Published May 5, 2017

One Step at a Time: Stage-Wise Analysis of a Series of Experiments

Tigist Mideksa Damesa, Jens Möhring, Mosisa Worku, and Hans-Peter Piepho*

ABSTRACT

Multienvironment trials can be analyzed using single-stage or stage-wise analysis. Single-stage analysis is fully efficient, meaning that the estimators can be expected to be as close as possible to the corresponding true genotypic values, and so is often deemed preferable to two-stage analysis. However, two-stage analysis is often favored in practice over single-stage analysis in the case of large datasets because of the larger computational burden of the latter and because the former allows separate analyses of individual trials in the first stage, accounting for any specifics of each trial. In this study we demonstrate the similarities of results of single-stage and two-stage analysis when information on mean estimates and the associated variance–covariance matrix is forwarded from the first stage to the second stage using four examples with maize (*Zea mays* L.) trial data from Ethiopia. A new fully efficient and an approximate two-stage method with diagonal weighting matrix are used for weighting in the second stage. We extend the method to three-stage analysis for multienvironment trials when sites are stratified by agro-ecological zones and demonstrate how to obtain best linear unbiased predictions of genotype effects per zone using the information from neighboring zones. Two macros that compute weights for use in the fully efficient and diagonal weighting approaches are provided.

Core Ideas

- Single-stage and two-stage analysis of multienvironment trials yield very similar results.
- Single-stage and two-stage analysis are identical when the same set of variance values is used.
- Modeling genotypes as random helps to exploit correlations between agro-ecological zones.
- In stage-wise analysis, genotypes need to be taken as fixed through all stages except the last.
- Fully efficient two-stage analysis is similar in spirit to meta-analysis.

Published in Agron. J. 109:845–857 (2017)
doi:10.2134/agronj2016.07.0395
Received 8 July 2016
Accepted 9 Feb. 2017
Supplemental material available online

Many trials are replicated in multiple environments to broaden the inference space. For example, plant breeding and variety trials are typically performed at multiple sites and in several years (Cochran, 1937; Comstock and Moll, 1963; Gauch, 1992; Talbot, 1997; Yates and Cochran, 1938). A joint analysis of such multienvironment trials (METs) can be done in a single stage by a linear mixed model (LMM) for the plot data (Smith et al., 2001, 2005). Such an analysis is commonly considered to be fully efficient because all sources of variation can be accounted for simultaneously in a single model and because the analysis provides best linear unbiased estimates (BLUEs) of all fixed effects as well as best linear unbiased predictions (BLUPs) of all random effects under that assumed single-stage model (Searle et al., 1992). An alternative method of analysis is to proceed in two stages. In the first stage genotype means are computed per trial, and in the second stage genotype means from all trials are subjected to a joint analysis. In principle, the stage-wise approach can also be extended to more than two stages (Piepho et al., 2012a).

In both cases, individual trials are first analyzed separately, paying due attention to all specifics of a trial, including outlier detection, the particular experimental design and randomization scheme used, and selection of a preferred analysis model among contending candidate models (purely randomization based, spatial, with covariate adjustments, etc.). In two-stage analysis, only the means and some measure of precision (standard errors, variance–covariance matrix of the means, or diagonal elements of the inverse of this matrix) are saved from the first stage and carried forward to the second stage. By contrast, in a single-stage analysis, the preferred analysis models identified for each individual trial are integrated into an overall model for analysis of the MET plot data, which is fitted in a single stage. The computational burden for single-stage analysis is typically larger than for stage-wise analysis because both the size of the dataset submitted to an analysis across environments and the complexity of the model are larger in single-stage analysis. How much of an advantage the alleviated computational burden by using

T.M. Damesa, J. Möhring, H.-P. Piepho, Biostatistics Unit, Institute of Crop Science, Univ. of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany; and M. Worku, International Maize and Wheat Improvement Center (CIMMYT) Nairobi, Kenya. *Corresponding author (hans-peter.piepho@uni-hohenheim.de).

Abbreviations: AIC, Akaike information criterion; BLUP, best linear unbiased prediction; BLUE, best linear unbiased estimation; CIMMYT, International Maize and Wheat Improvement Center; CS, compound symmetry; CSH, heterogeneous compound symmetry; GBLUP, genomic best linear unbiased prediction; LMM, linear mixed model; MET, multienvironment trial; TPE, target population of environments; UN, unstructured; UNR, unstructured correlation.

stage-wise analysis affords depends on several factors, including the size of the dataset, the designs and models used for the individual trials, and the complexity of the single-stage model. Moreover, stage-wise analysis is convenient for practical analysis because it facilitates a combined analysis of different trials with different design and modeling structures and allows for heterogeneity of variance between trials (Piepho and Eckl, 2014).

Researchers wanting to analyze MET are frequently faced with the question of whether to use a single-stage or stage-wise analysis. In this paper it will be argued that, although single-stage analysis can justly be regarded as the gold standard, a stage-wise analysis, if done properly, is perfectly valid and typically very close to a single-stage analysis.

Several papers have been written comparing single-stage and two-stage analysis (Möhring and Piepho, 2009; Piepho et al., 2012a; Schulz-Streeck et al., 2013; Welham et al., 2010). This in-depth treatment is not repeated here. Instead of giving very detailed theoretical background, we briefly review the key results, facts, and arguments justifying a stage-wise analysis and discuss the important practical implications. The main purpose of this study is to illustrate stage-wise analysis with typical examples using a new weighting method. This weighting method differs from previous weighting methods in that it carries the full variance–covariance matrix from the first stage to the next stage instead of using a diagonal weighting matrix. Also, it is simpler than an alternative approach, which is based on rotation (Piepho et al., 2012a), although the results are identical. To the best of our knowledge, a weighting method using the full variance–covariance matrix from the previous stage without rotation has not been used before in the context of series of trials. We provide two macros that can be used to get weights for stage-wise analysis by the new method and by a diagonal method that was suggested previously (Smith et al., 2001). Four examples illustrate the similarity between single-stage and stage-wise analysis.

STATISTICAL METHODS FOR TRIALS AT MULTIPLE SITES IN A SINGLE YEAR

Single-Stage Analysis

The randomization-based model for analysis of the series of experiments laid out as generalized lattice designs is (Caliński et al., 2005)

$$y_{ijkm} = \phi + g_i + s_j + (gs)_{ij} + r_{jk} + b_{jkm} + e_{ijkm} \qquad [1]$$

where ϕ is a general intercept, g_i is the fixed main effect of the i-th genotype, $s_j \sim N\left(0,\sigma_s^2\right)$ is the random main effect of the j-th site, $(gs)_{ij} \sim N\left(0,\sigma_{gs}^2\right)$ is the random interaction effect of the i-th genotype and the j-th site, $r_{jk} \sim N\left(0,\sigma_{r(j)}^2\right)$ is the random effect of the k-th replicate within the j-th site, $b_{jkm} \sim N\left(0,\sigma_{b(j)}^2\right)$ is the random effect of the m-th block nested within the k-th replicate at the j-th site, and $e_{ijkm} \sim N\left(0,\sigma_{e(j)}^2\right)$ is the residual plot error associated with the observation y_{ijkm}. The variances for replicates, blocks, and error are site specific, which is usually a realistic assumption (So and Edwards, 2011) and allows a two-stage analysis to be fully equivalent to a single-stage analysis (Piepho et al., 2012a).

Fully Efficient Two-Stage Analysis

The term "fully efficient two-stage analysis" refers to a two-stage analysis that forwards the full variance–covariance matrix of adjusted means obtained in the first stage to the next stage. For analysis of individual sites (first stage), it is convenient to rewrite Model 1 as

$$y_{ijkm} = \mu_{ij} + r_{jk} + b_{jkm} + e_{ijkm} \qquad [2]$$

where $\mu_{ij} = \phi + g_i + s_j + (gs)_{ij}$ is the conditional expected value of the i-th genotype ($i = 1,...,q$) at the j-th site ($j = 1,...,p$). We here regard μ_{ij} as a fixed effect for site-wise analysis [i.e., the analysis is conditional on the site-specific effects s_j and $(gs)_{ij}$]. Collecting expected values μ_{ij} at the j-th site into a vector $\mu_j = (\mu_{1j}, \mu_{2j},..., \mu_{qj})^T$ and plot observations into the vector y_j, we have $\text{var}\left(\hat{\mu}_j\right) = \Omega_j = \left(X_j^T \Sigma_j^{-1} X_j\right)^{-1}$, where $\hat{\mu}_j$ is the generalized least squares estimator of μ_j, given by $\hat{\mu}_j = \left(X_j^T \Sigma_j^{-1} X_j\right)^{-1} X_j^T \Sigma^{-1} y_j$; X_j is a full-rank treatment design matrix for μ_j at the j-th site; and $\Sigma_j = \text{var}(y_j)$ is a nonsingular variance–covariance matrix of the plot data at the j-th site, which depends on the experimental design and the variances $\sigma_{r(j)}^2$, $\sigma_{b(j)}^2$, and $\sigma_{e(j)}^2$.

In the second stage, we can fit the model

$$\hat{\mu}_{ij} = \mu_{ij} + f_{ij} = \phi + g_i + s_j + (gs)_{ij} + f_{ij} \qquad [3]$$

where f_{ij} is the residual of the i-th genotype in the j-th site and $\text{var}(f_j) = \Omega_j$ with $f_j = (f_{1j}, f_{2j},...f_{qj})^T$. In practice, Ω_j is replaced by its residual maximum likelihood estimate from the first stage. To fit the model in the second stage, we need the variance–covariance matrix of $f = \left(f_1^T, f_2^T,..., f_p^T\right)^T$ given by

$$\text{var}(f) = \begin{pmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_p \end{pmatrix} = \bigoplus_{j=1}^{p} \Omega_j = \Omega \qquad [4]$$

Plugging in the estimate of Ω_j from the first stage, we can then estimate the fixed genotype means across environments, $\theta_i = \phi + g_i$, and the variances σ_s^2 and σ_{gs}^2 at the second stage based on Model 3, thus providing estimates of all parameters of the single-stage Model 1, if analyses of both stages are taken together. Piepho et al. (2012a) show that both analyses are fully equivalent provided the same variance component values are used for all random effects. This is also why we denote this two-stage approach as fully efficient. For theoretical details the reader is referred to that paper. The fully efficient method described here is essentially the same as that in Piepho et al. (2012a), except that we omitted the rotation; all other results in that paper apply equally, especially those on the equivalence of single-stage and stage-wise analysis as derived from the mixed model equations.

Any numerical differences between resulting estimates of genotype means θ_i result from numerical differences in the variance component estimates under single-stage and two-stage analysis. We also note that we have used a simple model for the random genotype × environment effects, but the approach also works with more complex models, such as factor–analytic variance–covariance structures for genotype × environment effects (Piepho, 1997).

For illustration we use the PROC MIXED procedure of the SAS system to perform all analyses. To fit the Model 3 in the second stage, we can use the code in Box 1 (Piepho and Eckl, 2014, Electronic Appendix).

```
ods output lsmeans = mean_twostage_stagetwo_full_1
   covparms = cp_twostage_stagetwo_full_1;
proc mixed data = mean_twostage_stageone_full_1w;
class genotype site row;
model estimate = genotype;
random int genotype/sub = site;
repeated row/sub = site type = lin(1) ldata = mean_
   twostage_stageone_full_1w;
lsmeans genotype/diffs;
parms (1)(1)(1)/hold = 3;
run;
```

Box 1: SAS code for stage-two analysis of a fully efficient two-stage analysis.

In this code, mean_twostage_stageone_full_1w specified with the data option in PROC MIXED and the LDATA option to the REPEATED statement is a dataset containing the adjusted genotype-site means $\hat{\mu}_{ij}$ and the corresponding estimate of Ω from the first stage in a suitable format as detailed in the Appendix, GENOTYPE and SITE are variables representing the genotypes and sites, ROW is a sequential number indexing genotype-site means in the dataset mean_twostage_stageone_full_1w, and ESTIMATE is the response variable carrying the adjusted genotype-site means. The REPEATED statement (as well as the RANDOM statement) specifies SITE as a subject effect with the SUBJECT option, so the blocks of Ω are processed by sites, which entails savings in memory and computing time compared with a coding not making use of the SUBJECT option. Generally, where possible, it is important that the REPEATED statement and all RANDOM statements share the same subject effect so that levels of that effect are recognized as independent subjects. The smaller the size of the subjects and the more subjects there are, the better. In this example, the shared subject effect is SITE because correlations among observed data occur only within sites. We here exploit the fact that under the assumed LMM, observations from different sites are independent. Thus, inversions of variance–covariance matrices needed during residual maximum likelihood iterations can be performed by sites, which saves computing time. In the Supplemental Material we provide a macro (***%get_one_big_omega***) that assembles the estimate of Ω in a form suitable for use with the code in Box 1 based on site-wise first-stage analyses in which estimates of Ω_j are obtained using the COV option to the LSMEANS statement for estimating genotype means at each site.

Two-Stage Analysis with Diagonal Weight Matrix

An alternative approach to the fully efficient two-stage analysis described above was proposed by Smith et al. (2001), who suggested to fit the second-stage model assuming that $\text{var}(f_{ij}) = (\omega^{ij})^{-1}$, where ω^{ij} is the i-th diagonal element of Ω_j^{-1}. The rationale for this suggestion is that the mixed model equations for Model 3 depend linearly on $\Omega^{-1} = \bigoplus_{j=1}^{p} \Omega_j^{-1}$, which can be approximated by a diagonal matrix with diagonal elements equal to ω^{ij}. The SAS code in Box 2 can be used to perform this approximate analysis at the second stage.

```
ods output lsmeans = mean_twostage_stagetwosmith_1
   covparms = cp_twostage_stagetwosmith_1;
proc mixed data = mean_twostage_stageonesmith_1w;
class genotype site;
model estimate = genotype;
random int genotype/sub = site;
lsmeans genotype/cov;
weight weight_smith;
parms (1)(1)(1)/hold = 3;
run;
```

Box 2: SAS code for second stage of an approximate two-stage analysis, using the weights proposed by Smith et al. (2001).

In this code, all variables are as defined for Box 1, and weight_smith is the variable in the dataset mean_twostage_stageonesmith_1w holding the weights. In the Supplemental Material we provide an SAS macro ***%get_Smith_weights*** that can compute these weights based on the same site-wise first-stage analyses as under the fully efficient two-stage analysis. A brief description of this macro is available in the Appendix. Using the diagonal approximation in the second stage leads to savings in computing time compared with the fully efficient two-stage analysis.

STATISTICAL METHODS FOR TRIALS AT MULTIPLE SITES AND IN MULTIPLE YEARS

Again assuming a generalized lattice design, the first-stage model for the trial in the j-th site and h-th year is given by

$$y_{ijhkm} = \mu_{ijh} + r_{jhk} + b_{jhkm} + e_{ijhkm} \quad [5]$$

where

$$\mu_{ijh} = \phi + g_i + s_j + a_h + (gs)_{ij} + (ga)_{ih} + (sa)_{jh} + (gsa)_{ijh} \quad [6]$$

in which ϕ is a general intercept, g_i is the fixed main effect of the i-th genotype, $s_j \sim N\left(0, \sigma_s^2\right)$ is the random main effect of the j-th site, $a_h \sim N\left(0, \sigma_a^2\right)$ is the random main effect of the h-th year, $(gs)_{ij} \sim N\left(0, \sigma_{gs}^2\right)$ is the random two-way interaction of the i-th genotype and the j-th site, $(ga)_{ih} \sim N\left(0, \sigma_{ga}^2\right)$ is the random two-way interaction effect of the i-th genotype and the h-th year, $(sa)_{jh} \sim N\left(0, \sigma_{sa}^2\right)$ is the random two-way interaction effect of the j-th site and the h-th year, $(gsa)_{ijh} \sim N\left(0, \sigma_{gsa}^2\right)$ is the random three-way interaction effect of the i-th genotype, the j-th site, and the h-th year, $r_{jhk} \sim N\left(0, \sigma_{r(jh)}^2\right)$ is the random effect of the k-th replicate within the j-th site and the h-th year, $b_{jhkm} \sim N\left(0, \sigma_{b(jh)}^2\right)$ is the random effect of the m-th block nested within the k-th replicate at the j-th site and the h-th year, and $e_{ijhkm} \sim N\left(0, \sigma_{e(jh)}^2\right)$ is the error associated with the observation y_{ijhkm}. As before, the variances for replicates, blocks, and error depend on the site–year combination and hence are trial specific. When the experiment is laid out in randomized complete blocks, we drop the incomplete block effect. Complete blocks are then represented by the complete replicate effect. A stage-wise analysis computes genotype means

per trial (year–site combination) in the first stage and then fits Model 6 to these means across years and sites in Stage 2.

EXTENDING THE MODEL WHEN SITES ARE STRATIFIED INTO ZONES

If sites are stratified by zone, Model 5 for the observed data remains the same; however, the conditional expected value in Eq. [6] needs modification. Specifically, each effect involving site in Model 6 needs to be replaced by two effects, the one involving zone and the other one involving site nested within zone. Thus, Eq. [6] can be extended as

$$\mu_{ij(q)h} = \phi + g_i + z_q + (zs)_{j(q)} + a_h + (gz)_{iq} + (zgs)_{ij(q)} + (ga)_{ih} + (za)_{qh} + (zsa)_{jh(q)} + (zga)_{ihq} + (gzsa)_{ij(q)h} \quad [7]$$

where all effects involving sites (s) in Model 6 have been replaced by two effects [i.e., one involving zone (z) instead of site and the other involving site nested within zone (zs)]. Moreover, g_i is the main effect of the i-th genotype, and $(gz)_{iq}$ is the interaction of the i-th genotype and the q-th zone. To borrow strength across zones when estimating mean genotype yields for a specific zone (Kleinknecht et al., 2013; Piepho and Möhring, 2005; Piepho et al., 2016), we modeled g_i and $(gz)_{iq}$ as random [e.g., assuming that both g_i and $(gz)_{iq}$ have a normal distribution with a mean of zero and a constant or heterogeneous variance]. Thus, we may obtain estimates of genotype mean in zone q

$$\mu_{iq} = \phi + g_i + z_q + (gz)_{iq} \quad [8]$$

using BLUP, which is an estimation method for random effects in LMMs that minimizes the mean squared error under the assumed model, and it entails shrinkage, meaning that the estimate of genotype effects tend to fall back toward the mean of all genotypes. Therefore, BLUPs of good performers tend to be smaller than the corresponding BLUEs, whereas BLUPs of bad performers tend to be elevated compared with the corresponding BLUEs (Robinson, 1991; Searle et al., 1992). Moreover, in the case of correlated genetic effects, BLUP allows exploiting information from correlated observations. In our case, we consider the effect h_{iq} of the i-th genotype in the q-th zone:

$$h_{iq} = g_i + (gz)_{iq} \quad [9]$$

For a given genotype i, these effects are correlated between zones q due to the genotype main effect g_i shared between different zones. Thus, when estimating the effect of the i-th genotype in the q-th zone by BLUP, we are also making use of information on the same genotype from the other zones.

We can estimate effects in Model 7 in a single stage, in two stages, or in three stages. Two-stage analysis proceeds in the same way as described previously, with Model 7 fitted in the second stage. Three-stage analysis considers effects g_i and $(gz)_{iq}$ as fixed in the second stage to compute estimates of means in Model 8 and the associated variance–covariance matrix. In the third stage, Eq. [8] is fitted to these means taking $h_{iq} = g_i + (gz)_{iq}$ as random and using the variance–covariance matrix of adjusted means from the second stage for weighting. In our example, we have two zones, so we need to consider 2 × 2 variance–covariance structures of the form

$$\text{var}\begin{pmatrix} h_{i1} \\ h_{i2} \end{pmatrix} = \Gamma = \begin{pmatrix} \sigma^2_{g1} & \sigma_{g12} \\ \sigma_{g12} & \sigma^2_{g2} \end{pmatrix} \quad [10]$$

where σ^2_{g1} is the variance of the i-th genotype in Zone 1, σ^2_{g2} is the variance of the i-th genotype in Zone 2, and σ^2_{g12} is the covariance between effects of the i-th genotype in Zones 1 and 2. Equation [10] may be denoted as an unstructured variance–covariance model. Alternatively, we may impose a specific structure. Modeling g_i and $(gz)_{iq}$ as random with the assumptions $g_i \sim N\left(0, \sigma^2_g\right)$ and $(gz)_{iq} \sim N\left(0, \sigma^2_{gz}\right)$ results in a compound symmetry (CS) variance structure. The CS model has two parameters: a constant variance and a constant covariance. The CS variance–covariance structure of Model 9 can be written as

$$\Gamma = \begin{pmatrix} \sigma^2_g + \sigma^2_{gz} & \sigma^2_g \\ \sigma^2_g & \sigma^2_g + \sigma^2_{gz} \end{pmatrix} \quad [11]$$

An extension of this model is the heterogeneous compound symmetry (CSH) model, which has a different variance parameter for each diagonal element (zone). For two zones, it has the representation

$$\Gamma = \begin{pmatrix} \sigma^2_{g1} & \sigma_{g1}\sigma_{g2}\rho \\ \sigma_{g1}\sigma_{g2}\rho & \sigma^2_{g2} \end{pmatrix} \quad [12]$$

Because there are only two zones in our example, the CSH model is just a reparameterization of the unstructured model, and it also has the same specification here as the unstructured

Table 1. Variance component estimates for single-stage analysis, fully efficient two-stage analysis, and two-stage analysis with diagonal weights (Smith et al., 2001) (Example 1).

Variance parameter	Fully efficient two-stage	Smith et al. approximation two-stage	Single-stage
σ^2_s	10.4537	10.4227	10.4543
σ^2_{gs}	0.1272	0.1279	0.1053
$\sigma^2_{r(1)}$	0.1004	0.1004	0.0882
$\sigma^2_{r(2)}$	1.4012	1.4012	1.3829
$\sigma^2_{r(3)}$	0.0000	0.0000	0.0000
$\sigma^2_{r(4)}$	0.0141	0.0141	0.0144
$\sigma^2_{b(1)}$	0.2504	0.2504	0.3312
$\sigma^2_{b(2)}$	0.4645	0.4645	0.4747
$\sigma^2_{b(3)}$	0.0000	0.0000	0.0000
$\sigma^2_{b(4)}$	0.0720	0.0720	0.0695
$\sigma^2_{e(1)}$	1.2363	1.2363	1.3467
$\sigma^2_{e(2)}$	0.2020	0.2020	0.1936
$\sigma^2_{e(3)}$	1.0549	1.0549	1.1531
$\sigma^2_{e(4)}$	0.1126	0.1126	0.1112

model parameterized in terms of variances and correlations (UNR). All of these structures (CS, CSH, unstructured [UN], and UNR) are available in SAS.

A stage-wise analysis fits Eq. [7] across years and sites in Stage 2 to compute genotype means per zone by BLUE. The BLUE of genotype means at the second stage will be used for comparison with BLUP from two-stage and three-stage analysis. In Stage 3 of a three-stage analysis, the linear predictor (Model 8) is fitted to genotype–zone means using the random-effects specification in Models 9 and 10.

EXAMPLE 1: TRIALS CONDUCTED AT MULTIPLE SITES IN A SINGLE YEAR

The Dataset

Twenty-two different genotypes of nonquality protein maize (*Zea mays* L.) were evaluated in the Ethiopian preliminary national variety trials of maize (EVCDTH12; Evaluation of International Maize and Wheat Improvement Center [CIMMYT] drought-tolerant hybrids in the 2012 main rainy season). The trials aim to identify high-yielding, adapted hybrids for low-moisture stress areas. The experiment was conducted from 1 July to 25 Dec. 2012 in the low-moisture stress area at four sites (Dhera, Melkassa, Mieso, and Ziway). The experimental designs used at all sites were α-designs with 11 incomplete blocks in each replicate. Each trial had three replicates. The plot size was 7.5 m^2 with six planted rows. These data are available as dataset Example 1 in the Supplemental Material.

Results

Single-stage and two-stage analyses were performed. Variance component estimates for single-stage and two-stage analyses agree reasonably well (Table 1). The estimated means in Table 2 (columns 1 and 3) show that the fully efficient two-stage analysis carrying the full variance–covariance matrix of adjusted means forward from Stage 1 yields identical results to single-stage analysis provided the same variance component values are used as expected from theory (Piepho et al., 2012a). When variances are estimated separately in each type of analysis, adjusted genotype means from the single-stage analysis show correlations larger than 0.99 with those of the two-stage analysis (Table 3).

EXAMPLE 2: EXTENDING EXAMPLE 1 TO ALLOW FOR TRIAL-SPECIFIC ANALYSIS MODELS FOR POST-BLOCKING AND RESIDUAL ERROR

This example is presented to illustrate the performance when different analysis models (randomization-based baseline model; spatial models; and models with post-blocking for row, column, and column nested within replicate effects) are used for individual trials, using the dataset of Example 1. First, the baseline model is extended by effects for row, column, or column nested within replicate. Taking the optimal model from these candidate models for each site, local spatial trends are modeled by one-dimensional and two-dimensional autoregressive models. For the one-dimensional case we assume that a correlation exists

Table 2. Adjusted genotype estimates when (1) the full information of estimates and their corresponding variance–covariance matrix and (2) estimates and diagonal weights (Smith et al., 2001) are carried forward from the first stage to the second stage of the analysis. Analyses (3), (4), and (5) are single-stage analyses, where (3) and (4) use the variance–covariance matrix of mean estimates from (1) and (2), respectively. Analysis (5) is single-stage analysis where the variances components are estimated directly from the plot data (Example 1).

Genotype	(1) Fully efficient two-stage	(2) Smith et al. approximation two-stage	(3) Single-stage variance estimate from (1)	(4) Single-stage variance estimate from (2)	(5) Single-stage
			t ha^{-1}		
1	5.1348	5.1126	5.1348	5.1345	5.1484
2	5.5098	5.4315	5.5098	5.5093	5.5445
3	5.1470	5.1519	5.1470	5.1466	5.1871
4	4.5933	4.5839	4.5933	4.5934	4.5893
5	4.8448	4.8117	4.8448	4.8451	4.8167
6	4.6920	4.7045	4.6920	4.6924	4.6590
7	4.6628	4.6776	4.6628	4.6634	4.6361
8	4.4048	4.3884	4.4048	4.4055	4.3580
9	5.0554	5.0268	5.0554	5.0558	5.0330
10	4.8392	4.8026	4.8392	4.8392	4.8405
11	4.5420	4.5026	4.5420	4.5420	4.5386
12	4.8897	4.9100	4.8897	4.8900	4.8705
13	4.8500	4.8594	4.8500	4.8502	4.8415
14	4.3254	4.2796	4.3254	4.3259	4.2931
15	4.4482	4.4151	4.4482	4.4480	4.4721
16	4.3427	4.3281	4.3427	4.3424	4.3713
17	4.0719	4.0855	4.0719	4.0720	4.0677
18	4.9780	4.9162	4.9780	4.9784	4.9453
19	4.7108	4.7243	4.7108	4.7108	4.7148
20	4.8076	4.8187	4.8076	4.8070	4.8608
21	4.1284	4.1354	4.1284	4.1283	4.1334
22	4.5772	4.5810	4.5772	4.5764	4.6283

Table 3. Correlation among adjusted genotype means (above the diagonal: Pearson's product-moment correlation; below the diagonal: Spearman's rank correlation) when (1) the full information of estimates and their corresponding variance–covariance matrix and (2) estimates and diagonal weights (Smith et al., 2001) are carried forward from the first stage to the second stage of the analysis. Analyses (3), (4), and (5) are single-stage analyses, where (3) and (4) use the variance–covariance matrix of mean estimates from (1) and (2), respectively. Analysis (5) is single-stage analysis where the variances components are estimated directly from the plot data (Example 1).

Analysis	(1) Fully efficient two-stage	(2) Smith et al. approximation two-stage	(3) Single-stage variance estimates from (1)	(4) Single–stage variance estimate from (2)	(5) Single-stage
(1)	1	0.9971	1.0000	1.0000	0.9967
(2)	0.9966	1	0.9971	0.9971	0.9948
(3)	1.0000	0.9966	1	1.0000	0.9967
(4)	1.0000	0.9966	1.0000	1	0.9966
(5)	0.9898	0.9955	0.9898	0.9898	1.0000

within rows, within columns, or within columns nested within replicates. The two-dimensional AR(1) × AR(1) model is fitted assuming that correlation extends across the whole field. For all autoregressive models, the autocorrelation parameter was constrained to be non-negative (Piepho et al., 2015). The best model for each individual trial was selected using the Akaike information criterion (AIC) (Table 4).

Results

The estimates of the variance components using the fully efficient and diagonal weighting are quite similar with this example as well. However, compared with Example 1, the variance component estimates using single-stage analysis were somewhat more different from the stage-wise analysis (Supplemental Table S1). The correlations of adjusted genotype means using the different approaches (Supplemental Table S2) are slightly smaller than in Example 1 but are all greater than 0.98 (Table 5), indicating close similarity of single-stage and two-stage analysis.

EXAMPLE 3: TRIALS AT MULTIPLE SITES AND IN MULTIPLE YEARS

The Dataset

During the 1997 and 1998 main cropping seasons, 20 different maize varieties of East African and CIMMYT origin were tested at nine sites. These sites represent two of the maize-producing mega-environments (zones) in Ethiopia: the low (low-mid)-altitude subhumid zone (Zone 1) and the high-altitude subhumid zone (Zone 2) (Fig. 1). Randomized complete block designs with three replicates and two-row plots were used at all sites and in both years. Each row was 5.1 m in length. The space between rows was 0.75 m, and the distance between plants was 0.3 m. The recommended management was applied in each site. These data are available as dataset Example 3 in the Supplemental Material.

Results

Results demonstrate the similarity of single-stage and two-stage analysis for the multisite and multiyear dataset using the fully efficient two-stage analysis and the approximate two-stage

Table 4. Akaike information criterion (AIC) values for the baseline and different extended models for each site (Example 2). Akaike information criterion values for the model with the best fit are given in bold.

Model†	AIC from analysis of site			
	1	2	3	4
Baseline	173.7	134.3	153.4	79.9
Baseline model plus post-blocking				
Row	**162.4**	135.9	153.4	79.9
Col	175.7	134.3	153.4	79.9
Col(rep)	173.7	136.3	153.4	**74.6**
Row + col	164.4	135.9	153.4	79.9
Row + col(rep)	163.0	137.9	153.4	76.5
Best post-blocking model with spatial add-on component‡				
AR(1) along row	162.4	**130.1**	**152.7**	74.6
AR(1) along col	162.4	134.3	153.4	74.6
AR(1) along col(rep)	163.0	136.3	153.4	76.6
AR(1) along row + nugget	164.4	131.7	153.7	76.6
AR(1) along col + nugget	164.4	136.3	155.4	76.6
AR(1) along col(rep) + nugget	164.4	138.3	155.4	78.6
AR(1) × AR(1)	162.4	130.1	154.7	74.6
AR(1) × AR(1) + nugget	164.4	131.7	155.7	78.6

† AR(1), first-order auto-regressive correlations between plots along the mentioned experimental unit; AR(1) × AR(1): two-dimensional autoregressive variance-covariance structure, so correlations extended along both rows and columns; Baseline, baseline model with all randomization based effects including independent error effects; Row, col, col(rep), row+col and row+col(rep) are models extending the baseline model by post-blocking terms for row, column, or column nested within replicate; + nugget models include an additional independent error effect.

‡ All spatial models include effects of the best post-blocking model. If post-blocking was not effective, the baseline model was used for augmentation with a spatial error component.

Table 5. Correlation among adjusted genotype means (above the diagonal: Pearson's product-moment correlation; below the diagonal: Spearman's rank correlation) when (1) the full information of estimates and their corresponding variance–covariance matrix and (2) estimates and diagonal weights (Smith et al., 2001) are carried forward from the first stage to the second stage of the analysis. Analyses (3), (4), and (5) are single-stage analyses, where (3) and (4) use the variance–covariance matrix of mean estimates from (1) and (2), respectively. Analysis (5) is single-stage analysis where the variances components are estimated directly from the plot data (Example 2).

Approach	(1) Fully efficient two-stage	(2) Smith et al. approximation two-stage	(3) Single-stage variance estimates from (1)	(4) Single-stage variance estimate from (2)	(5) Single-stage
(1)	1.0000	0.9894	1.0000	0.9996	0.9819
(2)	0.9955	1.0000	0.9894	0.9917	0.9905
(3)	1.0000	0.9955	1.0000	0.9996	0.9819
(4)	0.9977	0.9921	0.9977	1.0000	0.9862
(5)	0.9842	0.9887	0.9842	0.9864	1.0000

Table 6. Correlation among adjusted genotype means (above the diagonal: Pearson's product-moment correlation; below the diagonal: Spearman's rank correlation) when (1) the full information of estimates and their corresponding variance–covariance matrix and (2) estimates and diagonal weights (Smith et al., 2001) are carried forward from the first stage to the second stage of the analysis. Analyses (3), (4), and (5) are single-stage analyses, where (3) and (4) use the variance–covariance matrix of mean estimates from (1) and (2), respectively. Analysis (5) is single-stage analysis where the variances components are estimated directly from the plot data for the multisite and multiyear maize trial dataset (Example 3).

Approach	(1) Fully efficient two-stage	(2) Smith et al. approx. two-stage	(3) Single-stage variance estimates from (1)	(4) Single-stage variance estimates from (2)	(5) Single-stage
(1)	1	1.0000	1.0000	1.0000	0.9999
(2)	1.0000	1	1.0000	1.0000	0.9999
(3)	1.0000	1.0000	1	1.0000	0.9999
(4)	1.0000	1.0000	1.0000	1	0.9999
(5)	1.0000	1.0000	1.0000	1.0000	1

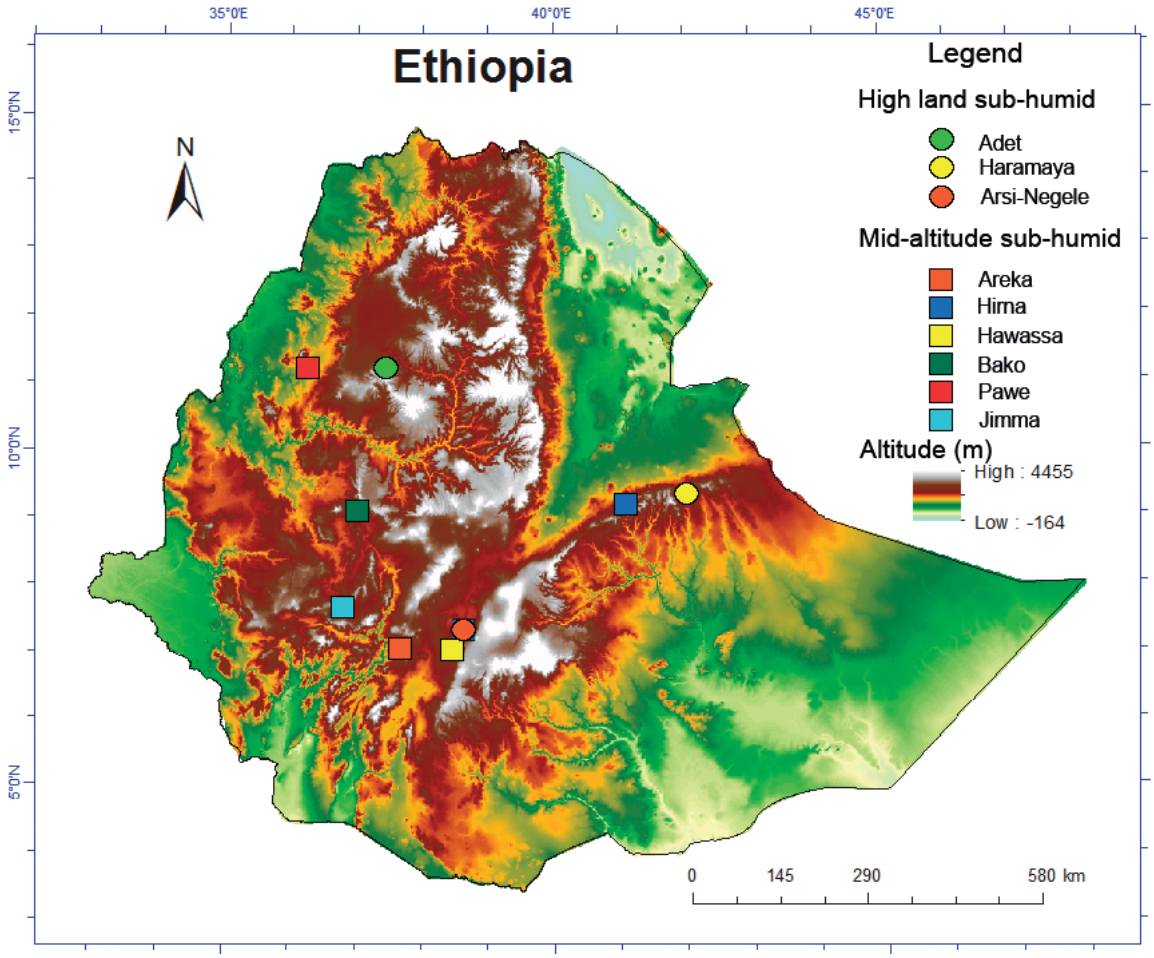

Fig. 1. Maize agro-ecologies of the low-mid–altitude subhumid and the high-altitude subhumid zones of Ethiopia.

Table 7. Akaike information criterion (AIC) and −2 residual log-likelihood (−2LL) values with different variance structures for fully efficient two-stage, three-stage, and single-stage analyses for the multisite and multiyear maize trial dataset (Example 4).

Covariance structure†	Two-stage analysis		Three-stage analysis		Single-stage analysis	
	AIC	−2LL	AIC	−2LL	AIC	−2LL
CS	984.1	972.1	73.8	69.8	2965.8	2901.8
CSH	985.4	971.4	75.1	69.1	–‡	–‡
UN	985.4	971.4	75.1	69.1	–‡	–‡
UNR	985.4	971.4	75.1	69.1	2967.2	2901.2

† CS, compound symmetry; CSH, heterogeneous compound symmetry; UN, unstructured; UNR, unstructured correlations.
‡ Did not converge from CSH and UN.

method of Smith et al. (2001). The variance parameter estimates are approximately equal for the three methods (Supplemental Table S3). Likewise, the estimated adjusted genotype means are quite similar for single-stage versus two-stage analysis (Supplemental Table S4), as also indicated by the correlations presented in Table 6.

EXAMPLE 4: EXTENDING THE MODEL FOR EXAMPLE 3 WITH SITES STRATIFIED INTO ZONES

Here we perform single-stage, two-stage, and three-stage analysis using the data of Example 3. The CS, CSH, UN, and UNR variance structures were imposed for the correlation between zones. Among the fitted variance–covariance structures, the CS model performed better (i.e., had a smaller AIC value) than the other models; therefore, we summarize the results to show the similarity of single-stage and stage-wise analysis using the CS variance structure. For CSH and UN, the single-stage analysis did not converge, so only results of CS and UNR are presented (Table 7).

Results

Because we only have two zones, the CSH, UN, and UNR models have equal variance–covariance and correlation values for the stage-wise analysis. The estimated genetic correlations between zones are large, which indicates a close relation of the two zones in terms of the adjusted genotype means (Tables 8–10).

Overall, the variance–covariance parameter estimates are very similar for the single-stage, two-stage, and three-stage analyses (Supplemental Table S5). There are also close similarities between the BLUPs of single-stage, two-stage, and three-stage analyses (Supplemental Table S6), as quantified by the Pearson correlations well above 0.96 in Table 11. The BLUEs are almost perfectly correlated. The Pearson correlations of BLUEs and BLUPs are smaller, with values between 0.96 and 0.98 (Table 11).

The differences and comparatively low correlations between BLUP and BLUE of genotype effects in Fig. 2, 3, and 4 imply that there is a considerable environmental variation and that BLUP borrows a substantial amount of information across zones. However, by contrast, BLUE cannot borrow information across zones. Note that BLUPs of genotype–zone means in three-stage analysis are compared with BLUEs of genotype-zone means computed at the second stage (i.e., a third stage is not needed with BLUE). In this example as well as in the above three examples, we found the estimated variance–covariance matrix for the random effects to be nonpositive definite in some stages of the analysis, which was due to some variance estimates being zero. Although a message to this effect is printed in the log window, this is no reason for concern; zero-variance component estimates are not uncommon. Effects with zero variance are effectively removed from the model, and the resulting analysis is fine.

DISCUSSION

Piepho et al. (2012a) have shown that two-stage and single-stage analyses yield fully equivalent results provided that (i) the same values are used for all relevant variance parameters and the full information on all effect estimates and their associated estimated variances and covariances are carried forward from

Table 8. Values of genotypic variance (on the diagonal), correlation (above the diagonal), and covariance (below the diagonal) for the fully efficient three-stage analysis, for compound symmetry (CS), heterogeneous compound symmetry (CSH), unstructured (UN), and unstructured correlation (UNR) variance structure for the multisite and multiyear maize trial dataset (Example 4).

	Covariance structure							
	CS		CSH		UN		UNR	
Zone	1	2	1	2	1	2	1	2
1	0.3325	0.7994	0.2615	0.8185	0.2615	0.8185	0.2615	0.8185
2	0.2658	0.3325	0.2577	0.3791	0.2577	0.3791	0.2577	0.3791

Table 9. Values of genotypic variance (on the diagonal), correlation (above the diagonal) and covariance (below the diagonal) for the fully efficient two-stage analysis, with compound symmetry (CS), heterogeneous compound symmetry (CSH), unstructured (UN), and unstructured correlations (UNR) variance structure for the multisite and multiyear maize trial dataset (Example 4).

	Covariance structure							
	CS		CSH		UN		UNR	
Zone	1	2	1	2	1	2	1	2
1	0.3318	0.7999	0.2611	0.8194	0.2611	0.8194	0.2611	0.8194
2	0.2654	0.3318	0.2575	0.3782	0.2575	0.3782	0.2575	0.3782

Table 10. Values of genotypic variance (on the diagonal), correlation (above the diagonal), and covariance (below the diagonal) for the single-stage analysis, with compound symmetry (CS) and unstructured correlations (UNR) variance structures for the multisite and multiyear maize trial dataset (Example 4).

	Covariance structure			
	CS		UNR	
Zone	1	2	1	2
1	0.3332	0.8008	0.2625	0.8197
2	0.2668	0.3332	0.2585	0.3789

the first to the second stage, (ii) the same model assumptions are used for all effects, and (iii) all effects for which estimates are carried forward are formally regarded as fixed in the first stage. These results naturally carry over to more than two stages, the requirement being that all effects for which estimates are carried forward in any stage are formally modeled as fixed up to that stage. This was illustrated in the present paper using MET data for maize in Ethiopia. Thus, providing the full equivalence of models, any discrepancies in genotype mean or effect estimates arise from differences in the variance–covariance parameter estimates. A further cause of differences between the analyses arises when the variance–covariance matrix of estimated effects from the first stage is approximated by a diagonal matrix (Smith et al., 2001) rather than carried forward in full, as was also illustrated in this paper.

In our study, the numerical differences are very small regarding the resulting genotype mean estimates, and this has also been found in other work by our group (Piepho and Eckl, 2014; Piepho and Möhring, 2005; Piepho et al., 2012a; Schulz-Streeck et al., 2013). Therefore, we believe that for the types of data we typically see, a stage-wise analysis is perfectly valid and acceptable for most practical purposes. The main advantage of stage-wise analysis is that analysis of individual trials with different designs can be done for all trials at the same time with their corresponding appropriate models, and the adjusted means and the associated variance–covariance matrix of adjusted means can be stored for later processing in a two-stage analysis. We found that even with our relatively simple examples there were convergence problems in single-stage analysis, particularly when different variance–covariance structures were imposed (Table 5). In addition, the time taken by stage-wise analysis was smaller than the time taken by single-stage analysis. For instance, with the CS variance

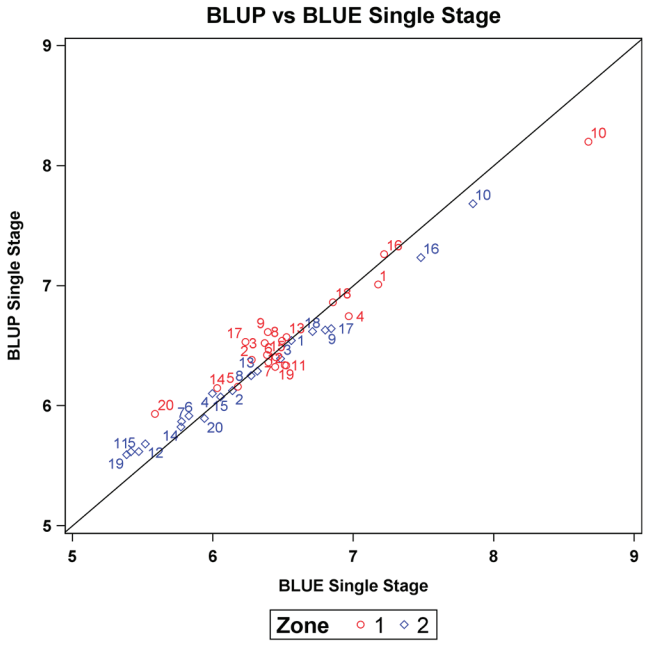

Fig. 2. Best linear unbiased prediction (BLUP) based on single-stage analysis (BLUP Single Stage) versus best linear unbiased estimation (BLUE) based on single-stage analysis (BLUE Single Stage) of genotype effects per each zone for the multisite and multiyear maize trial dataset (Example 4). Plotted number labels represent BLUP and BLUE of genotypes. The black diagonal line indicates no shrinkage.

structure, single-stage analysis for Example 4 took 33 h, two-stage analysis took approximately 2 min, and three-stage analysis took less than 1 min on a standard desktop computer (Windows 7, 64-bit operating system, 4GB RAM).

In general, two-stage analysis can be used provided the individual trials allow a separate analysis, as will be the case when designs with proper randomization and (partial) replication are used. Therefore, whenever single-stage analysis is inconvenient or computationally too demanding, stage-wise analysis can be recommended. The fully efficient weighting method is preferred because it carries all information forward to the next stage, but it is computationally more demanding than using diagonal weights. When computational resources are limited, diagonal weights can be used, and in our experience the loss of information compared with a fully efficient analysis is usually negligible (Möhring and Piepho, 2009).

Table 11. Correlation among adjusted genotype means using best linear unbiased prediction (BLUP) and best linear unbiased estimation (BLUE) (above the diagonal: Pearson's product-moment correlation; below the diagonal: Spearman's rank correlation). BLUEs are computed using (1) single-stage analysis (BLUE_1), (2) fully efficient two-stage analysis (BLUE_FE2), and (3) diagonal weights two-stage analysis (BLUE_Smith2), whereas BLUPs are computed based on (4) single-stage analysis (BLUP_1), (5) fully efficient two-stage analysis (BLUP_FE2), (6) fully efficient three-stage analysis (BLUP_FE3), (7) diagonal weights two-stage analysis (BLUP_Smith2), and diagonal weights three-stage analysis (BLUP_Smith3). Results are for the zoned multisite and multiyear maize trial dataset (Example 4).

Analysis	(1) BLUE_1	(2) BLUE_FE2	(3) BLUE_Smith2	(4) BLUP_1	(5) BLUP_FE2	(6) BLUP_FE3	(7) BLUP_Smith2	(8) BLUP_Smith3
(1)	1	0.9999	0.9999	0.9821	0.9821	0.9819	0.9826	0.9678
(2)	0.9981	1	1.0000	0.9816	0.9817	0.9815	0.9822	0.9672
(3)	0.9979	0.9998	1	0.9811	0.9812	0.9810	0.9818	0.9667
(4)	0.9503	0.9462	0.9452	1	0.9999	0.9999	0.9999	0.9971
(5)	0.9448	0.9407	0.9398	0.9994	1	1.0000	1.0000	0.9972
(6)	0.9448	0.9407	0.9398	0.9994	1.0000	1	0.9889	0.9972
(7)	0.9477	0.9435	0.9428	0.9996	0.9994	0.9994	1	0.9972
(8)	0.9180	0.9133	0.9124	0.9901	0.9916	0.9916	0.9914	1

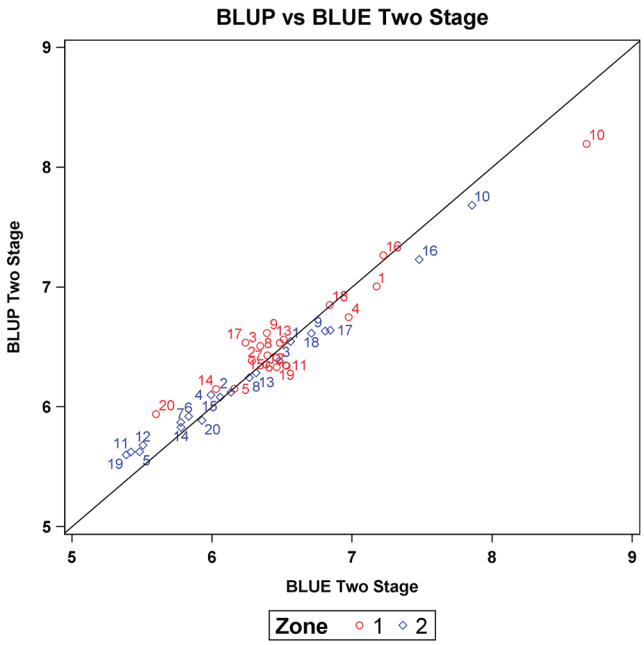

Fig. 3. Best linear unbiased prediction (BLUP) based on fully efficient two-stage analysis (BLUP Two Stage) versus best linear unbiased estimation (BLUE) based on fully efficient two-stage analysis (BLUE Two Stage) of genotype effects per each zone for the multisite and multiyear maize trial dataset (Example 4). Plotted number labels represent BLUP and BLUE of genotypes. The black diagonal line indicates no shrinkage.

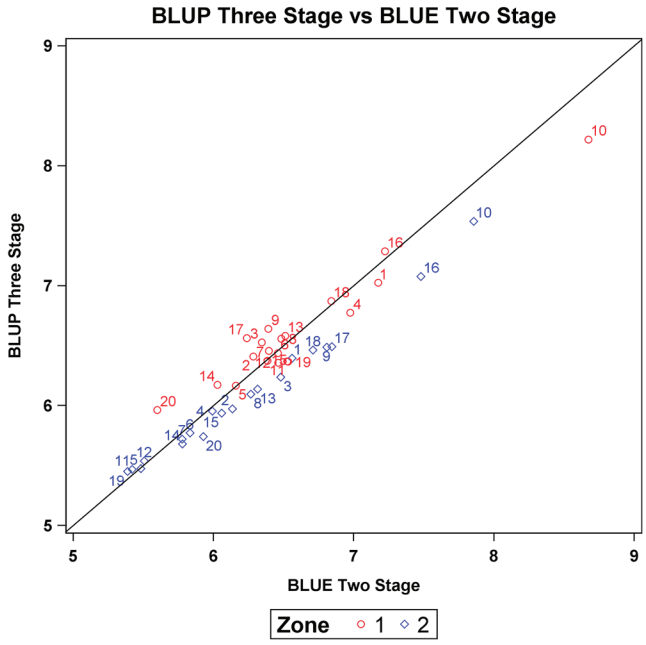

Fig. 4. Best linear unbiased prediction (BLUP) based on fully efficient three-stage analysis (BLUP Three Stage) versus best linear unbiased estimation (BLUE) based on fully efficient two-stage analysis (BLUE Two Stage) of genotype effects per each zone for the multisite and multiyear maize trial dataset (Example 4). Plotted number labels represent BLUP and BLUE of genotypes. The black diagonal line indicates no shrinkage.

A key question with the models we consider here is whether the genotype factor is fixed or random. This decision depends on the objectives of the experiment. For example, if the objective is selection of the best genotypes from a population of genotypes under study and it is reasonable to assume that genotype effects at least approximately follow a normal distribution, then genotype effects can be considered as random, and BLUP will be the best method of estimation to obtain ranks of the genotypes that are close to the true rankings of the genotype effects (Searle et al., 1992). On the other hand, if the objective of the analysis is to obtain significance tests for the difference between pairs of genotypes, then BLUE is an appropriate method (Smith et al., 2005). In variety testing in Ethiopia, it is customary to take genotypes as fixed and to compute adjusted means across environments. However, we have given an example where genotypes were taken as random to exploit correlations between zones for making zone-specific predictions. Such predictions (BLUPs) have been shown to be more accurate than zone-specific mean estimates (BLUEs) assuming fixed effects, which cannot borrow strength across zones (Kleinknecht et al., 2013). The approach does require a sufficient number of genotypes to estimate genotypic variances and covariances, and that the distribution of effects can reasonably be assumed to be approximately normal. Genotypes are also modeled as random in genomic prediction to permit estimation of effects of markers that may be much larger in number than the genotypes tested (Meuwissen et al., 2001). The assumption of genotypes as either fixed or random can be considered as an intrinsic part of the single-stage model for plot data. A salient feature of the stage-wise approach advocated here, however, is that, regardless of the status of genotypes as either fixed or random in the single-stage model, genotypes need to be formally taken as fixed through all stages of the analysis except the last, where genotypes are fixed or random depending on the status of genotypes in the single-stage model. Piepho et al. (2012a) show that this approach of stage-wise analysis leads to valid results that are identical to those of single-stage analysis when the same variance parameter values are used in single-stage and stage-wise analysis.

We frequently find in publications that a stage-wise analysis is conducted in which BLUPs of genotype means or effects are used in the first stage. This practice is problematic and should be discouraged. For example, when BLUP is also used in the second stage, this entails a double-shrinkage of effects, with one occurring in the first stage and the other occurring in the second stage (Smith et al., 2001). To correct for this problem, BLUPs obtained in the first stage would need to be unshrunken, and it is not clear how. Also, the resulting analysis is not equivalent to single-stage analysis when the same variance values are used in both. For these reasons, we recommend not using BLUP in the first stage of two-stage analysis.

Our view on the fixed versus random issue presented here is restricted to the modeling of genotypic effects. We admit that the view is somewhat pragmatic. In particular, we do not think the random assumption requires that the tested genotypes have been randomly sampled from a larger population. In support of our view, we would like to cite from the (decidedly non-Bayesian) textbook of Lee et al. (2006): "... even if the true model is the fixed-effects model (i.e., if there is no random sampling involved), the use of random-effect estimation has been advocated as shrinkage estimation. [...] Only when the number of random effects is small, for example three or four, will there be little gain from using the random-effect model (James and Stein, 1960)."

There are close ties between the analysis of series of trials and (network) meta-analysis (Madden et al., 2016; Piepho

et al., 2012c; Vargas et al., 2013). The two-stage approach corresponds to what is standard practice in meta-analysis of clinical trials (Whitehead, 2002). The result of clinical trials is usually stored in the form of effect size estimates and associated standard errors. This information may be summarized across trials using a mixed model with random effects for heterogeneity (i.e., treatment × trial interaction, using the standard errors of effect estimates to compute suitable weights for the combination of effect estimates). The resulting mean treatment effect estimates are either equivalent or very similar to estimates obtained by a single-stage analysis of individual patient data (Piepho et al., 2012c).

Instead of analyzing treatment differences, as is common practice in meta-analysis, one may proceed as in two-stage analysis of MET data and summarize treatment means by a suitable model in the second stage. This analysis, which is particularly helpful in meta-analyses comprising more than two treatments and trials with different treatment designs, is fully equivalent to analysis based on treatment differences if the site (study) main effect is taken as fixed rather than random so that all information on treatment comparison comes from comparisons within sites (studies) only (i.e., no inter-site [study] information is recovered) (Piepho et al., 2012c). This equivalence reinforces our assertion that a two-stage analysis, if done properly, is appropriate, with little difference from the corresponding single-stage analysis.

A key question in the analysis of series of trials is whether genotype effect estimates are to be obtained for the mean of a target population of environments (TPE) or for the individual environments where the trials were conducted. We would argue that in practical breeding programs the performance of genotypes in a specific test environment is rarely of interest. This is because varieties are needed that perform well on average across all environments in a given TPE. If this is what is required, then it is useful to assess the performance of contending genotypes in a random sample of environments from the TPE (Comstock and Moll, 1963; Yates and Cochran, 1938). The main error term for inferences about the means in the TPE is the genotype × environment interaction variance (Talbot, 1997), meaning that the difference between different approximations for the variance–covariance matrices of adjusted genotype means in Stage 2 is typically small (Möhring and Piepho, 2009). By contrast, when estimates for individual environments are of interest, which may be the case in research projects exploring the pattern and causes of genotype × environment interaction (corresponding to what is known as heterogeneity in meta-analysis), the only error term is that pertaining to the variance–covariance matrix of adjusted means. The genotype × environment interaction in this case is an effect to be predicted, not an error term. As a result, the difference between single-stage and two-stage results may be somewhat more relevant, and the advantage in efficiency of a single-stage analysis may be somewhat more pronounced (Welham et al., 2010).

In most applications, however, it is not an individual site and year that is of interest, but either a new year and a new site at which no trial has been conducted, such as a specific farmer's field, or a larger TPE to which a new variety is to be released. If predictions for an individual farmer's field are to be made, one may be tempted to use predictions of the closest trial site. Valid standard errors for these predictions cannot be obtained, however, because the interaction pattern between target site (farmer's field) and the nearest trial site and the corresponding interactions with years are unknown. If predictions are required for a whole TPE, however, valid inferences can be obtained, provided a random sample of sites and years from that TPE is available. From a breeder's perspective, prediction of the expected performance in a given TPE may be the most useful approach to analysis of MET because this helps to identify genotypes performing well on average (in the long run) in the TPE. By contrast, accurate predictions for an individual trial site and year are not usually of intrinsic interest because the trial environment does not usually represent conditions identical to any other environment in the TPE (Piepho et al., 2012a).

What is often more informative than predictions for individual environments is to subdivide a TPE into several agro-ecological zones, each represented by several environments, and then obtain predictions per zone. Modeling genotype × zone effects as random allows borrowing strength across zones (Atlin et al., 2000; Kleinknecht et al., 2013; Piepho et al., 2016). Realistic inferences are obtained at the zone level because several sites are used to assess the between-site sampling variation within zones. It needs to be borne in mind, however, that predictions are for zone means and not for individual sites within a zone. The random genotype × site interaction acts as the main error term for these zone-wise predictions, and as a result these predictions have a broader inference space than predictions for individual sites (McLean et al., 1991). Note that in this study the sites are assumed to be random samples. Therefore, all effects nested within sites (i.e., replicates and blocks, regardless of whether they are complete or incomplete) are considered as random (Piepho et al., 2012a).

The high correlation of random effects of the genotypes between the zones found in Example 4 indicates that, based on this study, the two zones are not very different agro-ecologically in terms of genotype means (Kleinknecht et al., 2013; Piepho and Möhring, 2005). At first sight, this result seems to contradict the Ethiopian maize breeder's perception, which is based on adjusted genotype means per zone rather than on estimates of the genotypic correlation between zones. It is a common finding in such studies, however, that the phenotypic correlations between zones are smaller than the corresponding genotypic correlations. Even if the current agro-ecology subdivision helps breeders in developing agro-ecologically adapted varieties, there exists variability within zones, which causes difficulties in selecting stable varieties. A further detailed examination of the agro-ecology within zones has been suggested by breeders (Worku et al., 2012). For a better delineation of zones, important agro-ecological factors should be taken into account (Atlin et al., 2000; Gauch, 1992).

Important examples where appropriate weighing in the second stage can be crucial are in genomic prediction (Meuwissen et al., 2001) and genome-wide association mapping (George and Cavanagh, 2015). For example, in genomic prediction it is common practice to compute genotype means across environments in one or several stages and then to submit these means to some standard routine for regularized regression on the markers, such as generic BLUP (GBLUP). The residual variance of such analyses comprises both true errors associated

with the genotype mean estimates and residual genotypic effects. The residual variance component may take on extreme values (e.g., it may move to a very small value in iterations). Such numerical problems may be tackled by explicitly modeling the unexplained polygenic effect and the error-of-a-mean effect by separate variance components, fixing the residual variance at the variance of a mean from the first stage of the analysis (Piepho et al., 2012b).

Our LMM approach for MET analysis can be readily extended to generalized linear mixed models (Stroup, 2015). In a generalized linear mixed model framework with other distributions and link functions, one can still obtain adjusted genotype means on the link scale, along with the variance–covariance matrix, in the first-stage analysis. With these results from the first stage, one can proceed exactly as described for LMM. Therefore, the only difference lies in the analysis of the first stage, where a different link and distribution function are used, and for this purpose PROC GLIMMIX is used instead of PROC MIXED (Madden et al., 2016).

ACKNOWLEDGMENTS

We would like to thank the staff at research stations where the field experiments were conducted and particularly the Low-Moisture Mega-Environment Maize Producing Coordinating unit (at Melkassa center) of the Ethiopian Institute of Agricultural Research (EIAR), for providing the data. We are also grateful to Dr. Gezahegn Bogale, staff member of EIAR at Melkassa, for making available the data and clarifying some of the questions related to the data. We greatly appreciate the help of Mengistie Kindu for producing the map of the study sites.

APPENDIX

We here briefly describe the two macros ***%get_one_big_omega*** and ***%get_Smith_weights***, which are available in the Supplementary Material as get_one_big_omega.sas and get_Smith_weights.sas, respectively, at the journal's website along with the full SAS code for performing all analyses reported in this paper for the Ethiopian maize datasets.

The macro %get_one_big_omega

This macro processes a dataset containing the variance–covariance matrices of adjusted genotype means from several trials and generates an SAS dataset containing the block-diagonal variance–covariance matrix Ω in a form ready for use with the LDATA = option in a REPEATED statement specifying a LIN(1) variance–covariance structure using the TYPE = option (see sample code in Box 1). The input dataset for this macro must be ordered by trials (sites, site-year, or zone–site-year combinations) and in a format as is generated when outputting adjusted means and associated variance–covariance matrices, computed with the MIXED procedure using the COV option on the LSMEANS statement and variables to identify trials (e.g., relevant combination of "site," "year," and "zone") as by-processing variables in a BY statement, via the output delivery system ODS. The macro generates a SAS dataset containing the following variables: PARM, a serial number for variance components of a specified linear variance–covariance structure (here PARM = 1 for all rows in the dataset); ROW, a sequential number for rows in the dataset; and COL1-COL*n*, where *n* is the number of genotype-trial means in the dataset. These latter variables carry the block-diagonal variance–covariance matrix Ω.

The macro %get_Smith_weights

This macro processes the same kind of input dataset as the ***%get_one_big_omega*** macro. It uses a call of the MIXED procedure to compute the inverses Ω_j^{-1}, from which the weights are then extracted and added as a column with variable name weight_smith in the output dataset. This weight variable can then be used in a two-stage analysis (see sample code in Box 2).

SUPPLEMENTAL INFORMATION

Additional supporting information can be found in the online version of this article:

get-one-big-omega.sas contains the SAS macro ***%get_one_big_omega***.

get-Smith-weights.sas contains the SAS macro ***%get_Smith_weights.***

Examples-get-one-big-omega.sas contains all the SAS codes used for performing fully efficient two-stage and three-stage analysis for the four examples considered in this study.

Examples-get-Smith-weights.sas contains all the SAS codes used for performing two-stage and three-stage analysis with diagonal weight matrix for the four examples considered in this study.

Supplemental Tables: These include Table S1, Table S2, Table S3, Table S4, Table S5, and Table S6, which are cited in the paper; they are tables of genotype means and their variance–covariance matrix and serve to show the similarity of the results for single-stage and stage-wise analysis (fully efficient and diagonal weights of Smith et al. [2001]) for Examples 2, 3, and 4.

REFERENCES

Atlin, G.N., R.J. Baker, K.B. McRae, and X. Lu. 2000. Selection response in subdivided target regions. Crop Sci. 40:7–13. doi:10.2135/cropsci2000.4017

Caliński, T., S. Czajka, Z. Kaczmarek, P. Krajewski, and W. Pilarczyk. 2005. Analyzing multi-environment variety trials using randomization-derived mixed models. Biometrics 61:448–455. doi:10.1111/j.1541-0420.2005.00334.x

Cochran, W.G. 1937. Problems arising in the analysis of a series of similar experiments. J.R. Stat. Soc. 4:102–118.

Comstock, R.E., and R.H. Moll. 1963. Genotype-environment interaction. In: W.D. Hanson and H.F. Robinson, editors, Statistical genetics and plant breeding. NAD-NRC Publ. No. 912. National Academy of Sciences-National Research Council, Washington, DC. p. 164–198.

Gauch, H.G. 1992. Statistical analysis of regional yield trials. Elsevier, Amsterdam.

George, A.W., and C. Cavanagh. 2015. Genome-wide association mapping in plants. Theor. Appl. Genet. 128:1163–1174. doi:10.1007/s00122-015-2497-x

James, W., and C. Stein. 1960. Estimation with quadratic loss. In: Proc. Fourth Berkeley Symp. Math. Statist. Probability. Univ. California Press, Berkeley. p. 361–380.

Kleinknecht, K., J. Möhring, K.P. Singh, P.H. Zaidi, G.N. Atlin, and H.P. Piepho. 2013. Comparison of the performance of BLUE and BLUP for zoned Indian maize data. Crop Sci. 53:1384–1391. doi:10.2135/cropsci2013.02.0073

Lee, Y., J.A. Nelder, and Y. Pawitan. 2006. Generalized linear models with random effects: Unified analysis via H-likelihood. Chapman & Hall/CRC, Boca Raton, FL. doi:10.1201/9781420011340

Madden, L.V., H.P. Piepho, and P.A. Paul. 2016. Statistical models and methods for network meta-analysis. Phytopathology 106:792–806. doi:10.1094/PHYTO-12-15-0342-RVW

McLean, R.A., W.L. Sanders, and W.W. Stroup. 1991. A unified approach to mixed linear models. Am. Stat. 45:54–64.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Möhring, J., and H.P. Piepho. 2009. Comparison of weighting in two-stage analyses of plant breeding trials. Crop Sci. 49:1977–1988. doi:10.2135/cropsci2009.02.0083

Piepho, H.P. 1997. Analyzing genotype-environment data by mixed models with multiplicative effects. Biometrics 53:761–766. doi:10.2307/2533976

Piepho, H.P., and T. Eckl. 2014. Analysis of series of variety trials with perennial grasses. Grass Forage Sci. 69:431–440. doi:10.1111/gfs.12054

Piepho, H.P., and J. Möhring. 2005. Best linear unbiased prediction for subdivided target regions. Crop Sci. 45:1151–1159. doi:10.2135/cropsci2004.0398

Piepho, H.P., J. Möhring, T. Schulz-Streeck, and J.O. Ogutu. 2012a. A stage-wise approach for analysis of multi-environment trials. Biometrics 54:844–860. doi:10.1002/bimj.201100219

Piepho, H.P., J. Möhring, M. Pflugfelder, W. Hermann, and E.R. Williams. 2015. Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. Comm. Biomet. Crop Sci. 10:3–16.

Piepho, H.P., M.F. Nazir, M. Qamar, A. Rattu, Riaz-ud-Din, M. Hussain, G. Ahmad, Fazal-e-Subhan, J. Ahmed, Abdullah, K.B. Laghari, I.A. Vistro, M.S. Kakar, M.A. Sial, and M. Imtiaz. 2016. Stability analysis for a country-wide series of wheat trials in Pakistan. Crop Sci. 56:2465–2475.

Piepho, H.P., J.O. Ogutu, T. Schulz-Streeck, B. Estaghvirou, A. Gordillo, and F. Technow. 2012b. Efficient computation of ridge-regression BLUP in genomic selection in plant breeding. Crop Sci. 52:1093–1104. doi:10.2135/cropsci2011.11.0592

Piepho, H.P., E.R. Williams, and L.V. Madden. 2012c. The use of two-way mixed models in multi-treatment meta-analysis. Biometrics 68:1269–1277. doi:10.1111/j.1541-0420.2012.01786.x

Robinson, G.K. 1991. That BLUP is a good thing: The estimation of random effects. Stat. Sci. 6:15–32. doi:10.1214/ss/1177011926

Schulz-Streeck, T., J.O. Ogutu, and H.P. Piepho. 2013. Comparisons of single-stage and two-stage approaches to genomic selection using ridge regression BLUP and component wise boosting. Theor. Appl. Genet. 126:69–82. doi:10.1007/s00122-012-1960-1

Searle, S.R., G. Casella, and C.E. McCulloch. 1992. Variance components. John Wiley & Sons, New York. doi:10.1002/9780470316856

Smith, A.B., B.R. Cullis, and A. Gilmour. 2001. The analysis of crop variety evaluation data in Australia. Aust. N. Z. J. Stat. 43:129–145. doi:10.1111/1467-842X.00163

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143:449–462. doi:10.1017/S0021859605005587

So, Y.S., and J. Edwards. 2011. Predictive ability assessment of linear mixed models in multienvironment trials in corn. Crop Sci. 51:542–552. doi:10.2135/cropsci2010.06.0338

Stroup, W.W. 2015. Rethinking the analysis of non-normal data in plant and soil science. Agron. J. 107:811–827. doi:10.2134/agronj2013.0342

Talbot, M. 1997. Resource allocation for selection systems. In: R.A. Kempton and P.N. Fox, editors, Statistical methods for plant variety evaluation. Chapman and Hall, London, UK. p. 162–174. doi:10.1007/978-94-009-1503-9_10

Vargas, M., E. Combs, G. Alvarado, G. Atlin, K. Mathews, and J. Crossa. 2013. META: A suite of SAS programs to analyze multi-environment breeding trials. Agron. J. 105:11–19. doi:10.2134/agronj2012.0016

Welham, S., B.J. Gogel, A.B. Smith, R. Thompson, and B.R. Cullis. 2010. A comparison of analysis methods for late-stage evaluation trials. Aust. N. Z. J. Stat. 52:125–149. doi:10.1111/j.1467-842X.2010.00570.x

Whitehead, A. 2002. Meta-analysis of controlled clinical trials. John Wiley & Sons, New York. doi:10.1002/0470854200

Worku, M., L. Wolde, B. Tadesse, D. Girma, A. Girum, W. Abera, et al. 2012. Status and future direction of maize research and production in Ethiopia. In: M. Worku, S. Twumasi-Afriyie, L. Wolde, B. Tadesse, G. Demisie, G. Bogale, et al., editors, Meeting the challenges of global climate change and food security through innovative maize research. Proceedings of the 3rd national maize workshop of Ethiopia, Addis Ababa, Ethiopia. 17–20 Apr. 2011. The International Maize and Wheat Improvement Center, Houston, TX. p. 17–23.

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci. 28:556–580. doi:10.1017/S0021859600050978