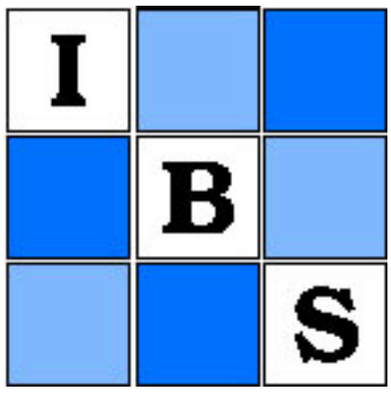

BIOMETRICS 44, 705–715
September 1988

Model Selection and Validation for Yield Trials with Interaction

Hugh G. Gauch, Jr.
Department of Agronomy, Cornell University, Ithaca, New York 14853, U.S.A.

SUMMARY

The additive main effects and multiplicative interaction (AMMI) model first applies the additive analysis of variance (ANOVA) model to two-way data, and then applies the multiplicative principal components analysis (PCA) model to the residual from the additive model, that is, to the interaction. AMMI analysis of yield trial data is a useful extension of the more familiar ANOVA, PCA, and linear regression procedures, particularly given a large genotype-by-environment interaction. Model selection and validation are considered from both predictive and postdictive perspectives, using data splitting and *F*-tests, respectively. A New York soybean yield trial serves as an example.

1. Introduction

Yield trials generate observations of yield, ordinarily replicated, for a number of genotypes grown in a number of environments (site–year combinations). Often the data are rather noisy, with a standard deviation for plot yields in excess of 25% of the mean. An additional challenging feature of these data is the frequent presence of important and complex genotype-by-environment (*GE*) interactions.

Plant breeders use yield trials to identify promising genotypes, and agronomists use them to make recommendations for farmers. The level of success in meeting these goals depends critically on two factors: (i) the accuracy of yield estimates, and (ii) the magnitudes of genotype-by-site, genotype-by-year, and genotype-by-site-by-year interactions (Talbot, 1984). In essence, these two factors reflect within-trial accuracy and between-trial predictability. This paper addresses only the first concern. Nothing is said regarding between-trial predictability, or agrotechnology transfer, other than to observe that success with within-trial accuracy is a necessary prelude to success with between-trial predictability.

The within-trial accuracy of a statistical model may be assessed by two fundamentally different criteria: *Postdictive success* concerns a model's fit to its own data, whereas *predictive success* concerns the fit between a model constructed using part of the data and validation data not used in modelling. In either case, the statistical setting is that of an incompletely specified model, where empirical considerations enter into the decision to include a given potential source in the model, or alternatively to relegate it to the model's residual (Bancroft, 1964). Whenever the data are noisy, postdiction and prediction are different tasks, and in general the model chosen by predictive criteria will be different and simpler than the model chosen by postdictive criteria. Because the several replicates of a yield trial constitute a noisy sample, a model chosen by predictive criteria may be expected to provide yield estimates that are closer to the true means (and to validation observations) than will a model chosen by postdictive criteria. This claim is readily subjected to experimental test. In what follows, the relative performances of statistical models chosen

Key words: AMMI; Analysis of variance; Biplot; Interaction; Model sequence; Postdiction; Prediction; Principal components analysis; Soybean; Validation.

by predictive and postdictive criteria are contrasted, and implications are drawn regarding statistical strategy.

Selection of a general strategy for statistical analysis of yield trials, or of a particular analysis for a given yield trial, requires several decisions: (1) The data may be conditioned prior to analysis in order to remove some problem, such as using a logarithmic transformation to restore normality or to remove interaction. (2) A model must be selected from an enormous roster of available models that partition the treatment $(GE - 1)$ degrees of freedom (df) in a variety of ways. Especially diverse are the possibilities for partitioning the interaction $(G - 1)(E - 1)$ df. (3) Model validation should justify the selection of a particular model, and reveal the model's accuracy and limitations. Combinations of these three choices generate an enormous number of possible statistical analyses.

Data structure, research purposes, and model selection interact in a complex fashion. Model validation must reflect the intended application. Subjective tradeoffs between the model's parsimony and accuracy must be judged primarily by subject-matter considerations. Any data conditioning must be evaluated by the desirability, from an agronomic viewpoint, of its implicit weighting of various features or values in the data, as well as by its statistical implications.

This paper considers several statistical options for analyzing yield trials. Emphasis is on model selection and validation. The analysis receiving primary attention is the biplot or AMMI (additive main effects and multiplicative interaction) model (Gollob, 1968; Bradu and Gabriel, 1978; Kempton, 1984; Gauch, Mimeo 85-7, Department of Agronomy, Cornell University, 1985) with validation of within-trial accuracy by measuring predictive success with a simple data-splitting method. Several alternative analyses are compared. A New York soybean yield trial serves as an example. Although discussed in the context of yield trials, the results are relevant more broadly to two-way layouts with replication that contain considerable noise and important interactions.

Validation in terms of predictive success is emphasized here, rather than postdictive success, because of two advantages: (1) As already mentioned, predictive success criteria tend to diagnose simpler, more parsimonious models. (2) Remarkable predictive success is possible if model diagnosis is effective.

The AMMI model is

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_{n=1}^{N} \lambda_n \gamma_{gn} \delta_{en} + \theta_{ge},$$

where

Y_{ge} is the yield of genotype g in environment e;
μ is the grand mean;
α_g are the genotype mean deviations (the genotype means minus the grand mean);
β_e are the environment mean deviations;
λ_n is the eigenvalue of principal components analysis (PCA) axis n;
γ_{gn} and δ_{en} are the genotype and environment PCA scores for PCA axis n;
N is the number of PCA axes retained in the model;
θ_{ge} is the residual.

If the experiment is replicated, an error term ϵ_{ger}, which is the difference between the Y_{ge} mean and the single observation for replicate r, may be added.

The least-squares fit for balanced data is obtained by first fitting the additive part by the ordinary analysis of variance (ANOVA), and then analyzing the interaction (namely, the nonadditive residual) by PCA (Gabriel, 1978).

The computations required to fit the AMMI model and to evaluate its predictive success are unproblematic, having a linear workload (twice as much data require only twice as much computation; Gauch, 1986). Furthermore, no complications are implied for the experimental design or for field work. Replication, or at least partial replication, is required, but this requirement is already met in most yield trials. In short, the above advantages do not have to be weighed against disadvantages in computational effort, experimental design, or field work.

2. Model Selection

Two steps in model selection are considered here: the preliminary choice of whether to condition the data by applying a transformation intended to remove or reduce the interaction, and the principal choice of a particular partitioning of the $(GE - 1)$ df for treatments (genotype and environment combinations).

If interaction is present, as indicated by a preliminary ANOVA, standard statistical practice recommends attempting to find a transformation to remove or reduce the interaction (Cox, 1984). As a typical example of data conditioning, Finlay and Wilkinson (1963) measured barley yields on a logarithmic scale in order to improve the statistical properties of the data. However, as Knight (1970) cautions, logarithmic transformation increases the influence of the small yield values at the expense of the large yields, in direct opposition to the usual agronomic perspective which focuses interest on the large yields (also see Finney, 1973). Likewise, the value of a crop is closely approximated by a linear expression of yield, but is badly approximated by a logarithmic transformation. Retention of the original yield units also simplifies the final step of reporting and interpreting the results. When agronomic and statistical considerations conflict, the former ordinarily merits the greater weight.

The initial partitioning of the $(GE - 1)$ df for treatments is straightforward given that the principal interest in yield trials is in the genotype means, with some interest also in environment means. This interest implies use of the traditional, additive ANOVA to produce an initial partitioning with $(G - 1)$ df for genotypes, $(E - 1)$ df for environments, and $(G - 1)(E - 1)$ df for interaction. Admittedly, an alternative partitioning, such as a multiplicative model using PCA, may be more statistically efficient for a particular yield trial (as has long been recognized; see Fisher and Mackenzie, 1923). However, the gain in statistical efficiency must be marked in order to outweigh the practical disadvantage of producing a yield trial model lacking genotype means.

The remainder of this section presupposes that we are dealing with the main case identified thus far: The yield trial data are given in untransformed units of yield, and the initial partition of variance is by the additive ANOVA. The remaining task, therefore, is that of finding an effective partitioning of the interaction.

There are at least five possibilities for partitioning the $(G - 1)(E - 1)$ df for interaction:

(1) *Composite* One choice is to regard the interaction as a single, composite source with its $(G - 1)(E - 1)$ df. This choice not to partition the interaction prevails if the analysis stops with the additive ANOVA model.

(2) *PCA of the interaction* Rather than applying PCA to the original yield data to generate a multiplicative model, one can apply PCA to the residual from the additive ANOVA model, that is, to the interaction. This is the AMMI model (Bradu and Gabriel, 1978; Gauch, Mimeo 85-7, Department of Agronomy, Cornell University, 1985). Frequently the first one or few PCA axes concentrate most of the interaction sum of squares SS into relatively few df. The number of df to assign to each PCA axis can be calculated by several different methods, as reviewed in the unpublished mimeo report by Gauch (1985). The simplest method (Gollob, 1968) closely approximates other methods based

on complex theory or extensive simulations, and is used here. PCA axis n is assigned $(G + E - 1 - 2n)$ df.

(3) *Regression* Finlay and Wilkinson (1963) popularized the regression of each genotype's yields on the environment means, partitioning the interaction into $(G - 1)$ df for genotype regressions and the residual $(G - 1)(E - 2)$ df for deviations from regression. At first sight, Finlay–Wilkinson regression and AMMI appear unrelated, but a simple reparameterization shows them to have the same form (Cornelius, 1978; Gauch, Mimeo 85-7, Department of Agronomy, Cornell University, 1985). They differ in the method of fitting, with the regression approach using staged fitting, whereas AMMI provides joint estimates of its parameters (Digby, 1979).

(4) *Single-df contrasts* Interest in single-df interaction contrasts can arise particularly in a case where one or several genotypes and environments are expected to behave differently from the others (perhaps because some are "controls"). Genotypes or environments or both may be grouped to simplify the data (Byth, Eisemann, and De Lacy, 1976). Contrasts may also be suggested by observing interaction PCA axes from AMMI which happen to have a simple structure (Bradu and Gabriel, 1974).

(5) *Single-df for nonadditivity* In the special case of concurrence (all regression lines in a Finlay–Wilkinson diagram intersect at a single point), the interaction can be explained as a constant times the matrix of products of genotype and environment main (additive) effects (Tukey, 1949). This joint regression constant requires 1 df. If the interaction happens to be of this particular structure, this model is very effective, but if not, it is irrelevant.

3. Model Validation

No single recommendation can be given regarding the above five models. Model preference for a given data set and research purpose may involve: (i) general statistical considerations, (ii) model validation by postdictive fitting success, and (iii) model validation by predictive success.

3.1 *General Statistical Considerations*

The composite test for interaction is problematic because its large number of df frequently makes the interaction mean square (MS) nonsignificant in an F-test even when the interaction SS is large. Furthermore, the mere declaration that the interaction is or is not significant is far too coarse a result to give agronomists or plant breeders effective insight into their research material.

Two misunderstandings about testing for interaction are commonplace in agronomic and plant-breeding research. First, when testing the main effects and interaction, the test of interaction should not be regarded as independent but rather as *preliminary* (Bancroft, 1964). Consequently, if genotype and environment effects are tested at the .05 level, the interaction should be tested at the different level of .25, contrary to the usual practice in agronomy of using .05 (or whatever) throughout. Second, the composite test of interaction lacks power, as some agronomists have emphasized (Freeman, 1973), in clear agreement with statistical considerations (Cox, 1984). Unfortunately, the misunderstanding is commonly expressed that failure of the composite test of interaction implies that there is no value or justification to any further analysis of the interaction [for example, Hill (1975) in his otherwise excellent paper]. However, the most extreme case possible is that the entire interaction SS can be captured by a single 1-df interaction contrast (or by 1 df for concurrence), causing this source's MS to equal the interaction SS. Consequently, only if the *GE* interaction SS is insignificant *when assigned* 1 *df* can one reject out of hand the possibility that efforts aimed at partitioning the interaction may be fruitful. Clearly, both

of these misunderstandings about the composite test tend to cause interaction to be discarded prematurely.

The Finlay–Wilkinson regression analysis is frequently effective, and it does have parameters that relate in a natural way to many agronomic and plant-breeding research purposes. However, as will be illustrated here, this model is sometimes entirely inappropriate. Furthermore, the fit is frequently mediocre, as indicated by a significant residual (Hill, 1975; also Finlay and Wilkinson's own example exhibits this problem).

Single-df interaction contrasts are problematic because they are so numerous. Most yield trials have hundreds of df in the interaction. However, classification of the genotypes and environments may help (Byth et al., 1976).

Finally, Tukey's concurrence model is problematic because it presumes a very specific interaction structure that is usually inappropriate.

Because of these problems, AMMI offers a welcome additional possibility. A particularly attractive feature is the automatic generation of a model sequence by including none to all of the interaction PCA axes. However, these five approaches are not just competitive; to some degree they are complementary. Each could be best for certain cases.

AMMI can also be useful for model diagnosis (Bradu and Gabriel, 1974, 1978). It may identify other models or subcases as most appropriate for a given data set. The biplot available from AMMI is a powerful diagnostic tool for identifying single-df interaction contrasts, concurrence, and Finlay–Wilkinson regression. If only the additive or only the multiplicative portions of the AMMI model are significant, then the ANOVA or PCA subcases are indicated. Furthermore, if the PCA axes in AMMI are ineffective in concentrating the interaction SS into a few axes, then the interaction is probably highly complex, and consequently it may be difficult or impossible to find any parsimonious or reduced model; hence, the ANOVA composite test for interaction *is* then appropriate for diagnosing the general means model.

Note that AMMI analysis may be helpful even in cases where its role is the diagnosis of some other model, either simpler or more complex. Indeed, unless one at least tries AMMI, in many cases it will be difficult to justify or to discover the more appropriate model.

Regardless of the approach that is applied to the interaction, residuals from the additive portion of the AMMI model should be examined to make sure that the interaction does not arise from one or a few outliers (Bradu and Gabriel, 1974).

3.2 *Postdictive Success*

The ability of a reduced, parsimonious model to fit its own data constitutes postdictive success, as assessed, for example, by *F*-tests and by noting the percentage of the treatment SS accounted for by a reduced model. Gauch (Mimeo 85-7, Department of Agronomy, Cornell University, 1985) provides 25 references concerning postdictive evaluation of the AMMI model, and recommends Gollob's (1968) simple *F*-test as a practical approach to be regarded as an approximate test and interpreted conservatively. No further comment on postdictive success is offered here because our interest focuses rather on predictive success.

3.3 *Predictive Success*

The ability of a model to predict validation data not used in constructing the model constitutes predictive success. To understand the significance of distinguishing predictive from postdictive success it is necessary to consider two basic concepts.

First, we assume that "data = pattern + noise" (Freeman, 1973), that is, that the data are noisy. This assumption hardly requires defense in the context of yield trials. Consequently, the intent of statistical analysis is to recover pattern, rather than the data: "One

wants to find as much as possible of the pattern while eliminating the maximum noise" (Freeman, 1973). Hence, if the treatment SS of a particular yield trial is composed of 70% pattern and 30% noise, then the goal of statistical analysis is *not* to recover 100% of the treatment SS, but rather only 70%, and more specifically that 70% that represents pattern. If an entire yield trial were duplicated, the resulting two data sets could be regarded as "data_1 = pattern + noise_1" and "data_2 = pattern + noise_2," where the pattern is stable, agronomically meaningful, and has predictive value, whereas the noise is idiosyncratic, uninterpreted, and of no predictive value. Presumably the pattern pertains to inherent, interesting features of genotypes and environments, whereas the noise pertains to stochastic, uncontrolled, and usually unexplainable variability among replicates.

Second, we observe that apposite multivariate models *selectively* recover pattern in their early df, and *selectively* recover noise in their late df. Simulated data having both pattern and noise were used to demonstrate this selectivity for a variant of PCA called reciprocal averaging (also called correspondence analysis, see Gauch, 1982; for PCA see Wold, 1978). Figure 1 shows a schematic representation of this phenomenon. The explanation is straightforward. Pattern involves sizable correlations among matrix values, whereas noise involves idiosyncratic deviations in individual matrix cells, so eigenanalysis (as in reciprocal averaging, PCA, or AMMI) initially concentrates on the pattern. After most of the pattern is exhausted, at first eigenanalysis extracts noise at a somewhat accelerated rate by exploiting chance correlations in the noise, but this is soon exhausted and the remainder of the SS is extracted at a slow rate. Remarkably, while pattern is processed in the early df, the extraction

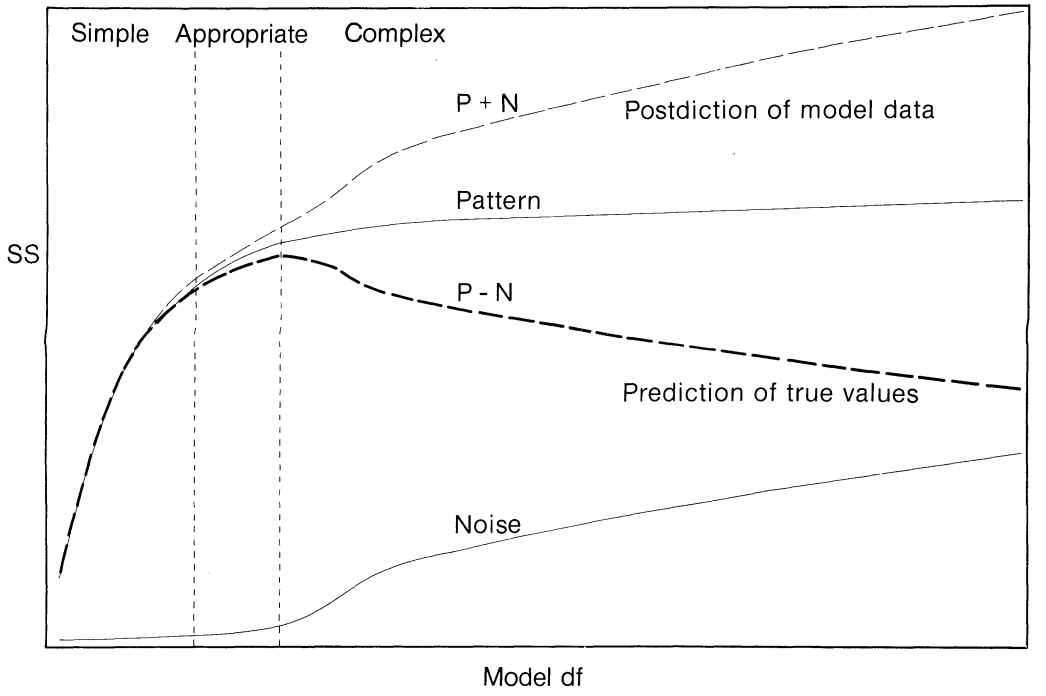

Figure 1. Retrieval of sum of squares as a function of the number of model df for pattern, noise, their sum ($P + N$), and difference ($P - N$). The total SS of $P + N$ pertains to the accuracy of a model for *postdiction* or fitting of the data used to construct the model. The difference $P - N$ pertains to the accuracy of a model for *prediction* of validation data not used to construct the model. Models to the right of the maximum in the $P - N$ curve are too complex (overfitted) and models to the left with a low SS are too simple (underfitted), but an intermediate number of df gives appropriate predictive models.

of noise is suppressed because the concentration on the pattern inhibits the exploitation of chance correlations in the noise. However, the major reason for selective focus on pattern in the early df is simply that the pattern can be described parsimoniously in the early df of the reduced model, whereas the inherent complexity of noise necessarily frustrates any substantial attempt at parsimony. The greater the dimensions of the matrix, and the greater the noise, the greater will be this selectivity. Hence, a large, noisy matrix of yield observations is ideal for demonstrating this selectivity. Of course, if the model is not even remotely apposite, then no selectivity can be expressed. Likewise, a more apposite model will exhibit greater selectivity. The AMMI model's relationship to Taylor series expansions implies its adequate appropriateness and effectiveness in a wide variety of modelling situations (Bradu, 1984).

With these two foundational concepts in mind, it is easy to grasp the significance of distinguishing predictive success from postdictive success. A postdictive perspective (i) does not distinguish pattern from noise, (ii) can only increase model accuracy as model df are accumulated, and (iii) regards the full, unreduced model as being most accurate. A predictive perspective (i) does distinguish pattern from noise, (ii) exhibits an initial increase and then a decrease in accuracy as model df are accumulated, and (iii) regards the full model (equivalent to the original data) as *less* predictively accurate than an appropriate reduced model. The predictive model can predict validation data *better* than the data used to construct it; the postdictive model can postdict its own data only *worse* (or equal in the trivial case of the full model).

The general goals in modelling are to achieve parsimony and accuracy (defined either predictively or postdictively). Because the "pattern + noise" curve of postdictive interest in Figure 1 only rises, these two goals are in continual conflict, leading to difficult compromises in model diagnosis. In contrast, the "pattern − noise" curve of predictive interest rises and then falls, so there is no reason of any kind for accumulating model df beyond the inflection, and hence diagnosis of an appropriate model is easier. From the viewpoint of predictive success, the postdictive model is inferior regarding both modelling goals: It is less parsimonious, and is less predictively accurate.

4. Example

A New York soybean trial with 15 genotypes grown in 15 environments (combinations of 5 sites and 3 years) is analyzed as an example; both a predictive and a postdictive perspective are applied. Only two complete replicates were available due to occasional problems in the field work. A partial third replicate consisting of 217 of the possible 225 yields served as validation data. The yield data and details of the field methods are available in reports from the Department of Agronomy, Cornell University (Mimeos 80-17, 81-23, and 82-9). A randomized complete block design was used. The AMMI model was computed using the MATMODEL program (Gauch, 1986).

Table 1 shows the analysis of variance for AMMI. Note that postdictive evaluation by F-tests finds three interaction PCA axes significant. Including the significant main effects, the diagnosed model contains 103 df. Figure 2 shows the postdiction %SS for the additive model (28 df) and for AMMI with 1 to 14 PCA axes (55 to 224 df, where 224 df is the full model, which equals the input data). This postdiction curve is analogous to the "$P + N$" curve in Figure 1. The salient feature is that credit is given for recovering treatment SS regardless of whether pattern or noise is recovered.

Figure 2 also shows the prediction sum of squared differences SSD between the 217 yield observations reserved as validation data and the corresponding yield predictions for each model in this model sequence. Because low SSD indicates predictive success, the SSD scale decreases in an upward direction so that success has a consistent direction in this figure.

Table 1

ANOVA for AMMI model of a soybean yield trial. Probability is indicated at .05, .01, and .001 levels. The environment MS is tested against the significant block MS.

Source	df	SS	MS	*F*
Environment	14	38,798	2,771.28	70.95***
Block	15	586	39.06	1.76*
Genotype	14	2,552	182.29	8.23***
$G \times E$	196	6,880	35.10	1.59***
PCA 1	27	2,348	86.95	3.91***
PCA 2	25	1,250	49.99	2.26***
PCA 3	23	1,010	43.91	1.98**
PCA 4	21	736	35.03	1.58
Residual	100	1,536	15.36	.69
Error	210	4,649	22.14	
Total	449	53,465		

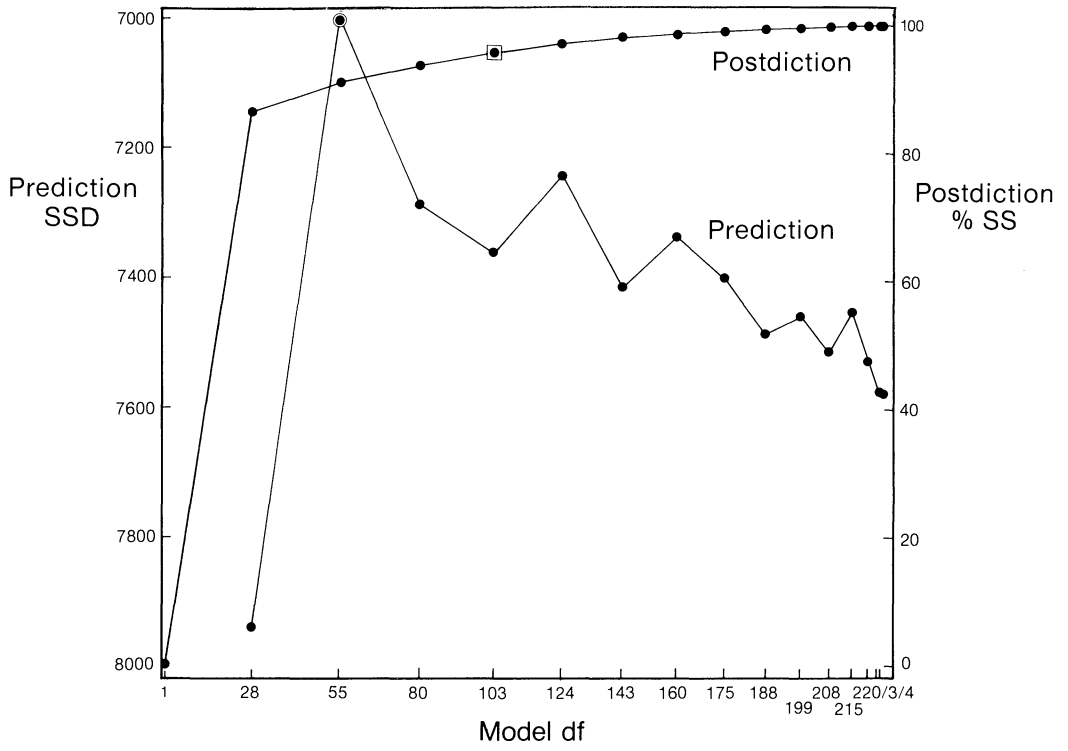

Figure 2. Prediction and postdiction of soybean yields. Sum of squared differences (SSD) between model predictions and validation data for AMMI models having 28 to 224 df (that is, 0 to 14 PCA axes) show maximum predictive accuracy for the model with 55 df (one PCA axis, circled point). Note that the model's input data, equivalent to the full model at 224 df, are *less* predictively accurate. By comparison, the postdiction percentage of sum of squares (SS) only increases. An *F*-test at the .05 level recommends the model with 103 df (three PCA axes, point enclosed by square).

The best predictive model is AMMI with one interaction PCA axis, and it contains 55 df. Beyond this point additional model df capture mostly noise, and this hinders rather than helps in the prediction of validation observations. Note that the predictive model is considerably more parsimonious, having about half the df of the postdictive model. The salient feature is that credit is given for recovering pattern but not for recovering noise.

Incidentally, the Finlay–Wilkinson analysis accounted for only 9% of the interaction SS. The concurrence model captured only 2%. Embarrassingly, the residuals had greater significance than either of these sources. For these data the classical analyses of interaction were not effective.

5. Discussion

A postdictive focus leads to a "tendency to overfit a particular past realization at the expense of the unrealized future" (Harrison and Stevens, 1976). This tendency is demonstrated here with yield trial data. These results may surprise some agronomists inclined to think that a subtle distinction between predictive and postdictive success could have significance only in other sciences, such as physics, offering much more accurate data. In fact, it is precisely the existence of noise that causes prediction and postdiction models to differ, so the noisier the data the greater the benefit of distinguishing between these two criteria of success.

Three refinements of the present procedure merit consideration. First, select the replicate for validation at random. However, this was precluded here by the incompleteness of the third replicate because MATMODEL does not fit AMMI to unbalanced data. Second, it might be worthwhile to remove each replicate in turn and average the validation results. However, this procedure is intermediate to cross validation and data splitting, and its merit thus requires careful evaluation. Third, it does ordinarily seem worthwhile to add a final step of combining all of the data in order to produce a more accurate final model. At this point no predictive model validation is possible, but it seems sensible to retain the model diagnosed by the previous data-splitting analysis. Only rarely would the additional data support an additional PCA axis, and if the resulting increase in accuracy is only marginal then parsimony considerations argue for retaining the original simpler model anyway. Alternatively, one could use statistical theory to estimate the gain in predictive accuracy from using more data, but on the whole, retention of the more conservative previous estimate would seem better.

When experimental sites are used consistently over a span of years, the present genotype-by-environment two-way layout may be viewed alternatively as a genotype-by-site-by-year three-way layout. The extension to three factors is straightforward in ANOVA (that is, in the additive portion of the AMMI model). However, eigenanalysis in the PCA portion of AMMI is defined only for the two-way layout, and cannot be extended to three-way. Consequently, it is not possible to generalize AMMI to three-way layouts. Nevertheless, a three-way problem can often be addressed effectively by AMMI as several two-way subproblems (Brebnor, Bradu, and Schneider, 1977; Gower, 1977; Gauch, Mimeo 85-7, Department of Agronomy, Cornell University, 1985).

Extensions of AMMI models to handle missing data or unbalanced designs are discussed by Freeman (1975), Digby (1979), deLigny et al. (1981), and Wold et al. (1983).

The present assessment of predictive success is minimal in the sense that its scope is *within* a yield trial, as assessed by reserving part of the data for validation. The larger question, which is sure to bring in additional problems and further reduction in predictive accuracy, is predictive success for new sites and new years. Under the rubric of "transferability of agrotechnology," several papers address this larger question (Laird and Cady, 1969; Cady and Allen, 1972; Wood and Cady, 1981; Allen and Jordan, 1982). In brief, one uses environmental characterization of the sites, together with a Bayesian utilization of prior information. Nevertheless, it is important to recognize that even this minimal within-trial analysis of predictive success leads to markedly different results than the customary postdictive analysis, and thus constitutes a significant step in the right direction.

ACKNOWLEDGEMENTS

This research was supported by the Rhizobotany Project, USDA-ARS, under Richard Zobel. I appreciate Rich's suggestions and encouragement. I also appreciate helpful discussions with James Berger, Dan Bradu, Jose Crossa, Ruben Gabriel, Rob Kempton, John Phillips, and Madison Wright, as well as insightful recommendations from the reviewers and editor. This paper is dedicated to the memory of my parents.

RÉSUMÉ

Le modèle comprenant les effets principaux additifs et l'interaction multiplicative (AMMI) consiste à appliquer tout d'abord le modèle d'analyse de variance additif (ANOVA) aux données à 2 facteurs, puis à appliquer le modèle d'analyse en composantes principales multiplicatif (PCA) à la partie résiduelle du modèle additif, c'est-à-dire à l'interaction. L'analyse AMMI des données d'essais de production est une extension très utile des procédures plus habituelles que sont l'ANOVA, la PCA et la régression linéaire, en particulier lorsque l'interaction génotype-milieu est importante. On s'intéresse à la sélection et la validation du modèle d'un point de vue prédictif, comme d'un point de vue "postdictif," par la partition des données et l'utilisation des tests de Fisher–Snedecor. On emploie comme exemple un essai de production de soja de l'état de New-York.

REFERENCES

Allen, D. M. and Jordan, D. C. (1982). The use of prior information for prediction. *Biometrics* **38,** 787–799.

Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics* **20,** 427–442.

Bradu, D. (1984). Response surface model diagnosis in two-way tables. *Communications in Statistics—Theory and Methods* **13,** 3059–3106.

Bradu, D. and Gabriel, K. R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *Journal of the American Statistical Association* **69,** 428–436.

Bradu, D. and Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics* **20,** 47–68.

Brebnor, L., Bradu, D., and Schneider, J. (1977). *A Model for Data from a Biological (Cancer) Research Project*. Pretoria: National Research Institute for Mathematical Science.

Byth, D. E., Eisemann, R. L., and De Lacy, I. H. (1976). Two-way pattern analysis of a large data set to evaluate genotypic adaptation. *Heredity* **37,** 215–230.

Cady, F. B. and Allen, D. M. (1972). Combining experiments to predict future yield data. *Agronomy Journal* **64,** 211–214.

Cornelius, P. L. (1978). Empirical models for the analysis of unreplicated lattice-split-plot cultivar trials. *Crop Science* **18,** 627–633.

Cox, D. R. (1984). Interaction. *International Statistical Review* **52,** 1–31.

deLigny, C. L., Nieuwdorp, G. H. E., Brederode, W. K., van Houwelingen, J. C., and Hammers, W. E. (1981). An application of factor analysis with missing data. *Technometrics* **23,** 91–95.

Digby, P. G. N. (1979). Modified joint regression analysis for incomplete variety × environment data. *Journal of Agricultural Science, Cambridge* **93,** 81–86.

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

Finney, D. J. (1973). Transformation of observations for statistical analysis. *Cotton Growing Review* **50,** 1–14.

Fisher, R. A. and Mackenzie, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *Journal of Agricultural Science, Cambridge* **23,** 311–320.

Freeman, G. H. (1973). Statistical methods for the analysis of genotype–environment interactions. *Heredity* **31,** 339–354.

Freeman, G. H. (1975). Analysis of interactions in incomplete two-way tables. *Applied Statistics* **24,** 46–55.

Gabriel, K. R. (1978). Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society, Series B* **40,** 186–196.

Gauch, H. G. (1982). Noise reduction by eigenvector ordinations. *Ecology* **63,** 1643–1649.

Gauch, H. G. (1986). *MATMODEL*. Ithaca, New York: Microcomputer Power.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33,** 73–115.

Gower, J. C. (1977). The analysis of three-way grids. In *Dimensions of Intra-Personal Space, Volume 2: The Measurement of Intra-Personal Space by Grid Technique*, P. Slater (ed.), 167–173. New York: Wiley.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B* **38,** 205–247.

Hill, J. (1975). Genotype–environment interactions—a challenge for plant breeding. *Journal of Agricultural Science, Cambridge* **85,** 477–493.

Kempton, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science, Cambridge* **103,** 123–135.

Knight, R. (1970). The measurement and interpretation of genotype–environment interactions. *Euphytica* **19,** 225–235.

Laird, R. J. and Cady, F. B. (1969). Combined analysis of yield data from fertilizer experiments. *Agronomy Journal* **61,** 829–834.

Talbot, M. (1984). Yield variability of crop varieties in the U.K. *Journal of Agricultural Science, Cambridge* **102,** 315–326.

Tukey, J. W. (1949). One degree of freedom for nonadditivity. *Biometrics* **5,** 232–242.

Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* **20,** 397–405.

Wold, S., Albano, C., Dunn, W. J., Esbensen, K., Hellberg, S., Johansson, E., and Sjöström, M. (1983). Pattern recognition: Finding and using regularities in multivariate data. In *Food Research and Data Analysis*, H. Martens and H. Russwurm (eds), 1–36. New York: Applied Science Publishers.

Wood, C. L. and Cady, F. B. (1981). Intersite transfer of estimated response surfaces. *Biometrics* **37,** 1–10.

Received November 1985; *revised May* 1986 *and March and November* 1987.