Available online at www.sciencedirect.com

ScienceDirect

Agriculture and Agricultural Science Procedia 9 (2016) 163 – 169

Agriculture and Agricultural Science Procedia

International Conference on Food, Agriculture, and Natural Resources, FANRes2015

AMMI Model for Yield Estimation in Multi-Environment Trials: A Comparison to BLUP

Halimatus Sa'diyah[a]*, Alfian Futuhul Hadi[b]

[a]*Agronomy department, Faculty of Agriculture, The University of Jember, Jl. Kalimantan 37KampusTegalboto, Jember 68121, Indonesia*
[b]*Mathematics department, Faculty of Mathematics and Natural Sciences, The University of Jember, Jl. Kalimantan 37KampusTegalboto, Jember 68121, Indonesia*

Abstract

Maximum information from the Multi-environment trials (MET) can be reached by seeking the best estimator of each genotype's mean yield in a given environment. AMMI (additive main-effects and multiplicative interaction) is popular for analyzing MET data with fixed effect. When the environment included in MET is the sample of large environment, then environment effects regarded as random may be preferable, so the model is called mixed model. The assessment of it may be viewed as a problem of prediction rather than estimation. The prediction of the outcome of random variables is commonly done by Best Linear Unbiased Prediction (BLUP).Both methods are compared using the experimental rice data set from the Indonesian Rice Consortium's research which aims to evaluate the phenotypic performance of rice (*Oryza sativa*). Applying postdictive success method resultedAMMI10 as the best model, and its Root Mean Square Error Prediction is smaller than BLUP. AMMI was found to outperform BLUP in this rice dataset.

Peer-review under responsibility of the organizing committee of IC-FANRes 2015

Keywords: Genotype-environmental experiment; BLUP; AMMI; random effect..

1. Introduction

The yield trial is one of the most common experiments in agricultural research. It is conducted by testing a number of genotypes in a number of environments, called multi-environment trials (MET). MET is commonly conducted to obtain information that supports recommendations of superior cultivars for cultivation. There are two factors included in MET, genotypes and environments. Environment can be a set of locations, sites, years, etc.

* Corresponding author. Tel.: +62-852-321-10100
E-mail address: sadiyah@unej.ac.id

Peer-review under responsibility of the organizing committee of IC-FANRes 2015
doi:10.1016/j.aaspro.2016.02.113

(Mattjik & Sumertajaya, 2000). In order to study about the best genotype, we have to extract maximum information from the data, and seek the best estimate of the mean yield of each genotype in a given environment. The common estimate is the arithmetic mean of a genotype across replicates in an environment. This mean is often called the cell mean. This method is simple, but not fully exploiting all information contained in the dataset.

Such models have been popularized as additive main effects and multiplicative interaction (AMMI) and a number of applications have been found (Gauch, 1988; 1992). AMMI model combines additive components for main effects (genotype and environment) and multiplicative components for genotype-environment interaction (GEI). It combines a univariate technique (ANOVA) for the main effects and a multivariate technique (PCA-principal component analysis) for GEI. Crossa (1990) suggests that the use of multivariate techniques permits a better use of information than the regression methods in the MET analysis.

AMMI analysis considers the main and interaction effects as fixed. Sometimes, this feature is not suitable in analyzing field data. A factor is commonly taken as random if the observed levels is the random sample from a population. Although the assumption of a truly random sample is often arguable for both environments and genotypes, it is frequently assumed that environments are random, because the environment included in MET is only the sample of large environment. This allows inferences which are not restricted to the observed environments (Piepho, 1998). When environment effects regarded as random and the genotype as fixed, so the model is called mixed model.

In the case of random environment effects, the assessment of the mean yield of a genotype in a certain environment may be viewed as a problem of prediction rather than one of estimation (Searle, Casella & McCulloch 1992). Random environment also implies random genotype-environmental interaction, so the prediction of yield involves prediction of a genotype's random interaction with a specific environment. The prediction of the outcome of random variables is commonly done by Best Linear Unbiased Prediction (BLUP), as originally suggested by Henderson (1975).

The purpose of this paper is to compare the accuracy of these two methods (i.e. AMMI and BLUP). The predictive accuracy was assessed based on the root mean square error prediction (RMSEP). The error is the difference between the yield estimation and the true yield measured in the original data. The smaller the RMSEP valueis, the better its ability to predict yield.

2. Methods

2.1. BLUP (Best Linear Unbiased Prediction) in Mixed Model

In the MET data set, *g* genotypes are tested in each of *e* environment. In each environment, the genotype can be arranged in randomized complete or incomplete block design. For further analysis in this paper, we consider the randomized complete block design (RCBD) in each environment, with the same number of replication or block in each environment. According to Yang (2007), the conventional ANOVA model for this situation is given by:

$$\gamma_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \gamma(\delta)_{jk} + \varepsilon_{ijk} \quad (1)$$

$i=1,\ldots,g;\ j=1,\ldots,e;\ k=1,\ldots r$

Where γ_{ijk} is measured response (i.e., yield) of the *k*-th replication of the *i*-th genotype in the *j*-th environment, μ is the overall mean, $\tau_i$is the effect of the *i*-th genotype, $\delta_j$is the effect of the *j*-th environment, $(\tau\delta)_{ij}$ is the interaction effect of the *i*-th genotype with the *j*-th environment, $\gamma(\delta)_{jk}$is the effect of the *k*-th replication in the *j*-th environment, ε_{ijk} is the random error.

There are three possible versions of Eq. (1): (i) Random model with all effect being random except μ;(ii) fixed model with all effects being fixed except $\gamma(\delta)_{jk}$ and ε_{ijk}; (iii) mixed model with either of genotypic and environmental effect is fixed whereas the other is random (Yang 2007). When it is considered that the genotypic effect is fixed and the environmental effect is random, μ dan τ_i are fixed effects whilesδ_j, $(\tau\delta)_{ij}$, $\gamma(\delta)_{jk}$and ε_{ijk} are independently and normally distributed with zero mean and variances σ_δ^2, $\sigma_{\tau\delta}^2$, $\sigma_{\gamma(\delta)}^2$, and σ_ε^2 respectively. Eq. (1) can be written in the standard linear mixed model (Littell et al., 2006; Yang, 2007),

$y = X\beta + Zu + \varepsilon$

where **y** is an ger×1 vector of observations $\mathbf{y} = [\boldsymbol{y_{111}}, \boldsymbol{y_{211}}, \dots, \boldsymbol{y_{ger}}]'$ **;** $\boldsymbol{\beta}$ is a (g+1)×1 vector of unknown fixed effects $\beta = [\boldsymbol{\mu}, \boldsymbol{\tau_1}, \boldsymbol{\tau_2}, \dots, \boldsymbol{\tau_g}]'$; **u** is an (e+ge+er)×1 vector of random effects,

$$\boldsymbol{u} = \begin{bmatrix} \delta_1, \delta_2, \dots, \delta_e, (\tau\delta)_{11}, (\tau\delta)_{21}, \dots, \\ (\tau\delta)_{ge}, \gamma(\delta)_{11}, \gamma(\delta)_{12}, \dots, \gamma(\delta)_{er} \end{bmatrix},$$

X is a (*g*+1)×1 design matrix, ε is a *ger*×1 vector of random errors $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon_{111}}, \boldsymbol{\varepsilon_{211}}, \dots, \boldsymbol{\varepsilon_{ger}}]'$**.** Random vector **u** and **ε**are assumed to be normally and independently distributed with zero mean vectors and variance-covariance matrixes G and R respectively, such that

$$\begin{bmatrix} \boldsymbol{u} \\ \varepsilon \end{bmatrix} \sim N\left\{ \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \right\}$$

Thus, E(**y**)= Xβ and Var(**y**)=$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$**.**The simple form of **G** and **R** suggested by Yang (2007):

$$\boldsymbol{G} = \begin{bmatrix} \sigma_\delta^2 I_e & 0 & 0 \\ 0 & \sigma_{\tau\delta}^2 I_{ge} & 0 \\ 0 & 0 & \sigma_{\gamma(\delta)}^2 I_{er} \end{bmatrix}$$

and R=$\sigma_\varepsilon^2 I_{ger}$, where I_e, I_{ge}, I_{er} and I_{ger} are the identity matrixes of orders *e*, *g*×*e*, *er* and *ger*, respectively.

In the statistical terminology, estimation of random effects is referred to as prediction (Searle et al., 1992). The corresponding procedure for mixed linear models is BLUP. The basic idea is to estimate the effects in the linear model and then to weight some or all of the effects by an estimate of the pattern-to-noise ratio associated with the respective effect.

The BLUPs of the *ij*-th cell means from a balanced data set using the development of Cornelius and Crossa (1999) is:

BLUP(μ_{ij}) =BLUE(μ_i)+BLUP(δ_j)+BLUP$[(\tau\delta)_{ij}]$

Where BLUE(μ_i)==simple mean of the *i*-th genotype$(\bar{\gamma}_{i..})$,

BLUP$(\delta_j) = \frac{rg\sigma_\delta^2}{E(MS_\delta)}(\bar{\gamma}_{.j.} - \bar{\gamma}_{...})$ and

BLUP$[(\tau\delta)_{ij}]$=$\frac{r\sigma_{\tau\delta}^2}{E(MS_\delta)}(\bar{\gamma}_{.j.} - \bar{\gamma}_{...}) + \frac{r\sigma_{\tau\delta}^2}{E(MS_{\tau\delta})}(\bar{\gamma}_{ij.} - \bar{\gamma}_{i..} - \bar{\gamma}_{.j.} + \bar{\gamma}_{...})$

With $E(MS_\delta) = \sigma_e^2 + g\sigma_{\gamma(\delta)}^2 + r\sigma_{\tau\delta}^2 + rg\sigma_\delta^2$ and E$(MS_{\tau\delta})$=$\sigma_e^2 + r\sigma_{\tau\delta}^2$. $E(MS_\delta)$ and E$(MS_{\tau\delta})$ are expected mean squares for environmental and GEI factors from the ANOVA table, respectively, and Thus,

BLUP$(\mu_{ij}) = \bar{\gamma}_{i..} + h_\delta(\bar{\gamma}_{.j.} - \bar{\gamma}_{...}) + h_{\tau\delta}(\bar{\gamma}_{ij.} - \bar{\gamma}_{i..} - \bar{\gamma}_{.j.} + \bar{\gamma}_{...})$

Where$h_\delta = [rg\sigma_\delta^2 + r\sigma_{\tau\delta}^2]/(MS_\delta)$ and $h_{\tau\delta} = r\sigma_{\tau\delta}^2/E(MS_{\tau\delta})$are the shrinkage factor for environmental and GEI effects, respectively.

2.2. AMMI (Additive Main-effects and Multiplicative Interactions)

The AMMI analysis is the technique to analyze the two-way experimental data, which main effects are additive and the interaction effect is multiplicative. AMMI is assuming that all effects (except error) are fixed. The mean $\bar{\boldsymbol{y}}_{\boldsymbol{ij.}}$ of genotype *i* in environment *j* is the arithmetic mean of r replicates (e.g. Cornelius and Crossa, 1999) by averaging Eq(1) across replication within an environment:

$$\bar{\gamma}_{ij.} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \gamma(\delta)_{jk} + \bar{\varepsilon}_{ij.} Where \bar{\varepsilon}_{ij.} = \sum_{k=1}^{r}(\gamma(\delta)_{jk} + \varepsilon_{ijk})/r \quad (2)$$

The arithmetic mean $\bar{y}_{ij.}$is an estimate of the yield of the *i*th genotype in *j*-th environment. The arithmetic mean is a Best Linear Unbiased Estimator (BLUE) of $\mu + \tau_i + \delta_j + (\tau\delta)_{ij}$ (Searle 1987). The better estimate can be obtained by AMMI (Gauch, 1992). First, AMMI fit additive main effects for genotypes and environments by an ANOVA procedure using BLUE and then apply principle component analysis (PCA) using SVD (Singular Value Decomposition) to the remaining residuals after the fitting of main effects (Piepho, 1994). The interaction plus mean error $(\tau\delta)_{ij}+\bar{\varepsilon}_{ij.}$ can be decomposed into *s* PCA axes:

$$\bar{y}_{ij.} = \mu + \tau_i + \delta_j + \sum_{s=1}^{S} l_s\, a_{is} t_{js} + \theta_{ij}$$

Where l_s is the singular value for PCA axis s, a_{is} the eigenvector of genotype for axis s and t_{js} is the eigenvector of environment, and θ_{ij} is residual that remains if not all axes are used (Crossa, 1990; Gauch, 1992). There are at most min (g - 1, e - 1) axes. The models are denoted as AMMIO, AMMI1,..., AMMIF. Depending on the number of PCA axes retained, using the postdictive success method (Gauch & Zobel, 1988; Dias & Krzanowski, 2003). AMMIO means no PCA axis is fitted, while AMMIF means all axes is used in the model i.e. the cell means model is used.

AMMI may be viewed as a procedure to separate pattern in interaction effect $(\tau\delta)_{ij}$ from noise. It can be achieved by PCA, where the interaction effect is decomposed to many PC axes. In the frequent case, most of the patterns are recovered by the first few components, while most of the noise in later axes. PCA provides a useful low-dimensional representation of the data. The variation captured can be expressed as a percentage of the original total variation (Gauch, 2006).

3. Result and Discussion

We will now compare the predictive accuracy of AMMI to that of BLUP, employing the data taken from the research of Indonesian Rice Consortium. The trial aims in evaluating the phenotypic performance of rice from the latest generation in the different environment. There are 14genotypesevaluated at 20 sites (The environments in this trial are sites). There are 3 genotypesfrom BATAN, 4 genotypes from ICRR, 2 genotypesfrom Biogen, and 2 genotypesfrom IPB, with 3 commercial varieties (Ciherang, Inpari1, Cimelati). It used randomized complete block design (RCBD), 3 plots of replication for each genotypesin each environment.

Table 1. ANOVA degrees of freedom, means squares (MS) and p-value of rice data set

Source of variation	Degrees freedom	MS	P-value
Environment	19	81.11	0.001
Block(environment)	40	1.21	0.001
Genotype	13	7.89	0.001
G×E	247	0.81	0.001
Error	520	0.18	

As described, the grand mean of this rice data is 5.759. The mean yield of the genotype range from 5.132 (genotype 2) to 6.219 (genotype 12). While the mean yield of the environment range from 3.868 (environment 19) to 9.511 (environment 4). A combined analysis of variance for grain yield of the 14rice genotypes tested across 20 environments is presented in Table 1. The main effect differences among genotypes, environments, and the G×E interaction effects were highly significant ($P \leq 0.001$). Presence of the G×E interaction indicates that the phenotypic expression of one genotype might be superior to another genotype in one environment but inferior in a different environment (Falconer & Mackay, 1996), and indicates that AMMI and BLUP procedures that will be applied in this paper is appropriate.

Table 2. The significant IPC axes of rice data

Source	Df[a]	Sum square	Mean square	F	p-value
IPC1	31	46.36	1.50	8.52	0.000
IPC2	29	43.00	1.48	8.45	0.000
IPC3	27	35.16	1.30	7.42	0.000
IPC4	25	19.88	0.80	4.53	0.000
IPC5	23	15.05	0.65	3.73	0.000
IPC6	21	12.12	0.58	3.29	0.000
IPC7	19	8.70	0.46	2.61	0.000
IPC8	17	6.36	0.37	2.13	0.005
IPC9	15	5.64	0.38	2.14	0.007
IPC10	13	4.00	0.31	1.75	0.046

[a]Degrees of freedom of each IPC are computed using the method suggested in Gollob (1968)

Table 2 shows the significant principle components of the interaction matrix (IPC) resulting from postdictive success method. After the estimates of main effects using its BLUE have been calculated, they use to achieve the GEI matrix. There are 280 combinations that serve as reference points to conduct PCAusing SVD method that is applied to the GEI matrix. Decomposing the GEI matrix results the multiplicative term. The postdictive success method through the multiplicative terms results 10 IPC which are significant at 0.05 level. Based on those results, the best model is built from 10 significant IPC, and the modelis called AMMI10. The singular value and its percentage is shown in Table 3. The first singular value as the largest, recover 23.09 % of variation. AMMI10 model used the first 10 singular value in the model, so it recovers 97.74 % variation of the G×E interaction.

Table 3. Singular value and percent

Singular value of fixed GEI	Percent	Cumulative percent
3.9313	23.09	23.09
3.7861	21.42	44.5
3.4234	17.51	62.01
2.5742	9.9	71.91
2.2396	7.49	79.41
2.0097	6.03	85.44
1.7025	4.33	89.77
1.4563	3.17	92.94
1.3714	2.81	95.75
1.1546	1.99	97.74
0.8837	1.17	98.91
0.7535	0.85	99.76
0.404	0.24	100

Table 4 shows the variance components estimates of random effects for sites, block (environments), G×E interaction and errors using the MIVQUE0 method that suggested by Hartley, Rao & LaMotte (1978). Here, the environment accounts 80.3% of the total variation, and the other is lower than 10% of the total variation. This result shows that grain yield was significantly affected by changes in environment. The highly significant environment effect and its high variance component could be attributed to the large differences among the test locations in fertility or both amount and distribution of annual rainfall. The high variations due to environmental differences is expected in MET conducted through several years (Yan & Kang, 2003). The G×E interaction is not too strong, and it reduces the limitation in selection. Because the strong G×E interaction for quantitative traits such as seed yield can

severely limit gain in selecting superior genotypes for improved cultivar development (Baker, 1996). A cultivar grown in different environments will show significant fluctuations in yield performance relative to other cultivars. These changes are influenced by the different environmental conditions and are referred to as genotype-by-environment interaction or G×E (Dias & Krzanowski, 2003; Gauch, 2013).

Table 4. Variance Component Estimates for environments (E), block(environments)and G× E interaction

variance component	Estimate	%
Environments	1.88720	80.32211
block(environments)	0.07424	3.159767
environments ×genotype	0.21250	9.044324
Residual	0.17560	7.473803

Variance component estimates in Table 4 is used to estimate the shrinkage of random effect in order to achieve the estimate of BLUP. It results the shrinkage of 0.985 and 0.784 for environmental and interaction effects, respectively. Evaluation through the estimation using best AMMI model and EBLUP results the RMSEP of AMMI10 is 0.4146693, and 1.7982 for EBLUP. It means that AMMI estimation is closer to the true value of yield than BLUP prediction. So, in Rice data set, AMMI is better procedure to estimate the yield. This RMSEP is reasonable according to the Stroup and Mulitze (1991) who emphasized that in variety trials, BLUP is typically more efficient than BLUE if the number of random effects (sites) more than 200, provided that the distribution of treatment effects is reasonably symmetric. Piepho (1994) suggested that AMMI and BLUP applied together routinely. The assessment will show which model is better in a given situation.

4. Conclusion

Predictive accuracy as assessed by the RMSEP showed that AMMI10 as the best AMMI model is better than BLUP model. The RMSEP is 0.4146693and 1.7982 for AMMI10 and BLUP model respectively. It means that AMMI's estimation is closer to the true value of yield than BLUP prediction. So, in this rice data set, AMMI is better procedure to estimate the mean yield of genotype in each environment

Acknowledgements

Indonesian Rice Consortium and ICRR for providing the secondary data as in this paper.

References

Cornelius, P.L., Crossa, J., 1999. Prediction Assessment of Shrinkage Estimators of Multiplicative Models for Multi-Environment Cultivar Trials. Crop Sci. 39, 998-1008.

Crossa, J., 1990. Statistical Analysis of Multilocation Trials. Adv. Agron. 44, 55-85.

Dias, C.T.S., Krzanowski, W.J., 2003. Model Selection and Cross Validation in Additive Main Effect and Multiplicative Interaction Models. Crop Sci. 43, 865-873.

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. Longman, Essex, U.K

Gauch, H.G., 1988. Model Selection and Validation for Yield Trials with Interaction. Biometrics 44, 705-715.

Gauch, H.G., 1992. Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier. Amsterdam:

Gauch, H.G., Zobel, R.W., 1988. Predictive and Postdictive Success of Statistical Analysis of Yield Trials. Theor. Appl. Genet. 76, 1–10.

Gollob, H.F., 1968. A Statistical Model which Combines Features of Factor Analytic and Analysis of Variance Techniques. Psychometrika 33, 73-115.

Hartley, H.O., Rao, J.N.K., LaMotte, L., 1978. A Simple Synthesis-Based Method of Variance Component Estimation. Biometrics 34, 233-244.

Henderson, C.R., 1975. Best Linear Unbiased Prediction under a Selection Model. Biometrics 31,423-447.

Littell, R.C., Miliken, G.A., Stroup, W.W., Wolfinger, R.D., Schabenberger, O., 2006. SAS for Mixed Model. Second edition. North Caroline: SAS Institute Inc.

Mattjik, A.A., Sumertajaya, I.M., 2000. Perancangan Percobaan dengan Aplikasi SAS dan Minitab. IPB Press. Bogor.
Piepho, H.P., 1994. Best Linear Unbiased Prediction (BLUP) for Regional Yield Trials: a Comparison to Additive Main Effect and Multiplicative Interaction (AMMI) Analysis. Theor. Appl. Genet. 89, 647-654.
Piepho, H.P., 1998. Empirical Best Linear Unbiased Prediction in Cultivar Trials Using Factor-Analytic Variance Covariance Structures. Theor. Appl. Genet. 97: 195-201.
Searle, S.R., 1987. Linear Models for Unbalanced Data. Wiley, New York
Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York
Stroup, W.W., Mulitze, D.K., 1991. Nearest Neighbour Adjusted Best Linear Unbiased Prediction. Am. Stat. 45, 194-200.
Yan, W., Kang, M.S., 2003. GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists. CRC Press. Boca Raton, FL.
Yang, R.C., 2007. Mixed Model Analysis of Crossover Genotype - Environment Interaction. Crop Sci. 47, 1051-1062.