RESEARCH ARTICLE

A comparison of empirical BLUP with different considerations of residual error variance for genotype evaluation of multi-location trials

Renhe Zhang and Xiyuan Hu

Northwest A&F University, College of Agronomy, Taicheng Lu 3, Yangling 712100, Shaanxi, China

Abstract

The empirical best linear unbiased prediction (eBLUP) is usually based on the assumption that the residual error variance (REV) is homogenous. This may be unrealistic, and therefore limits the accuracy of genotype evaluations for multi-location trials, where the REV often varies across locations. The objective of this contribution was to investigate the direct implications of the eBLUP with different considerations about REV based on the mixed model for evaluation of genotype simple effects (*i.e.* genotype effects at individual locations). A series of 14 multi-location trials from a rape-breeding program in the north of China were simultaneously analyzed from 2012 to 2014 using a randomized complete block design at each location. The results showed that the model with heterogeneous REV was more appropriate than the one with homogeneous REV in all of the trials according to model fitting statistics. Whether the REV differences across locations were accounted for in the analysis procedure influenced the variance estimate of related random effects and testing of the variance of genotype-location (G-L) interactions. Ignoring REV differences by use of the eBLUP could result not only in an inflation or deflation of statistical Type I error rates for pair-wise testing but also in an inaccurate ranking of genotype simple effects for these trials. Therefore, it is suggested that in application of the eBLUP for evaluation of genotype simple effects in multi-location trials, the heterogeneity of REV should be accounted for based on mixed model approaches with appropriate variance-covariance structure.

Introduction

Best linear unbiased prediction (BLUP), as originally suggested by Henderson (1975) and verified by Harville (1976), has a clearly understood theoretical basis. It is sought such that the correlation between the true and predict effect is maximized and the mean squared error of prediction is minimized among all linear unbiased predictor, provided the assumed model holds and the parameters of the model are known (Searle *et al*., 1992; Mrode, 2005). If parameters are

estimated, this optimality no longer holds, but it can be hoped that the performance of the so-called eBLUP (empirical BLUP) is not far from optimal (Piepho, 1998). A lot of studies (Cornelius *et al*., 1994; Piepho, 1994, 1998; Piepho & Möhring, 2005) have shown that the predictive accuracy of the eBLUP based on a two-way analysis of variance (ANOVA) model was better than that of least-squares estimators based the same models and other models, such as the additive main effects multiplicative interaction (AMMI) models (Piepho, 1998). Therefore, eBLUP has recently gained

increasing acceptance and use for genotype evaluation in plant breeding trials (Smith *et al.*, 2005; Piepho *et al.*, 2008; Kleinknecht *et al.*, 2013).

In the analysis of yield trial data from multiple-location trials it is common to assume a mixed linear model, where genotypes are fixed while locations and interactions are random (Cochran & Cox, 1957; Shukla, 1972; Steel & Torrie, 1980; Kelly *et al.*, 2007). In this context, the genotype simple effects at given locations can be evaluated using the eBLUP. The eBLUP based on mixed model has an advantage of its applicability for unbalanced data. Another salient feature of the eBLUP based on mixed model is that it is not only possible to consider the correlation (or variance-covariance) structure of genotype-location (G-L) interaction but also to account for residual error variance (REV) heterogeneity between the trials conducted in different locations with different levels of precision and eventually to consider spatial variation of error terms. Apart from these, *t*-tests using the eBLUP can be constructed as a worthwhile alternative method for the hypothesis test about genotype effects based on the mixed model framework (Littell *et al.*, 1996, 2006), although BLUP is originally developed for ranking and selection (Robinson, 1991). Some authors have examined the usefulness of the eBLUP *t*-tests based on the mixed model (Forkman & Piepho, 2013; Hu, 2015).

In China, in traditional analysis of the variety trial data with random locations, if there is evidence of variety-location interaction, the variety simple effect difference at specific locations is tested by analyzing each location separately. Such approach is not only inconsistent with mixed model theory but also can limit the power and precision of inference at each location (Littell *et al.*, 1996), because with random locations the appropriate method of inference for variety simple effects at specific locations is BLUP, which permits location-specific inference using information from the entire trial for all locations simultaneously (Atlin *et al.*, 2000a; Piepho & Möhring, 2005; Leiser *et al.*, 2012; Windhausen *et al.*, 2012; Kleinknecht *et al.*, 2013). In China and also in some other countries, the statistical test about difference of experiment effects being random or associated with random effects is usually not done or done not based on BLUP in practice (Littell *et al.*, 1996; Smith *et al.*, 2005).

The usual application of eBLUP as well as eBLUP *t*-tests, as considered in most previous studies, assumed that the REV was homogeneous. However, data from multiple-location trials are often characterized by strong heterogeneous error variation across environments (Piepho, 1995; Casanoves *et al.*, 2005; Hu *et al.*, 2013; Singh *et al.*, 2013). The implication of the heterogeneity of the REV for evaluation of genotype effects by use of the eBLUP has not yet been examined. The objective of this contribution was to compare the evaluation difference by use of the eBLUP with different considerations about REV, *i.e.* the eBLUP based on the mixed model with homogeneous REV and that with heterogeneous REV, in view of ranking and pair-wise test of genotype simple effects based on diverse data sets from realistic multi-location trials and hence to convince the practitioner of using the appropriate procedure for genotype effect evaluation, where the variance heterogeneity of residual error effects would be accounted for. The study contains three consecutive steps: (1) fitting the mixed model to each data set using restricted maximum likelihood (REML) under two different considerations about REV, one assumed homogeneous REV and the second assumed heterogeneous REV, and comparing the appropriateness of the model with the different considerations about REV; (2) examining the influence of the two considerations on estimate and testing of related variances; (3) comparing the difference between the eBLUP with different considerations of REV in ranking and difference testing of genotype simple effects.

Material and methods

Trials and data analysis

The data sets used in this study came from multi-location trials in a rape (*Brassica napus* L.) breeding program in northern China conducted from 2012 to 2014. There were four trial groups (A-B-C-D) in these regions for different production types during each year. Therefore, in total there were 12 data sets (3 years × 4 groups). Some 12–13 genotypes were tested at 10–12 locations each year. The genotypes were totally different each year apart from a control variety. All trials at each location were laid out as a randomized complete block design with three replicates. All trial plots were 20 m^2, planting density was 2.8×10^5 plants/ha and yield data was expressed in kilograms of seed per plot. The details of the data set structure are described in Table 1.

Each of the 12 year-group combinations was treated as an independent data set and separately analyzed based on the mixed model, in which genotype effects were fixed, and block, location and G-L interaction effects were random, respectively. Assuming location as random has an advantage that wide space inference about genotype main effects (*i.e.* genotype effects averaged over the entire population represented by locations) is applicable to entire target population of locations, not only observed but also unobserved, and

Table 1. Structure of data sets of rape evaluation trials for north China.

Year	Groups	Number of	
		varieties	locations
2012	A	13	12
	B	13	12
	C	13	12
	D	12	12
2013	A	12	10
	B	12	10
	C	12	10
	D	12	10
2014	A	12	11
	B	12	10
	C	12	10
	D	12	11

has been adopted by organizers of variety trials in many countries (Littell *et al.*, 1996). Two procedures for evaluation of the simple effect of genotypes at specific locations or sets of locations (*i.e.* location-specific genotype effects) were used. The first was the eBLUP based on a two-way ANOVA mixed model with homogeneous REV structure across locations, and the second was the eBLUP based on the same model with heterogeneous REV structure across locations.

The models and procedures considered here were implemented in the context of the mixed linear models using PROC MIXED of the SAS System, vers. 9.2 (SAS Inst., 2011). The *t*-test of the eBLUP was constructed using the program statement "ESTIMATE" of SAS PROC MIXED. The denominator degrees of freedom of *t*-test were determined using the Kenward-Roger method (Kenward & Roger, 1997) as implemented in the SAS System. This approximation uses the basic idea of Satterthwaite (1941). Its extension relative to the Satterthwaite method of Giesbrecht & Burns (1985) and Fai & Cornelius (1996) is an asymptotic correction of the estimated standard error of model effects due to Kackar & Harville (1984) in small and/or unbalanced data structures.

Assessment of REV and G-L interaction

The Akaike Information Criterion (AIC) (Oman, 1991) was used to evaluate models with homogeneous and heterogeneous REV. The smaller the AIC the better is the performance of the model. Since REML was used, only models with the same fixed-effects structure can be compared. AIC is preferred over the Bayesian Information Criterion (BIC), because the latter has a penalty that involves sample size in terms of independent observational units, and the concept of "effective" sample size is not well defined for mixed models, where random effects give rise to possibly complex dependencies among observations (Raman *et al.*, 2011). In fact, there is no established definition of BIC for mixed models (Pauler, 1998).

Since the model with homogeneous REV is a reduced model compared to a model with heterogeneous REV based on we also used the likelihood-ratio test (LRT) to assess the relative goodness of fit of the two models. With the same principle, whether the variance of G-L interaction, *i.e.* the effect of G-L interaction, significantly existed was also identified using the LRT.

Results

Appropriateness of procedures with different considerations of REV

The AIC value of the model with heterogeneous REV was substantially smaller than that with homogeneous REV for all data sets (Table 2), which implies that the REV of the trials varied across the locations and that the analysis procedures with heterogeneous REV were more appropriate than their homogeneous REV versions. This can be further verified given that the model with heterogeneous residual variances fitted the data significantly ($p < 0.001$ in the LRT, Table 3) better than the model with homogeneous residual variances in all of the trials.

Table 2. Akaike Information Criterion (AIC) values of the model with homogeneous (Hom) and heterogeneous (Het) residual error variance (REV) in fitting the data sets. Smaller AIC values indicate better fitting models.

Year	Group A		Group B		Group C		Group D	
	Hom	Het	Hom	Het	Hom	Het	Hom	Het
2012	604.2	568.3	572.6	514.4	706.5	618.0	560.6	501.9
2013	457.7	428.1	495.4	482.2	521.3	473.6	492.0	455.2
2014	704.1	576.1	596.3	586.2	705.6	630.2	514.6	483.3

Table 3. Likelihood-ratio test (LRT) results of the model with heterogeneous residual error variance (REV) compared to that with homogeneous REV in fitting the trail data sets.

Year	Group A			Group B			Group C			Group D		
	df	χ^2	*p*-value	*df*	χ^2	*p*-value	*df*	χ^2	*p*-value	*df*	χ^2	*p*-value
2012	11	57.9	< 0.0001	11	80.2	< 0.0001	11	110.5	< 0.0001	11	80.7	< 0.0001
2013	9	47.6	< 0.0001	9	31.2	< 0.0001	9	65.7	< 0.0001	9	54.8	< 0.0001
2014	10	148.6	< 0.0001	9	28.1	0.0002	9	92.8	< 0.0001	10	51.3	< 0.0001

df: degrees of freedom.

Estimate and test of variances under different considerations of REV

As well known, the eBLUP is based on the estimate of variances, and only when the variance of G-L interaction statistically significantly exists in multi-location trials an evaluation of the simple effects of genotypes at specific locations is just meaningful. Therefore, a comparatively investigation of the estimate and test of the G-L interaction variance under different considerations about residual error effects may be valuable.

In Table 4 are the percentage differences of the estimates of the involved variance of the models with homogeneous REV from their heterogeneous REV versions. There was some discrepancy of variance estimates between the two considerations about REV. And this discrepancy was large for the block variance, ranging from -57.0% (for the trial of group B in 2012) to 380.0% (for the trial of group C in 2013), intermediate for the G-L interaction variance, ranging from -0.7% (for the trial of group B in 2014) to 13.1% (for the trial of group A in 2013), and very small for location variance, ranging from -0.8% (for the trial of group C in 2013) to 1.7% (for the trial of group C in 2012). This suggest that whether considering the REV variation across locations had impact mainly on estimate of the variance for block and G-L interaction effects and slightly on estimate of the variance for location effects.

The *p*-value of the LRT about the variance of G-L interaction effects was smaller than 0.0001 under both considerations of the REV in all of the considered trials (Table 5), which is extremely small compared to $\alpha = 0.01$ and means that the variance of G-L interaction effects existed highly significantly in these trials. However, the

χ^2 value obviously showed difference between the two considerations of the REV. This suggests that whether or not considering the variation of REV also would influence the test about the variance of G-L interaction effects. Just because of the extremely small *p*-value of the LRT their difference did not showed at $\alpha = 0.01$ level in these special cases.

Ranking of genotypes using eBLUP with different considerations of REV

As shown above, the variance of G-L interaction was highly significantly in all of the trials analyzed in this study. Therefore, evaluations of genotype simple effects at specific locations were necessary. One evaluation lies in ranking genotypes. As a detailed example for showing genotype ranking differences between the eBLUP with two different considerations of REV, Table 6 shows the ranking result of genotype simple effects at different locations using the eBLUP with homogeneous REV and that with heterogeneous REV, respectively, for the trial of group B in 2013. There was some rank discrepancy between the two eBLUP versions. For example, at location L1, genotype 12 ranked first by the eBLUP with homogeneous REV and second by the eBLUP with heterogeneous REV; at location L2, genotype 12 ranked fifth by the eBLUP with homogeneous REV and third by the eBLUP with heterogeneous REV. For this trial the proportion of locations with rank discrepancy of the genotype simple effect between the two eBLUP versions reached 60.0% (6 locations out of 10, *i.e.* locations L1, L2, L4, L6, L8 and L10). At the locations with rank discrepancy of the genotype simple effect, the proportion of genotypes with rank discrepancy between the two eBLUP versions reached from 33.3% (4 genotypes out of 12, *i.e.* genotypes

Table 4. The percentage difference of the variance estimates based on the model with homogeneous residual error variance (REV) compared to its heterogeneous version for the trials.

Variance	2012				2013				2014			
	A[1]	B	C	D	A	B	C	D	A	B	C	D
Location	0.1	0.4	1.7	0.6	-0.3	-0.2	-0.8	-0.3	0.3	-0.1	0.9	0.0
Genotype-location interaction	1.8	12.6	8.1	-0.5	13.1	7.3	0.6	6.9	0.1	-0.7	4.9	1.3
Block	-28.6	-57.0	-42.8	-13.5	-9.4	17.8	380.0	-7.7	-51.6	78.1	-45.3	351.6

[1] Letters are related to groups.

Table 5. Likelihood-ratio test (LRT) results for the variance of genotype-location (G-L) interaction effects obtained from the model with homogeneous and heterogeneous residual error variance (REV) for the trials.

Year	Group	Homogeneous REV			Heterogeneous REV		
		df	χ^2	*p*-value	*df*	χ^2	*p*-value
2012	A	2	421.3	< 0.0001	2	286.9	< 0.0001
	B	2	191.5	< 0.0001	2	135.0	< 0.0001
	C	2	171.1	< 0.0001	1	146.3	< 0.0001
	D	1	200.0	< 0.0001	1	201.4	< 0.0001
2013	A	2	186.8	< 0.0001	2	146.5	< 0.0001
	B	1	136.5	< 0.0001	1	90.2	< 0.0001
	C	1	49.9	< 0.0001	2	77.2	< 0.0001
	D	2	167.7	< 0.0001	1	162.3	< 0.0001
2014	A	2	151.5	< 0.0001	1	238.2	< 0.0001
	B	2	245.4	< 0.0001	1	221.0	< 0.0001
	C	2	147.6	< 0.0001	1	178.9	< 0.0001
	D	2	308.5	< 0.0001	2	306.1	< 0.0001

df: degrees of freedom. The *df* theoretically should be 1 in the LRT for variance of G-L interaction effects in all cases. Because the variance for block in some cases was estimated null, when the G-L interaction effects were dropped in the analysis, the *df* of the LRT in these cases became 2.

2, 6, 10 and 12 at location L1, and genotypes 2, 3, 7 and 9 at location L8) to 58.3% (7 genotypes out of 12, *i.e.* genotypes 1, 4, 6, 7, 9, 11 and 12 at location L2, and genotypes 3, 4, 5, 6, 7, 8 and 10 at location L10). For all trials, the proportion of locations and genotypes with rank discrepancy between the two eBLUP versions is summarized in Table 7. It is to observe that there was rank

discrepancy between the two eBLUP versions in all of the trials. The proportion of locations with rank discrepancy of the genotype simple effect between the two eBLUP versions reached from 18.2% (for the trial of group D in 2014) to 100% (for the trial of group A in 2014). At the locations with rank discrepancy of the genotype simple effect, the proportion of genotypes with rank discrepancy of the genotype simple effect between the two eBLUP versions reached from 15.4% (for the trial of group A in 2012) to 58.3% (for the trial of group B in 2013).

Testing of genotype simple effects using eBLUP with different considerations of REV

We also tested genotype simple effects when there is variance of G-L interaction. To illustrate the difference for pair-wise testing of genotype simple effects between the two eBLUP versions, the ratio of the number of genotype pairs with significant ($\alpha = 0.05$) differences based on the eBLUP with heterogeneous REV compared to its homogeneous version is given in Table 8. With exception of L1–L2 locations for the trials of groups A, C and D in 2012, group C in 2013, as well as groups C and D in 2014, where the number of genotype pairs with significant differences was the same (*i.e.* the ratio of the number of genotype pairs with significant differences between the two eBLUP versions was unity), there was a substantial discrepancy (*i.e.* the mentioned ratio was not unity) of the number of genotype pairs with significant differences between the two eBLUP versions at most locations for these trials and at all locations for the other six trials.

Table 6. Yield rank of genotypes at different locations (L1-L10) for the trial of group B in 2013 according to the genotype effect estimates based on the eBLUP with homogeneous (Hom) and heterogeneous (Het) residual error variance (REV). The genotypes with different rank in Hom to that in Het are in bold type.

Rank	L1		L2		L3		L4		L5		L6		L7		L8		L9		L10	
	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het	Hom	Het
1	12[1]	**2**	2	2	3	3	2	2	2	2	2	2	2	2	12	12	9	9	2	2
2	2	**12**	3	3	2	2	3	3	3	3	3	**8**	3	3	2	**3**	2	2	12	12
3	3	3	6	**12**	12	12	12	**10**	12	12	8	**3**	9	9	3	**2**	5	5	3	**6**
4	5	5	8	8	9	9	10	**12**	10	10	10	10	12	12	5	5	7	7	6	**3**
5	7	7	12	**6**	8	8	5	5	5	5	9	9	6	6	9	**7**	3	3	8	**10**
6	9	9	4	**7**	5	5	9	9	4	4	5	**11**	7	7	7	**9**	10	10	10	**4**
7	8	8	7	**9**	4	6	6	**11**	11	11	12	**5**	8	8	8	8	6	6	4	**8**
8	6	**10**	9	**4**	6	4	8	**6**	8	8	11	**12**	4	4	10	10	8	8	5	**7**
9	10	**6**	5	5	7	7	11	**8**	9	9	7	7	5	5	6	6	12	12	7	**5**
10	4	4	10	10	10	10	7	7	6	6	4	4	10	10	11	11	4	4	9	9
11	11	11	1	**11**	11	11	4	4	7	7	6	6	11	11	4	4	1	1	11	11
12	1	1	11	**1**	1	1	1	1	1	1	1	1	1	1	1	1	11	11	1	1

[1] Numbers are related to genotypes.

Table 7. Percentage of locations and genotypes with rank discrepancy of the genotype simple effect between the eBLUP with homogeneous and that with heterogeneous residual error variance (REV) for the trials.

Year	% of locations				% of genotypes			
	A[1]	B	C	D	A	B	C	D
2012	33.3	33.3	58.3	41.7	15.4–15.4	15.4–30.8	15.4–30.8	16.7–50.0
2013	50.0	60.0	70.0	40.0	16.7–50.0	16.7–58.3	16.7–16.7	16.7–16.7
2014	100.0	40.0	50.0	18.2	16.7–33.3	16.7–16.7	16.7–33.3	16.7–16.7

[1] Letters are related to groups.

We also examined other statistics, *e.g.*, estimates of genotype simple effect difference, standard errors of simple effect difference estimates, degrees of freedom, as well as *t*-values in the *t*-test, between the two eBLUP versions (results not shown). There were differences in all of these statistics between the two eBLUP versions. This suggests that whether the heterogeneity of REV is accounted for by use of the eBLUP has an impact on the *t*-test about genotype simple effect in various aspects, which together resulted in the discrepancy of the number of genotype pairs with significant differences between the two eBLUP versions.

Discussion

In this work, the models with heterogeneous REV fitted the data better than their homogeneous REV versions for all of the considered trials according to both the information criterion AIC and the LRT. This further illustrates that the heterogeneity of REV across locations generally existed in multi-location trials, and

that assuming a homogeneous REV is generally not realistic and makes the procedure with consideration of heterogeneous REV a more appropriate choice. Previous work (Hu *et al.*, 2013) has showed that failing to take into account REV variations across locations by use of best linear unbiased estimation (BLUE) could result in an inflation or deflation of statistical Type I error rates for pair-wise difference test of genotype simple effects depending on specific locations. By use of the eBLUP in the present study, the ratio of the number of genotype pairs with significant differences between the two eBLUP versions was mostly not unity. The ratios smaller and larger than 1 indicate an inflation and deflation of statistical Type I error rates (Hu *et al.*, 2013), respectively, for pair-wise testing of genotype simple effects by use of the eBLUP with homogenous REV in comparison with that with heterogonous REV. The reasons for this discrepancy understandably are error variations across locations and the eBLUP with homogenous REV failing to consider this variation. Apart from this, the present study also showed that whether the heterogeneity of REV was accounted for in

Table 8. The ratio of the number of genotype pairs with significant (α = 0.05) differences based on the eBLUP with heterogeneous residual error variance (REV) compared to its homogeneous version for different locations (L1-L12) of the trials.

Location	2012				2013				2014			
	A[1]	B	C	D	A	B	C	D	A	B	C	D
L1	1.46	1.15	1.57	0.95	0.80	1.14	0.86	1.22	0.89	0.86	0.44	1.65
L2	0.94	0.70	0.50	0.96	0.82	0.71	0.76	0.73	2.22	0.83	0.57	1.00
L3	0.98	0.96	1.09	0.88	0.90	0.93	0.54	0.56	2.27	0.95	1.27	1.26
L4	0.85	1.00	2.86	1.02	1.11	0.80	1.54	1.78	0.64	0.85	2.07	0.84
L5	1.00	1.07	1.77	1.55	1.38	1.21	0.89	0.89	0.29	1.20	1.00	0.80
L6	2.12	1.96	0.94	0.15	0.67	0.56	1.65	1.24	1.46	0.71	1.47	0.92
L7	0.86	1.07	1.18	1.23	1.14	1.04	1.00	1.18	1.59	0.81	1.26	0.87
L8	1.29	1.13	1.45	1.22	1.18	1.62	1.10	1.18	1.90	1.13	1.43	0.90
L9	0.71	0.84	0.76	0.92	1.38	0.76	1.00	1.06	1.46	1.17	0.82	1.00
L10	0.71	0.89	0.92	1.00	0.92	1.12	1.55	1.46	0.61	1.13	1.09	1.20
L11	1.10	1.32	1.06	1.27					0.78			1.02
L12	1.02	1.03	1.24	1.25								

[1] Letters are related to groups.

the analysis procedures impacted the variance estimate of random effects, testing of the variance of G-L interaction effects, as well as the ranking of genotype simple effects by use of the eBLUP. In this context, it is to say that accounting for the heterogeneity of REV is more essential by use of the eBLUP than that by use of BLUE, because by the latter the heterogeneity of REV influences merely the pair-wise t-test of genotype simple effects, and by the former it influences not only the pair-wise t-test but also the ranking of genotype simple effects.

Mixed model equations developed by Henderson (1975) are a useful tool to analyze trials with heterogeneous REV (Henderson, 1975; Harville, 1976, 1977; McLean et al., 1991; Marx & Stroup, 1993). Solutions to the mixed model equations give BLUE for fixed effects and BLUP for random effects (Searle et al., 1992). Generally, when a REML-based mixed model package such as MIXED is employed, the user needs not worry about how to account for the heterogeneity of REV. This will be account for automatically on the basis of the mixed model with a heterogeneous structure for residual error effects. Besides, the mixed model framework also allows analysis procedures to be assessed using likelihood-based criteria (Wolfinger, 1993). This study used AIC and the LRT for assessing the appropriateness of the analysis procedure with different consideration about REV. This may be preferable in practice to the computer-intensive cross-validation (Piepho, 1998). Therefore, the mixed model should be routinely used for genotype evaluation in multi-location trials.

This paper exclusively focused on the ANOVA-type mixed model, which implies a simple variance-covariance for G-L interaction effects. There are other complex structures, e.g. the factor-analytic variance-covariance structure (Piepho, 1998). The complex variance-covariance, if viewed from a mixed-model perspective, implies heterogeneities of the variance-covariance for G-L interaction effects. There are studies on the impact of the heterogeneity of variance-covariance for G-L interaction effects on estimate of genotype effects (Piepho, 1994, 1998) in multi-location trials. An analysis procedure simultaneously accounting for the heterogeneity of variances of both G-L effects and residual error effects and a simulation study on the precision and efficiency of this procedure would be worthwhile. This will be the subject of further research.

Most of the studies on eBLUP are exclusively focused on the estimate of genotype simple effects. This paper examined the impact of the heterogeneity of REV not only on the ranking but also on the statistical hypothesis testing of genotype simple effects for multi-location trials. The latter is especially important for

the analysis of late-stage variety evaluation trials or some agronomy trials, where the number of varieties or treatments is fewer and hypothesis testing is more relevant. For example, in the trials for commercial release and recommendation of variety to farmers (e.g. on-farm trials) in China, the statistical hypothesis testing of genotypes is in routine use. The trials used in this study are only some examples of these scenarios. It also should be mentioned that an evaluation of genotype main effects usually is one of the important objectives in multi-location trials. The genotype main effects usually are considered fixed and they are evaluated using BLUE. For information on the impact of the heterogeneity of REV on the evaluation of genotype main effects, readers are referred to Hu et al. (2013).

Conventionally, hypothesis tests are defined for fixed parameters only. Just as BLUP is not estimate, the hypothesis test based on BLUP is not a true one as conventionally defined (Littell et al., 1996). Distribution theory associated with BLUP is not nearly as well-understood as it is with conventional estimable functions, and there are no exact methods for statistical inference on random effects (Littell et al., 1996). These notwithstanding, t-tests based on BLUP can be very useful in assessing variety simple effects at specific locations (Littell et al., 1996).

In addition to yield comparison of genotypes, there is question regarding the stability of genotypes in many multi-location trials. By assessing the genotype simple effects using eBLUP, the stability issue can be also addressed using mixed models with random effects for G-L interaction (Littell et al., 1996, 2006). For information on the impact of the heterogeneity of REV on the evaluation of genotype stability, readers are referred to Hu et al. (2014).

In China and as showed in this paper, genotypes are modeled as fixed and locations as random. In contrast, in Australia genotypes are generally modeled as random and locations as fixed (Smith et al., 2001, 2005). Which of them, especially assuming genotypes as fixed or random, being reasonable, is still a controversial topic among statisticians. Piepho (1994) showed that the predictive accuracy of eBLUP based on a two-way ANOVA model differed only slightly depending on whether genotypes, environment, or both, were regarded as random and that the most important assumption was that interactions are random. This paper mainly investigated the properties of eBLUP of interaction effects. Based on this, the conclusion about eBLUP from this work is also applicable to the case as in Australia because fixed genotypes and random locations also imply random G-L interaction and the prediction of the output of random variables is commonly done by BLUPs.

In multi-environment trials, the presence of genotype-environment (G-E) interaction is a constant concern since the performance of a variety can vary significantly when the G-E interaction effect is accentuated, and since it is difficult to evaluate the differences among the genotypes in all environments, making the selection process laborious. Thus, the G-E interaction imposes real difficulties to the breeder's work; however, it is also an excellent opportunity to explore its positive effects through specific recommendations in mega-environments (Annicchiarico & Perenzin, 1994; Annicchiarico & Piano, 2005). This paper has been restricted to the problem of obtaining good estimates of genotypes in trial environments. Clearly, the estimate or therefore the recommendation is only for environments under trial, not for 'new' environments. At times, the main interest is in estimate for new environments not under trial. For example, the farmer's interest is in an appropriate estimate of genotypes in their own fields which are not exactly the same locations as trial, there may be G-L interaction. This problem was dealt with by Annicchiarico & Perenzin (1994), Weber & Westermann (1994), Piepho *et al*. (1998) and Annicchiarico *et al*. (2005). Even in presence of G-E interaction, it is usually required to find the stable high-performing genotypes across environments. In this case, the best we can do is that variety effects can be estimated across environments by considering the main effects across environments and treating different environments as a sample from a target population of environments. Information on this issue is in the papers by Atlin *et al*. (2000b) to find. There may be scope to improve predictions by making use of a stratification of the target population of environments into ecological zones according to similarity in agroclimatic conditions and production constraints, such as in the paper by Kleinknecht *et al*. (2013). But each zone would still be represented by a random sample of locations and estimation would focus on a genotype's zone mean rather than on the location mean.

It may also be worth making a clear distinction between locations and years because G-L interactions are reproducible but genotype-year interactions are not (Annicchiarico *et al*. 2000, 2006). Predicting G-L-year means is much less meaningful than predicting G-L means across years. On repeatability of G-E interactions and genotype recommendation for the following growing season, readers can refer to relevant literature (Annicchiarico *et al*., 2000, 2006; Yan & Rajcan, 2003; Annicchiarico & Piano, 2005; Annicchiarico, 2007; Ma & Stützel, 2014).

In summary, we have found heterogeneity of REV in all of the considered rape cultivar trials. Whether the REV differences across locations were accounted for in the analysis procedure influenced the variance estimate needed for the eBLUP, testing of the variance of G-L interaction, and hence influenced the evaluation of genotype simple effects by use of the eBLUP. In application of the eBLUP for evaluation of genotype simple effects, the heterogeneity of REV can be accounted for based on the mixed model with appropriate variance-covariance structure.

References

Annicchiarico P, 2007. Wide-versus specific-adaptation strategy for lucerne breeding in northern Italy. Theor Appl Genet 114: 647-657. https://doi.org/10.1007/s00122-006-0465-1

Annicchiarico P, Perenzin M, 1994. Adaptation patterns and definition of macro-environments for selection and recommendation of common wheat genotypes in Italy. Plant Breed 113: 197-205. https://doi.org/10.1111/j.1439-0523.1994.tb00723.x

Annicchiarico P, Piano E, 2005. Use of artificial environments to reproduce and exploit genotype×location interaction for lucerne in northern Italy. Theor Appl Genet 110: 219-227. https://doi.org/10.1007/s00122-004-1811-9

Annicchiarico P, Pecetti L, Boggini G, Doust MA, 2000. Repeatability of large-scale germplasm evaluation results in durum wheat. Crop Sci 40: 1810-1814. https://doi.org/10.2135/cropsci2000.4061810x

Annicchiarico P, Bellah F, Chiari T, 2005. Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programs: a case study. Crop Sci 45: 1741-1749. https://doi.org/10.2135/cropsci2004.0524

Annicchiarico P, Bellah F, Chiari T, 2006. Repeatable genotype×location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. Eur J Agron 24: 70-81. https://doi.org/10.1016/j.eja.2005.05.003

Atlin GN, Baker RJ, McRae KB, Lu X, 2000a. Selection response in subdivided target regions. Crop Sci 40: 7-13. https://doi.org/10.2135/cropsci2000.4017

Atlin, GN, Lu X, McRae KB, 2000b. Genotype×region interaction for yield in two-row barley in Canada. Crop Sci 40: 1-6. https://doi.org/10.2135/cropsci2000.4011

Casanoves F, Macchiavelli R, Balzarini M, 2005. Error variation in multienvironment peanut trials: Within-trial spatial correlation and between-trial heterogeneity. Crop Sci 45: 1927-1933. https://doi.org/10.2135/cropsci2004.0547

Cochran WG, Cox GM, 1957. Experimental designs, 2nd ed. Wiley, NY.

Cornelius PL, Crossa J, Seyedsadr MS, 1994. Tests and estimators of multiplicative models for variety trials. Proc

1993 Kansas State Univ. Conf. on Applied Statistics in Agriculture. Manhatten, KS, USA. pp: 156-169.

Fai AHT, Cornelius PL, 1996. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. J Sta Comput Sim 54: 363-378. https://doi.org/10.1080/00949659608811740

Forkman J, Piepho HP, 2013. Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. J Agr Sci 151: 381-395. https://doi.org/10.1017/S0021859612000445

Giesbrecht FG, Burns JC, 1985. Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results. Biometrics 41: 477-486. https://doi.org/10.2307/2530872

Harville D, 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. Ann Stat 4: 384-395. https://doi.org/10.1214/aos/1176343414

Harville, D, 1977. Maximum likelihood approaches to variance component estimation and related problems. J Am Statist Assoc 72: 320-338. https://doi.org/10.1080/01621459.1977.10480998

Henderson CR, 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447. https://doi.org/10.2307/2529430

Hu XY, 2015. A comprehensive comparison between ANOVA and BLUP to valuate location-specific genotype effects for rape cultivar trials with random locations. Field Crops Res 179: 144-149. https://doi.org/10.1016/j.fcr.2015.04.023

Hu XY, Yan SW, Shen KL, 2013. Heterogeneity of error variance and its influence on genotype comparison in multi-location trials. Field Crops Res 149: 322-328. https://doi.org/10.1016/j.fcr.2013.05.011

Hu XY, Yan SW, Li SL, 2014. The influence of error variance variation on analysis of genotype stability in multi-environment trials. Field Crops Res 156: 84-90. https://doi.org/10.1016/j.fcr.2013.11.001

Kackar AN, Harville DA, 1984. Approximation for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc 79: 853-861.

Kelly AM, Smith AB, Eccleston JA, Cullis BR, 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci 47: 1063-1070. https://doi.org/10.2135/cropsci2006.08.0540

Kenward MG, Roger JH, 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53: 983-997. https://doi.org/10.2307/2533558

Kleinknecht K, Möhring J, Singh KP, Zaidi PH, Atlin GN, Piepho HP, 2013. Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Sci 53: 1384-1391. https://doi.org/10.2135/cropsci2013.02.0073

Leiser WL, Rattunde HF, Piepho HP, Weltzien E, Diallo A, Melchinger AE, Parzies HK, Haussmann BIG, 2012. Selection strategy for sorghum targeting phosphorus limited environments in West Africa: Analysis of multi-environment experiments. Crop Sci 52: 2517-2527. https://doi.org/10.2135/cropsci2012.02.0139

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, 1996. SAS system for mixed models. SAS Inst, Inc., Cary, NC, USA.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, 2006. SAS system for mixed models, 2nd ed. SAS Institute, Inc., Cary, NC, USA.

Ma D, Stützel H, 2014. Prediction of winter wheat cultivar performance in Germany: at national, regional and location scale. Eur J Agron 52: 210-217. https://doi.org/10.1016/j.eja.2013.09.005

Marx DB, Stroup WW, 1993. Analysis of spatial variability using PROC MIXED. Proc of the 1993 Kansas State Univ Conf on Appl Stat in Agr. Kansas State Univ, Manhattan, KS, USA. https://doi.org/10.4148/2475-7772.1371

McLean RA, Saunders WL, Stroup WW, 1991. A unified approach to mixed linear models. Am Statist 45: 54-64.

Mrode R, 2005. Linear models for the prediction of animal breeding values, 2nd ed. CAB Int., Oxford, UK. https://doi.org/10.1079/9780851990002.0000

Oman SD, 1991. Multiplicative effects in mixed model analysis of variance. Biometrika 78: 729-739. https://doi.org/10.1093/biomet/78.4.729

Pauler DK, 1998. The Schwarz criterion and related methods for normal linear models. Biometrika 85: 13-27. https://doi.org/10.1093/biomet/85.1.13

Piepho HP, 1994. Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects multiplicative interaction (AMMI) analysis. Theor Appl Genet 89: 647-654. https://doi.org/10.1007/BF00222462

Piepho HP, 1995. Detecting and handing heteroscedasticity in yield trial data. Commun Statist Simul 24: 243-274. https://doi.org/10.1080/03610919508813240

Piepho HP, 1998. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. Theor Appl Genet 97: 195-201. https://doi.org/10.1007/s001220050885

Piepho HP, Möhring J, 2005. Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci 45: 1151-1159. https://doi.org/10.2135/cropsci2004.0398

Piepho HP, Denis JB, van Eeuwijk FA, 1998. Predicting cultivar differences using covariates. J Agr Biol Env Stat 3: 151-162. https://doi.org/10.2307/1400648

Piepho HP, Möhring J, Melchinger AE, Büchse A, 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161: 209-228. https://doi.org/10.1007/s10681-007-9449-8

Raman A, Ladha JK, Kumar V, Sharma S, Piepho HP, 2011. Stability analysis of farmer participatory trials

for conservation agriculture using mixed models. Field Crops Res 121: 450-459. https://doi.org/10.1016/j.fcr.2011.02.001

Robinson GK, 1991. That BLUP is a good thing: The estimation of random effects: Statist Sci 6: 15-51. https://doi.org/10.1214/ss/1177011926

SAS Inst., 2011. SAS/STA software: changes and enhancement through release 9.2. SAS Inc., Cary, NC, USA.

Satterthwaite FE, 1941. Synthesis of variance. Psychometrika 6: 309-316. https://doi.org/10.1007/BF02288586

Searle SR, Casella G, McCulloch CE, 1992. Variance components. Wiley, NY. https://doi.org/10.1002/9780470316856

Shukla GK, 1972. Some statistical aspects of partitioning genotype-environmental components of variability. Heredity 29: 237-245. https://doi.org/10.1038/hdy.1972.87

Singh M, Tadesse W, Sarker A, Maalouf F, Imtiaz M, Capettini F, Nachit M, 2013. Capturing the heterogeneity of the error variances of a group of genotypes in crop cultivar trials. Crop Sci 53: 811-818. https://doi.org/10.2135/cropsci2012.11.0637

Smith AB, Cullis BR, Gilmour AR, 2001. The analysis of crop variety evaluation data in Australia. Aus N Z J Stat 43: 129-245. https://doi.org/10.1111/1467-842X.00163

Smith AB, Cullis BR, Thompson R, 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agr Sci Cambridge 143: 449-462. https://doi.org/10.1017/S0021859605005587

Steel RGD, Torrie JH, 1980. Principles and procedures of statistics. A biometrical approach, 2nd ed. McGraw Hill, NY.

Weber WE, Westermann T, 1994. Prediction of yield for specific locations in German winter-wheat trials. Plant Breed 113: 99-105. https://doi.org/10.1111/j.1439-0523.1994.tb00711.x

Windhausen VS, Wagener S, Magorokosho C, Makumbi D, Vivek B, Piepho HP, Melchinger AE, Atlin GN, 2012. Strategies to subdivide a target population of environments: Results from the CIMMYT-led maize hybrid testing programs in Africa. Crop Sci 52: 2143-2152. https://doi.org/10.2135/cropsci2012.02.0125

Wolfinger R, 1993. Covariance structure selection in general mixed models. Commun Stat-Simul C 22: 1079-1106. https://doi.org/10.1080/03610919308813143

Yan W, Rajcan I, 2003. Prediction of cultivar performance based on single- versus multiple-year tests in soybean. Crop Sci 43: 549-555. https://doi.org/10.2135/cropsci2003.0549