The Journal of Agricultural Science

cambridge.org/ags

Allowing for the structure of a designed experiment when estimating and testing trait correlations

Hans-Peter Piepho

Biostatistics Unit, University of Hohenheim, 70593 Stuttgart, Germany

Abstract

Crop scientists occasionally compute sample correlations between traits based on observed data from designed experiments, and this is often accompanied by significance tests of the null hypothesis that traits are uncorrelated. This simple approach does not account for effects due to the randomization layout and treatment structure of the experiments and hence statistical inference based on standard procedures is not appropriate. The present paper describes how valid inferences accounting for all relevant effects can be obtained using bivariate mixed linear model procedures. A salient feature of the approach is that the bivariate model is commensurate with the model used for univariate analysis of individual traits and allows bivariate correlations to be computed at the level of effects. Heterogeneity of correlations between effects can be assessed by likelihood ratio tests or by graphical inspection of bivariate scatter plots of effect estimates. if heterogeneity is found to be substantial, it is crucial to focus on the correlation of effects, and usually, the main interest will be in the treatment effects. If heterogeneity is judged to be negligible, the marginal correlation can be estimated from the bivariate model for an overall assessment of association. The proposed methods are illustrated using four examples. Hints are given to alternative routes of analysis accounting for all treatment and design effects such as regression with groups and analysis of covariance.

Crops and Soils Research Paper

Cite this article: Piepho H-P (2018). Allowing for the structure of a designed experiment when estimating and testing trait correlations. *The Journal of Agricultural Science* **156**, 59–70. https://doi.org/10.1017/S0021859618000059

Received: 16 May 2017
Revised: 20 December 2017
Accepted: 24 January 2018
First published online: 22 February 2018

Key words:
Analysis of variance; bivariate mixed model; block effects; designed experiment; likelihood ratio test; residual maximum likelihood; treatment effects; variance components

Author for correspondence:
Hans-Peter Piepho, E-mail:
hans-peter.piepho@uni-hohenheim.de

Introduction

Papers based on designed experiments appearing in agronomy and crop science journals occasionally report bivariate correlations among several traits (response variables) computed from the observed raw data. Almost invariably, these bivariate correlations are estimated and tested for significance, as Pearson's product–moment correlations, without any regard for the randomization layout or treatment structure of the experimental design. This practice is in contrast to the univariate analysis of variance (ANOVA) usually presented for the individual traits, which decomposes total variance into sources of variation due to treatments and the randomization layout. So, univariate and bivariate analyses presented in many research papers are not consistent with one another, and arguably the bivariate analyses presented violate the assumptions made in univariate ANOVA.

To illustrate the problem, consider data from eight apple trees for each of 13 different rootstocks (Andrews & Herzberg 1985, p. 357–360) (Example 1). Two of the traits measured were trunk girth at 4 years (mm × 100) and weight of tree above ground at 15 years (lb × 1000). The one-way ANOVA for both traits using rootstock as the treatment factor is shown in Table 1. The bivariate raw data are plotted in Fig. 1. With data from this kind of experiment, many research scientists would not hesitate to compute and test pairwise sample correlations r between traits from the observed raw data. A pairwise sample correlation (r) can be regarded as a descriptive statistic and as such is often worth reporting. Here, the sample correlation between trunk girth and tree weight is $r = 0.7490$ (for the benefit of readers wishing to reproduce these results using their software, results in the current tutorial paper are printed with more digits than would normally be required). When trying to interpret and further investigate this correlation, however, or testing its significance, it becomes necessary to account for the effects involved in the two traits being correlated and make meaningful distributional assumptions accounting for the presence of these effects. This all boils down to thinking carefully about a suitable bivariate model for the two traits (response variables) that are to be correlated. The usual test of significance, for example, assumes that pairs of observations on the two traits are independent, but in the case at hand pairs of data from the same rootstock are likely to be more similar than pairs from different rootstocks because they share effects of the same rootstock, thus violating the assumptions of the usual test. Note that in a univariate ANOVA, comparing the rootstocks (Table 1) requires fitting a treatment effect for rootstocks, so such an effect should also be accounted for in any bivariate analysis.

© Cambridge University Press 2018

CAMBRIDGE UNIVERSITY PRESS

Table 1. Univariate analyses of variance for trunk girth at 4 years (mm × 100) and tree weight at 15 years (lb × 1000) of six rootstocks of apple (*Malus domestica* L.) each tested on eight trees (Example 1)

		Trunk girth (mm × 100)			Tree weight (lb × 1000)		
M	Degrees of freedom	Mean square	*F*-value	*P*-value	Mean square	*F*-value	*P*-value
Rootstocks	12	0.1062	16.44	<0.001	1.3289	22.05	<0.001
Error	91	0.006462			0.0603		

In the present paper, it will be demonstrated that it is possible and useful to reconcile univariate and bivariate analyses of traits of interest and perform them in a consistent manner using mixed model procedures (Schabenberger & Pierce 2002; Littell *et al.* 2006). It will be shown how to account for treatment effects in assessing the correlation among response variables. In the next section, a bivariate model is introduced that is commensurate with the linear model used for univariate analyses (Hill & Thompson 1978; Singh & El-Bizri 1992; Booth *et al.* 2009). Based on this model, estimation of correlations and associated significance tests are described. The methods are illustrated using the rootstock example (Example 1). Three further examples serve to show a range of settings, where simple correlation might be contemplated initially, but closer investigation leads to alternative analyses, providing more insights.

Materials and methods

This section considers linear models for univariate and bivariate analysis. Several equations are given to explain how analysis of simple correlation is different from a bivariate analysis accounting for treatment and design effects and why the latter is preferable. In the development, it is assumed for simplicity that a single treatment factor has been tested according to a completely randomized design (CRD). Extension to blocked experiments is straightforward and will be illustrated using further examples.

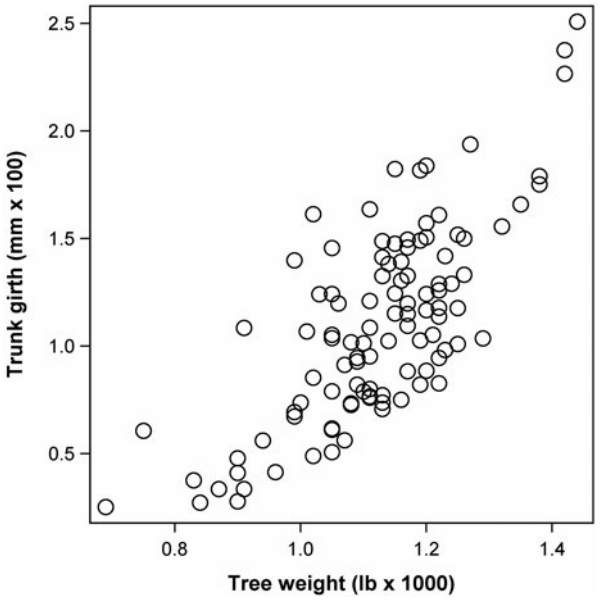


Fig. 1. Scatter plot of trunk girth at 4 years (mm × 100) *v.* tree weight at 15 years (lb × 1000) for apple (*Malus domestica* L.) rootstock data from an experiment with 13 rootstocks (Example 1).

Defining a bivariate model

The linear model underlying a univariate one-way ANOVA for an individual trait is:

$$y_{ij} = \mu + \tau_i + e_{ij} \quad (1)$$

where y_{ij} is the response of the jth replicate of the ith treatment, μ is a general intercept, τ_i is the effect of the ith treatment, and e_{ij} is the residual error term associated with the response y_{ij}. All effects are assumed to be fixed, except the error term e_{ij}, which is assumed to follow a normal distribution with zero mean and constant variance σ_e^2, i.e.

$$e_{ij} \sim N\left(0, \sigma_e^2\right) \quad (2)$$

where $N\left(0, \sigma_e^2\right)$ represents a normal distribution, the first argument denoting the mean (0 in this case) and the second argument denoting the variance (σ_e^2 in this case). The expected value, or mean, of the response y_{ij} for a given treatment i is $E(y_{ij}) = \mu + \tau_i$. Thus, the response has an assumed normal distribution with mean $\mu + \tau_i$, comprised the fixed effects of the model, and variance σ_e^2, which in short-hand notation can be written as

$$y_{ij} \sim N\left(\mu + \tau_i, \sigma_e^2\right) \quad (3)$$

The natural starting point for bivariate analysis is the model used for univariate ANOVA. It seems reasonable to require that the model holding for univariate ANOVA of each individual trait is also implied by any bivariate model contemplated for the joint analysis of two traits. In other words, the bivariate model should be a natural extension of the univariate model. Thus, model (1) may be assumed to hold for both traits simultaneously. All the effects will need to be trait-specific, of course, so all effects are indexed by traits. For two traits 1 and 2, the two univariate models then are written as:

$$y_{ij1} = \mu_1 + \tau_{i1} + e_{ij1} \quad (4)$$

and

$$y_{ij2} = \mu_2 + \tau_{i2} + e_{ij2} \quad (5)$$

The errors are assumed to have zero mean and trait-specific variances:

$$\text{var}(e_{ij1}) = \sigma_{e(1)}^2 \quad (6)$$

and

$$\text{var}(e_{ij2}) = \sigma_{e(2)}^2 \quad (7)$$

In addition, now that a bivariate analysis is considered, the correlation among traits also needs to be accounted for. Errors

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

for observations on the two traits made on the same plot can be assumed to be correlated with covariance:

$$\text{cov}(e_{ij1}, e_{ij2}) = \sigma_{e(1,2)} \tag{8}$$

The correlation is defined as:

$$\text{corr}(e_{ij1}, e_{ij2}) = \rho_{e(1,2)} = \frac{\sigma_{e(1,2)}}{\sigma_{e(1)}\sigma_{e(2)}} \tag{9}$$

For bivariate analysis it is convenient to collect effects of the same type into vectors of length two, allowing Eqns (4) and (5) to be written more compactly as

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \end{pmatrix} + \begin{pmatrix} e_{ij1} \\ e_{ij2} \end{pmatrix} \tag{10}$$

A short-hand notation for the distributional assumption for the errors then is

$$\begin{pmatrix} e_{ij1} \\ e_{ij2} \end{pmatrix} \sim BVN\left[\varphi = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2_{e(1)} & \sigma_{e(1,2)} \\ \sigma_{e(1,2)} & \sigma^2_{e(2)} \end{pmatrix}\right] \tag{11}$$

where $BVN(\varphi, \Sigma)$ represents a bivariate normal distribution, with the first argument (φ) denoting the mean vector (here, means for both e_{ij1} and e_{ij2} are zero) and the second (Σ) the variance-covariance matrix with the two variances $\sigma^2_{e(1)}$ and $\sigma^2_{e(2)}$ on the diagonal and the covariance $\sigma_{e(1,2)}$ on the off-diagonal. With this assumption, the distributional model for the observed data can be stated as:

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix} \sim BVN\left[\begin{pmatrix} \mu_1 + \tau_{i1} \\ \mu_2 + \tau_{i2} \end{pmatrix}, \begin{pmatrix} \sigma^2_{e(1)} & \sigma_{e(1,2)} \\ \sigma_{e(1,2)} & \sigma^2_{e(2)} \end{pmatrix}\right] \tag{12}$$

It is important to note that the bivariate distributional model (12) is consistent with the assumption that the univariate distributional model (3) holds for each of the two traits because both comprise the same set of parameters (effects and variances).

It may now be asked: what is the implicit assumption for the distribution of the bivariate observations (y_{ij1}, y_{ij2}), if simple sample correlations r are computed from the observed data and subjected to a significance test? The answer is that we are effectively assuming that the paired observations are a simple random sample from the same (!) population, implying this bivariate model:

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix} \sim BVN\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2_{e(1)} & \sigma_{e(1,2)} \\ \sigma_{e(1,2)} & \sigma^2_{e(2)} \end{pmatrix}\right] \tag{13}$$

In particular, this is the assumption needed for the classical t-test of the null hypothesis that the two traits are uncorrelated, based on the test statistic $t = r\sqrt{n-2}/\sqrt{1-r^2}$ where n is the sample size, or the equivalent F-statistic $F = t^2$ (Mead *et al.* 1993). So, when simple correlations are computed and tested, the treatment effects implicitly drop out of the model! From the development of the bivariate model (12) as a natural extension of the univariate model (1), however, it is clear that the simple model (13) violates the distributional assumption made about each individual trait in a univariate ANOVA. Specifically, under model (12), each observation vector (y_{ij1}, y_{ij2}) has its own expected value ($\mu_1 + \tau_{i1}$, $\mu_2 + \tau_{i2}$), as opposed to (13), where all observation vectors have the same expectation (μ_1, μ_2). For this reason, model (13) is not tenable for data from a designed experiment, yet many researchers readily assume that model, presumably often without realizing, when computing pairwise correlations from plot data and performing significance tests on them. Subsequently, a method to estimate bivariate correlations will be outlined that takes treatment effects into account. Before doing so, the related problem of assessing the importance of the variability of treatment effects for single traits will be considered. This is done to motivate the approach proposed for bivariate analysis.

Defining and estimating variance and covariance for treatment effects

When computing a marginal correlation for the bivariate observation vector (y_{ij1}, y_{ij2}) in a designed experiment, the effects of the model must be taken into account. Rather than thinking directly in terms of a correlation of the observed data, one may consider correlations at the level of each of the effects in the model. Here is where the bivariate formulation is given in Eqn (10) comes in handy. For a CRD, there are two bivariate vectors of effects that vary across the data, i.e. treatment effects (τ_{i1}, τ_{i2}) and residual errors (e_{ij1}, e_{ij2}). The important point here is that a bivariate correlation for each of these bivariate effects should be expected.

As a prelude to defining a general procedure for estimating bivariate correlations in designed experiments, taking all effects into account, this section considers the problem of estimating a variance of treatment effects for a single trait and the related problem of estimating a covariance of treatment effects for two traits. It will be argued that it is both reasonable and convenient for this particular purpose to consider treatment effects as random and estimate variances and covariances using residual maximum likelihood (REML) (Patterson & Thompson 1971; Searle *et al.* 1992), even in situations where the treatment effect would normally be regarded as fixed. To justify the approach, a variance estimator for a balanced one-way layout as in Example 1 will first be introduced under the assumption of fixed treatment effect. Next, it will be stressed that this estimator coincides with that obtained under a random-effects model. Subsequently, the equivalence of estimators under the fixed and random assumptions will also be demonstrated for the covariance.

The one-way ANOVA has two sources of variation, corresponding to the two effects in the model, i.e. treatments τ_i and error e_{ij}. Table 2 shows the sums of squares of a one-way ANOVA with t treatments and r replications per treatment. It may be of interest to quantify the importance of treatment effects in relation to the total variation. A possible measure is the 'sample variance', or finite-population variance, of the fixed treatment effects, given by:

$$Q(\tau) = \frac{\sum_{i=1}^{t}(\tau_i - \bar{\tau}_\bullet)^2}{t-1} \tag{14}$$

where $\bar{\tau}_\bullet$ denotes the arithmetic mean of the treatment effects τ_i.

This is a natural measure to quantify the variability of the fixed treatment effects (Winer *et al.* 1991; Gelman 2005). $Q(\tau)$ can be estimated from the two ANOVA mean squares as

$$\hat{Q}(\tau) = \frac{MS_{treat} - MS_{error}}{r} \tag{15}$$

where $MS_{treat} = SS_{treat}/(t-1)$, $MS_{error} = SS_{error}/[t(r-1)]$, with treatment and error sums-of-squares a given in Table 3. The error variance is estimated by $\hat{\sigma}^2_e = MS_{error}$.

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

Table 2. Variance parameter estimates for the apple (*Malus domestica* L.) rootstock data in Table 1 (Example 1) under the hypotheses $H_A: \rho_{y(1,2)} \neq 0$

Parameter	Effect	Estimate	S.E.
$\sigma^2_{\tau(1)}$	Treatment	0.01247	0.005422
$\sigma^2_{\tau(2)}$	Treatment	0.1586	0.06781
$\rho_{\tau(1,2)}$	Treatment	0.8908	0.06978
$\sigma^2_{e(1)}$	Error	0.006462	0.000958
$\sigma^2_{e(2)}$	Error	0.06026	0.008933
$\rho_{e(1,2)}$	Error	0.4572	0.08292
$\sigma^2_{y(1)}$ [a]	Total	0.01893	
$\sigma^2_{y(2)}$ [a]	Total	0.2188	
$\sigma_{y(1,2)}$ [a]	Total	0.04863	
$\rho_{y(1,2)}$ [a]	Total	0.7556	

Trait 1 = trunk girth at 4 years (mm × 100), trait 2 = tree weight at 15 years (lb × 1000).
[a]These values were read off the output of the statistical package and are slightly different from the values one would obtain by plugging the estimates of variances and covariances for the effects shown in this table, which are rounded to the forth decimal place, into Eqns (26)–(28) for the marginal variances, covariance and correlation.

So far, the treatment effect has been regarded as fixed. Since the interest here is in quantifying variability, one may alternatively consider the treatment effect as random and define a variance of treatment effects, $\sigma^2_\tau = \text{var}(\tau_i)$. Using the expected sums of squares in Table 2, the treatment variance σ^2_τ can be estimated from the two ANOVA mean squares as:

$$\hat{\sigma}^2_\tau = \frac{MS_{treat} - MS_{error}}{r} \tag{16}$$

This is the ANOVA or methods-of-moments estimator of the treatment variance σ^2_τ (Winer *et al.* 1991, p. 92; Searle *et al.* 1992, p. 59). Assuming normality of the treatment effects and non-negativity of the variance estimator, (16) also coincides with the REML estimator of σ^2_τ (Searle *et al.* 1992, p. 92). The important point to make here is that the estimator of $Q(\tau)$ in (15), assuming fixed treatment effects, is identical to the estimator of σ^2_τ in (16), assuming random effects, i.e. $\hat{Q}(\tau) = \hat{\sigma}^2_\tau$.

The same idea can be applied to define and estimate a covariance between treatment effects. Just as one can define a 'sample variance' $Q(\tau)$ of fixed treatment effects for a single trait, one can also define a 'sample covariance' of fixed treatment effects for two traits as

$$Q(\tau_1, \tau_2) = \frac{\sum_{i=1}^{t} (\tau_{i1} - \bar{\tau}_{\bullet 1})(\tau_{i2} - \bar{\tau}_{\bullet 2})}{t-1} \tag{17}$$

This covariance can be estimated by the method of moments using two sums of cross-products:

$$SP_{treat} = r \sum_{i=1}^{t} (\bar{y}_{i\bullet 1} - \bar{y}_{\bullet\bullet 1})(\bar{y}_{i\bullet 2} - \bar{y}_{\bullet\bullet 2}) \tag{18}$$

and

$$SP_{error} = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij1} - \bar{y}_{i\bullet 1})(y_{ij2} - \bar{y}_{i\bullet 2}) \tag{19}$$

The method-of-moments estimator for (17) and the covariance of errors, $\sigma_{e(1,2)}$, are:

$$\hat{Q}(\tau_1, \tau_2) = \frac{MP_{treat} - MP_{error}}{r} \tag{20}$$

and

$$\hat{\sigma}_{e(1,2)} = MP_{error} \tag{21}$$

where $MP_{treat} = SP_{treat}/(t-1)$ and $MP_{error} = SP_{error}/[t(r-1)]$. When treatments are taken as random and there is a covariance between random treatment effects, $\sigma_{\tau(1,2)} = \text{cov}(\tau_{i1}, \tau_{i2})$, the same estimators are obtained, i.e. $\hat{\sigma}_{\tau(1,2)} = \hat{Q}(\hat{\tau}_1, \hat{\tau}_2)$ and $\hat{\sigma}_{e(1,2)} = MP_{error}$, and these estimators coincide with the REML estimators based on a model with random treatment effects (see the following subsection), provided the variance estimates obtained from (15) are non-negative (Searle *et al.* 1992; Thimm 2002, p. 378). It is noted that sums of cross-products can be obtained conveniently from ANOVA means squares for the two traits and their sum or difference (Searle *et al.* 1992; Piepho

Table 3. One-way ANOVA table with expected sums of squares (E(SS)) and mean squares for models with fixed and random treatment effects

			E(SS)	
Source	Degrees of freedom[a]	Sum of squares (SS)	τ_i fixed[b]	τ_i random[c]
Treatments	$t-1$	$SS_{Treat} = r\sum_{i=1}^{t}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	$(t-1)[\sigma^2_e + rQ(\tau)]$	$(t-1)(\sigma^2_e + r\sigma^2_\tau)$
Error	$t(r-1)$	$SS_{Error} = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \bar{y}_{i\bullet})^2$	$t(r-1)\sigma^2_e$	$t(r-1)\sigma^2_e$
Corrected total	$rt-1$	$SS_{Total} = SS_{Treat} + SS_{Error} = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \bar{y}_{\bullet\bullet})^2$	$(rt-1)\sigma^2_e + (t-1)rQ(\tau)$	$(rt-1)\sigma^2_e + (t-1)r\sigma^2_\tau$

[a]t = number of treatments; r = number of replications per treatment.
[b]$Q(\tau) = \frac{\sum_{i=1}^{t}(\tau_i - \bar{\tau}_\bullet)^2}{t-1}$; see, e.g. Winer *et al.* (1991, p. 85).
[c]See, e.g., Winer *et al.* (1991, p. 92) or Searle *et al.* (1992, p. 60).

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

et al. 2014). For example,

$$SP_{error} = \frac{[SS_{error}(y_1 + y_2) - SS_{error}(y_1) - SS_{error}(y_2)]}{2} \quad (22)$$

where $SS_{error}\,(y_1)$, $SS_{error}\,(y_2)$ and $SS_{error}\,(y_1 + y_2)$ denote the error sum-of-squares for the two traits and their sum, respectively.

The practical upshot of all this is that the same estimator of the variance is obtained as well as the covariance of treatment effects, no matter whether treatment effects are modelled as fixed or random. This suggests that for the purpose of estimating covariance and correlation for treatment effects, treatment effects can be taken as random and fit a bivariate model even in situations where the treatment effect would normally be taken as fixed.

The development here is for the balanced case. Residual maximum likelihood is generally applicable also to unbalanced data, and one of its advantages over an ANOVA approach is the uniqueness of the estimators (Searle *et al.* 1992). Hence, it is desirable to generally use REML to implement the proposed procedures, which involves the explicit assumption of random effects, which will be examined below.

Decomposing the correlation using random effects

In the statement of model (1), the treatment effects were taken as fixed, and there are usually good reasons for this assumption. For example, the treatments are not usually randomly sampled from a larger population of treatments (Searle *et al.* 1992; Gelman 2005). The same can often be said about the blocks, for example in a field experiment where blocks are placed adjacent to each other and their position is generally purposely selected (Dixon 2016). Nevertheless, based on the reasoning in the previous section, it is proposed here to take effects as random in such settings for the purpose of defining and estimating covariances and correlations among effects.

There are two possible reactions for the reader at this juncture: (1) one insists on the fixed-effects assumption and concludes that correlations can only be defined by the residual errors (e_{ij1}, e_{ij2}). In this case, a study of marginal correlations at the level of the observed data (y_{ij1}, y_{ij2}), or at the level of treatment effects (τ_{i1}, τ_{i2}) becomes basically meaningless, or (2) one regards the random assumption mainly as a convenient device allowing for estimates of correlation to be obtained and tested for all effects and a marginal correlation to be defined at the level of the observed data. The following considerations are based on this pragmatic view. Before proceeding, however, it should be mentioned that an alternative route to analysis observing all effects but allowing some of them to be fixed is to consider a regression analysis that uses one variate as the response and the other as a regressor variable. This option will be touched upon in Examples 3 and 4, but the focus will remain on correlation.

Assuming that all effects are random corresponds to the assumption of bivariate normality for the vectors of error and treatment effects in model (10). For the errors (e_{ij1}, e_{ij2}), this assumption has been specified in Eqn (11). For treatments the assumption can be stated as:

$$\begin{pmatrix} \tau_{i1} \\ \tau_{i2} \end{pmatrix} \sim BVN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\tau(1)} & \sigma_{\tau(1,2)} \\ \sigma_{\tau(1,2)} & \sigma^2_{\tau(2)} \end{pmatrix}\right] \quad (23)$$

where $\sigma_{\tau(1,2)} = \mathrm{cov}(\tau_{i1}, \tau_{i2})$, $\sigma^2_{\tau(1)} = \mathrm{var}(\tau_{i1})$, and $\sigma^2_{\tau(2)} = \mathrm{var}(\tau_{i2})$. With this assumption, correlations may be defined for the treatment effects as:

$$corr(\tau_{i1}, \tau_{i2}) = \rho_{\tau(1,2)} = \frac{\sigma_{\tau(1,2)}}{\sigma_{\tau(1)}\sigma_{\tau(2)}} \quad (24)$$

Moreover, the marginal (total) correlation for the observed data (y_{ij1}, y_{ij2}) may be defined as:

$$\mathrm{corr}(y_{ij1}, y_{ij2}) = \rho_{y(1,2)} = \frac{\sigma_{y(1,2)}}{\sigma_{y(1)}\sigma_{y(2)}} \quad (25)$$

where the marginal (total) variances and the covariance of the response variables are defined as the sums of variances and covariances of the corresponding effects:

$$\sigma^2_{y(1)} = \mathrm{var}(y_{ij1}) = \sigma^2_{\tau(1)} + \sigma^2_{e(1)} \quad (26)$$

$$\sigma^2_{y(2)} = \mathrm{var}(y_{ij2}) = \sigma^2_{\tau(2)} + \sigma^2_{e(2)} \quad (27)$$

and

$$\sigma_{y(1,2)} = \mathrm{cov}(y_{ij1}, y_{ij2}) = \sigma_{\tau(1,2)} + \sigma_{e(1,2)} \quad (28)$$

The bivariate mixed model just described may be estimated by REML using a suitable mixed model package. Here, the GLIMMIX procedure of SAS is used. If the assumption of random effects is tenable, correlations may be tested for significance using likelihood-ratio tests (LRT). The random assumption will always be tenable for the error effects, but not necessarily for the other effects in the model. For this reason, the focus here is on the error effects and treatment effects taken as fixed, keeping them in the model throughout. Generally, the test statistic T is computed as the difference in -2 log L_R between the reduced model under the null hypothesis (H_0) and the full model under the alternative (H_A), where L_R is the maximized residual likelihood. Here, the H_0 is that the errors (e_{ij1}, e_{ij2}) are uncorrelated. In order to obtain the likelihood under H_0, the model under H_0 needs to be fitted to both traits simultaneously, constraining the covariance (or correlation) to be zero. Under the alternative H_A, the covariance is allowed to differ from zero. Under the null hypothesis H_0, the test statistic has an approximate chi-squared (χ^2) distribution with degrees of freedom (d.f.) equal to the number of parameters equated to zero under H_0 (Schabenberger & Pierce 2002). Thus, under the null hypothesis $H_0 : \rho_{e(1,2)} = 0$, we set $\sigma_{e(1,2)} = 0$, so the test statistic T has 1 d.f. One can also test the correlations of other effects by an LRT in the same way, provided the random assumption is tenable. Using the same general approach, one may also perform an LRT of the null hypothesis that error and treatment effects are equally correlated. If this null hypothesis is rejected, interpretation should focus on the correlations for effects. Otherwise, the marginal correlation provides a meaningful measure of association.

An attractive feature of the sample correlation coefficient r is that there is a simple link to the scatterplot of the data. When a bivariate mixed model is fitted, such plots can also be generated for each of the effects using their best linear unbiased predictors (BLUP). In case of the error effects, their estimators correspond to the residuals. The bivariate scatter plots for the effects can further be used to check that the association is approximately linear, which is a prerequisite for the bivariate model to be appropriate,

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

and to inspect the degree of heterogeneity between effects in their bivariate correlations.

Extensions of the bivariate model

In the development of the bivariate mixed model, a CRD has been assumed so the model only had a treatment effect and a residual error. It is straightforward to extend the model to allow for more complex designs and covariance structures. For example, for blocked designs, the model is augmented by adding a block effect for each trait, which has a trait-specific variance and a bivariate correlation (Booth *et al.* 2009). It is also possible to allow for heterogeneity between treatments in the variance of error effects. Both of these extensions will be illustrated with examples.

Application to four examples

This section will look at four examples. Example 1 serves to demonstrate that treatment and error effects can have rather different correlations, in which case it is essential to inspect correlations for each type of effect. Example 2 is used to demonstrate how block effects can be incorporated into the model and that block effects can have a correlation that is rather different from that of treatment effects. Example 3 shows that there may be heterogeneity between treatments in the correlation as well as the variance of residual errors and that a regression analysis may be more informative than a simple correlation analysis. In Example 4, an experiment with subsampling on the plots is considered to demonstrate that the general approach also works with more complex mixed models and that the marginal correlation can provide a meaningful summary of the data when the residual variances are large in comparison with the variances of other effects.

Example 1 (continued)

Parameter estimates for the bivariate model for the rootstock data are shown in Table 3. The marginal correlation between trunk girth at 4 years and tree weight at 15 years is estimated as $\hat{\rho}_{y(1,2)} = 0.7556$, which is rather similar, but not identical to, the sample correlation of $r = 0.7490$. These correlations agree with the visual impression from the scatter plot of tree weight against trunk girth in Fig. 1. By contrast, the correlation estimates for the treatment and error effects are $\hat{\rho}_{\tau(1,2)} = 0.8908$ and $\hat{\rho}_{e(1,2)} = 0.4572$, respectively, which are rather different in value from each other and from the marginal correlation. Notably, there is a relatively high correlation between treatments, but a low correlation for errors within treatments. Taking treatment effects as fixed, the LRT of $H_0 : \rho_{e(1,2)} = 0$ for residual errors is significant at the 5% level ($T = 21.34$, d.f. = 1, $P < 0.001$). The difference in correlations is also apparent from a bivariate plot of BLUPs of random treatment effects (Fig. 2) and of residuals (Fig. 3). These scatter plots show no indication of departure from linearity. A further LRT shows that treatment and error correlations are significantly different ($T = 7.16$, d.f. = 1, $P = 0.0074$). The fact that correlations are rather different for error and treatment effects underscores the importance of accounting for treatment effects when estimating correlations and suggests that the marginal correlation $\rho_{y(1,2)}$ is not a very meaningful summary of the data for this example. It is more informative to focus interpretation on the treatment and error effects.

The observant reader may be given to wonder why r and $\hat{\rho}_{y(1,2)}$ are different, for they seem to be estimating the same thing. The answer is that they are clearly different estimators, even if one assumes that they are estimating the same population correlation. The usual estimator r falsely assumes independence of the bivariate observations, while $\hat{\rho}_{y(1,2)}$ accounts for the correlation among bivariate observations induced by the presence of block and treatment effects. The usual estimator r is biased on account of these correlations between bivariate observations. This problem is reminiscent of, for example, the bias in the sample variance estimator in the presence of serial correlation among univariate observations (Anderson 1971, p. 448).

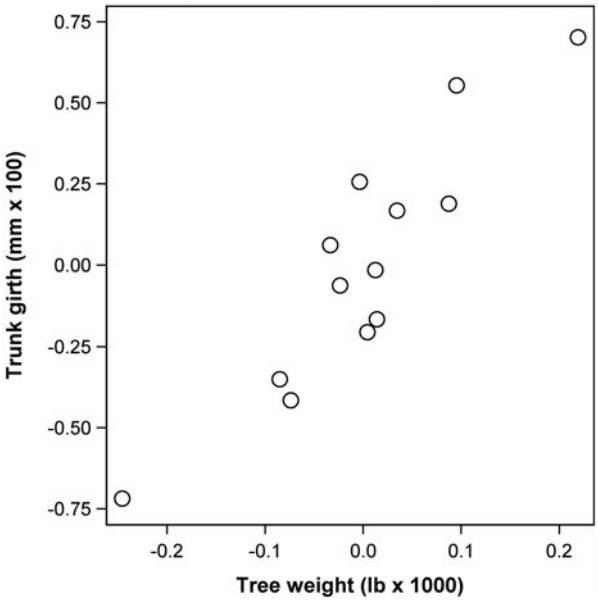


Fig. 2. Scatter plot of BLUPs of treatment effects trunk girth at 4 years (mm × 100) *v.* tree weight at 15 years (lb × 1000) for apple (*Malus domestica* L.) rootstock data from an experiment with 13 rootstocks (Example 1).

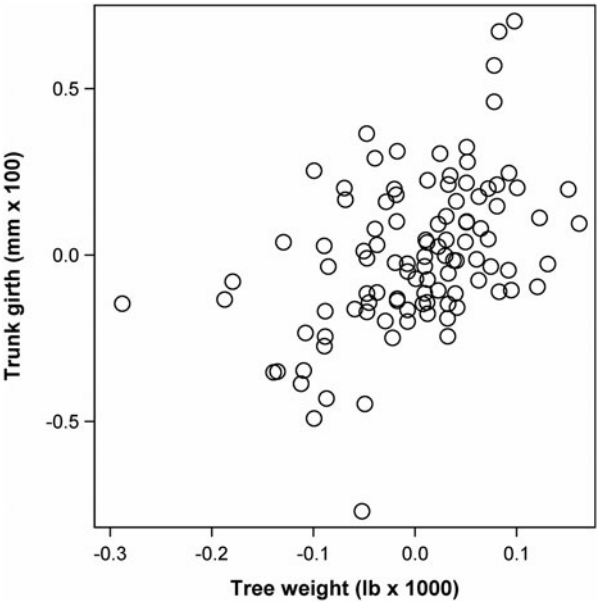


Fig. 3. Scatter plot of residuals for trunk girth at 4 years (mm × 100) *v.* tree weight at 15 years (lb × 1000) for apple (*Malus domestica* L.) rootstock data from an experiment with 13 rootstocks (Example 1).

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

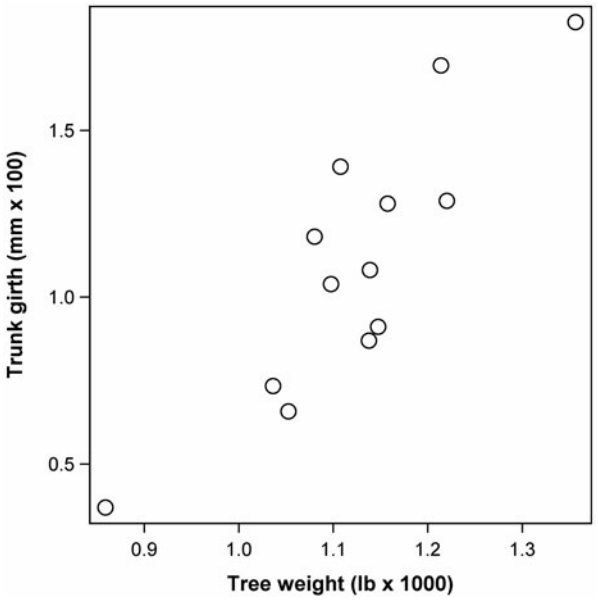


Fig. 4. Scatter plot rootstock means for trunk girth at 4 years (mm × 100) *v.* tree weight at 15 years (lb × 1000) for apple rootstock data from an experiment with 13 apple (*Malus domestica* L.) rootstocks (Example 1).

One might consider correlating observed treatment means rather than fitting a bivariate mixed model with correlated treatment effects (Fig. 4). The sample correlation of treatment means is $r = 0.8675$ ($P < 0.001$), which is lower than the estimated correlation of random treatment effects under the bivariate mixed model ($\hat{\rho}_{\tau(1,2)} = 0.8908$). The reason for this dilution is that sample means (i.e. $\bar{y}_{i\bullet} = \mu + \tau_i + \bar{e}_{i\bullet}$) are functions not only of the treatment effects but also of the treatment averages of the residual errors, which have a considerably lower correlation ($\hat{\rho}_{e(1,2)} = 0.4572$) than the treatment effects. This suggests that it is preferable to fit a bivariate mixed model and inspect the bivariate scatter plot of the BLUPs of treatment effects.

For comparison, tree weight at 15 years was jointly analysed with trunk girth at 15 (not 4!) years (Table 4). The correlations for both treatment and error effects are now rather high and also quite similar ($\hat{\rho}_{\tau(1,2)} = 0.9337$ and $\hat{\rho}_{e(1,2)} = 0.8616$). An LRT reveals no significant difference between the two correlations ($T = 2.885$, d.f. = 1, $P = 0.0894$), so the marginal correlation ($\hat{\rho}_{y(1,2)} = 0.9137$) is a meaningful measure of association that is also close to the sample correlation ($r = 0.9127$).

Example 2

A field experiment with 64 breeding lines of oats (*Avena sativa* L.) was laid out as a square lattice with three replicates (Friedrich Utz, personal communication). Two of the traits assessed were yield (tonnes/ha × 10) and plant height (cm) (Fig. 5). Three outlying observations for yield were set to missing after careful inspection of residuals (it should be stressed here that outlier removal requires great care and observations should be dropped only with good justification). The sample correlation between the two traits is $r = 0.5750$ based on $n = 189$ plots with data on both traits. The following mixed model was fitted:

$$y_{ijk} = \mu + \tau_i + r_k + b_{jk} + e_{ijk} \quad (29)$$

Table 4. Variance parameter estimates for the apple (*Malus domestica* L.) rootstock data in Table 1 (Example 1) under the hypotheses $H_A : \rho_{y(1,2)} \neq 0$

Parameter	Effect	Estimate	Standard error
$\sigma^2_{\tau(1)}$	Treatment	0.4221	0.1780
$\sigma^2_{\tau(2)}$	Treatment	0.1585	0.06779
$\rho_{\tau(1,2)}$	Treatment	0.9337	0.03944
$\sigma^2_{e(1)}$	Error	0.1128	0.01672
$\sigma^2_{e(2)}$	Error	0.06026	0.008933
$\rho_{e(1,2)}$	Error	0.8616	0.02701
$\sigma^2_{y(1)}$ [a]	Total	0.5349	
$\sigma^2_{y(2)}$ [a]	Total	0.2188	
$\sigma_{y(1,2)}$ [a]	Total	0.3126	
$\rho_{y(1,2)}$ [a]	Total	0.9137	

Trait 1 = trunk girth at 15 years (mm × 100), trait 2 = tree weight at 15 years (lb × 1000).
[a]These values were read off the output of the statistical package and are slightly different from the values one would obtain by plugging the estimates of variances and covariances for the effects shown in this table, which are rounded to the forth decimal place, into Eqns (26)–(28) for the marginal variances, covariance and correlation.

where y_{ijk} is the response of the ith genotype in the jth incomplete block nested within the kth replicate. In addition to the treatment and error effects also present in (1), model (29) has an effect for the kth replicate (r_k) and for the jth incomplete block nested within the kth replicate (b_{jk}). The bivariate model then involves correlations not only among the treatment and error effects for both traits, but also among the replicate and incomplete block effects.

The fitted variance parameters are shown in Table 5. The estimated marginal correlation of $\hat{\rho}_{y(1,2)} = 0.6000$ is slightly larger than the sample correlation. The correlation of treatment effects is rather smaller ($\hat{\rho}_{\tau(1,2)} = 0.4796$), as is the correlation of error effects ($\hat{\rho}_{e(1,2)} = 0.2795$), whereas the correlation of block effects ($\hat{\rho}_{b(1,2)} = 0.9118$) is considerably larger than both the

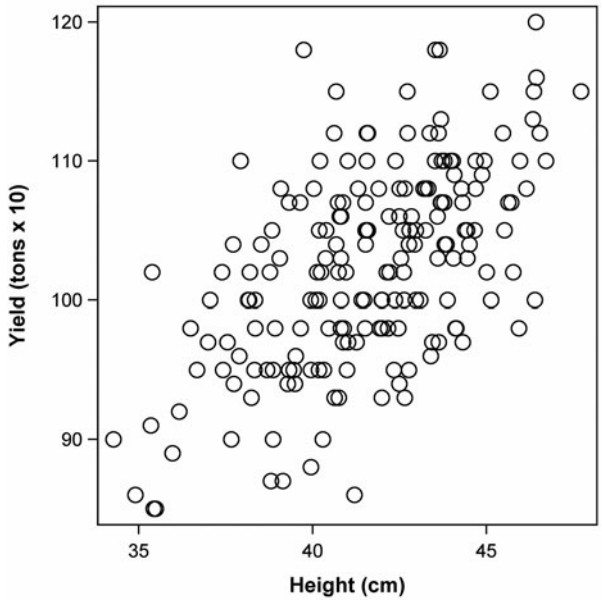


Fig. 5. Scatter plot of yield (kg/ha) *v.* height (cm) for oat (*Avena sativa* L.) data (Example 2).

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

Table 5. Variance parameter estimates for oats (*Avena sativa* L.) data (Example 2) under the hypotheses $H_A : \rho_{y(1,2)} \neq 0$

Parameter	Effect	Estimate	Standard error
$\sigma^2_{\tau(1)}$	Treatment	22.9713	4.7077
$\sigma^2_{\tau(2)}$	Treatment	2.8592	0.6688
$\rho_{\tau(1,2)}$	Treatment	0.4796	0.1207
$\sigma^2_{r(1)}$	Replicate	13.9498	15.6364
$\sigma^2_{r(2)}$	Replicate	1.7517	2.0340
$\rho_{r(1,2)}$	Replicate	0.7661	0.3275
$\sigma^2_{b(1)}$	Block	12.3524	4.2842
$\sigma^2_{b(2)}$	Block	1.9970	0.7245
$\rho_{b(1,2)}$	Block	0.9118	0.06440
$\sigma^2_{e(1)}$	Error	8.5396	1.1615
$\sigma^2_{e(2)}$	Error	1.9457	0.2769
$\rho_{e(1,2)}$	Error	0.2795	0.08932
$\sigma^2_{y(1)}$ [a]	Total	8.5537	
$\sigma^2_{y(2)}$ [a]	Total	57.8131	
$\sigma_{y(1,2)}$ [a]	Total	13.3422	
$\rho_{y(1,2)}$ [a]	Total	0.6000	

Trait 1 = yield (tonnes × 10), trait 2 = height (cm).
[a]These values were read off the output of the statistical package and are slightly different from the values one would obtain by plugging the estimates of variances and covariances for the effects shown in this table, which are rounded to the forth decimal place, into Eqns (26)–(28) for the marginal variances, covariance and correlation.

model-based marginal correlation and the sample correlation. The bivariate scatter plots of BLUPs for effects of treatment (Fig. 6) and blocks (Fig. 7) as well as the residual plot (Fig. 8), confirm these divergent correlations. The example reinforces the observation that correlations can be quite different for different effects

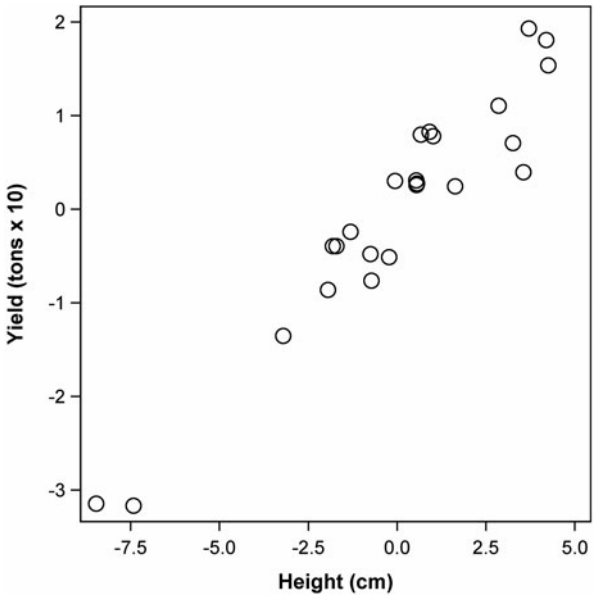


Fig. 7. Scatter plot of BLUPs of block effects for yield (kg/ha) *v.* height (cm) for oat (*Avena sativa* L.) data (Example 2).

and it is therefore important to study correlations at the level of effects. In this application, the main interest will be in the correlation among the treatment effects because these can be used to infer if indirect selection for yield via plant height would be a more promising approach than direct selection (Falconer & Mackay 1996). The relatively weak correlation for treatment effects suggests that indirect selection is not worthwhile in the case at hand.

Example 3

In an experiment with three rice cultivars (IR1514A, IR8 and Peta), the agronomic yield (t/ha) was assessed on seven plots

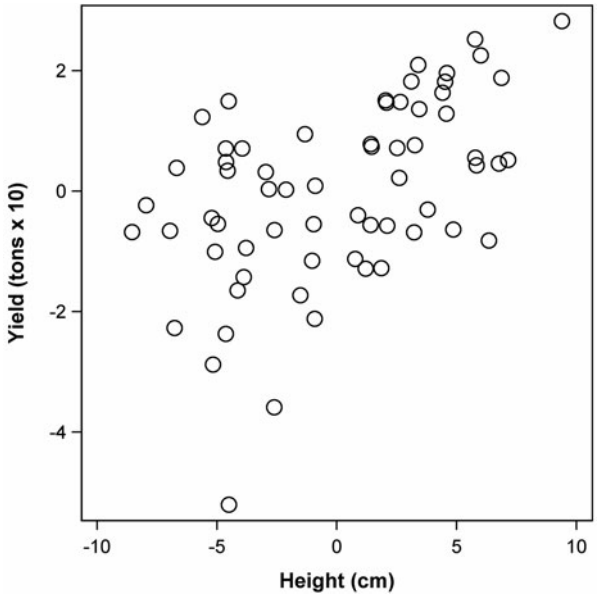


Fig. 6. Scatter plot of BLUPs of treatment effects for yield (kg/ha) *v.* height (cm) for oat (*Avena sativa* L.) data (Example 2).

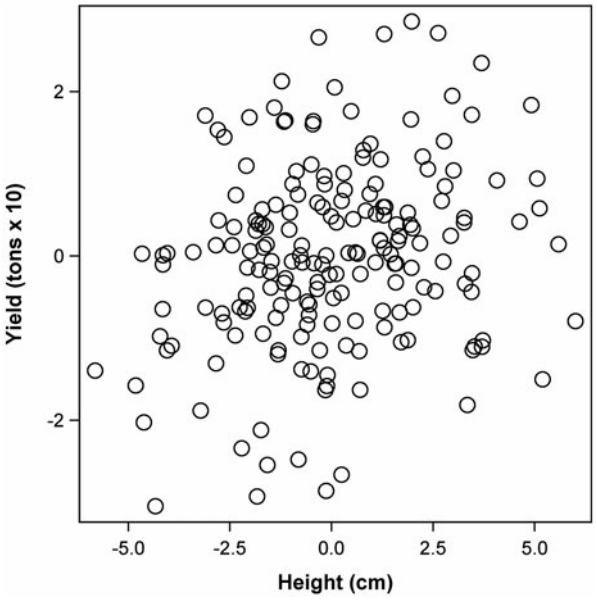


Fig. 8. Scatter plot of residuals for yield (kg/ha) *v.* height (cm) for oat (*Avena sativa*) data (Example 2).

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

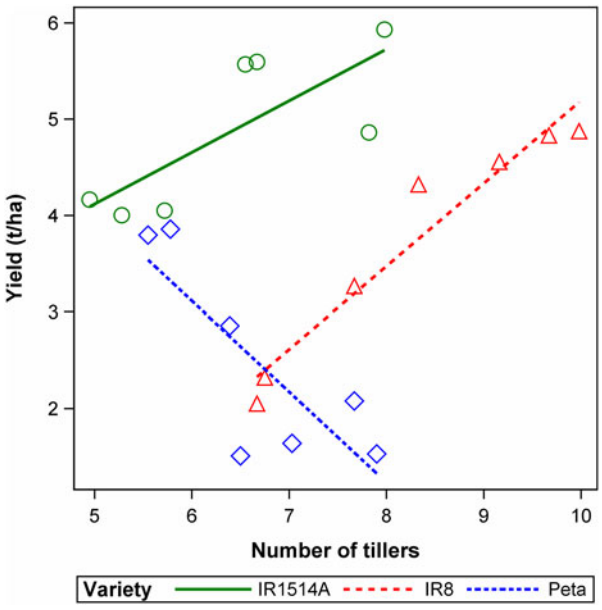


Fig. 9. Plot of yield (t/ha) *v.* number of tillers for rice (*Oryza sativa* L.) data for three varieties of rice, along with fitted regression lines for each variety (Gomez & Gomez 1984, p. 377) (Example 3). Colour online.

per variety (Gomez & Gomez 1984, p. 377, Table 9.4). On each plot, the number of tillers per hill was assessed. It is assumed here that the experiment was completely randomized, because Gomez & Gomez (1984) provide no information on any blocks in their table with the raw data, so our basic model is the one-way model in Eqn (1). Inspection of residuals for both traits indicated that approximate normality could be assumed. The sample correlation between the two traits is $r = 0.1973$ ($P = 0.3912$). Figure 9 shows a plot of yield *versus* the number of tillers, along with fitted regression lines per variety for emphasis of the general pattern. Obviously, there are strong correlations for the individual varieties, so the simple correlation ignoring the factor variety is grossly misleading. Moreover, Fig. 9 suggests that residual correlations differ markedly between varieties, meaning that the bivariate mixed model needs to allow for heterogeneity of residual variances and correlations. Heterogeneity of variance for a random effect is straightforward to implement in a mixed model. For example, in SAS one may use the option GROUP = variety on the REPEATED statement in order to allow the error variance to differ between varieties. Allowing for heterogeneity, the correlation of treatment effects is estimated as $\hat{\rho}_{\tau(1,2)} = -0.1631$ (S.E. = 0.7734). The fit of the error variance parameters of the heterogeneous model is shown in Table 6. None of the correlations are anywhere near the simple correlation of $r = 0.1973$. Correlations for the errors are large in absolute value and positive for varieties IR1514A and IR8, but negative for variety Peta. This heterogeneity is significant by an LRT with a chi-squared value of $T = 26.73$ on 6 d.f. ($P < 0.001$). Due to the significant heterogeneity, and because correlations for errors are rather different from the correlation between treatment effects, it would be inappropriate here to compute a single marginal correlation for the observed data. This is also clear from graphical inspection of the scatter of points in Fig. 9. Instead, it seems more reasonable to focus on the treatment-specific residual correlations shown in Table 5 or fit variety-specific regression lines as shown in Fig. 3 (Gomez & Gomez 1984). In fact, this 'regression with groups' with groups corresponding to varieties may be more informative for this dataset, but that analysis will not be explored in detail because our focus is on correlation. It is stressed, however, that correlation and regression analysis have close links. Specifically, a regression model can be derived from a bivariate mixed model by conditioning on one of the variates (Searle *et al.* 1992, p. 464). In the example, the slope for the regression of yield of a variety of tiller number can be derived as the ratio of covariance between the two traits over the variance of tiller number for that variety. It should also be pointed out that in general estimating correlation for an effect requires a reasonable number of levels for that effect. Three effects of treatments in this example are clearly a rather small number to make any substantive statement about correlation. This is why the focus was on the within-variety correlation.

Example 4

An experiment with chia (*Salvia hispanica* L.), laid out according to a randomized complete block design (RCBD), was performed in order to derive a prediction model for leaf area as a function of leaf width (cm) and leaf length (cm) (Mack *et al.* 2017). The experiment comprised three nitrogen fertilizer treatments (0, 20 and 40 kg N/ha). On each plot, one chia plant was selected randomly and ten leaves were sampled per plant. For each leaf, the length, width and area were measured. The objective of the experiment was to identify a suitable prediction model for leaf area for individual leaves. As a preliminary step, it was therefore of interest to estimate the marginal correlation between leaf width and leaf length. The Pearson correlation between these two traits is $r = 0.9697$. The raw data are plotted in Fig. 10.

Since there are subsamples (leaves) taken per plot, the linear model needs to be extended to comprise both a plot error as

Table 6. Residual variance parameter estimates by variety for rice (*Oryza sativa* L.) data of Gomez & Gomez (1984, p. 377) (Example 3) under a model with heterogeneity between varieties

	Variety					
	IR1514A		IR8		Peta	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\sigma^2_{e(1)}$	1.3688	0.7731	1.8521	1.0964	0.7862	0.4473
$\sigma^2_{e(2)}$	0.6649	0.3772	1.4543	0.8583	1.0692	0.6105
$\rho_{e(1,2)}$	0.7665	0.1667	0.9702	0.0242	−0.8061	0.1421

Trait 1 = number of tillers, trait 2 = yield of rice (t/ha).

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

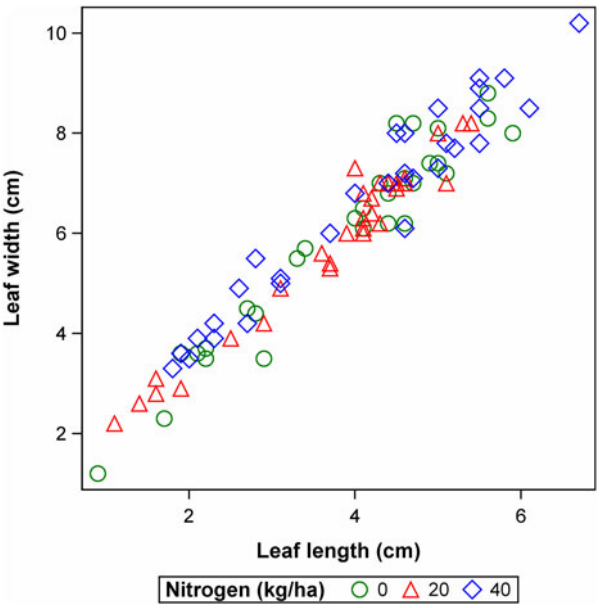


Fig. 10. Scatter plot of leaf length (cm) *v.* leaf width (cm) for chia (*Salvia hispanica* L.) leaves from an RCBD experiment with three nitrogen treatments (0, 20 and 40 kg N/ha) (Mack *et al.* 2017) (Example 4). Colour online.

well as a leaf error. For a single trait, the model is:

$$y_{ijk} = \mu + b_j + \tau_i + e_{ij} + f_{ijk} \tag{30}$$

where y_{ijk} is the response of the kth leaf for ith treatment in the jth block, μ is a general intercept, b_j is the effect of the jth block, τ_i is the effect of the ith treatment, and e_{ij} is the error term associated with the ijth plot, and f_{ijk} is the residual leaf error associated with the response y_{ijk}. This univariate model is extended to the bivariate model for leaf length and leaf width. Fitting this model, the block variance for leaf length was estimated to be zero (results not shown), and the statistical package reported a message indicating that the final Hessian was not positive definite. This reflects the fact that a correlation cannot be computed if one of the associated variances is zero. Thus, the random block effect for leaf length was dropped from the model. This was implemented by defining a dummy variable z with $z = 0$ for leaf length and $z = 1$ for leaf width. In coding the model using the statistical package, a block effect was then multiplied with this variable, such that the effect was 'switched off' for leaf length (Piepho *et al.* 2006). The fitted model parameters for the bivariate model and the corresponding model under the null hypothesis that all effects are uncorrelated is shown in Table 7. The treatment and plot effects have an estimated correlation of 1.000, whereas the leaf effects have a correlation of $\hat{\rho}_{f(1,2)} = 0.9736$. These very high correlations also translate into a very large marginal correlation of $\hat{\rho}_{y(1,2)} = 0.9686$, again a meaningful summary in this case because effect correlations are so similar. The leaf effect correlation is highly significant ($T = 248.73$, $P < 0.001$), whereas the correlations for plot effects are not significant ($T = 1.21$, $P = 0.2713$). Also, the variances for leaf effects are rather larger than those for treatment and plot effects. Thus, the variance and covariance of leaf effects dominate the corresponding marginal variance and covariance for the observed data. This is also why the leaf error correlation, as well as the marginal correlation, are both rather close to the sample correlation of $r = 0.9697$. But whether a large sample correlation is also associated with large correlations of all the effects can only be verified by fitting a bivariate model as illustrated here.

Table 7. Variance parameter estimates for chia (*Salvia hispanica* L.) data (Mack *et al.* 2017) (Example 4) under the hypotheses $H_A : \rho_{y(1,2)} \neq 0$

Parameter	Effect	Estimate	Standard error
$\sigma^2_{b(1)}$	Block	0[a]	–
$\sigma^2_{b(2)}$	Block	0.008856	0.02211
$\sigma_{b(1,2)}$	Block	0[a]	–
$\rho_{b(1,2)}$	Block	0[a]	–
$\sigma^2_{\tau(1)}$	Treatment	0.1137	0.2135
$\sigma^2_{\tau(2)}$	Treatment	0.02437	0.06402
$\rho_{\tau(1,2)}$	Treatment	1.0000	–
$\sigma^2_{e(1)}$	Plot	0.01181	0.06370
$\sigma^2_{e(2)}$	Plot	0.03219	0.07053
$\rho_{e(1,2)}$	Plot	1.0000[b]	–
$\sigma^2_{f(1)}$	Leaf	3.6137	0.5460
$\sigma^2_{f(2)}$	Leaf	1.7052	0.2595
$\rho_{f(1,2)}$	Leaf	0.9736	0.005685
$\sigma^2_{y(1)}$[c]	Total	3.7480	
$\sigma^2_{y(2)}$[c]	Total	1.7618	
$\sigma_{y(1,2)}$[c]	Total	2.4889	
$\rho_{y(1,2)}$[c]	Total	0.9686	

Trait 1 = leaf length (cm), trait 2 = leaf width (cm).
[a]Fixed at zero at convergence.
[b]Fixed at unity at convergence.
[c]These values were read off the output of the statistical package and are slightly different from the values one would obtain by plugging the estimates of variances and covariances for the effects shown in this table, which are rounded to the forth decimal place, into Eqns (26) to (28) for the marginal variances, covariance and correlation.

Concluding remarks

The present paper has considered the question of how to estimate and test correlations among traits in designed experiments. The common practice of merely computing sample correlations from the raw data should be discouraged because such correlations do not account for effects due to treatments and blocks and hence are difficult to interpret and the significance test for these correlations is invalid. The advice given in the current paper, therefore, is that the simple correlation not been used at all when the data have structure, as is the case in designed experiments. The main recommendation here is to use methods that allow accounting for all effects one would consider in a univarate analysis of a single trait, thus making the univariate and bivariate analyses commensurate. If one is primarily interested in correlations, then fitting a bivariate mixed model is suggested as a viable option. The approach allows decomposing correlation according to effects in the model, providing detailed insight into the structure of correlation. If the effects display a correlation of about equal magnitudes, then a marginal correlation accounting for these effects is a meaningful summary measure.

The first step of the analysis is to inspect the correlations for each of the effects of the bivariate model. This can be done using REML estimates of the correlations and use scatter plots

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

of BLUPs of effects and residuals as illustrated with Examples 1 and 2. If correlations of effects are judged to be similar, either based on a non-significant LRT or from graphical inspection of scatter plots, the marginal correlation provides a meaningful summary of the data. Example 4 provides a case where the marginal correlation is the primary focus because good predictions are sought for leaf area at the level of individual observations. If the marginal correlation is high, which was the case here, one may fit a regression model for prediction (Mack *et al.* 2017). Conversely, when correlations are deemed to be substantially heterogeneous between effects, it is necessary to study correlations at the level of individual effects. Most of the time, the research will mainly be interested in the treatment effects. Example 2 provides a pertinent example, where treatments correspond to genotypes and genetic correlation is a key quantitative genetic parameter. In Example 3, by contrast, there was heterogeneity of within-treatment correlation, so the study of the within-treatment error correlations proved to be most interesting. So, which of the effects is of primary interest in case of heterogeneity of correlations between effects needs to be decided on a case-by-case basis.

The proposed approach entails modelling block and treatment effects as random for the purpose of estimating correlations for these effects. If one is prepared to estimate a correlation for an effect, then the assumption of random effects will seem quite natural. Indeed, without assuming random effects, the notion of a correlation or covariance for effects, or of observations composed of these effects, may seem difficult to envisage. Yet, the choice of both treatments and blocks in a designed experiment may be such that the assumption of a random sampling of effects is deemed hard to justify. For this reason, one may insist that these effects should be modelled as fixed (Dixon 2016). It was shown for the case of a CRD, however, that variances and covariance for treatment effects can be defined also for fixed effects and that estimates of variances and covariances coincide under the fixed and random-effects assumptions. This equivalence holds more generally, as pointed out by Gelman (2005, p. 13), who denotes a fixed-effects variance such as $Q(\tau)$ in (14) as finite-population variance, and the random-effects equivalent such as σ^2_τ as super-population variance. This fact lends support to the approach proposed here, i.e. to take effects as random for the purpose of estimating correlations, even in settings where one would otherwise hesitate to model block and treatment effects as random. There is, of course, a long-standing debate in the statistical literature as to whether blocks should be modelled as fixed or random. This is not the place to review that debate, but it should be pointed out that one camp insists that block effects should always be modelled as random, even for an RCBD (Nelder 1965; Casella & Berger 1990, p. 530; Giesbrecht & Gumpertz 2004, p. 86; Booth *et al.* 2009). The assumption of random treatment effects is less common, except in breeding experiments where treatments correspond to genotypes which can be regarded as a (usually large) sample from a common breeding population and the main interest is in making inferences about that population, including estimation of genetic correlation between traits, heritability of traits, etc. (Falconer & Mackay 1996). But note that even with a relatively small number of levels of an effect and lacking truly random sampling, the random assumption can be beneficial for inference about a treatment effect (James & Stein 1961; Lee *et al.* 2006, p. 151; Forkman & Piepho 2013). Certainly, if a Bayesian framework is adopted, there is no conceptual difficulty in assuming a joint probability distribution on all unknown parameters, including effects for treatments (Gelman 2005).

The bivariate model used in the current paper has been considered by Booth *et al.* (2009), who used it to derive an analysis-of-covariance model by conditioning on the observed values of one of the traits, regarded as the concomitant variable (covariate). The definition of the marginal correlation in (25) is not entirely novel either. It is very similar to the one proposed by Hill & Thompson (1978; also see Singh & El-Bizri 1992), also for an RCBD, for the specific context of genetic experiments where treatments correspond to genotypes. The difference is that for a blocked design the marginal correlation in (25) includes the block effects, while Hill & Thompson (1978) exclude the block effect from the definition of marginal correlation because it is irrelevant for selection decisions. Depending on the objective one may decide to either include or omit the block effect from the correlation. The same holds for any other design, where one needs to decide if design effects for blocks should be included in the definition of total correlation. The great advantage of the REML framework employed here is its generality and complete flexibility in terms of the effects that are included in the overall model and in the definition of total correlation. In plant breeding and quantitative genetic experiments, where the focus is on genetic correlation, block effects are not relevant in the definition of marginal correlations, but they need to be fitted nonetheless to represent the experimental design (Piepho & Möhring 2011).

Before fitting a bivariate random-effects model, it is advisable to fit univariate random-effects models to the individual traits. If a variance for an effect is estimated to be zero for at least one of the traits, a covariance cannot be fitted, because a covariance requires both variances to be non-zero and a correlation cannot be defined otherwise. It is then reasonable to constrain the bivariate model to have a zero covariance for such effects. If the variance of an effect is zero for only one trait, however, it needs to be kept in the model, but the covariance explicitly constrained to be zero. Most mixed models have covariance structures for this case, known as 'diagonal model' in some packages. In SAS, the structure is denoted as UN(1). Alternatively, one can use dummy coding as illustrated with Example 4. Diagonal covariance structures can also be used to fit two univariate models simultaneously as a joint bivariate model assuming independence for all effects.

The proposed methods were illustrated using experiments laid out according to a CRD, an RCBD (with sub-sampling per plot) or a lattice design, but the framework is broadly applicable to any experimental design. All that needs to be done is taking each effect of the bivariate model as random, assuming a bivariate normal distribution. Multivariate linear mixed models can also be fitted to more than two traits simultaneously and when data are unbalanced (Holland 2006; Piepho & Möhring 2011) or when spatial covariance needs to be modelled (Ganesalingam *et al.* 2013; de Faveri *et al.* 2017), but occasionally convergence problems may hamper such analyses. Good starting values are crucial to larger problems. One option to obtain starting values for the variances is to fit univariate models separately for each trait. Covariances can be obtained from univariate analyses of differences or sums of two traits (see Eqn (22) and accompanying text; Searle *et al.* 1992; Piepho *et al.* 2014).

The present paper has focused on correlation in bivariate data, stressing the importance of accounting for effects of blocks and treatments. This is also important when regression is used instead of correlation, which may be more informative. Figure 3 for Example 3 shows fitted regression lines for each variety, providing variety-specific predictions of yield as a function of tiller numbers. The regression lines very clearly show that the associations

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press

between yield and tiller number are rather different between varieties. In this example, fitting a single regression line through all points, ignoring the varieties, obviously makes as little sense as computing a simple correlation coefficient. If the regressor variable shows no treatment effect and slopes are identical across treatments, regression analysis accounting for block and treatment effects in the response is known as analysis of covariance (Milliken & Johnson 2002; Pearce 2006). When there is an interaction between treatment and covariate, i.e. slopes differ between treatments as is the case in Example 3, a regression with groups is appropriate.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0021859618000059.

Acknowledgements. I sincerely thank Philip Dixon (Department of Statistics, Iowa State University) for providing a preprint of his interesting article and for very inspiring discussions on the modelling of block effects in designed experiments during my visit at Iowa State University, which eventually prompted me to finally sit down and write this paper, which I had wanted to write for a long time. Berhard Ludwig (Universität Kassel) is thanked for very helpful comments on an earlier version of the paper. Further, I wish to thank Friedrich Utz (University of Hohenheim) for providing the oats data used in Example 2. Thanks are also due to Laura Mack of the Anton & Petra Research Ehrmann-Stiftung Research Training group 'Water – People – Agriculture' at the University of Hohenheim (Germany) for providing the Chia trial data, and to the team at Ihinger Hof (University of Hohenheim, Germany) for invaluable technical support in conducting the trial. The German Research Foundation (DFG) is thanked for financial support (grant PI 377/18-1). Last but not least, I thank the reviewer for very helpful and constructive comments.

References

Anderson TW (1971) *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.

Andrews DF and Herzberg AM (1985) *Data: A Collection of Problems From Many Fields for the Student and Research Worker*. New York: Springer.

Booth JG, Federer WT, Wells MT and Wolfinger RD (2009) A multivariate variance components model for analysis of covariance in designed experiments. *Statistical Science* **24**, 223–237.

Casella G and Berger RL (1990) *Statistical Inference*. Belmont: Duxbury Press.

De Faveri J, Verbyla AP, Cullis BR, Pitchford WS and Thompson R (2017) Residual variance–covariance modelling in analysis of multivariate data from variety selection trials. *Journal of Agricultural, Biological, and Environmental Statistics* **22**, 1–22.

Dixon P (2016) Should blocks be fixed or random? In *Annual Conference on Applied Statistics in Agriculture*. Manhattan, KS: Kansas State University. Available at http://newprairiepress.org/agstatconference/2016/proceedings/4 (Accessed 24 January 2018).

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th edn. Harlow, UK: Longmans Green.

Forkman J and Piepho HP (2013) Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. *Journal of Agricultural Science, Cambridge* **151**, 381–395.

Ganesalingam A, Smith AB, Beeck CP, Cowling WA, Thompson R and Cullis BR (2013) A bivariate mixed model approach for the analysis of plant survival data. *Euphytica* **190**, 371–383.

Gelman A (2005) Analysis of variance: why is it more important than ever. *Annals of Statistics* **33**, 1–53.

Giesbrecht FG and Gumpertz ML (2004) *Planning, Construction, and Statistical Analysis of Comparative Experiments. Wiley Series in Probability and Statistics*. New York: Wiley.

Gomez KA and Gomez AA (1984) *Statistical Procedures for Agricultural Research*. New York: John Wiley and Sons.

Hill WG and Thompson R (1978) Probabilities of nonpositive definite between group or genetic covariance matrices. *Biometrics* **34**, 429–439.

Holland JB (2006) Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS proc MIXED. *Crop Science* **46**, 642–654.

James W and Stein C (1961) Estimation with quadratic loss. In Neyman J (ed.). *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, CA: University of California Press, pp. 361–379.

Lee Y, Nelder JA and Pawitan Y (2006) *Generalized Linear Models with Random Effects. Unified Analysis via H-Likelihood*. Boca Raton, FL: Chapman & Hall/CRC.

Littell RC, Milliken GA, Stroup WW, Wolfinger R and Schabenberger O (2006) *SAS System for Mixed Models*, 2nd edn. Cary, NC: SAS Institute.

Mack L, Capezzone F, Munz S, Piepho HP, Claupein W, Phillips T and Graeff-Hönninger S (2017) Nondestructive leaf area estimation for chia. *Agronomy Journal* **109**, 1960–1969.

Mead R, Curnow RN and Hasted AM (1993) *Statistical Methods in Agriculture and Experimental Biology*, 2nd edn. London, UK: Chapman & Hall.

Milliken GA and Johnson DE (2002) *Analysis of Messy Data. Volume III: Analysis of Covariance*. Boca Raton, FL: CRC Press.

Nelder JA (1965) The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London A* **283**, 147–178.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

Pearce SC (2006) Analysis of covariance. In Kotz S, Balakrishnan N, Read CB and Vidakovic B (eds). *Encyclopedia of Statistical Sciences 1*, 2nd edn. New York: Wiley, pp. 126–132.

Piepho HP and Möhring J (2011) On estimation of genotypic correlations and their standard errors by multivariate REML using the MIXED procedure of the SAS system. *Crop Science* **51**, 2449–2454.

Piepho HP, Williams ER and Fleck M (2006) A note on the analysis of designed experiments with complex treatment structure. *HortScience* **41**, 446–452.

Piepho HP, Müller BU and Jansen C (2014) Analysis of a complex trait with missing data on the component traits. *Communications in Biometry and Crop Science* **9**, 26–40.

Schabenberger O and Pierce PJ (2002) *Contemporary Statistical Models for the Plant and Soil Sciences*. Boca Raton, FL: CRC Press.

Searle SR, Casella G and McCulloch CE (1992) *Variance Components*. New York: John Wiley & Sons.

Singh M and El-Bizri KS (1992) Phenotypic correlation: its estimation and testing significance. *Biometrical Journal* **34**, 165–171.

Thimm NH (2002) *Applied Multivariate Analysis*. New York: Springer.

Winer BJ, Brown DR and Michels KM (1991) *Statistical Principles in Experimental Design*, 3rd edn. New York: McGraw-Hill.

https://doi.org/10.1017/S0021859618000059 Published online by Cambridge University Press