J. Ind. Soc. Agril. Statist.
60(2), 2006 : 118-125

Assessment of Sensitivity with Incomplete Data

B.M.K. Raju, V.K. Bhatia[1] and V.V. Sumanth Kumar[2]
National Council of Educational Research and Training, New Delhi
(Received : February, 2006)

SUMMARY

This paper examines a methodology appearing in literature to evaluate varietal sensitivity while simultaneously dealing with incomplete data as well as random environments. Conditional mean given the environment has been used as regression variable in this method. The limitations and weak points of this methodology are discussed with the help of a real data. Two methods, Zero Substituted BLUP and Ignored BLUP are proposed to evaluate varietal sensitivity when the data are incomplete and environmental effect is treated as random. The proposed methods are shown superior to regression on conditional mean method.

Key words : Regression on conditional mean, Joint regression, Mixed model, Zero Substituted BLUP, Ignored BLUP.

1. INTRODUCTION

The existence of interaction reflecting differences among varieties in their ability to maintain performance over a wide range of environmental conditions has long been recognized (Finlay and Wilkinson 1963). This ability, which is an important property of a crop variety, is usually referred to as the sensitivity or adaptability of a variety. The sensitivity of a variety to environmental change is usually measured using the very popular Joint regression technique, which consists in regression of observed yield on the environmental mean yield. The regression coefficient obtained in the context may be interpreted as linear sensitivity.

If all the cells in the two-way table of mean yields are filled, the computation of linear sensitivity is straightforward. But in practice, the main problematic feature of Multi-Environment Trials (MET) data is its unbalancedness. If data sets from several locations are combined, some genotypes might not have been tested in all the locations. Similarly, variety sortment changes over the years. As new varieties become available for evaluation and the older ones become obsolete, the data set gets further unbalanced.

If yields of some of the genotypes are not available or are unreliable, the orthogonality of the design does not exist and bias is introduced in the observed varietal means. The comparison based on these means is likely to favour the varieties which happen to be exposed to better than average environmental conditions. The main objective of any analysis is to effect such compensation for the environments in which particular varieties are not present. While attempting to address the issue of providing adjustments to varietal effects in an incomplete data situation, Digby (1979) suggested Modified Joint Regression analysis wherein the adjustments offered to the varietal effects made use of the varietal sensitivities. He achieved this using multiplicative model and solving the parameters of the same using an iterative procedure.

When the environments consist of years or site year combinations, treating environment effect as fixed limits the reliability of stability and sensitivity estimates. This is true even for locations as environments when our interest lies in estimation of variety performance within the region sampled by locations. Hence environmental effect may be treated as random, which amounts to assuming the environments in the data as a random sample from the population of all possible environments

[1] *Indian Agricultural Statistics Research Institute, New Delhi-12*
[2] *NRC-Groundnut, Junagarh, Gujarat*

of the testing domain. Estimation of random effects is often called prediction. Treating the environment effect as random will have some desirable consequences in the sense that estimates of varietal effects get corrected for possible random environmental contributions and thus for selection bias. This thus leads to the problem of prediction rather than estimation.

Dealing with the above two aspects (incomplete data and random environments) simultaneously, Jukka (1998) proposed a methodology to evaluate varietal sensitivity when site-year combination is treated as environment. In fact this methodology can be visualised as an extension of Digby's Modified Joint Regression (1979) to the context of random environmental effects.

Jukka (1998) proposed mathematically well defined conditional expectation of yield given the environment be used as a measure for natural environment. Variety testing is a two-stage process in which the experimenter first chooses a set of environments and then collects yield recordings from the environments chosen. Let S_{jk} represents an environment specified by the j-th year and the k-th site (j =1,2,...P; k =1,2,...Q), and $S_{jk}(w)$ represent the population of all possible yield recordings that could have arisen in this environment. Then effect of that environment, e_{jk} may be estimated from the yields of varieties grown in that environment. Let y_{ijk} be the yield of i-th variety grown in the j-th year at the k-th site, then one may write as

$$e_{ij} = E[y_{ijk} / y_{ijk} \in S_{jk}(w)] \qquad (1)$$

Biologically, e_{jk} is the expected value of an observation y_{ijk} arising from an environment S_{jk}. It is a random variable because S_{jk} is randomly chosen. Mathematically, e_{jk} is the conditional expectation of y_{ijk} given the environment S_{jk} (Rao 1973, p. 96). The defined conditional mean becomes an unobservable random variable with a constant realized value for each trial occurring in the data. An observed trial mean can always be used as an unbiased estimator of the respective realized conditional mean. For balanced data, it is also the maximum likelihood estimator for the latter. So, the observed mean can be included as a regressor in the mixed model commonly used in analysis of variety trial data.

Replacing the observed mean with the defined conditional mean e_{jk} and following Patterson and Nabugoomu (1992), the model for the data may be written as

$$y_{ijk} = \alpha_i + \beta_i e_{jk} + \phi_j + \gamma_k + (\alpha\phi)_{ij} + (\alpha\gamma)_{jk} + \varepsilon_{ijk} \qquad (2)$$

where

(y_{ijk} and e_{jk} are as defined earlier)

α_i is the fixed mean deviation of i-th variety

β_i is the fixed coefficient of linear sensitivity associated with i-th variety

ϕ_j is a random effect due to j-th year

γ_k is a random effect due to k-th site

$(\alpha\phi)_{ij}$ is a random effect due to interaction between i-th variety and j-th year

$(\alpha\gamma)_{ik}$ is a random effect due to interaction between i-th variety and k-th site

$(\phi\gamma)_{jk}$ is a random effect due to interaction between j-th year and k-th site

ε_{ijk} is a random residual error

All the random terms are assumed to be mutually independent and the single random effects are independently and identically distributed normal variates with zero expectation.

Jukka (1998) further reported that preliminary analyses indicated that when the model (2) was fitted to the data, the random terms ϕ_j, γ_k and $(\phi\gamma)_{jk}$ were found to be negligible, as could be expected, because these terms were due to the same environmental sources of variation as the regression $\beta_i e_{jk}$. Hence the model chosen for the final detailed analysis of the data by Jukka (1998) is

$$y_{ijk} = \alpha_i + \beta_i e_{jk} + (\alpha\phi)_{ij} + (\alpha\gamma)_{ik} + \varepsilon_{ijk} \qquad (3)$$

Estimation of realized values of e_{jk} simultaneously with other unknown elements appearing in the model requires iterative methods. For this Jukka (1998) introduced a procedure that combines the Expectation and Maximisation (EM) algorithm for the incomplete data with standard maximum likelihood estimation of mixed model parameters. Details of Algebra that lead to best linear predictor for environment effect and final estimators of variety parameters and the iterative

algorithm proposed by Jukka (1998) to solve the parameters may be seen in Appendix.

To examine the methodology proposed by Jukka (1998) with regard its ability to produce reliable estimates of sensitivity, a real data reported by Raju (2002a) has been used. The data set consists of mean pod yields of 15 groundnut genotypes grown in 20 site-year combinations (environments) derived from trials conducted in years 1990 and 1991 in different agro-climatic zones of Andhra Pradesh in Randomised Block Design (RBD) with 3 blocks. Unbalanced data were artificially created by deleting the values in 20 cells taken at random in the two-way table of genotype by environments (Raju 2002a).

Balanced as well as unbalanced groundnut data sets were analysed using SAS code provided in the Appendix of Jukka (1998). The estimates of variance components of the model (3) with balanced as well as unbalanced situation are presented in Table 1. In case of balanced groundnut data, with environmental means as starting values, the convergence was achieved in the first round itself. The convergence criteria taken here was that the sum of squares of deviations of estimates of e_{jk} from the estimates obtained in the preceding round is less than or equal to 0.01. In the unbalanced situation, it was observed that the difference in the estimates of variance components recorded after the first iterative round (correspond to regression on observed environmental means) and the final estimates recorded after the last iterative round (correspond to regression on conditional environmental means) is very large, especially for the component 'variety × site interaction' where the inflation is (nearly) 10 times of the first iterative round estimates. Further, the 'variety × year interaction' component also got doubled as compared to first iterative round (correspond to regression on observed environmental mean).

Table 1. Estimated variance components

Variance component for	Balanced data	Unbalanced data	
		Regression on observed environmental mean	Regression on conditional environmental mean
Variety × site interaction	47312	41315	433951
Variety × year interaction	15977	19757	44089
Residual	86510	87223	91552

Observed environmental mean versus realized conditional mean for the 20 environments resulted from unbalanced groundnut data are presented in Fig. 1. One can see the estimated realized conditional means

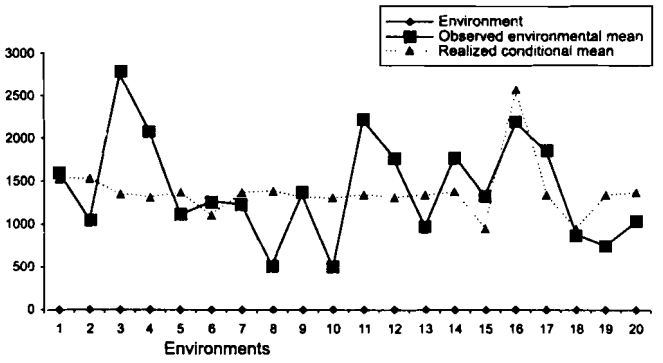

Fig. 1. Observed environmental mean versus realized condtional mean for the 20 environments

(obtained after convergence) differed a great deal from the observed environmental means (starting values). The estimated variety parameters (intercept and slope) corresponding to observed environmental mean and realized conditional mean are presented in Table 2. In this table, one can see a dramatic change in the estimates of variety parameters obtained from regression on conditional mean to that of regression on observed mean. Of course a rise in sensitivity is associated with a reduction in $\hat{\alpha}_i$ (corrected downwards); on the other hand, reduction in sensitivity be compensated by rising $\hat{\alpha}_i$ (corrected upwards). But the change is too dramatic for the varieties 4, 6, 10, 11, 13 and 15.

In Table 2 one may also compare the estimated variety parameters from unbalanced data obtained by regression on conditional mean with estimates obtained from balanced data using same method. It is unrealistic to visualise that the loss of 1 or 2 observations for a variety leads to a completely distorted picture for the results. Further one can see that the estimated variety parameters obtained from regression on observed mean are closer to the estimates obtained in the case of balanced data as compared to estimates obtained from regression on conditional mean. For ease of comparison Table 2 also reports absolute deviations in the estimated variety intercepts and slopes between balanced and unbalanced situations. The proximity of sensitivities obtained from unbalanced data with regression on observed mean as

Table 2. Estimated variety intercepts and slopes

Variety	Balanced data		Unbalanced data				Absolute difference in the estimated variety intercepts and slopes between balanced and unbalanced data situations			
	$\hat{\alpha}_i$	$\hat{\beta}_i$	Regression on observed mean		Regression on conditional mean		Regression on observed mean		Regression on conditional mean	
			$\hat{\alpha}_i$	$\hat{\beta}_i$	$\hat{\alpha}_i$	$\hat{\beta}_i$	$D(\hat{\alpha}_i)$	$D(\hat{\beta}_i)$	$D(\hat{\alpha}_i)$	$D(\hat{\beta}_i)$
1	69.91	1.007	71.47	0.993	-116.01	1.089	1.56	0.014	185.92	0.082
2	-37.43	0.949	-22.80	0.930	36.59	0.860	14.63	0.019	74.02	0.089
3	-252.60	1.184	-226.06	1.168	99.39	0.942	26.55	0.016	352.00	0.242
4	0.81	0.952	62.18	0.885	1176.07	0.035	61.37	0.067	1175.26	0.917
5	-135.70	1.036	-141.93	1.044	148.36	0.834	6.23	0.008	284.06	0.202
6	105.82	1.118	82.33	1.129	-1026.70	1.928	23.49	0.011	1132.56	0.810
7	-40.02	1.069	-16.21	1.053	181.55	0.931	23.81	0.016	221.57	0.138
8	-19.16	0.892	-15.21	0.902	32.19	0.859	3.95	0.010	51.35	0.033
9	-40.23	0.994	-74.52	1.017	17.24	0.961	34.29	0.023	57.47	0.033
10	-41.40	1.016	41.36	0.972	1211.49	0.132	82.76	0.044	1252.89	0.884
11	3.03	0.968	-56.23	0.994	-721.82	1.489	59.26	0.026	724.85	0.521
12	186.30	0.953	159.01	0.957	-170.22	1.177	27.29	0.004	356.52	0.224
13	208.77	0.919	254.80	0.892	965.52	0.419	46.03	0.027	756.75	0.500
14	145.36	0.926	117.84	0.944	220.11	0.865	27.52	0.018	74.75	0.061
15	-152.90	1.019	-122.45	0.996	-658.33	1.436	30.45	0.023	505.43	0.417

D - Represents the absolute deviation

compared to the sensitivities obtained from unbalanced data with regression on conditional mean may be quantified with the help of two statistics viz., correlation coefficient and mean squared deviation. The correlation coefficient between the sensitivities obtained from unbalanced data with regression on observed mean and the sensitivities obtained from balanced data is seen to be of the magnitude 0.95 whereas correlation between the sensitivities obtained from unbalanced data with regression on conditional mean and the sensitivities obtained from balanced data is 0.34. Mean squared deviation (MSD) in sensitivity values are 0.0007 and 0.2109 for regression on observed mean and regression on conditional mean respectively.

Problems with the iterative algorithm of Jukka (1998)

1. It is empirically observed that the convergence for the EM algorithm is very slow. The unbalanced groundnut data considered here for the study took 600 iterations to converge. For some other datasets, the convergence was not even attained.
2. The convergence is more influenced by the structure of unbalancedness rather than the degree of unbalancedness. Here by 'structure' we refer to position of missing cells in two-way table. That means sometimes convergence is not attained with lesser number of missing cells in a certain structure and is attained with larger number of missing cells in a different structure.
3. For some data sets the convergence was found non-monotonic. It converged for some iterations and began to diverge; again it starts converging.
4. For any iterative algorithm, it is expected that starting values do not affect the end results except that influencing the number of iterations required to attain convergence. The algorithm introduced by Jukka (1998) is very much subjective to the starting values of e_{jk}. To illustrate this fact, the methodology was applied to the unbalanced groundnut data with starting values other than the observed environmental means. The results obtained are remarkably different from the results

shown in Fig. 1 and Table 2. Details of arbitrarily chosen starting values for the 20 environments and the realized conditional means obtained after convergence and corresponding estimates of variety parameters may be seen in Raju (2002b).

Nabugoomu *et al.* (1999) also worked on the multiplicative model for estimation of variety means and sensitivity coefficients when locations are treated as random variables and the same has been compared with least squares analysis. Their work was also based on iterative procedure (but does not involve EM algorithm) and it was proposed as an approximate iterative algorithm leading to approximation to exact REML (Residual Maximum Likelihood) solution.

In the light of above empirical observations, there is a need to develop a methodology involving no iterative algorithms (thereby avoiding the problems of convergence) and capable of producing reliable (robust) estimates for variety parameters. Two alternative methodologies have been proposed on the lines of Joint Regression when the data are incomplete and the environment component is treated as random.

2. DEVELOPMENT OF ZERO SUBSTITUTED BLUP AND IGNORED BLUP METHODS

The model considered by Jukka (1998) and Patterson *et al.* (1999) is a multiplicative model wherein sensitivity coefficients and environment effects are to be solved (estimated) simultaneously which eventually calls for implementation of iterative algorithms, which may sometimes prove to be inconclusive. Therefore it is suggested here to consider first the classical additive ANOVA model used for two-way crossed classification with interaction, which may be written as (assuming homogenous plot error variance across the environments)

$$Y_{ijkr} = \mu + \alpha_i + e_{jk} + (\alpha e)_{ijk} + \varepsilon_{ijkr} \qquad (4)$$

$i = 1, 2, \ldots, K; j,k = 1, 2, \ldots, N; r = 1, 2, \ldots, R$

where

Y_{ijkr} is the yield of r-th replicate of i-th variety in jk-th environment

μ is the grand mean

α_i is the main effect of i-th variety such that $\sum_i \alpha_i = 0$

e_{jk} is the main effect of jk-th environment (j-th year and k-th site) such that $\sum_{jk} e_{jk} = 0$

$(\alpha e)_{ijk}$ is the interaction effect of i-th genotype in jk-th environment such that

$$\sum_i (\alpha e)_{ijk} = 0 \text{ and } \sum_{jk} (\alpha e)_{ijk} = 0$$

ε_{ijkr} is random error associated with Y_{ijkr}

The environment and interaction effects here may be estimated from the above additive model and they may be used in computing the coefficients of sensitivity in the second step of analysis. The following two alternatives are suggested on these lines.

2.1 Zero Substituted BLUP Method

To the data from an incomplete genotype × environment table, no model with full interaction terms, like $E(Y_{ijkr}) = \mu + \alpha_i + e_{jk} + (\alpha e)_{ijk}$ can be fitted without confounding or aliasing parts of the interactions. For the missing cells interaction parameters will be undefined. To circumvent this problem, Eeuwijk (1995) while fitting factorial regression model to an incomplete data, as an approximate method, suggested treating appropriate terms in (4) as random. If we reasonably assume that the environment effect and interaction effect as random, we can obtain the Best Linear Unbiased Predictions (BLUPs) of all the environmental effects and the interaction effects except the interaction effects corresponding to missing cells in the two-way genotype × environment table. This can be achieved by substituting the REML estimates of variance components in the Henderson's mixed model. Patterson and Thompson (1971) also gave a method, which consists of maximising the likelihood, not of all the data, but of a set of selected error contrasts. The interactions for the missing cells can now be estimated by the expected value for that term (zero) so that the two-way table of interactions be complete. The idea of Eeuwijk (1995) may be extended to fit Joint Regression to incomplete data, when environmental effect is treated as random.

The complete two-way table of Y_{ijk}'s may be predicted from the fitted mixed model

$$\tilde{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \tilde{e}_{jk} + (\tilde{\alpha e})_{ijk}$$

One may now proceed with regressing the predicted y_{ijk}'s on the BLUP estimates of environmental

effects to obtain the estimates of linear sensitivities as if the data set is balanced.

2.2 Ignored BLUP Method

As done in the Zero Substituted BLUP method, first the mixed model

$$Y_{ijkr} = \mu + \alpha_i + e_{jk} + (\alpha e)_{ijk} + \varepsilon_{ijkr}$$

is fitted, where Y_{ijkr} is the yield of r-th replicate of i-th variety in jk-th environment, μ and α_i are fixed, e_{jk} and $(\alpha e)_{ijk}$ are random. BLUP estimates for e_{jk} as well as $(\alpha e)_{ijk}$ are obtained by substituting the REML estimates of variance components in Henderson's mixed model. Now, the incomplete two-way table of y_{ijk}'s will be predicted from the fitted model

$$\tilde{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \tilde{e}_{jk} + (\tilde{\alpha e})_{ijk}$$

These predictions are done only for the y_{ijk}'s present in the data. The missing cells in the two-way table are left intact. Now for each variety the predicted y_{ijk}'s are regressed on to the BLUP estimates of e_{jk}'s to obtain the estimate of linear sensitivity. The difference with the former methodology stems from the fact that only the y_{ijk}'s existing in the data are predicted and used in the regression and the missing cells are ignored.

3. EVALUATION OF ZERO SUBSTITUTED BLUP AND IGNORED BLUP METHODS

BLUP estimates of 20 environments obtained with the incomplete data are presented in Table 3, which need to be compared with realized conditional means. For this purpose realized conditional means after shifting the origin to zero are also presented in Table 3. The difference observed was of very high magnitude. Here BLUP estimate of environment does not contain or utilize any information regarding the sensitivity of the variety where as the realized conditional mean gets adjusted for the varietal sensitivity. The difference in conditional mean also arises due to the fact that some additional information on sites and years has been used.

As could be expected, the estimates of α_i obtained from fitting the additive model and the estimates of α_i obtained after regressing $\tilde{y}_{ijk}$ on $\tilde{e}_{jk}$ are found to be same with Zero Substituted BLUP method. But with Ignored BLUP, the estimates of α_i obtained after regression are different from that of the estimates obtained from fitting the additive model. One can visualize the difference in the estimates of α_i derived from Zero Substituted BLUP and Ignored BLUP methods in Table 4. Further, while fitting factorial regression model to an incomplete data, Raju and Bhatia (2003) have shown that substitution of zero for the interaction effects of missing cells leads to sensitivity coefficients that are biased towards zero. Consequently, it can be easily shown that Zero Substituted BLUP method for Joint Regression leads to sensitivity coefficients that are biased towards one. Shrinkage towards one in the coefficients of β_i in Zero Substituted BLUP as compared to Ignored BLUP method is evident from Table 4.

Table 3. Realized conditional mean vs BLUP estimate for the environmental effect

Environment	Origin shifted realized conditional mean	BLUP estimate of environmental effect
1	158.50	177.54
2	156.42	– 352.52
3	– 22.85	1336.86
4	– 61.03	655.04
5	– 12.31	– 302.68
6	– 263.54	– 135.60
7	– 10.37	– 184.30
8	9.01	– 895.99
9	– 42.64	– 47.06
10	– 59.46	– 902.46
11	– 34.70	794.24
12	– 63.26	342.23
13	– 31.40	– 438.90
14	9.85	361.50
15	– 433.57	– 81.97
16	1202.09	762.53
17	– 32.47	449.65
18	– 433.68	– 525.88
19	– 31.40	– 646.32
20	– 3.15	– 365.92

Consider the statistics Pearson's correlation coefficient and Mean Squared Deviation (MSD) to quantify the ability of a method (suggested for incomplete data) to produce the sensitivity estimates obtained from the balanced data by regressing the yields on BLUP estimates of environments. Pearson's correlation coefficient of sensitivities obtained from

Table 4. Estimated variety parameters with Zero substitute BLUP and ignored BLUP methods

Variety	Zero Substituted BLUP		Ignored BLUP	
	$\hat{\alpha}_i$	$\hat{\beta}_i$	$\hat{\alpha}_i$	$\hat{\beta}_i$
1	1502.40	1.030	1503.14	1.031
2	1320.40	1.009	1321.24	1.012
3	1420.38	1.086	1420.17	1.087
4	1295.39	1.058	1298.75	1.064
5	1309.47	1.102	1307.83	1.103
6	1692.65	1.103	1692.65	1.103
7	1474.49	1.016	1474.38	1.016
8	1289.77	0.965	1287.53	0.959
9	1380.83	1.078	1380.07	1.078
10	1402.78	1.031	1400.89	1.034
11	1362.41	0.981	1361.90	0.979
12	1471.20	0.950	1468.16	0.945
13	1473.44	0.887	1474.25	0.887
14	1456.61	0.965	1455.83	0.965
15	1277.60	0.932	1281.27	0.922

Jukka (1998) method, Zero Substituted BLUP method and Ignored BLUP method with balanced data sensitivities are – 0.117, 0.951 and 0.953 respectively. MSDs computed are 3.767, 0.021 and 0.019 respectively for Jukka (1998) method, Zero Substituted BLUP method and Ignored BLUP method.

Another measure of association, a non-parametric measure, Spearman rank correlation coefficient has been considered. The reason being, ultimately plant breeder is interested in ranking of genotypes with respect to their sensitivities with the objective of evaluating the stabilities of genotypes under study. Let us assume that sensitivity rank order obtained from balanced data by regressing the yield observations on to the BLUP estimates of environments as true rank order. Spearman rank correlation coefficient for Jukka (1998) method, Zero Substituted BLUP and Ignored BLUP methods with the true rank order were – 0.1571, 0.9464 and 0.95 respectively.

From the above results, the following conclusions may be drawn: Regression on conditional mean is inferior to Zero Substituted BLUP and Ignored BLUP methods. The choice between Zero Substituted BLUP and Ignored BLUP is less critical. However, Ignored BLUP may be preferred over Zero Substituted BLUP as the estimates from the latter are biased towards unity.

REFERENCES

Digby, P.G.N. (1979). Modified joint regression analysis for incomplete variety × environment data. *J. Agri. Sci.*, Cambridge, **93**, 81-86.

Eeuwijk, F.A. Van (1995). Linear and bilinear models for the analysis of multi-environment trials. **I.** An inventory of models. *Euphytica*, **84**, 1-7.

Finlay, K. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant-breeding programme. *Autr. J. Agril. Res.*, **14**, 742-754.

Jukka Ofversten (1998). Assessing sensitivity of agricultural crop varieties. *J. Ind. Soc. Agril. Statist.*, **3(1)**, 37-47.

Nabugoomu, F., Kempton, R.A. and Talbot, M. (1999). Analysis of series of trials where varieties differ in sensitivity to locations. *J. Ag. Biol. Environ. Stat.*, 311-325.

Patterson, H.D. and Nabugoomu, F. (1992). REML and the analysis of crop variety trials. *Proceedings of* XVI *International Biometric Conference.* **I**, 1992 Hamilton, New Zealand, 77-93.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika,* **58(3),** 545-554.

Raju, B.M.K. (2002a). A study on AMMI model and its Biplots. *J. Ind. Soc. Agril. Statist.*, **55(3)**, 297-322 .

Raju, B.M.K. (2002b). On Some Statistical Aspects of Assessing Sensitivity of Crop Varieties. Ph.D. Thesis, P.G. School, Indian Agricultural Research Institute, New Delhi.

Raju, B.M.K. and Bhatia, V.K. (2003). Bias in the estimates of sensitivity from incomplete G × E tables. *J. Ind. Soc. Agril. Statist.*, **56(2)**, 177-189.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley, New York.

APPENDIX-I

Details of Jukka (1998) Method

Derivation of estimators of parameters

Let

$$\eta_{ijk} = \phi_j + \gamma_k + (\alpha\phi)_{ij} + (\alpha\gamma)_{ik} + (\phi\gamma)_{jk} + \varepsilon_{ijk}$$

From the assumptions of model given in (2) of Section 1

$$E[\eta_{ijk}] = 0 \qquad \text{and}$$

$$E[y_{ijk}/e_{jk}] = E[\alpha_i + \beta_i e_{jk} + \eta_{ijk}/e_{jk}] = \alpha_i + \beta_i e_{jk} \qquad (5)$$

From (1) of Section 1 and (5) of this Appendix

$$e_{jk} = \alpha_i + \beta_i e_{jk}$$

Writing $X_{ijk} = \alpha_i + \beta_i e_{jk}$ yields $X_{ijk} = e_{jk}$

Arraying the quantities X_{ijk} into a vector X and the quantities e_{jk} into a vector E, it can be written that $X = TE$

where T is a design matrix defined by the incidence of the quantities e_{jk} through X.

Here both X and E are random variables, so writing V = Var(X) yields

$$Cov(E, X') = var(E)T' = (T'V^{-1}T)^{-1}T'$$

For the given X, the best linear predictor (BLP) for E is

$$\tilde{E} = (T'V^{-1}T)^{-1}T'V^{-1}X \qquad (6)$$

So, if X and Y are known one could easily obtain $\tilde{E}$. On the other hand if E are known, the model (2) would be a standard mixed model and the standard maximum likelihood methods would give the best linear estimators, $\hat{\alpha}_i$ and $\hat{\beta}_i$ for each α_i and β_i. The best linear estimator for X_{ijk} would then be $\hat{X}_{ijk} = \hat{\alpha}_i + \hat{\beta}_i e_{jk}$ and arraying these quantities as the preceding X_{ijk}'s one would obtain the best linear estimator $\hat{X}$ for X.

Iterative Algorithm

Using the preceding arguments, an iterative procedure can be constructed as

Step 1 : Insert a starting value for each e_{jk} in model (2)

Step 2: Considering e_{jk} values as fixed, use standard mixed model analysis to estimate the parameters α_i and β_i as well as the variance components associated with the model (2)

Step 3 : Using the results of step 2, compute the quantities $\hat{X}_{ijk} = \hat{\alpha}_i + \hat{\beta}_i e_{jk}$, form $\hat{X}$ and compute $\hat{V} = var(\hat{X})$

Step 4 : Use $\hat{X}$ and $\hat{V}$, respectively in place of X and V in (6) and solve for $\tilde{E}$

Step 5 : Use the elements of $\tilde{E}$ as updated values for the e_{jk}'s and return to step 1

This procedure is easily seen to be an EM algorithm in which step 2 and 3 constitute E step and step 4 is the M step. It therefore produces consistent and asymptotically efficient maximum likelihood estimators. This iterative algorithm may be evaluated till convergence in $\tilde{E}$.