METHODOLOGY FOR IDENTIFYING WIDE ADAPTABILITY IN CROPS

Khalid M. Nor and Foster B. Cady

BU-642-M

March 1978

Abstract

When several varieties are grown over a range of environments, various regression measures of wide adaptability can be calculated. Methodology for identifying widely adaptable crop varieties is developed for analyzing experiments where yield and environmental (site variables) data are available. The analysis, based on a beta response model, is straightforward and easy to apply. Motivated by problems associated with using the site mean yield as an environmental index, emphasis is on the development of a technique for formulating a physical environmental index, combining statistical and subject matter knowledge about the crop and the environment. The methodology is illustrated with maize data from several New York environments.

Additional Key Words: Stability, Genotype by environment interaction, Combining experiments, Regression analysis, Multivariate methods.

METHODOLOGY FOR IDENTIFYING WIDE ADAPTABILITY IN CROPS

Khalid M. Nor and Foster B. Cady

BU-642-M March 1978

Wide adaptation refers to a crop's ability to survive and reproduce in many and diverse environments. Recent references to "wide adaptability" and "stability" in connection with crop improvement actually mean more than natural wide adaptation. There is a definite expectation of the crop's performance in relation to the environmental worth; specifically, that crop yield increases with improvement in environmental factors related to yield. The overall objective here is to develop a feasible and straightforward methodology to identify varieties that are adaptable to a wide range of environments.

Yield is an important criterion in evaluating adaptability. The measure of yield commonly used in making varietal recommendations is an average yield calculated from field trials on a limited number of environments (sites) over several locations and years. Modern agricultural technology demands more than a variety with satisfactory average yield. Specifically desired is a variety also possessing the ability to respond with increasingly better environment. A measure of yield response to environment is a measure of wide adaptability. Such a measure could be used as an additional criterion to evaluate varieties for recommendation, to lower risk and enhance profit potential for the grower, to account for variability in yields over sites, and to transfer agrotechnology to other environments without extensive specific site experimentation.

The merits of a regression approach were shown by Yates and Cochran (1938). The basic problem in regressing yield on environment is the metric or measure of the environment. In the stability parameters developed by Finlay and

Wilkinson (1963) and Eberhart and Russell (1966), the crop itself is used as an index of the environment, specifically the site means, calculated by averaging all the varieties at a site. Use and interpretation of the regression coefficients is stressed by Eagles et al. (1977).

Another class of procedures involves analysis of variance techniques. The magnitude of the contribution of each environment to the total variety by environment interaction components was identified by Horner and Frey (1958), while Plaisted and Peterson (1958) utilized the variance components directly. Intermediate between these two approaches is the regression approach, utilizing the variety-by-environment interaction as given by Perkins and Jinks (1968).

In this paper a quantitative measure of wide adaptability, the beta response model, is developed from a multivariate regression approach. The objective is to characterize a crop's wide adaptability to a range of environments as a specific relationship between the crop's yield in different environments and physical indices of those environments. The beta response model is proposed to formalize this concept and the slope parameters are used in the identification of widely adaptable varieties.

Crop yield (the effect) is the output variable which depends on the physical factors of the environment (the cause), the input variables. Desired is an environmental index so that the site information can be described and compared by a single valued function of the physical environmental factors. Techniques are developed to calculate the physical environmental indices.

BETA RESPONSE METHODOLOGY

Beta response methodology is the fitting of varietal population mean yields, the response variables, to physical environmental measures, the input variables. Yield is assumed to be expressible as a linear function of the physical environmental measure. Each physical environmental measure is computed based on the physical factors of the environment. Physical factors include measurable climatic and soil factors which affect the crop.

One motivation for the development of the beta response technique is to preserve the separation of identities between the crop, the effect, and the environment. The environment is the sum total of all participating environmental factors which cause and contribute to the yield expression of the crop. Yield is an output of the environment. This separation of identities between crops and environments is consistent with a definition given by Billings (1952). He states that the environment includes all the external forces and substances affecting the growth, structure and reproduction of the crop. The crop, or components such as yield, is not an external force. This implies that environmental worth should be obtained from measures of external forces only and not from measures of the crop.

The beta response approach begins by insisting on the separation of identities between crops and environments and ends up by linking them both together with a seal of adaptability. It is an attempt to link together two quantitative measures of a different nature, crop yield and physical environmental indices, which are functionally related in the true sense of cause and effect. Beta response is a description of a particular aspect of this linkage which is felt important to the cause of wide adaptability in crops.

The first step in the development of the methodology are three definitions.

<u>Definition 1</u> The average beta response model is the regression of the average yield for varieties at g environments on an index of physical environmental measures. The resulting regression slope is called the average beta response coefficient.

The varietal beta response model is the regression of each varietal mean at each environment on the index of physical environmental measures. The resulting regression slope is called the varietal beta response coefficient.

<u>Definition 2</u> A physical environmental index is a quantity derived from the environment based on physical environmental measures. The quantity should contain information reflecting the variability of the environments. A function of crop production factors of the environment reflects the raw state of the environmental productive capacity with respect to the performance of the evaluated crop.

<u>Definition 3</u> A crop variety from a population of v varieties grown in g environments is defined as widely adaptable if the varietal beta response coefficient is equal to the average beta response coefficient.

The functional relationship assumed in the beta response concept between the $i\underline{th}$ variety ($i = 1, 2, \cdots, v$) and the $j\underline{th}$ environment ($j = 1, 2, \cdots, g$) is a linear relation of the form

$$Y_{ij} = \alpha_i + \beta_i I_j + \epsilon_{ij} \tag{1}$$

where Y_{ij} = the $i\underline{th}$ varietal response to the $j\underline{th}$ environment, α_i = the $i\underline{th}$ varietal intercept, β_i = the $i\underline{th}$ varietal beta response coefficient, I_j = the $j\underline{th}$ environmental index, and ϵ_{ij} = the error term assumed to be from a normal distribution with a mean of zero and a variance of σ^2 .

By definition 3, any variety within the population whose beta response coefficient, β_i, is equal to the average beta response coefficient, $\bar{\beta}$, is considered widely adaptable.

The environmental index to be proposed utilizes physical variables such as weather and soil variables which have been identified as important environmental variables. In general, an index should contain the maximum information about the environment based on the selected physical variables. This is equivalent to quantifying the productivity and worth of the environment with respect to the crop by reducing the complexities of the environmental factors into a scalar quantity. Used here is a first principal component approach. Other physical indices are given by Nor (1977), including one based on line or discriminant function techniques.

Principal components methodology is developed in detail by Morrison (1976) for the general case of p variables. In the maize example to be presented, three environmental variables are used, x_1 = rainfall, x_2 = temperature, and x_3 = growing degree days. For each environment the data are available for a six-year period. Consequently, the data matrix for each environment has a dimension of six rows and three columns. Added over the six environments, the 36 x 3 data matrix $\underset{\sim}{X}$ is assumed to have a multivariate distribution with a mean vector $\underset{\sim}{\mu}$ and a variance-covariance matrix $\underset{\sim}{V}$. Then, the first principle component of $\underset{\sim}{X}$ is that linear combination $a_1x_1 + a_2x_2 + a_3x_3$, where the a_1, a_2 and a_3 are the elements of the eigenvector associated with the greatest eigenroot of the pooled sample variance-covariance matrix. The eigenvector is calculated with the SAS-PROC FACTOR program (SAS, 1976), using the pooled sample correlation matrix to avoid the difficulty of different scales of measurement. The a_1, a_2 and a_3 weights are then multiplied by the standardized values of the three environmental variables for each of the six environments to give the environmental index "score" to be used in the beta response calculations. Whereas standardization of the variables to obtain the correlation matrix was based on the six years of replications for each location, standardization of variables to obtain the score (index) was based on the six locations. An index thus contains the information about the location (with respect to its past and present) in the correlation matrix, and about the growing season for that location in relation to other locations in the standardized variate.

A Numerical Illustration

Based on 1976 commercial hybrids maize field trials, 18 ($v = 18$) early maturity hybrids were selected which were common to the following six ($g = 6$) environments in New York State: Boonville, Candor, Canton, Chazy, Delevan, and Morrisville.

For each of the locations, three ($p = 3$) environmental variables were measured for six ($r = 6$) consecutive years from 1971 to 1976. The six years constitute the replications. The variables are

x_1: The total rainfall in inches for the period May 1 to August 31.

x_2: The mean daily temperature for the period May 1 to September 30.

x_3: Total growing days, the sum of all the deviations of the daily average temperature from base $50^{\circ}F$, for the period May 1 to September 30.

Physical environmental variables are not limited to weather variables. Soil variables such as fertility and water holding ability may also be used. In fact, all available factors of corn production in the location are candidates for environmental index formation.

Details of the environmental index computations are given by Nor (1977). The resulting values for the six environments were Boonville (-2.52), Delevan (-0.85), Morrisville (-0.06), Canton (0.51), Candor (0.74) and Chazy (1.70). The 18 varietal regressions were calculated by regressing the varietal means, the replicate averages at each environment, on the environmental index values. The resulting regression statistics are given in Table 1. The analysis of variance table shows that the F statistic for testing the null hypothesis of equal varietal slopes is equal to 1.32, a value expected on chance alone between 5 and 10% of the time. Even though the evidence is not overwhelming that the varieties should be differentiated, they have been ranked in the table with low rank identifying those varieties most widely adaptable. Further guide is provided by calculating the estimated standard error for the difference between any varietal beta response, b_i, and the common slope, $\bar{b}$, i.e., the square root of the variance

$$\mathrm{Var}(b_i - \bar{b}) = (V-1)s^2/V\epsilon(I_j - \bar{I})^2$$
$$= (17)(288.59)/(18)(10.39) = (5.12)^2 \quad .$$

Table 1. Beta response analysis of maize trials

Table 1a: The estimated varietal beta response coefficient, V.R.C.; Absolute difference from the estimated average beta response coefficient, A.D.; and Adaptability rank, A.R. The A.R. using the site mean as an index is given in parentheses.

Entry	V.R.C.	A.D.	A.R.	Entry	V.R.C.	A.D.	A.R.
1	13.58	4.31	15 (12)	10	6.37	2.90	12 (14)
2	3.59	5.68	17 (15)	11	8.41	0.86	4 (3)
3	5.03	4.24	14 (17)	12	11.67	2.40	7 (9)
4	8.15	1.12	5 (2)	13	11.97	2.70	11 (10)
5	19.61	10.34	18 (18)	14	9.41	0.14	2 (4)
6	4.74	4.53	16 (6)	15	10.73	1.46	6 (8)
7	5.43	3.84	13 (7)	16	8.67	0.60	3 (11)
8	11.69	2.42	8 (13)	17	9.13	0.14	1 (1)
9	11.87	2.60	10 (5)	18	6.72	2.55	9 (16)

Table 1b: Part of the joint regression analysis of variance: $Y_{ij} = \alpha_i + \beta_i X_j + \epsilon_{ij}$, $i = 1, 2, \cdots, 18$, $j = 1, 2, \cdots, 6$

Source of Variation	d.f.	Sums of Squares	Mean Squares	F
Combined Regressions	18	18725.52		
Common Slope	1	12250.27		
H:β_i's equal	17	6475.25	380.90	1.32
Combined Residuals	72	20778.29	288.59	

Other statistics:

(i) The standard error of a V.R.C. (using the combined residual mean square) = 5.27 .

(ii) The estimated average beta response coefficient: $\bar{b} = 9.27$.

(iii) The standard error of the difference between a V.R.C. and $\bar{b}$ = 5.12 .

An environmental index describes the relative contribution and interplay of factors in the environment. Crop yields, usually in the form of site means, have been and are widely used as biological environmental indices. One motivating force for the present study is the failure of the biological index, the site mean, to be consistent with regression methodology assumptions. When an individual variety is regressed on the site mean index, the variety yield is the dependent variable (a random variable) but is also part of the independent variable. Consequently, the independent variable is also a random variable and the two variables are not statistically independent. This invalidates the use of the ordinary least squares slope coefficient. In practice, the problem is less important if the number of varieties going into each site mean is large, but does give an impetus for developing a methodology which will more nearly fulfill basic assumptions of the regression analysis.

The state of the art, especially with respect to physical environmental indices, during the 1960's is well summarized by the following two quotes:

> "The lack of a quantitative measure of complex natural environments, more than any other single factor, has held up the study and exploitation of adaptation in plant introduction and breeding programmes. The use in the present study of an average performance value (yield) of a large group of varieties has provided an abstract measure of the environment."
>
> K. W. Finlay and G. N. Wilkinson (1963)

> "An index independent of the experimental varieties and obtained from environmental factors such as rainfall, temperature, and soil fertility would be desirable. Our present knowledge of the relationship of these factors and yield does not permit the computation of such an index. Until we can measure such factors in order to formulate a mathematical relation with yield, the average of the varieties in a particular environment must suffice."
>
> S. A. Eberhart and W. A. Russell (1966)

In the past decade important site variables have been identified for explaining yield variability of a crop grown in different environments (Laird and Cady 1969; Voss et al. 1970; Sopher et al. 1973). Improved field instrumentation for measuring site variables are being incorporated with better understanding of the soil, plant and weather interacting effects. Available now is the methodology to associate specific site variables with physiological growth periods of a crop (Meisinger et al. 1978). Consequently, the acquisition of basic site variable data can now become part of the experimental design and planning for camping out experiments in several environments. Fortunately this improvement in knowledge of environmental factors has been associated with the increased understanding and availability of multivariate statistical methods.

The adaptability rankings, based on the first principal component approach used in this paper for formulating a physical environmental index, are in general agreement with rankings resulting from the use of the site means, as shown in Table 1. The Spearman rank correlation coefficient is equal to 0.69 with 16 degrees of freedom. This is encouraging and demonstrates the feasibility of using environmental factor information as a quantitative independent measure of a site's productive capacity. Differences in the rankings do exist; entries 6 and 16 are the most notable. However, a comparison between the methodologies is not appropriate without more extensive field evaluation and verification.

ACKNOWLEDGMENTS

Appreciation is expressed to Professor William Pardee, Department of Plant Breeding and Biometry, Cornell University, for use of the maize variety data.

LITERATURE CITED

Billings, W. D. 1952. The environmental complex in relation to plant growth and distribution. Quarterly Review of Biology 27:251-265.

Eagles, H. A., P. N. Hinz, and K. J. Frey. 1977. Selection of superior cultivars of oats by using regression coefficients. Crop Science 17:101-105.

Eberhart, S. A., and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Science 6:36-40.

Finlay, K. W., and G. N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research 14:742-754.

Horner, T. W., and K. J. Frey. 1957. Methods for determining natural areas for oat varietal recommendations. Agronomy Journal 43:313-315.

Laird, R. J., and F. B. Cady. 1969. Combined analysis of yield data from fertilizer experiments. Agronomy Journal 61:829-834.

Meisinger, J. J., S. Dallyn, F. B. Cady, and D. R. Bouldin. 1978. Relationships between potato yields and weather variables. Submitted to the Agronomy Journal.

Morrison, D. F. 1976. Multivariate statistical methods. 2nd edition. McGraw-Hill, New York.

Nor, K. M. 1977. Beta response as a measure of wide adaptability in crops. Ph.D. thesis, Cornell University. Ithaca, New York.

Perkins, J. M., and J. L. Jinks. 1968. Environmental and genotype-environmental components of variability. Heredity 23:523-535.

Plaisted, R. L., and L. C. Peterson. 1959. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. American Potato Journal 36:381-385.

Sopher, C. D., R. J. McCracken, and D. D. Mason. 1973. Relationships between drouth and corn yields on selected South Atlantic coastal plain soil. Agronomy Journal 65:351-354.

Voss, R. E., J. J. Hanway, and W. A. Fuller. 1970. Influence of soil, management, and climatic factors on the yield response by corn (Zea mays L.) to N, P, and K fertilizer. Agronomy Journal 62:736-740.

Yates, F., and G. Cochran. 1938. The analysis of groups of experiments. Journal of Agricultural Science 28:556-580.