is 27,386.1 for set one, and 22,540.3 for set two. A single component taken out in the

way described in section 2 accounts for 68 and 64 per cent of these totals. Two components account for 81 and 74 per cent of the totals. In contrast Perkins (1972) using regression on the first three principal components of the climatic variables and the sowing date and the squares of these four variables accounted for 92 and 94 per cent of the variation. In the set of data for nine seasons the variate which contributed least to the regression was discarded, as otherwise a perfect fit would have been obtained. (It should be noted when interpreting these results that the rainfall value for 1970 may be misprinted in Perkins' paper.)

TABLE 3

*The correlations of the heights of the eight lines with the first and second components from the proposed analysis and the first and second principal components of the environmental variables*

Set	Line	Correlations			
		Proposed analysis		Principal components	
		Component 1	Component 2	1	2
1 (10 Seasons)	12	0.72	0.39	0.29	0.46
	15	0.94	0.07	0.39	0.38
	35	0.94	0.20	0.32	0.37
	42	0.94	0.13	0.35	0.38
	2	0.72	-0.19	0.71	-0.17
	21	0.73	0.27	0.46	0.37
	30	0.70	-0.63	0.52	-0.24
	34	0.72	-0.19	0.44	0.17
2 (9 Seasons)	1	0.71	0.08	0.17	0.56
	5	0.87	0.40	0.47	0.28
	13	0.80	0.00	0.24	-0.12
	38	0.76	0.09	0.44	0.13
	11	0.89	0.16	0.45	0.24
	22	0.89	0.14	0.65	0.13
	23	0.66	-0.60	0.33	0.29
	27	0.84	-0.20	0.34	0.35

The proposed analysis is therefore much more economical in terms of the number of parameters used. Perkins uses 30 and 27 parameters, respectively, to derive the principal components. The regression equations then require 64 and 56 coefficients. In contrast the proposed analysis involves 19 and 18 parameters if one component is fitted or 38 and 36 if two components are fitted.

Some of the differences between the two models are illustrated by table 3 which gives the correlations between the heights of the lines and the components. The correlations are much larger for the first component in the proposed analysis than they are for the first principal component of the environmental variables.

In Perkins' original analysis sowing date was not included in the environmental variables when calculating the component, but it is included here to permit direct comparison of the two methods as they might be applied in situations where there is little *a priori* knowledge of the likely effect of environmental variables.

The contributions of the environmental variables to the components are illustrated by table 4 which shows the correlations between the environmental

variables and the components. The main difference is that sowing date makes a substantial contribution to the first component in the proposed analysis but not in the principal components analysis. Otherwise the first components from the two analyses show a similar pattern of positive correlation with humidity and rainfall and negative correlation with hours of sunshine and temperature.

TABLE 4

*The correlations of the environmental variables with the first and second components from the proposed analysis and the first and second principal components of the environmental variables. The climatic variables are averaged over May, June, July and August*

Set	Environmental variable	Correlations			
		Proposed analysis		Principal components	
		Component 1	Component 2	1	2
1 (10 Seasons)	Relative humidity at 6 a.m.	0.38	-0.50	0.86	-0.20
	Relative humidity at 12 noon	0.36	-0.41	0.82	-0.33
	Daily rainfall	0.34	-0.12	0.79	-0.06
	Daily sunshine in hours	-0.54	0.36	-0.94	0.04
	Maximum air temperature	-0.58	0.08	-0.91	-0.25
	Minimum air temperature	-0.22	-0.22	-0.64	-0.66
	Sowing date	0.38	0.67	-0.14	0.85
2 (9 Seasons)	Relative humidity at 6 a.m.	0.40	0.07	0.85	-0.20
	Relative humidity at 12 noon	0.40	0.33	0.82	-0.34
	Daily rainfall	0.27	-0.01	0.80	-0.10
	Daily sunshine in hours	-0.53	-0.02	-0.94	0.06
	Maximum air temperature	-0.60	0.04	-0.90	-0.25
	Minimum air temperature	-0.28	-0.02	-0.64	-0.67
	Sowing date	0.42	-0.08	-0.09	0.87

*Acknowledgments.*—I would like to thank Dr M. S. Davies, Botany Department, University College, Cardiff, and Dr R. W. Snaydon, Agricultural Botany Department, University of Reading for permission to use the data in the first example.

#### 4. REFERENCES

- DAVIES, M. S., AND SNAYDON, R. W. 1974. Physiological differences among populations of *Anthoxanthum odoratum* L. collected from the Park Grass Experiment, Rothamsted. *J. Appl. Ecol.*, 11, 699-707.
- FINLAY, K. W., AND WILKINSON, G. N. 1963. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.*, 14, 742-754.
- FRIPP, Y. 1972. Genotype-environmental interaction in *Schizophyllum commune*. II. Assessing the environment. *Heredity*, 28, 223-238.
- GITTENS, R. 1963. A study of the structure of some grassland and heath communities in Anglesey. Ph.D. Thesis, University of Wales.
- HARDWICK, R. C., AND WOOD, J. T. 1972. Regression methods for studying genotype-environment interactions. *Heredity*, 28, 209-222.
- JINKS, J. L., AND CONNOLLY, V. 1973. Selection for specific and general response to environmental differences. *Heredity*, 30, 33-40.
- PERKINS, J. M. 1972. The principal component analysis of genotype-environmental interactions and physical measures of the environment. *Heredity*, 29, 51-70.
- PERKINS, J. M. 1974. Orthogonal and principal components analysis of genotype-environmental interactions for multiple metrical traits. *Heredity*, 32, 189-209.
- PERKINS, J. M., AND JINKS, J. L. 1968. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23, 339-356.

- RAO, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya, A*, 26, 329-358.
- SNAYDON, R. W. 1970. Rapid population differentiation in a mosaic environment. I. The response of *Anthoxanthum odoratum* populations to soils. *Evolution*, 24, 257-269.
- YATES, F., AND COCHRAN, W. G. 1938. The analysis of groups of experiments. *J. Agric. Sci. Camb.*, 28, 556-580.