METHODS PAPER

On two-stage analysis of multi-environment trials

Arūnas Verbyla

Abstract Two-stage analysis methods are often used in multi-environment trials (MET) for plant variety selection, when a single-stage approach is not feasible or too time consuming. In any two-stage analysis, the estimated effects taken to stage two must be unbiased for the effects of interest, and this means using best linear unbiased estimates based on a model with fixed genetic effects. The error (or weights) associated with the estimates must also be taken to stage two. These weights are functions of unknown variance parameters that need to be estimated at stage one. These parameters may be better estimated if genetic effects are taken as random, but resulting predicted genetic effects are biased. The bias can be removed by so-called de-regression in animal sciences. The proper weights involve a block diagonal matrix with blocks corresponding to environments, whereas diagonal weights were originally proposed in animal sciences. Two MET experiments, one fully replicated and one with partial replication of varieties, were used to compare one-stage and two-stage approaches. The results were similar, but using a full weight matrix for two-stage methods was superior to using diagonal weights. A small simulation study for trials with partial replication showed that fitting random genetic effects, de-regressing, and using a full weight matrix, was very similar to a one-stage analysis, and was superior to starting with fixed genetic effects at stage one. The use of diagonal weights was found to be very poor.

Introduction

The advent of the digital age has allowed the development of genetic analyses for complex data. Despite advances in hardware, scalable software remains an issue. This is particularly true of mixed model software, which is largely built on methods from twenty-five or more years ago. The underlying algorithms that are used restrict the speed of analysis, and the size of the problem can prevent a full single stage analysis.

Crop improvement programs usually involve trials being run in multiple environments (METs) because genotype or variety by environment (GE or VE) interaction is likely to be substantial, and recommendations on the use of specific varieties may therefore need to be environment specific. In analyzing such sets of trials, it is preferable to use single stage methods, usually based on mixed models; see

A. Verbyla
QAAFI, University of Queensland, St. Lucia, QLD 4072, Australia

A. Verbyla (✉)
AV Data Analytics, Echunga, SA 5153, Australia
e-mail: a.verbyla@uq.edu.au; ari.verbyla@gmail.com

the first paragraph of the introduction in Piepho et al. (2012). Single stage methods have been proposed by many authors and examined for optimality, for example see Smith et al. (2001b) and Welham et al. (2010). These methods typically rely on factor analytic models and can be computationally expensive. Two-stage methods however, in which each trial is analyzed separately, and the results combined in some manner (Smith et al. 2001a; Piepho et al. 2012; Gogel et al. 2018) are simpler to use and faster. However, there is a loss of information and hence efficiency, provided the single stage approach is based on the "correct" model. The issue of two-stage analysis with weighting has been discussed by a number of authors (Mohring and Piepho 2009; Damesa et al. 2017, 2019; Buntaran et al. 2019; Endelman 2023).

Both Piepho et al. (2012) and Gogel et al. (2018) show that a two-stage approach that uses fixed genetic effects at the first stage of analysis, is exactly equivalent to the single stage approach if variance parameters are known, and a known full weight matrix is carried forward to the second stage. Variance parameters are not known, and for the case of fixed genetic effects at the first stage of analysis, Gogel et al. (2018) quantify the information loss in using a two-stage approach through dissection of the residual likelihood used for estimating the unknown variance parameters.

Genetic effects are usually assumed to be random. In animal breeding, this is necessary because replication is not possible (except for clones). For genomic prediction in animal breeding, Garrick et al. (2009) develop so-called de-regression. Random genetic effects are predicted or estimated at the first stage, and then transformed or de-regressed for use in genomic prediction. De-regression is in fact returning the best linear unbiased predictions (BLUPs) to best linear unbiased estimates (BLUEs), as is shown in this paper. Garrick et al. (2009) consider individual BLUPs, rather than the full vector, and a diagonal weight matrix is used at the second stage of analysis for genomic prediction. This is a special case of the more general mixed effects two stage analysis presented in this paper. This approach is called random plus de-regression in this paper.

For partially replicated trials (Cullis et al. 2006), Gogel et al. (2018) state on page 6 of their paper, that using single stage analysis is preferred. It might be conjectured that using random plus de-regression at the first of a two-stage process, might be preferable

to starting with fixed genetic effects as in Piepho et al. (2012) and Gogel et al. (2018). This approach is examined and formalised in this paper, and ultimately uses the approach given by Smith et al. (2001a) for practical application.

Two examples are used in this paper to illustrate the approaches and to compare various two-stage methods and a one-stage method. The two data sets examined were a barley breeding MET with 24 trials (environments) and replicated lines (at least 2), and a MET for 4 trials in a wheat variety trial in which partial replication was used. The approaches that use either fixed or random plus de-regression genetic effects are presented. The results of analysis of both data sets are presented. In animal breeding, animals cannot be replicated, and genetic effects need to be assumed random. A relationship matrix is required for analysis. With partially replicated designs (Cullis et al. 2006) of low replication, it seems intuitively clear that using fixed genetic effects at the first stage will be problematic, even with an optimal design. A small simulation study is presented to illustrate that at the first stage, it is preferable to use random genetic effects, and then convert these to best linear unbiased estimates through de-regression, for use at the second stage. Discussion and conclusions complete the paper.

Examples

Barley stage 3 variety selection trials

Multi-environment trials were conducted by the South Australian Barley Program in 2006 and 2007. A total of 24 Stage 3 trials were grown over the two years. The designs for all trials were randomised complete blocks, with two or three blocks and hence two or three replicates of the varieties. The details of each trial are given in Table 1. The number of common lines within the two years is approximately 180, while between years there are approximately 50 lines for each pair of trials. A total of 322 lines were used in the trials. No pedigree or genomic information is available. The trait of interest is yield. Note the mean yields (in tonnes/hectare) are highly variable reflecting the multi-environment nature of the trials. The aim of the trials was promotion of lines to stage 4 trials, and hence best linear unbiased predictions of

Table 1 Barley trial information: name of the trial, layout (number of rows and columns), number of varieties at each site, number of missing values and the mean yield for each site

Site	Number of				Mean
	Rows	Columns	Varieties	Missing values	Yield
06S3BRE	60	6	180	1	1.98
06S3BRI	46	12	184	4	0.25
06S3CAL	46	12	184	0	0.66
06S3CLI	46	12	184	2	0.91
06S3MRC	6	60	180	2	0.41
06S3MUR	46	12	184	0	2.00
06S3NAR	46	12	184	0	0.81
06S3NTR	46	12	184	0	3.06
06S3PAR	46	12	184	2	1.58
06S3WEE	46	12	184	4	0.73
06S3YEE	46	12	184	2	1.10
07S3BRE	60	6	180	0	2.36
07S3BRI	45	12	180	0	1.81
07S3CAL	45	12	180	0	1.19
07S3CLI	45	12	180	0	2.27
07S3COR	45	12	180	0	3.22
07S3GRA	45	12	180	12	2.50
07S3MRC	60	6	180	25	0.31
07S3MUR	45	12	180	1	3.11
07S3NAR	45	12	180	0	0.49
07S3SWA	45	12	180	4	1.11
07S3TAR	45	12	180	3	2.63
07S3WEE	45	12	180	0	1.70
07S3YEE	45	12	180	4	1.01

variety by environment effects are required, to enable selection of varieties.

Multi-site p-rep yield trials

A series of wheat variety trials were conducted at 4 sites over 2 years, sites 1 and 2 in year 1 and sites 3 and 4 in year 2. The layouts (rows,columns and blocks), the number of varieties in each trial and the proportion that are replicated, the number of missing values and mean yield (the trait of interest, tonnes/hectare) are given in Table 2. A simple initial randomised block design of 2 blocks (Reps) was used for all trials. Additional blocking with 2 blocks within each Rep, was included in the first year, while in the second year, further blocking was based on the level of phenology (flowering time), classified into 3 classes. Within each Rep, there were 6 blocks, each phenology class being replicated. As pointed out by a reviewer, blocking on phenology might be an issue, as randomisation of lines is compromised by the likely presence of a genetic component for phenology traits, and hence estimated variance components may be biased.

The trials were designed using partial replication of lines, although the replication rate is fairly high. There were 943 lines in total with 25 standard or commercial lines used in the trials. The concurrence of lines across sites varied from 482 to 616.

As for the barley trials, the aim was selection of varieties, using best linear unbiased prediction.

Simulation study: multi-site p-rep

A small simulation study was conducted to investigate two-stage methods for p-rep or partially replicated designs. In the simulation study, data was generated for four sites, and three levels of partial replication were examined, namely 50%, 20% and

Table 2 p-rep trial information: name of the trial, layout (number of rows, columns, Reps, blocks and the configuration for each Rep and block), number of varieties at each site, level of partial replication (P-rep), number of missing values and the mean yield for each site

Site	Number of						Missing values	Mean yield
	Rows	Columns	Rep	Blocks	Varieties	P-rep		
1	28	36	2 (28 × 18)	4 (14 × 18)	735	0.37	6	2.80
2	20	56	2 (20 × 28)	4 (20 × 14)	801	0.39	14	4.40
3	60	16	2 (30 × 16)	12 (5 × 16)	641	0.49	6	2.49
4	84	12	2 (42 × 12)	12 (7 × 12)	666	0.51	2	2.71

10% replication of the total of 480 lines. The simulated trials consisted of 24 columns and 30, 24 and 22 rows for the three replication rates (50, 20 and 10%) respectively. The replicated lines were assigned to 6 blocks (of 4 columns) in a balanced incomplete block design for all scenarios, using find.BIB from the crossdes R package (https://CRAN.R-project.org/package=crossdes). In the case of 50% replication, allocation of replicated lines was balanced over sites. For the lower replication rates, different replicated lines appeared at each site. Non-replicated lines were allocated to each block at random. An optimal row(-column) BIB design was then constructed using odw (Butler and Cullis 2022) using the tabu+rw option for each site. For a discussion of design of plant breeding experiments, including p-rep designs, see Piepho et al. (2021).

For each simulation, the model used consisted of fixed site means, random Block, Row and Column effects and a spatial ar1 × ar1 (Gilmour et al. 1997) residual model. The parameters used are given in Table 3.

The variance matrix for genetic effects for sites was (with correlations below the diagonal)

$$\mathbf{G}_e = \begin{bmatrix} 0.150 & 0.027 & 0.120 & 0.060 \\ 0.100 & 0.500 & 0.082 & 0.383 \\ 0.800 & 0.300 & 0.150 & 0.150 \\ 0.200 & 0.700 & 0.500 & 0.600 \end{bmatrix} \qquad (1)$$

All designs were generated using the variance components and parameters for each site so that the designs are optimal for the simulations.

The aim of the simulation study was to compare 5 approaches, namely, the one-stage, the two-stage analysis using fixed or random plus de-regression

variety effects at the first stage, and subsequently using the full weight matrix and diagonal weights at the second stage. Using an unstructured variance matrix for the site genetic variance matrix resulted in convergence issues. Thus a factor analytic model (of order 1) was used for the analysis in all the simulations across all methods.

The approaches were compared using mean square error of prediction (MSEP) of the predicted random variety effects for the five approaches, with the simulated true variety effects for each scenario. The mean and standard deviation of the MSEP were also calculated as an overall summary over the 100 simulations for each scenario. Thus for simulation i, the 1920 (480 varieties ×4 sites) effects were used to find the MSEP, namely

$$\text{MSEP}_i = \frac{\sum_{j=1}^{1920}(u_{ij} - \hat{u}_{ij})^2}{1920} \qquad (2)$$

and subsequently the mean and standard deviation of the MSEP was calculated over the 100 simulations. In addition, counts of the ranking of the 5 approaches on MSEP_i (smallest 1 to worst 5) over the 100 simulations were found.

Methods

Models

Piepho et al. (2012) and Gogel et al. (2018) provided details of the one-stage analysis for MET data. Suppose we have t environments. Let $\mathbf{y}_j$ be the $n_j \times 1$ vector of observations for environment j, $j = 1, 2, \ldots, t$. A mixed model forms the basis of analysis, and for each environment

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\tau}_j + \mathbf{Z}_{gj}\mathbf{u}_{gj} + \mathbf{Z}_{oj}\mathbf{u}_{oj} + \mathbf{e}_j \qquad (3)$$

where the four terms on the right-hand side of (3) represent fixed effects, genetic effects (typically a random effect), other random effects (hence o) and the residual. The fixed effects typically include a mean effect for the environment, and may also include global trends for that environment. The "other" random effects $\mathbf{u}_{oj}$ typically contain design (blocking) effects, and where required global and extraneous effects. Lastly, the residual effects often allow for

Table 3 Parameter values for all simulations, Mean(s) are fixed effects, and the remaining effects are random.

Effect	Site 1	Site 2	Site 3	Site 4
Mean	3	4	5	2
Block	0.03	0.02	0.02	0.04
Column	0.06	0.05	0.08	0.02
Row	0.03	0.10	0.04	0.02
Variance	0.15	0.25	0.15	0.20
Column cor	0.40	0.50	0.60	0.30
Row cor	0.20	0.30	0.40	0.50

spatial variation for that site. All these aspects are discussed by Gilmour et al. (1997), and the examples provide details on terms that appear in the (3) and the models developed below.

The genetic effects $\mathbf{u}_{gj}$ will vary in dimension in most cases, unless all varieties are the same for each environment or trial. Let n_{gj} be the number of varieties in environment j.

For a succinct presentation, the models (3) are put together into a single model. Thus if $\mathbf{y} = (\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_t^T)^T$ is the vector of responses across all environments, and the full data set has $n = \sum_{j=1}^{t} n_j$ observations, the combined mixed model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{g2}\mathbf{u}_{g2} + \mathbf{Z}_o\mathbf{u}_o + \mathbf{e} \tag{4}$$

where the design matrices are all block diagonal, with the diagonal blocks being the design matrices in (3), and the vectors are concatenations of the vectors of effects in (3). The effects $\mathbf{u}_o$ and residual vector $\mathbf{e}$ are independent, with $\mathbf{u}_o \sim N(\mathbf{0}, \mathbf{G}_o)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$. The variance matrices $\mathbf{G}_o$ and $\mathbf{R}$ are block diagonal, so that non-genetic and residual effects are also environment specific. The labelling for the genetic effects includes a subscript 2, and this denotes the form that would be appropriate in a two-stage analysis.

One-stage analysis for MET data

In a one-stage analysis, the model specification is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_o\mathbf{u}_o + \mathbf{e} \tag{5}$$

where the non-genetic effects match those in (4), but the genetic effects are such that all varieties are represented at all environments. This allows prediction of all varieties at all environments, even if particular varieties were not in a particular trial. In (5), the genetic effects are assumed $\mathbf{u}_g \sim N(\mathbf{0}, \mathbf{G}_g)$, where $\mathbf{G}_g = \mathbf{G}_e \otimes \mathbf{I}_{n_g}$, where $\mathbf{G}_e$ is the covariance matrix for the environments. Thus here, the lines are assumed unrelated, because no pedigree or genomic information is available. The methods can be modified easily, if such information is available. In that case, multiple terms would be required in plant trials with replicated lines.

Note that

$$\mathbf{Z}_g = \mathbf{Z}_{g2}\mathbf{D} \tag{6}$$

for a block diagonal matrix $\mathbf{D}$, and where this matrix pads out $\mathbf{Z}_{g2}$ with zero columns for those varieties not observed at a particular site.

Let $\boldsymbol{\epsilon} = \mathbf{Z}_o\mathbf{u}_o + \mathbf{e}$. The variance matrix is $\operatorname{var}\boldsymbol{\epsilon} = \mathbf{V} = \mathbf{Z}_o\mathbf{G}_o\mathbf{Z}_o^T + \mathbf{R}$. The mixed model (5) can then be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \boldsymbol{\epsilon} \tag{7}$$

The mixed model equations for estimation of fixed effects and prediction of random effects are

$$\begin{bmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{V}^{-1}\mathbf{Z}_g \\ \mathbf{Z}_g^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{Z}_g^T\mathbf{V}^{-1}\mathbf{Z}_g + \mathbf{G}_g^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \tilde{\mathbf{u}}_g \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{Z}_g^T\mathbf{V}^{-1}\mathbf{y} \end{bmatrix} \tag{8}$$

Define

$$\mathbf{H} = \mathbf{V} + \mathbf{Z}_g\mathbf{G}_g\mathbf{Z}_g^T \tag{9}$$

$$\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^{-1} \tag{10}$$

The conventional estimate of $\boldsymbol{\tau}$ and best linear unbiased prediction of $\mathbf{u}_g$ are given by

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}^T\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^{-1}\mathbf{y} \tag{11}$$

$$\tilde{\mathbf{u}}_g = \mathbf{G}_g\mathbf{Z}_g^T\mathbf{P}\mathbf{y} \tag{12}$$

Let $\mathbf{X} = [\mathbf{X}_e \ \mathbf{X}_o]$ and $\boldsymbol{\tau} = [\boldsymbol{\tau}_e^T \ \boldsymbol{\tau}_o^T]^T$ be a partition of the fixed effects design matrix and parameter vector into site or environment means (indexed by e) and other fixed effects (indexed by o). The examples present terms that appear in the other fixed effects. Note that

$$\mathbf{X}_e = \mathbf{Z}_g\mathbf{D}_e \tag{13}$$

where $\mathbf{D}_e$ is a diagonal matrix whose diagonal blocks are simply a vector of unities of length the number of varieties in the corresponding environment. The matrix $\mathbf{D}_e$ appears in the model for the second stage in a two-stage analysis. The other fixed effects can be eliminated from the mixed model equations to form the reduced mixed model equations

$$\begin{bmatrix} \mathbf{X}_e^T\mathbf{S}_o\mathbf{X}_e & \mathbf{X}_e^T\mathbf{S}_o\mathbf{Z}_g \\ \mathbf{Z}_g^T\mathbf{S}_o\mathbf{X}_e & \mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g + \mathbf{G}_g^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}}_e \\ \tilde{\mathbf{u}}_g \end{bmatrix} = \begin{bmatrix} \mathbf{X}_e^T\mathbf{S}_o\mathbf{y} \\ \mathbf{Z}_g^T\mathbf{S}_o\mathbf{y} \end{bmatrix} \tag{14}$$

where

$$\mathbf{S}_o = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_o(\mathbf{X}_o^T\mathbf{V}^{-1}\mathbf{X}_o)^{-1}\mathbf{X}_o^T\mathbf{V}^{-1} \qquad (15)$$

allows for the elimination of $\boldsymbol{\tau}_o$. Note that using (9), (15) and the Sherman–Morrison–Woodbury formula, (10) can be written as

$$\mathbf{P} = \mathbf{S}_o - \mathbf{S}_o\mathbf{Z}_g(\mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g + \mathbf{G}_g^{-1})^{-1}\mathbf{Z}_g\mathbf{S}_o \qquad (16)$$

which will be used to connect one and two-stage methods.

Note that in Gogel et al. (2018), their Eq. (3) is not correct as it is missing a term on the right-hand side.

Two-stage analysis of MET data

There are two important components of a two-stage analysis. The first component is that the estimated effects taken forward to the second stage of analysis must be unbiased for the effects of interest. This means that at the end of stage one, best linear unbiased estimates and hence fixed effects estimates are required. The second component is a weight matrix (or weights). This is required because estimates are taken as the response in the stage two model, and estimates have an error associated with them. The weight matrix depends on non-genetic variance parameters and these are estimated at the first stage of analysis. It might be better to estimate the non-genetic variance components in an extra step in the first stage of analysis, using a model with random genetic effects. But the resulting predicted genetic effects are biased, and hence an additional step to convert (de-regress) these to fixed effects is required, that is, random plus de-regression as noted earlier.

The following sub-sections consider the stage one analysis starting with fixed genetic effects, and alternatively starting with random genetic effects, then de-regressing to obtain fixed effect estimates.

Fixed genetic effects at stage 1

Piepho et al. (2012) and Gogel et al. (2018) present the two-stage approach assuming fixed genetic effects at stage 1. In particular they show that for known variance matrices, and a full weight matrix carried from stage 1 to stage 2, the same form for estimates of $\boldsymbol{\tau}_e$ and predictions of $\mathbf{u}_g$..

Equation (4) is the basis for two-stage analysis. If $\mathbf{u}_{g2}$ are taken as fixed effects in (4), estimation is based on the normal equations

$$\begin{bmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{V}^{-1}\mathbf{Z}_{g2} \\ \mathbf{Z}_{g2}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{Z}_{g2}^T\mathbf{V}^{-1}\mathbf{Z}_{g2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}} \\ \hat{\mathbf{u}}_{g2} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{Z}_{g2}^T\mathbf{V}^{-1}\mathbf{y} \end{bmatrix} \quad (17)$$

with solutions

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}_{g2}\hat{\mathbf{u}}_{g2}) \qquad (18)$$

$$\hat{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{V}^{-1}\mathbf{Z}_{g2})^{-1}\mathbf{Z}_{g2}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\tau}}) \qquad (19)$$

Substituting (18) into (19), rearranging, and defining

$$\mathbf{S} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1} \qquad (20)$$

we find

$$\hat{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2})^{-1}\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{y} \qquad (21)$$

Equation (21) is a key results for the following sub-section. Notice that $\hat{\mathbf{u}}_{g2}$ is an unbiased estimate of $\mathbf{u}_g$, and this is a key property that is needed to proceed to the second stage of analysis. Taking forward a biased estimate, would introduce bias into the second stage analysis.

The final details for the first stage are presented after the discussion of random plus de-regression genetic effects at the first stage.

Random genetic effects plus de-regression at stage 1

Now suppose $\mathbf{u}_{g2}$ is taken as a random effect at stage 1. A two-stage analysis ignores genetic covariance between environments. Indeed each environment is analyzed separately or in a single analysis where all fixed, random, and genetic effects have block diagonal design matrices, and in the latter two cases variance matrices that are block diagonal so that each environment has it's own effects without covariance or correlation across environments. For our examples, $\mathbf{G}_{g2d} = \mathrm{diag}\left(\sigma_{gj}^2\mathbf{I}_{n_{gj}}\right)$, recalling n_{gj} is the number of varieties at environment j. In particular, the genetic random effects in the combined model (4) are specified as $\mathbf{u}_{g2} \sim N(\mathbf{0}, \mathbf{G}_{g2d})$, with the d

making the block diagonal form explicit. So our mixed model at stage 1 is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau}_d + \mathbf{Z}_{g2}\mathbf{u}_{g2} + \boldsymbol{\epsilon} \tag{22}$$

and in the mixed model Eq. (8), $\tilde{\mathbf{u}}_g$ and $\mathbf{G}_g$ are replaced by $\tilde{\mathbf{u}}_{g2}$ and $\mathbf{G}_{g2d}$. The solutions of the mixed model equations are

$$\hat{\boldsymbol{\tau}}_d = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}_{g2}\tilde{\mathbf{u}}_{g2}) \tag{23}$$

$$\tilde{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{V}^{-1}\mathbf{Z}_{g2} + \mathbf{G}_{g2d}^{-1})^{-1}\mathbf{Z}_{g2}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\tau}}_d) \tag{24}$$

and similar to the case of fixed genetic effects, substituting (23) into (24) we find

$$\tilde{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2} + \mathbf{G}_{g2d}^{-1})^{-1}\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{y} \tag{25}$$

where $\mathbf{S}$ is given in (20). Note that (21) and (25) only differ by the term $\mathbf{G}_{g2d}^{-1}$.

Now,

$$\mathrm{E}(\tilde{\mathbf{u}}_{g2}) = (\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2} + \mathbf{G}_{g2d}^{-1})^{-1}\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2}\mathbf{u}_{g2}$$

and hence is biased for $\mathbf{u}_{g2}$. This in turn means $\hat{\boldsymbol{\tau}}_d$ is also biased, using (23), and the effects that would be taken to the second stage of analysis (combined estimated site means and the predicted genetic effects) would also be biased. However, the bias can be removed. Thus if we define

$$\hat{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2})^{-1}(\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2} + \mathbf{G}_{g2d}^{-1})\tilde{\mathbf{u}}_{g2} \tag{26}$$

this estimate is unbiased for $\mathbf{u}_{g2}$. If this estimate is used in (23), $\hat{\boldsymbol{\tau}}_d$ is unbiased for $\boldsymbol{\tau}_d$, and the effects taken forward to the second stage of analysis will be unbiased for the genetic by environment means. In the animal breeding literature, this is termed de-regression (Garrick et al. 2009).

Note that substituting (25) into (26),

$$\hat{\mathbf{u}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{Z}_{g2})^{-1}\mathbf{Z}_{g2}^T\mathbf{S}\mathbf{y}$$

which is identical to (21), found if the genetic effects are taken as fixed. Thus, de-regression results in an estimate of the same form as for fixed genetic effects.

The practical effect of the above derivation is that when a stage 1 analysis is undertaken with random genetic effects, it is necessary to de-regress both the predicted random genetic effects and the fixed effects of interest. A practical way to do this is presented

in Smith et al. (2001a). The stage 1 model with random genetic effects is fitted. The resulting estimated non-genetic variance parameters (and hence their structures) are then fixed and the model with genetic effects taken as fixed effects rather than random is then fitted.

Stage 1 fitted model

The final details of the first stage of analysis using either fixed or random plus de-regression genetic effects are as follows. It is simpler to fit genetic by environment means at the end of first stage, rather than site means plus genetic by environment effects. This is true for both fixed and random plus de-regression cases, noting that in the latter, the fitting is achieved by fixing variance parameters at their values when genetic effects were fitted as random. This means that $\mathbf{X}_e\boldsymbol{\tau}_e$ is omitted from the model (7), and the model fitted is

$$\mathbf{y} = \mathbf{X}_o\boldsymbol{\tau}_o + \mathbf{Z}_{g2}\boldsymbol{\tau}_{g2} + \boldsymbol{\epsilon} \tag{27}$$

where $\boldsymbol{\tau}_{g2}$ are the fixed effects for genetic by environment means. With $\mathbf{S}_o$ defined in (15), and in a similar fashion to (21), we find under (27)

$$\hat{\boldsymbol{\tau}}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{Z}_{g2})^{-1}\mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{y} \tag{28}$$

and importantly the conditional distribution of $\hat{\boldsymbol{\tau}}_{g2}$ given $\boldsymbol{\tau}_{g2}$ is

$$\hat{\boldsymbol{\tau}}_{g2}|\boldsymbol{\tau}_{g2} \sim \mathrm{N}\left(\boldsymbol{\tau}_{g2}, (\mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{Z}_{g2})^{-1}\right) \tag{29}$$

This distribution forms the basis of the second stage of analysis.

Stage 2 analysis

The second stage model is

$$\hat{\boldsymbol{\tau}}_{g2} = \mathbf{D}_e\boldsymbol{\tau}_e + \mathbf{D}\mathbf{u}_g + \mathbf{e}_{g2} \tag{30}$$

where $\mathbf{u}_g$ is the full set of genetic effects as in the one-stage model (5), and $\mathbf{D}_e$ and $\mathbf{D}$ are defined (13) and (6). Note that $\mathbf{u}_g \sim \mathrm{N}(\mathbf{0}, \mathbf{G}_g)$ and $\mathbf{e}_{g2} \sim \mathrm{N}(\mathbf{0}, \mathbf{V}_{g2})$ where

$$\mathbf{V}_{g2} = (\mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{Z})_{g2}^{-1} \tag{31}$$

The distribution of $\mathbf{e}_{g2}$ has no unknown parameters, and the variance matrix is block diagonal (the blocks correspond to the environments) and is fully known (found at the first stage of analysis). With these specifications the mixed model equations based on (30) are

$$\begin{bmatrix} \mathbf{D}_e^T \mathbf{V}_{g2}^{-1} \mathbf{D}_e & \mathbf{D}_e^T \mathbf{V}_{g2}^{-1} \mathbf{D} \\ \mathbf{D}^T \mathbf{V}_{g2}^{-1} \mathbf{D}_e & \mathbf{D}^T \mathbf{V}_{g2}^{-1} \mathbf{D} + \mathbf{G}_g^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}}_e \\ \tilde{\mathbf{u}}_g \end{bmatrix} = \begin{bmatrix} \mathbf{D}_e^T \mathbf{V}_{g2}^{-1} \hat{\boldsymbol{\tau}}_{g2} \\ \mathbf{D}^T \mathbf{V}_{g2}^{-1} \hat{\boldsymbol{\tau}}_{g2} \end{bmatrix} \tag{32}$$

Substituting for the weight matrix $\mathbf{V}_{g2}^{-1}$ in (32) using (31),

$$\begin{bmatrix} \mathbf{D}_e^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \mathbf{D}_e & \mathbf{D}_e^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \mathbf{D} \\ \mathbf{D}^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \mathbf{D}_e & \mathbf{D}^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \mathbf{D} + \mathbf{G}_g^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}}_e \\ \tilde{\mathbf{u}}_g \end{bmatrix} = \begin{bmatrix} \mathbf{D}_e^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \hat{\boldsymbol{\tau}}_{g2} \\ \mathbf{D}^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \hat{\boldsymbol{\tau}}_{g2} \end{bmatrix} \tag{33}$$

Noting (13), (6) and

$$\mathbf{D}_e^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \hat{\boldsymbol{\tau}}_{g2} = \mathbf{X}_e^T \mathbf{S}_o \mathbf{y} \qquad \mathbf{D}^T \mathbf{Z}_{g2}^T \mathbf{S}_o \mathbf{Z}_{g2} \hat{\boldsymbol{\tau}}_{g2} = \mathbf{Z}_g^T \mathbf{S}_o \mathbf{y}$$

we see that (33) is identical to (14). Thus the *form* of estimates and predictions is the same for one and two-stage methods of analysis. Note that in Gogel et al. (2018), the derivation of $\hat{\boldsymbol{\tau}}_e$ is incorrect as a term is missing on the right-hand-side of their equation.

Lastly, the difference between one-stage and two-stage methods lies in the estimation of variance parameters in the linear mixed models. This difference is presented in the following sub-section.

Estimation of variance parameters

In all approaches, there is a need to estimate unknown parameters in $\mathbf{V}$ and $\mathbf{G}_g$. The score equations are compared for the one and two-stage approaches. If $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ are the parameters in $\mathbf{V}$ and $\mathbf{G}_g$ respectively, in all the equations presented below, the dot above a matrix means the derivative with respect to a parameter, either γ_j or ϕ_j, and j also appears in the subscript for derivative of the matrix.

Genetic variance parameters

The usual approach for estimation of variance parameters is to use residual maximum likelihood or REML (Patterson and Thompson 1971). For the one-stage approach, the residual log-likelihood is

$$\ell(\boldsymbol{\phi}, \boldsymbol{\gamma}; \mathbf{y}) = -\frac{1}{2} \left\{ \log(|\mathbf{H}|) + \log\left(\left|\mathbf{X}^T \mathbf{H}^{-1} \mathbf{X}\right|\right) + \mathbf{y}^T \mathbf{P} \mathbf{y} \right\} \tag{34}$$

where $|\cdot|$ denotes the determinant of the matrix argument and $\mathbf{H}$ and $\mathbf{P}$ were given in (9) and (10). The REML estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ are found by maximizing (34). This involves finding the score or derivative of the log-likelihood with respect to each parameter. For γ_j, the score is given by

$$U(\gamma_j) = -\frac{1}{2} \mathrm{tr}\left(\mathbf{P} \mathbf{Z}_g \dot{\mathbf{G}}_{gj} \mathbf{Z}_g^T\right) + \frac{1}{2} \mathbf{y}^T \mathbf{P} \mathbf{Z}_g \dot{\mathbf{G}}_{gj} \mathbf{Z}_g^T \mathbf{P} \mathbf{y} \tag{35}$$

and using (12) this reduces to

$$U(\gamma_j) = -\frac{1}{2} \mathrm{tr}\left(\mathbf{Z}_g^T \mathbf{P} \mathbf{Z}_g \dot{\mathbf{G}}_{gj}\right) + \frac{1}{2} \tilde{\mathbf{u}}_g^T \mathbf{G}_g^{-1} \dot{\mathbf{G}}_{gj} \mathbf{G}_g^{-1} \tilde{\mathbf{u}}_g \tag{36}$$

For the two-stage approach, the residual log-likelihood is based on (30). If

$$\mathbf{H}_2 = \mathbf{V}_{g2} + \mathbf{D} \mathbf{G}_{g2} \mathbf{D}^T \tag{37}$$

and

$$\begin{aligned} \mathbf{P}_2 &= \mathbf{H}_2^{-1} - \mathbf{H}_2^{-1} \mathbf{D}_e (\mathbf{D}_e^T \mathbf{H}_2^{-1} \mathbf{D}_e)^T \mathbf{D}_e^T \mathbf{H}_2^{-1} \\ &= \mathbf{Q}_{22} (\mathbf{Q}_{22}^T \mathbf{H}_2 \mathbf{Q}_{22})^{-1} \mathbf{Q}_{22}^T \end{aligned} \tag{38}$$

where $\mathbf{Q}_{22}$ is such that $\mathbf{Q}_{22}^T \mathbf{D}_e = \mathbf{0}$, the residual log-likelihood is

$$\ell(\boldsymbol{\gamma}; \hat{\boldsymbol{\tau}}_g) = -\frac{1}{2} \left\{ \log(|\mathbf{H}_2|) + \log\left(\left|\mathbf{D}_e^T \mathbf{H}_2^{-1} \mathbf{D}_e\right|\right) + \hat{\boldsymbol{\tau}}_g^T \mathbf{P}_2 \hat{\boldsymbol{\tau}}_g \right\} \tag{39}$$

The score for γ_j is then given by

$$U_2(\gamma_j) = -\frac{1}{2} \mathrm{tr}\left(\mathbf{P}_2 \mathbf{D} \dot{\mathbf{G}}_{gj} \mathbf{D}^T\right) + \frac{1}{2} \hat{\boldsymbol{\tau}}_g^T \mathbf{P}_2 \dot{\mathbf{G}}_{gj} \mathbf{P}_2 \hat{\boldsymbol{\tau}}_g \tag{40}$$

Now using $\mathbf{G}_g \mathbf{P}_2 \hat{\boldsymbol{\tau}}_g = \tilde{\mathbf{u}}_g$, the sum of squares in (40) equals

$$\hat{\boldsymbol{\tau}}_g^T \mathbf{P}_2 \dot{\mathbf{G}}_{gj} \mathbf{P}_2 \hat{\boldsymbol{\tau}}_g = \tilde{\mathbf{u}}_g^T \mathbf{G}_{g2}^{-1} \dot{\mathbf{G}}_{gj} \mathbf{G}_g^{-1} \tilde{\mathbf{u}}_g \tag{41}$$

The trace term requires some manipulation as follows. Firstly, using (37), (38), and the Sherman-Morrison-Woodbury formula,

$$\mathbf{P}_2 = \mathbf{P}_{V_{g2}} - \mathbf{P}_{V_{g2}} (\mathbf{D}^T \mathbf{P}_{V_{g2}} \mathbf{D} + \mathbf{G}_g^{-1})^{-1} \mathbf{D}^T \mathbf{P}_{V_{g2}}$$

where

$$\mathbf{P}_{V_{g2}} = \mathbf{Q}_{22}(\mathbf{Q}_{22}^T\mathbf{V}_{g2}\mathbf{Q}_{22})^{-1}\mathbf{Q}_{22}^T$$

Now $\mathbf{V}_{g2}{}^{-1} = \mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{Z}_{g2}$ and hence using (13) and (6), $\mathbf{D}_e^T\mathbf{Z}_{g2}^T\mathbf{S}_o\mathbf{Z}_{g2}\mathbf{D}_e = \mathbf{X}_e^T\mathbf{S}_o\mathbf{X}_e$ and $\mathbf{D}^T\mathbf{P}_{V_{g2}}\mathbf{D} = \mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g$, so that the term in the trace

$$
\begin{aligned}
\mathbf{D}^T\mathbf{P}_2\mathbf{D} &= \mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g - \mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g(\mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g + \mathbf{G}_g^{-1})^{-1}\mathbf{Z}_g^T\mathbf{S}_o\mathbf{Z}_g \\
&= \mathbf{Z}_g^T\mathbf{P}\mathbf{Z}_g
\end{aligned}
\tag{42}
$$

using (16). Substituting (41) and (42) into (40), the score is then equal to (35), the score for the one-step analysis.

While the form of the score is identical for the one and two-step approaches, estimates of unknown parameters in $\mathbf{V}$ (and hence $\mathbf{R}$ and $\mathbf{G}_o$) have to be used in calculating the effects and variance parameters. These estimates differ in the two two-stage approaches and differ with the one stage approach. This is because the score equations are updated in an iterative process for the one-stage analysis, whereas in the two-stage analysis, the non-genetic variance parameters are fixed from the first stage analysis and define the weight matrix $\mathbf{V}_{g2}{}^{-1}$. This matrix also differs between the two-stage approaches using fixed or random genetic effects at the first stage.

Non-genetic variance parameters

Using the residual log-likelihood (34), the score equation for a non-genetic variance parameter ϕ_j for the one-stage approach is

$$U(\phi_j) = -\frac{1}{2}\mathrm{tr}\left(\mathbf{P}\dot{\mathbf{H}}_j\right) + \frac{1}{2}\mathbf{y}^T\mathbf{P}\dot{\mathbf{H}}_j\mathbf{P}\mathbf{y} \tag{43}$$

For the two-stage approach using fixed genetic effects, the residual log-likelihood used for estimating non-genetic variance parameters at stage 1 is

$$\ell(\boldsymbol{\phi};\mathbf{y}) = -\frac{1}{2}\left\{\log(|\mathbf{V}|) + \log\left(\left|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right|\right) + \mathbf{y}^T\mathbf{S}\mathbf{y}\right\} \tag{44}$$

where $\mathbf{S}$ is given by (20). The score for ϕ_j then follows as

$$U_f(\phi_j) = -\frac{1}{2}\mathrm{tr}\left(\mathbf{S}\dot{\mathbf{V}}_j\right) + \frac{1}{2}\mathbf{y}^T\mathbf{S}\dot{\mathbf{V}}_j\mathbf{S}\mathbf{y} \tag{45}$$

For random genetic effects at stage 1, the residual log-likelihood is

$$\ell(\boldsymbol{\phi},\boldsymbol{\gamma};\mathbf{y}) = -\frac{1}{2}\left\{\log(|\mathbf{H}_d|) + \log\left(\left|\mathbf{X}^T\mathbf{H}_d^{-1}\mathbf{X}\right|\right) + \mathbf{y}^T\mathbf{P}_d\mathbf{y}\right\} \tag{46}$$

where $\mathbf{H}_d$ and $\mathbf{P}_d$ are of the same form as (9) and (10) but with $\mathbf{G}_g$ replaced by $\mathbf{G}_{g2d}$. The score for ϕ_j is

$$U_r(\phi_j) = -\frac{1}{2}\mathrm{tr}\left(\mathbf{P}_d\dot{\mathbf{H}}_{dj}\right) + \frac{1}{2}\mathbf{y}^T\mathbf{P}_d\dot{\mathbf{H}}_{dj}\mathbf{P}_d\mathbf{y} \tag{47}$$

These are all of different forms and hence the estimates of non-genetic variance parameters will differ across the methods. As the score equations under the random genetic effects scenario are closer to the one-stage equations, it might be the case that the estimates in that case will be "better" than under a fixed genetic effects scenario.

De-regression and diagonal weight matrices

The original development of de-regression by Garrick et al. (2009) considered adjusting individual predicted genetic random effects. This also meant that individual weights were formed, and hence a diagonal weight matrix resulted. For very large problems, it may only be feasible to use a diagonal weight matrix. Smith et al. (2001a) take the diagonal elements of the inverse variance matrix $\mathbf{V}_{g2}^{-1}$ as the weights. These elements are conditional variances and hence provide an adjustment for other effects. The use of diagonal weights was investigated to some extent by Gogel et al. (2018) and they show that in some cases using diagonal weights shows considerable disagreement with a one-stage and two-stage approach using the full weight matrices. This is further examined in the examples and the simulation study.

Results

All analyses were conducted using the asreml package in R (Butler et al. 2018), which is a commercial product (VSN International, https://vsni.co.uk/software/asreml). A function that automates the stage 1 analysis (stage1) is available from the author, as is data and code for the analyses of the two examples.

Barley stage 3 variety selection trials

To investigate two-stage methods of analysis, the one-stage analysis of the 24 site multi-environment trials was conducted first. Each trial was analyzed separately (needed for two-stage methods as well) to examine what fixed and random effects were needed to account for extraneous and global variation (Gilmour et al. 1997). A subset of linear row and column, and random row and column effects were needed across sites. These constitute terms that appear in $\mathbf{X}_o\boldsymbol{\tau}_o$ and $\mathbf{Z}_o\mathbf{u}_o$, the latter also including random block effects. All analyses included the separable ar1 by ar1 spatial model. The necessary extraneous and global effects were then included in a full multi-site analysis.

The models fitted for the random genetic effects were a diagonal (diag) structure over sites, and then factor analytic models of order 1 to 9, labelled fa1 to fa9 in Table 4. This table also includes the residual log-likelihood, sequential residual likelihood ratio statistics with the corresponding degrees of freedom and p-value for each test.

The likelihood ratio tests suggests the fa7 model is required and this model was used in the comparisons given below.

There were four two-stage methods used for comparison. For all four methods, linear row and column, and random row and column effects were included in the model as well as random block effects. The spatial

ar1 by ar1 model was fitted for the residuals. This enabled an automated first stage analysis.

Firstly, fixed genetic effects were fitted for each site. The estimated fixed genetic effects were then taken forward, with a full weight matrix and a diagonal weight matrix. The same approach was used for random genetic effects at the first stage, with the de-regressed genetic effects taken forward to the second stage analyses, again with a full weight matrix and a diagonal set of weights.

The second stage analyses followed the one-stage process by starting with a diagonal variance matrix for sites for random site by genetic effects, and then fitted a sequence of factor analytic models.

Table 5 gives the models fitted for four two-stage approaches, fixed or random plus de-regression effects at the first stage and either a full weight matrix or diagonal weights carried forward. Using sequential residual likelihood ratio tests, a factor analytic model of order 8 (fa8) was selected for all two-stage approaches. Note that if the standard approach was used, model fitting would have stopped at the fa7 model for three of the approaches and the fa6 selected as the best model. This highlights the fact that important factors need not enter in order of importance, and that a sequential approach might miss factors that are important. Discovering that the fa8 improved the fit occurred by chance and not design. The two-stage models used for comparison were the fa8 in all cases compared to the fa7 used for the one-stage approach.

As variety selection is the aim, the predicted variety effects are of interest. These effects for one-stage model and the four two-stage models are plotted against each other for a selection of sites. In Fig. 1, the plots of the BLUPs for each pair of methods and their Pearson correlation are presented for Site 06S3NTR. For this site, the full weight matrix is slightly better than the diagonal form, for both fixed and random plus de-regression stage 1 analyses and the fixed first stage models are marginally better than their random counterparts.

Figure 2 provides the pairwise plots of BLUPs and their Pearson correlations for site 06S3WEE. For this site using random plus de-regression genetic effects at the first stage leads to a better stage 2 correspondence with the one-stage results than when fixed genetic effects were used at the first stage.

An examination of the 24 sites showed that there is no consistent pattern of which two-stage method,

Table 4 Barley Multi-environment trial, one-stage analysis: The sequence of models for the site genetic variance matrix is presented together with the residual log-likelihood for each model, residual likelihood ratio statistics (REMLRS) for sequential models, the degrees of freedom (df) and p-value for the test

Model	Residual	Sequential		
	Log-likelihood	REMLRS	Df	P-value
Diag	10188.48			
fa1	11293.31	2209.65	23	0.000
fa2	11614.14	641.68	23	0.000
fa3	11710.59	192.89	22	0.000
fa4	11740.97	60.77	18	0.000
fa5	11761.47	41.00	19	0.002
fa6	11783.62	44.30	18	0.000
fa7	11802.13	37.00	17	0.003
fa8	11813.70	23.16	14	0.058
fa9	11822.51	17.61	15	0.284

Table 5 Barley Multi-environment trial, second stage analysis using fixed (Fixed) or random plus de-regression (Random) genetic effects at the first stage with either a full weight matrix (Full) or diagonal weights (Diagonal): The sequence of models for the site genetic variance matrix is presented together with the residual log-likelihood for each model, residual likelihood ratio statistics for sequential models (REMLRS), the degrees of freedom (df) and p-value for the test

Genetic effects at Stage 1	Weights	Model	Residual Log-likelihood	Sequential REMLRS	Df	p-value
Fixed	Full	Diag	3516.79			
		fa1	4618.35	2203.11	24	0.000
		fa2	4931.46	626.21	23	0.000
		fa3	5027.21	191.51	22	0.000
		fa4	5052.05	49.67	19	0.000
		fa5	5083.57	63.04	19	0.000
		fa6	5102.87	38.60	17	0.002
		fa7	5118.53	31.33	18	0.026
		fa8	5130.70	24.33	14	0.042
		fa9	5141.10	20.80	13	0.077
Fixed	Diagonal	Diag	3499.94			
		fa1	4597.72	2195.56	24	0.000
		fa2	4902.71	609.98	23	0.000
		fa3	4993.90	182.38	22	0.000
		fa4	5018.80	49.80	19	0.000
		fa5	5043.98	50.37	19	0.000
		fa6	5065.76	43.55	17	0.000
		fa7	5077.98	24.45	17	0.108
		fa8	5092.08	28.20	15	0.020
		fa9	5102.16	20.16	15	0.166
Random	Full	Diag	3544.55			
		fa1	4646.01	2202.93	24	0.000
		fa2	4962.73	633.43	23	0.000
		fa3	5058.92	192.38	22	0.000
		fa4	5088.79	59.75	18	0.000
		fa5	5110.45	43.33	19	0.001
		fa6	5134.47	48.02	19	0.000
		fa7	5145.94	22.94	15	0.085
		fa8	5161.88	31.89	15	0.007
		fa9	5172.73	21.69	13	0.060
Random	Diagonal	Diag	3512.36			
		fa1	4610.34	2195.97	24	0.000
		fa2	4917.27	613.87	23	0.000
		fa3	5008.60	182.66	22	0.000
		fa4	5035.72	54.23	21	0.000
		fa5	5059.56	47.69	20	0.000
		fa6	5081.60	44.08	19	0.001
		fa7	5095.90	28.61	18	0.053
		fa8	5110.28	28.76	17	0.037
		fa9	5118.45	16.34	16	0.429

fixed or random plus de-regression is better overall. It depends on the site. However, the differences were generally small. The full weight matrix approach is generally better than using diagonal weights, but the BLUPs were highly correlated with the corresponding full weight matrix case.

Two-stage methods do not estimate the unknown variance parameters in the same manner as an

Fig. 1 Barley Multi-environment Trial, Site 06S3NTR: Five models in a scatterplot matrix. A factor analytic model of seven factors is fitted in a one-stage analysis (label fa7), an fa8 was used two-stage models with fixed (label fixed) or random plus de-regression (label random) at stage one, and in each case a full weight matrix and diagonal weights were used at stage two
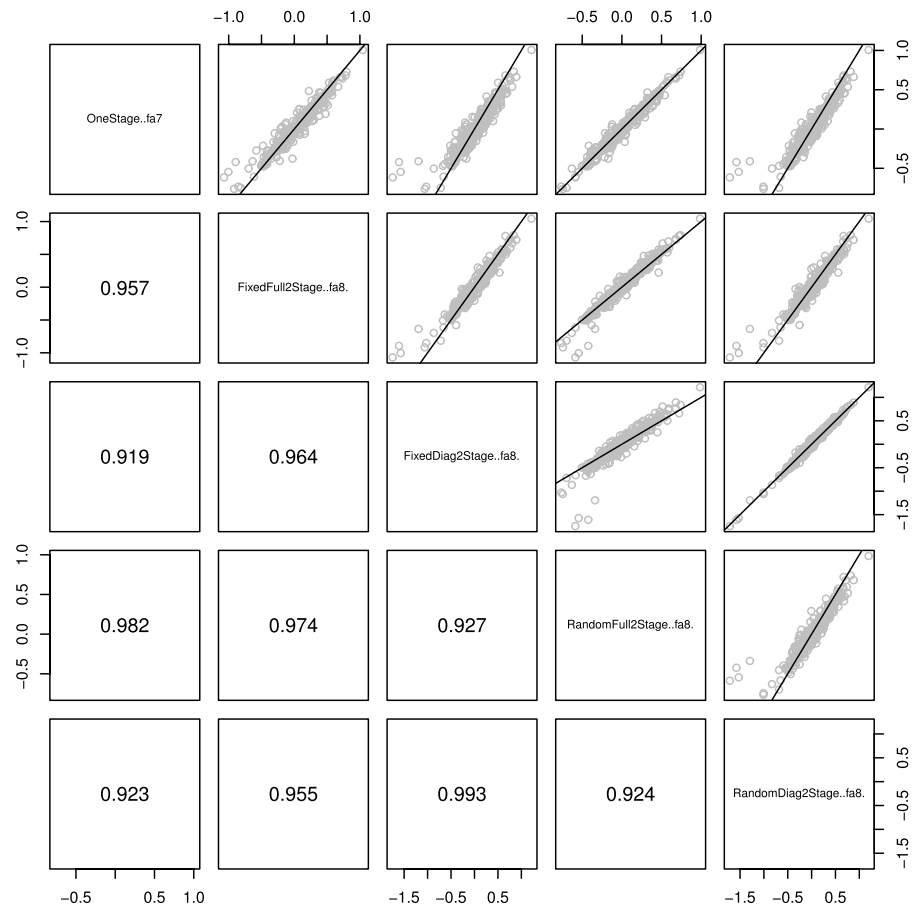
iterative full analysis. The estimated variance components were compared graphically for the different methods.

For the genetic variance model, the full estimated 24×24 variance matrix was formed for the four approaches that involved weights from the first stage of analysis. The estimated correlations between the sites were found for each model and Fig. 3 is a scatterplot matrix of these estimated correlations, together with Pearson correlations for each pair of estimated correlations. It is clear that the correspondence of two-stage methods with the one-stage approach is very strong with high correlations. Using the fixed and random plus de-regression full weight matrices in particular are very close with the fixed approach being marginally better.

Figure 4 is a similar plot for the genetic variances. Again the Pearson correlations are very high, but under the fixed genetic effects stage 1 model

(using both the full weight matrix and diagonal weights) and the random plus de-regression genetic effects stage 1 model with diagonal weights, there is one over-estimated genetic variance. This is site 06S3NTR for which the BLUPs were presented in Fig. 1. It was clear from those plots that only the stage 1 approach using random genetic effects with a full weight matrix at stage 2 captured the full set of genetic effects well, when compared to a one-stage analysis. Note that the ranking of top lines is not affected in this example.

Non-genetic parameters were matched across the 3 approaches where these parameters are estimated (one-stage, and stage 1 analyses with fixed or random plus de-regression genetic effects) and are presented graphically in Fig. 5. The correlations with the one-stage approach are higher for the random plus de-regression stage 1 analysis, although the difference appears minor.

Fig. 2 Barley Multi-environment Trial, Site 06S3WEE: Five models in a scatterplot matrix. A factor analytic model of seven factors is fitted in a one-stage analysis (label fa7), and fa8 was used for two-stage models with fixed (label fixed) or random plus de-regression (label random) at stage one, and in each case a full weight matrix and diagonal weights were used at stage two
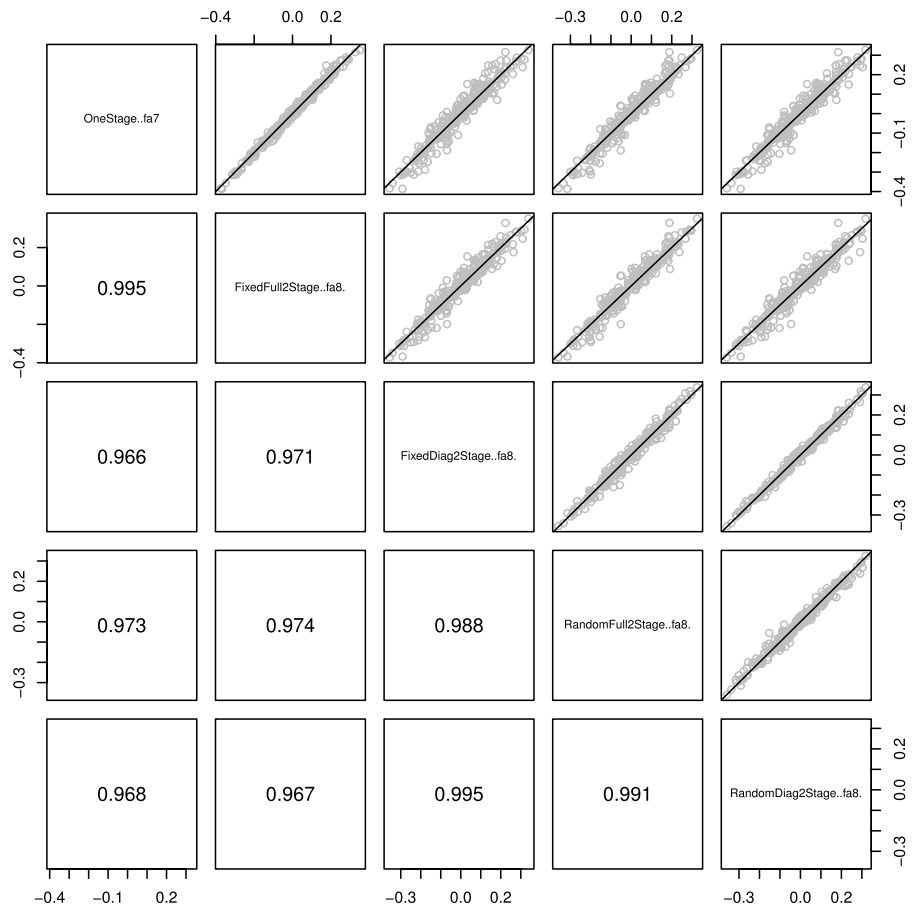
Multi-site p-rep yield trials

The same process was used for the p-rep trials as for the replicated Barley trials. Thus the five methods, one-stage, two-stage with fixed or random plus de-regression genetic effects at the first stage, and full and diagonal weights at the second stage, were examined.

The p-rep MET data were analyzed firstly using a one-stage approach. The Site factor (or trials) was included in the fixed effects in the model ($\mathbf{X}_e\boldsymbol{\tau}_e$), as was the phenology group factor for sites 3 and 4 ($\mathbf{X}_o\boldsymbol{\tau}_o$). The phenology factor was included so that shrinkage of variety effects would be relative to the membership of each variety to a phenology group. Random Rep, Block, and row and column effects were included for each trial ($\mathbf{Z}_o\mathbf{u}_o$) and separable autoregressive processes of order 1 for row and column components were used in the residual model (Gilmour et al. 1997) also for each trial. The genetic

model allowed for a separable structure for sites and lines, with a fully unstructured variance matrix used for the site component. Lines were assumed independent as no pedigree or genomic information was available.

For the two-stage analyses, the stage 1 model had a mean for the individual site being analyzed, as well as the phenology factor for sites 3 and 4, and random Rep, Block, row and column effects as for the one-stage model. The genetic effect for line was fixed or random with a scaled identity variance matrix. The residual model was a separable autoregressive processes of order 1 for row and column components as for the one-stage analysis.

For the second stage analyses, a full-weight matrix and the diagonal approximation as described above were used. The site by line model was the same as for the one-stage approach.

Figure 6 presents a scatterplot matrix of predicted genetic effects for all lines in the trials, for the five

Fig. 3 Barley Multi-environment Trial, comparison of estimated genetic correlations
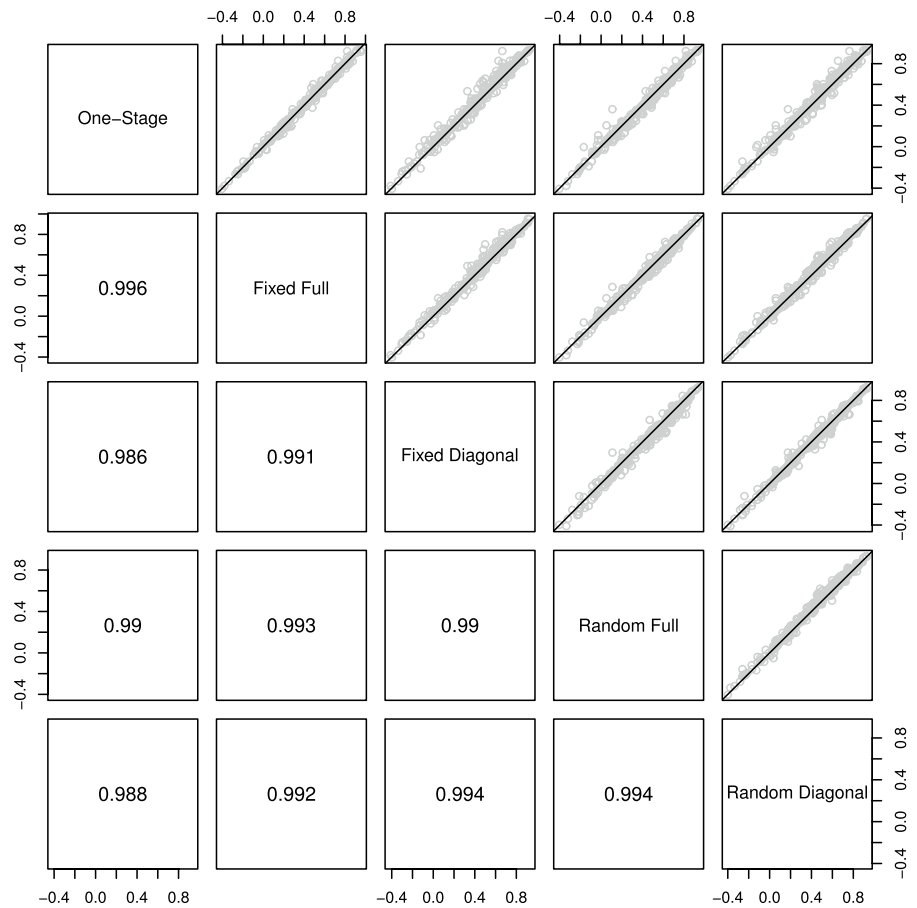
models for trial site 2 (similar results were found for other trials). The plotting panels represent the predicted line effects for pairs of models. The Pearson correlations are very high with the full weight matrix providing the highest correlation. Although not presented, the prediction standard errors are also highly correlated, although there is bias when a diagonal weight matrix is used.

The correlations and variances of the estimated genetic variance matrix were extracted for the three approaches, one-stage, two-stage with a full weight matrix, and two-stage using a diagonal weight matrix. Because of the smaller size of this set of METs, the scatterplot matrix of the estimated correlations and variances (chosen because they are of the same order) are presented in a single figure, rather than splitting into two as for the replicated MET. Figure 7 shows that using the full weight matrix results in a very high Pearson correlation of the components of the estimated genetic variance matrix with those found using

the one-stage method. The diagonal weight matrix option is also very good.

Figure 8 is a plot of the estimated non-genetic variance parameters for the one-stage and the stage 1 analysis with fixed or random plus de-regression genetic effects. The correspondence is good even though the larger effects under the full model appear to be under-estimated by the two-stage approach.

Simulation study: multi-site p-rep

Table 6 presents a summary of results from simulations described in the Materials sub-section. The setup is given in Table 3 and (1). One hundred simulations of a four site MET with three levels of partial replication were examined.

The mean square error of prediction MSEP (2) was calculated for each simulation for each scenario, and then the mean and standard deviation of the 100 values calculated. Note these values were multiplied by

Fig. 4 Barley Multi-environment Trial, comparison of estimated genetic variances
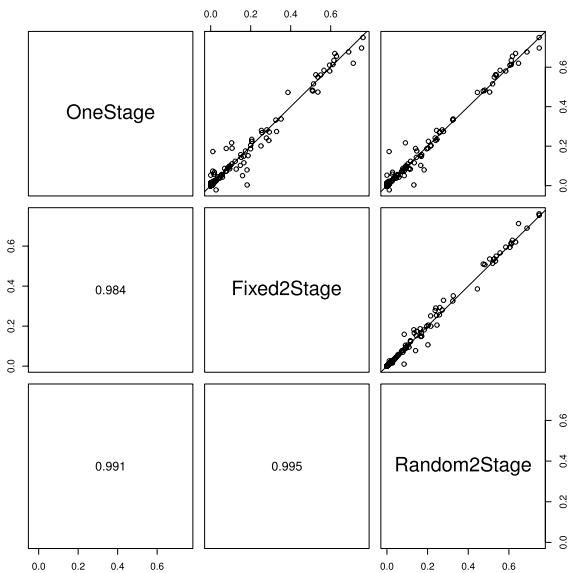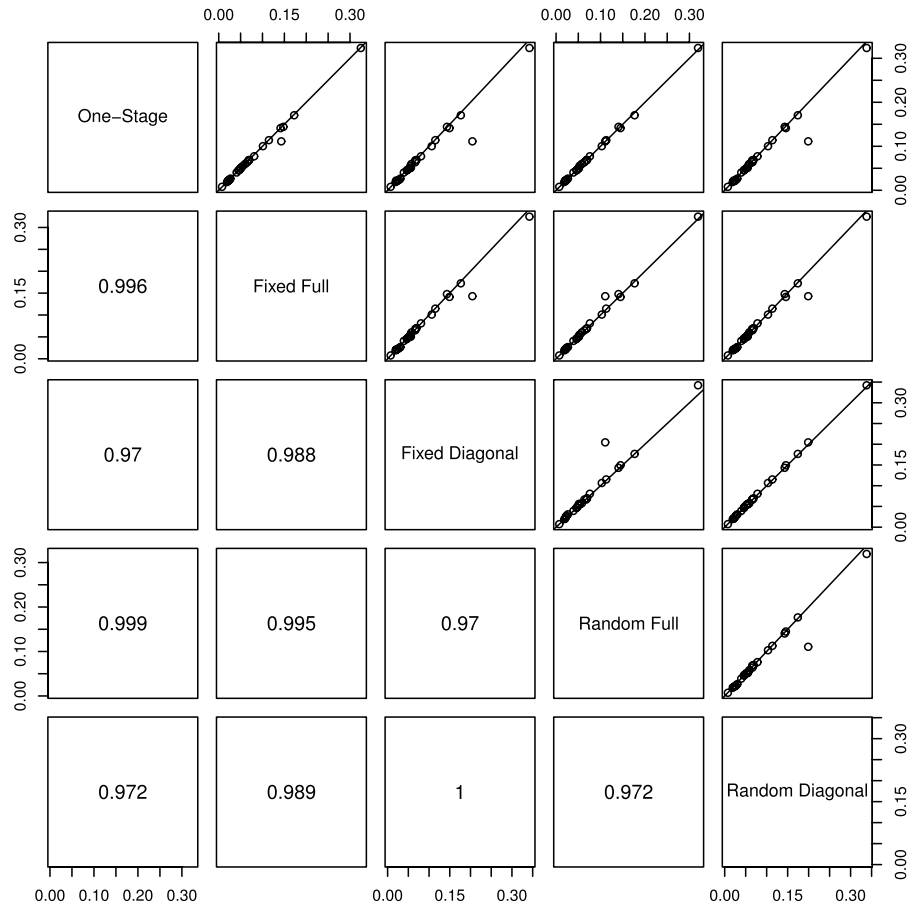

Fig. 5 Barley Multi-environment Trial, comparison of estimated non-genetic variance and correlation parameters

1920, the number of effects, for scaling the results. The values are presented in Table 6. For each simulation, the five methods were ranked according the MSEP and the counts of rankings in each class (1, smallest, to 5, largest) were accumulated over the 100 simulations for each method. These are also presented in Table 6.

For a partial replication of 0.5 or 50%, the mean MSEP and standard deviation of the MSEP are similar for one-stage, and two-stage with fixed or random plus de-regression genetic effects at stage 1. There is an ordering but the differences are small. However, the rankings of the methods show the one-stage and two-stage plus de-regression random genetic effects methods are generally better. Using diagonal weights is poor whether fixed or random plus de-regression genetic effects at stage 1 are used, but random seems better.

The change as the partial replication decreases can be seen for 0.2 and 0.1. For 0.2 partial replication,

Fig. 6 For trial site 2, a scatterplot matrix for pairs of all predicted genetic line effects across five different models. The models were a one-stage, two stage fixed or random plus de-regression genetic effects at stage one, with full weight matrix and diagonal weights at stage two. The lower panels show the Pearson correlations between the predicted pairs of effects
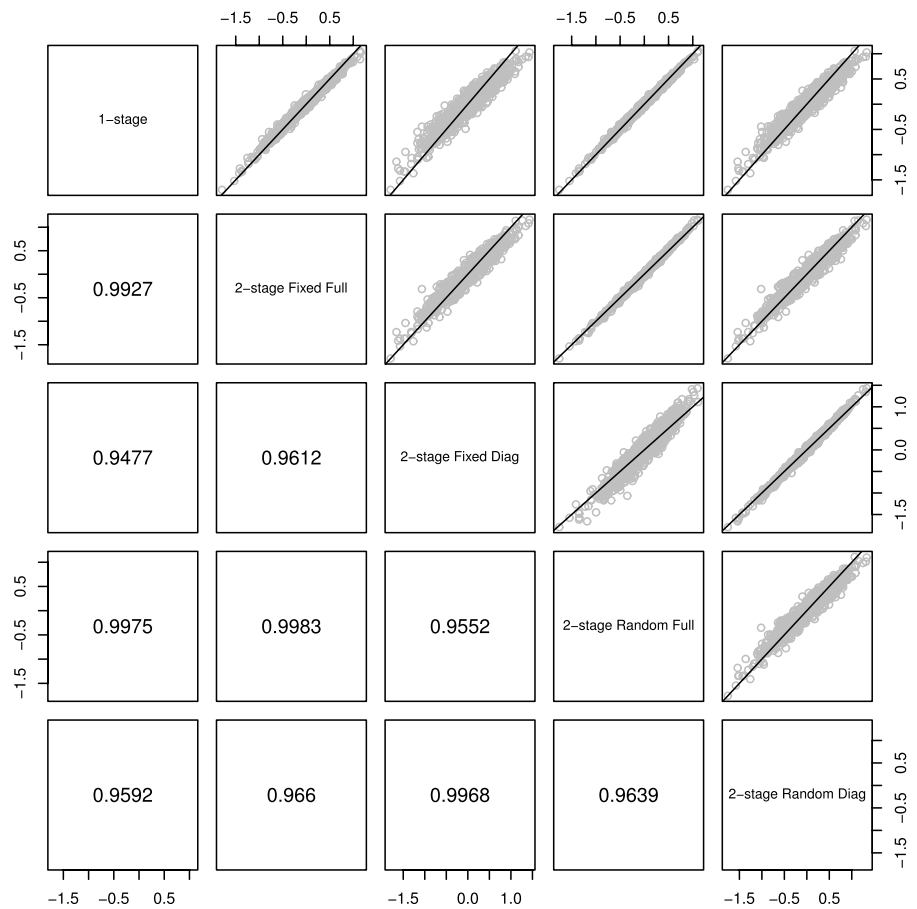
the mean and standard deviation of MSEP are very similar for one-stage and two-stage with random plus de-regression genetics effects at stage 1. The rankings favour one-stage but there are cases where random plus de-regression two stage is ranked higher. Only once does starting with fixed genetic effects at stage 1 result in the lowest MSEP. Again diagonal weights are poor, with random plus de-regression rather than fixed genetic effects in the stage 1 analysis heavily favoured.

For 0.1 or 10% partial replication the results are stronger and again random plus de-regression at stage 1 in a two-stage analysis compares favourably with a one-stage approach. Note that the model for the fixed genetic effects at stage 1 had to be modified to ensure analysis of simulations did not fail by removing the ar1 by ar1 spatial model for the residuals. The anomaly for diagonal weights where fixed appears better than random plus de-regression may be a function of the different models fitted.

The clear message from the simulations is that for p-rep designed METs, if a two-stage approach is to be used, a random plus de-regression genetic effects formulation at stage 1 should be used, and furthermore a full weight matrix should be used in the second stage analysis. The small simulation study shows the results will then be similar to those of a one-stage analysis.

Discussion and conclusions

This paper examined one-stage versus two-stage analysis of multi-environment trials. These methods were examined because there are situations when one-stage analysis is not feasible (computational requirements or time taken to fit the model is prohibitive) or feasible but with a compromised model.

In a one-stage analysis the model, genetic effects are random. In a two-stage analysis, fixed or random genetic effects can be used at the first stage of

Fig. 7 p-rep Multi-environment Trial, comparison of estimated genetic variance matrix parameters
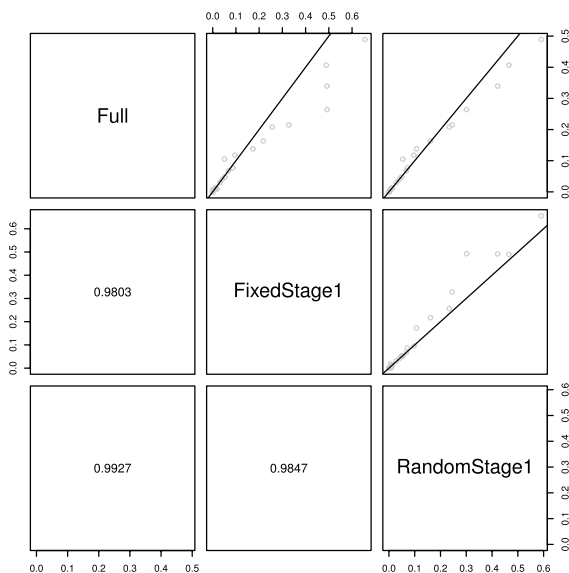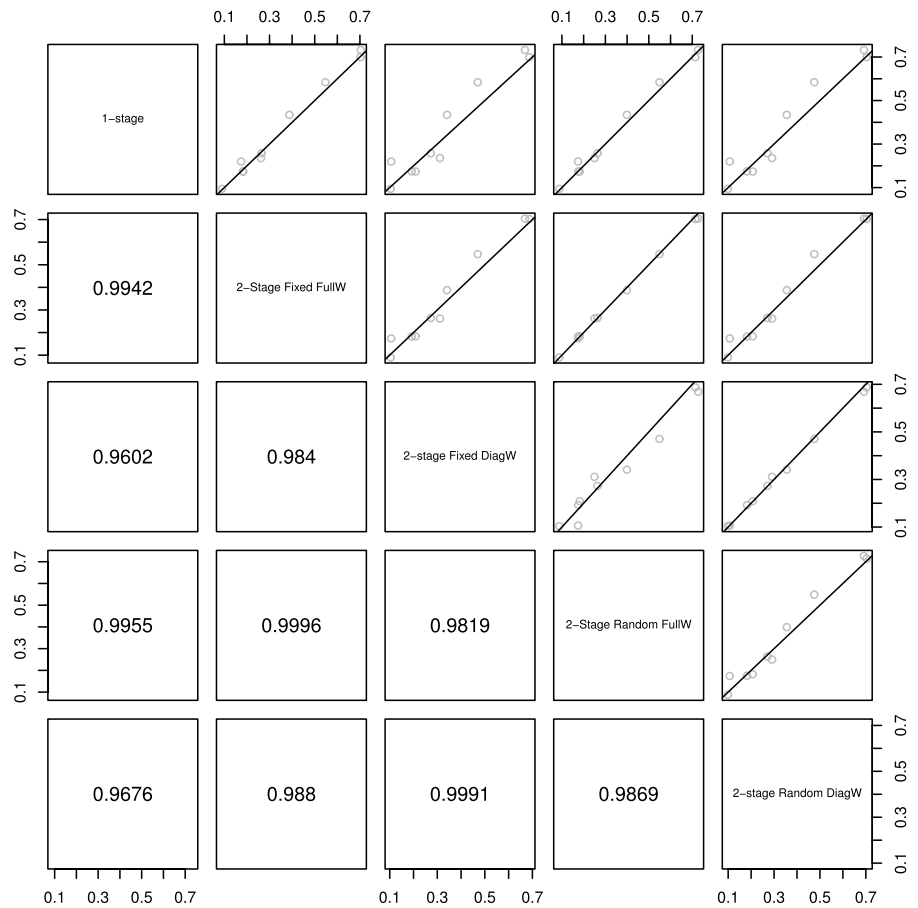

Fig. 8 p-rep Multi-environment Trial, comparison of estimated non-genetic variance and correlation parameters

analysis. Using fixed genetic effects results in unbiased estimated variety means, and together with a weight matrix, these means can be used at the second stage of analysis. If random genetic effects are used at the first stage, the predictions are biased, but de-regression leads again to unbiased variety means for the second stage and effectively fixed effects estimates, and again these estimates can be used at the second stage with a weight matrix. It might be expected that non-genetic effects that are estimated at the first stage, might be better estimated if genetic effects are initially random, and in the examples this appears to be the case. It might be expected that predicted genetic effects at the second stage would also be better. In the examples, this is also true, but the predicted genetic effects did not differ greatly if the full weight matrix was used. Diagonal weights were not as good.

With partial replication, fitting fixed genetic effects might be problematic with low replication. In animal

Table 6 Comparison of five methods (Method) of analysis, one-stage, two-stage with fixed or random plus de-regression genetic effects at stage 1, and the latter two with a full or diagonal weight matrix at the second stage. Three levels of partial replication (p-rep), 0.5, 0.2 and 0.1. For 0.1, using fixed effects at stage 1 required a simpler model to be fitted (denoted by *).

Mean square error of prediction (MSEP) times 1920 was calculated for each of 100 simulations. The mean and standard deviation (Std dev) calculated for all methods. The counts of the rank (Ranking) from smallest (1) to largest (5) MSEP of each of the 100 simulations for all methods is presented

P-rep	MSEP			Ranking				
	Method	Mean	Std dev	1	2	3	4	5
0.5	One-stage	164.79	6.48	64	24	12	0	0
	Two-stage random	165.89	6.57	27	61	12	0	0
	Two-stage fixed	167.81	7.32	9	15	76	0	0
	Two-stage random/diag	209.39	10.87	0	0	0	80	20
	Two-stage fixed/diag	214.73	12.99	0	0	0	20	80
0.2	One-stage	169.80	8.88	66	34	1	0	0
	Two-stage random	170.15	8.90	34	65	1	0	0
	Two-stage fixed	206.79	30.54	0	1	99	0	0
	Two-stage random/diag	311.39	23.66	0	0	0	97	3
	Two-stage fixed/diag	386.89	46.28	0	0	0	3	97
0.1	One-stage	213.41	11.46	67	33	0	0	0
	Two-stage random	214.73	10.84	33	67	0	0	0
	Two-stage fixed*	297.40	36.26	0	0	99	1	0
	Two-stage random/diag	469.51	45.53	0	0	0	22	78
	Two-stage fixed/diag*	429.92	87.49	0	0	1	77	22

breeding, genetic effects must be fitted as random, because there is no replication. A small simulation study showed that using fixed genetic effects at stage 1 is not as good as random plus de-regression genetic effects, with model fitting becoming problematic for low replication (10%) with fixed genetic effects, even with good design. One and two stage analyses with random genetic effects plus de-regression were very close in mean square error of prediction and ranking across methods showed these approaches were superior to using fixed genetic effects. Diagonal weights were very poor.

In animal breeding it is necessary to use a pedigree based or genomic relationship matrix to ensure a useful analysis is possible. Partially replicated trials would benefit with inclusion of a relationship matrix both from a design and analysis perspective (Cullis et al. 2020).

The conclusion from this paper is that if a two-stage analysis needs to be conducted, random genetic effects should be used at the first stage, then de-regressed, and subsequently a full weight matrix used in the second stage of analysis. It may be necessary in large data problems to use diagonal weights, but in some circumstances these weights will be far from optimal.

Data availability The data for the examples is available from the author.

Code availability The full code for analyses used in the R environment is available from the author.

Declarations

Conflict of interest The author has no non-financial interests to declare.

References

Buntaran H, Piepho HP, Schmidt P et al (2019) Cross-validation of stagewise mixed-model analysis of Swedish variety trials with winter wheat and spring barley. Crop Sci 60:2221–2240

Butler DG, Cullis BR, Gilmour AR, et al (2018) ASReml-R Reference Manual Version 4. University of Wollongong

Butler D, Cullis B. (2022) On model Based Design of Comparative Experiments In R. Working paper 09/22, NIASRA, The University of Wollongong.

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381–393

Cullis BR, Smith AB, Cocks NA et al (2020) The design of early-stage plant breeding trials using genetic relatedness. J Agric Biol Environ Stat 25:553–578

Damesa TM, Möhring J, Worku M et al (2017) One step at a time: Stage-wise analysis of a series of experiments. Agron J 109:845–857

Damesa TM, Hartung J, Gowda M et al (2019) Comparison of weighted and unweighted stage-wise analysis for genome-wide association studies and genomic selection. Crop Sci 59:2572–2584

Endelman JB (2023) Fully efficient, two-stage analysis of multi-environment trials with directional dominance and multi-trait genomic selection. Theor Appl Genet 136:65

Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol 41:55

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2:269–293

Gogel B, Smith A, Cullis B (2018) Comparison of a one- and two-stage mixed model analysis of Australia's national variety trial southern region wheat data. Euphytica 214:44

Mohring J, Piepho HP (2009) Comparison of weighting in two-stage analysis of plant breeding trials. Crop Sci 49:1977–1988

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58:545–554

Piepho HP, Mohring J, Schulz-Streeck T et al (2012) A stage-wise approach for the analysis of multi-environment trials. Biom J 58:844–860

Piepho HP, Williams ER, Michel V (2021) Generating row-column field experimental designs with good neighbour balance and even distribution of treatment replications. J Agron Crop Sci 207:745–753

Smith AB, Cullis BR, Gilmour AR (2001) The analysis of crop variety evaluation data in Australia. Aust N Z J Stat 43:129–145

Smith AB, Cullis BR, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147

Welham SJ, Gogel BJ, Smith AB et al (2010) A comparison of analysis methods for late-stage variety evaluation trials. Aust N Z J Stat 52:125–149