Conference paper

Performance of Empirical BLUE and Empirical BLUP in Swedish Crop Variety Testing

H. Buntaran, H. P. Piepho, J. Hagman, J. Forkman

The Eleventh Working Seminar on Statistical Methods in Variety Testing, June 25-29, 2018, Słupia Wielka, Poland.

Published in

Biuletyn Oceny Odmian 35, 15–17.

PERFORMANCE OF EMPIRICAL BLUE AND EMPIRICAL BLUP IN SWEDISH CROP VARIETY TRIALS

Harimurti Buntaran[1,2], Hans-Peter Piepho[3], Jannie Hagman[2], Johannes Forkman[1,2]

[1]Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, 750 07, Uppsala, Sweden
[2]Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Box 7043, 750 07, Uppsala, Sweden
[3]Institute of Crop Science, University of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany

1. Introduction

The official Swedish variety trials aim at providing information and advice for farmers to decide which variety that is the best or the most suitable with regards to their field conditions. The multi-environment trials (MET) are conducted every year to accomplish this aim. Therefore, reliable statistical methods are necessary to provide accurate prediction.

The current statistical method follows the tradition established in the UK by Patterson and Silvey (1980) and has not been changed for many years. Based on to this tradition, the effects of varieties are modelled as fixed, and so the estimation is "best linear unbiased estimation" (BLUE). On the other hand, Smith et al. (2001) recommended to model the effects of varieties as random, and so the method is known as "best linear unbiased prediction" (BLUP). Robinson (1991) gave several arguments for BLUP being preferable to BLUE: i) BLUP produces smaller expected mean square errors, ii) BLUP is more appropriate in plant variety trials when the goal is to predict the future variety performance by having the correct ranking of the varieties, and iii) BLUP is suitable for small-area estimation. Hence, there is a scope for improvement to give better accuracy for the cultivar performance and ranking in different environments in Sweden.

In this study, we present a cross-validation study of different linear mixed models that utilizes either empirical BLUE (E-BLUE) and empirical BLUP (E-BLUP) to improve the prediction accuracy of variety × regions. The term E-BLUP is used because the variance components need to be estimated.

2. Materials and methods

We used Swedish variety trials in spring-barley and winter wheat yield. The datasets were balanced. The trials were laid out as split-plot trials, since there were two fungicide treatments, i.e., treated and untreated. Within each fungicide treatment, varieties were arranged in an alpha design. The trials were performed in three different Swedish agricultural regions: North, Middle, and South.

The current analysis procedure is done with an unweighted two-stage analysis (Möhring and Piepho, 2009). In the first stage, the experiment is analysed using a linear mixed model with varieties, fungicide treatments, and variety-by-fungicide treatment interactions as fixed effects. The effects of replicates and incomplete blocks are modelled as random. In the second stage, the linear mixed model is applied for each region, fungicide treatment, and year. In this work, we used the second stage data and focused on the fungicide-treated subsets of the datasets.

Tables 2 and 3 list the E-BLUP and E-BLUE models for single-year series and five-year series, respectively. The notation is based on Piepho et al. (2003), where the fixed-effects factors are given before the colon, and the random-effects factors are given after the colon. The letter V is the variety ($c = 1, \ldots, C$), R is the region ($j = 1, \ldots, J$), L is the trial, which was always nested within R ($t = 1, \ldots, T$), and Y is the year (Y = 1, …, m). We modelled the covariance structure of the trials as $\mathbf{G}_{\mathrm{L}} = \bigoplus_{j=1}^{J} \mathbf{T}_j$, where $\mathbf{T}_j$ is an L_j-by-L_j diagonal matrix where all diagonal elements are $\sigma^2_{L_j}$ or, in other words, region-specific variance. This structure was used in all models, except the E-BLUP models that did not include regions (SYR 3 single-year series and MYR 2 five-year series), and the E-BLUE models in the single-year series. The model with heterogeneous residuals structure (only in SYR 1 and SYF 2) are region-specific, i.e., $\mathbf{R} = \oplus_{j=1}^{J} \mathbf{R}_j$, where $\mathbf{R}_j$ is the residual matrix for *j*-th region.

Table 2. List of E-BLUP and E-BLUE models for single-year series

Model		Fixed terms	Random terms	Covariance structure
SYR 1	(Basic BLUP Models)	R	V + L + V.R	$\mathbf{G}_L = \oplus_{j=1}^{J} \mathbf{T}_j$, $\mathbf{R} = \oplus_{j=1}^{J} \mathbf{R}_j$
SYR 2				$\mathbf{G}_L = \oplus_{j=1}^{J} \mathbf{T}_j$, $\mathbf{R} = \sigma^2\mathbf{I}$
SYR 3	(BLUP No Region)	-	V + L	$\mathbf{R} = \sigma^2\mathbf{I}$
SYF 1	(BLUE Region Random)	V	R + L + V.R	$\mathbf{G}_L = \oplus_{j=1}^{J} \mathbf{T}_j$, $\mathbf{R} = \sigma^2\mathbf{I}$
SYF 2	(BLUE currently used model)*	V + R + V.R	L	$\mathbf{G}_L = \oplus_{j=1}^{J} \mathbf{T}_j$, $\mathbf{R} = \oplus_{j=1}^{J} \mathbf{R}_j$

Table 3. List of E-BLUP and E-BLUE models for five-year series

Model	Fixed term	Random terms
MYR 1 (Basic BLUP Model)	R	V + L + Y +V.R + V.Y + Y.R + V.R.Y
MYR 2 (BLUP No Region)	-	L + V + Y + V.Y
MYF 1 (BLUE Region Random)	V	R + L + Y + V.R + V.Y + Y.R + V.R.Y
MYF 2 (BLUE current model)*	V	L + Y + V.Y

For the single-year series cross-validation, we employed a 2-fold cross-validation for all models. Each fold was used as validation set once. The trials in each region were divided equally (50/50) into training and validation set. For the five-year series cross-validation, we modified the leave-one-out cross-validation for the practice of predicting the variety performance based on five-year datasets. The five-year datasets in a row were assigned as the training set and the sixth year were assigned as the validation set. The main interest of variety trials is predicting differences among tested varieties. Thus, the performance of cross-validation should be assessed by mean squared error of prediction differences (MSEP) of variety differences in the training set ($z_{ct} - z_{c't}$), compared to the observed differences in the validation set ($y_{ct} - y_{c't}$):

$$MSEP = \frac{\sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{c \neq i}^{C} [y_{ct} - y_{c't} - (z_{ct} - z_{c't})]^2}{TC(C-1)}$$

We ranked the performance of the models based on the average of MSEP.

3. Results

Tables 3 and 4 list the mean of MSEP for single-year and five-year series, respectively. Clearly, the E-BLUE model (current practice), performed comparatively poorly for winter wheat and spring barley datasets. As it was expected, the E-BLUP models performed the best, i.e., SYR 1 with heterogeneous residuals for the single-year series, and MYR 1 for the five-year series. The SYR 1 with heterogeneous residuals structure for the single-year series was still computationally feasible. However, for the five-year series, this residual structure was not computationally feasible, and so we did not employ such structure.

Our study confirmed the previous simulation studies to use E-BLUP as a routine procedure (Forkman and Piepho, 2013, Kleinknecht et al., 2011, Piepho and Möhring, 2006). The empirical datasets hardly satisfied the assumption of normality. However, E-BLUP per se does not require normality (Searle et al., 1992) and our cross-validation results revealed that the E-BLUP performed better than the E-BLUE.

We demonstrated the potential of borrowing strength across regions from random effects of variety-by-region interaction, thereby increasing the accuracy of region-based yield prediction. In conclusion, the model that should be applied for the routine analysis of single-year series is the model SYR 1, i.e., the model R : V + L + V.R, with heterogeneous residuals. For the five-year series, the model MYR 1, i.e., the model R : V + L + Y +V.R + V.Y + Y.R + V.R.Y, is recommended.

Table 3. Mean of MSEP for single-year series of winter wheat ($N = 8$) and spring barley ($N = 5$).

Ranking	Model	Winter wheat	Spring Barley
		Mean	Mean
1	SYR 1	6781	1751
2	SYR 2	6846	1766
3	SYF 1	7093	1783
4	SYR 3	7245	1814
5	SYF 2* (current method)	7407	1959

Table 4. Mean of MSEP for five-year series winter wheat ($N = 6$) and spring barley ($N = 6$).

Ranking	Model	Winter wheat	Spring Barley
		Mean	Mean
1	MYR 1	854685	276814
2	MYR 2	859878	278789
3	MYF 2* (current method)	938231	307994
4	MYF 1	1940205	611592

Key words: BLUP, cross-validation, multi-environment trials, statistical models

References

Forkman, J. & Piepho, H. P. 2013. Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. The Journal of Agricultural Science, 151**,** 381–395.

Kleinknecht, K., Laidig, F., Piepho, H.-P. & Möhring, J. 2011. Best linear unbiased prediction (BLUP): Is it beneficial in official variety performance trials? Biuletyn Oceny Odmian, 33**,** 21–33.

Möhring, J. & Piepho, H. P. 2009. Comparison of weighting in two-stage analysis of plant breeding Trials. Crop Science, 49**,** 1977–1988.

Patterson, H. D. & Silvey, V. 1980. Statutory and recommended list trials of crop varieties in the United Kingdom. Journal of the Royal Statistical Society. Series A (General), 143**,** 219–252.

Piepho, H. P. & Möhring, J. 2006. Selection in cultivar trials—is it ignorable? Crop Science, 46**,** 192–201.

Piepho, H. P., Büchse, A. & Emrich, K. 2003. A hitchhiker's guide to mixed models for randomized experiments. Journal of Agronomy and Crop Science, 189**,** 310–322.

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. Statistical Science, 6**,** 15–32.

Searle, S. R., Casella, G. & Mcculloch, C. E. 1992. Variance Components. John Wiley & Sons, New York

Smith, A., Cullis, B. R. & Gilmour, A. 2001. The analysis of crop variety evaluation data in Australia. Australian & New Zealand Journal of Statistics, 43**,** 129–145.