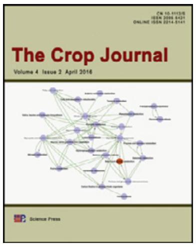
HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect



Estimation of genetic parameters and their sampling variances for quantitative traits in the type 2 modified augmented design

Frank M. You[a,*], *Qijian Song*[b], *Gaofeng Jia*[a,c], *Yanzhao Cheng*[a], *Scott Duguid*[a], *Helen Booker*[c], *Sylvie Cloutier*[d]

[a]Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB R6M 1Y5, Canada
[b]Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD 20705, USA
[c]Crop Development Centre, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada
[d]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada

ARTICLE INFO

Article history:
Received 11 November 2015
Received in revised form 16 December 2015
Accepted 2 February 2016
Available online 11 February 2016

Keywords:
Broad-sense heritability
Genetic correlation
Sampling variance
Modified augmented design

ABSTRACT

The type 2 modified augmented design (MAD2) is an efficient unreplicated experimental design used for evaluating large numbers of lines in plant breeding and for assessing genetic variation in a population. Statistical methods and data adjustment for soil heterogeneity have been previously described for this design. In the absence of replicated test genotypes in MAD2, their total variance cannot be partitioned into genetic and error components as required to estimate heritability and genetic correlation of quantitative traits, the two conventional genetic parameters used for breeding selection. We propose a method of estimating the error variance of unreplicated genotypes that uses replicated controls, and then of estimating the genetic parameters. Using the Delta method, we also derived formulas for estimating the sampling variances of the genetic parameters. Computer simulations indicated that the proposed method for estimating genetic parameters and their sampling variances was feasible and the reliability of the estimates was positively associated with the level of heritability of the trait. A case study of estimating the genetic parameters of three quantitative traits, iodine value, oil content, and linolenic acid content, in a biparental recombinant inbred line population of flax with 243 individuals, was conducted using our statistical models. A joint analysis of data over multiple years and sites was suggested for genetic parameter estimation. A pipeline module using SAS and Perl was developed to facilitate data analysis and appended to the previously developed MAD data analysis pipeline (http://probes.pw.usda.gov/bioinformatics_ tools/MADPipeline/index.html).

[*] *Corresponding author*. Tel.: +1 204 822 7525; fax: +1 204 808 7507.
E-mail address: Frank.You@agr.gc.ca (F.M. You).
Peer review under responsibility of Crop Science Society of China and Institute of Crop Sciences, CAAS.

http://dx.doi.org/10.1016/j.cj.2016.01.003

1. Introduction

In the early stages of breeding programs, a considerable number of test lines and a limited seed supply constrain the use of complete experimental designs with replications. Augmented designs, a class of unreplicated experimental designs, are a potential solution to this problem [1–3]. The augmented design usually has control lines arranged in a standard design such as a Latin square with several replications in soil-homogeneous blocks. Then the blocks are augmented to accommodate unreplicated test lines. Since control lines are in a standard design, the block effects can be estimated to adjust the observations of the test lines, and the error effects within control lines can be used to test the significance of performance differences among lines. Lin and Poushinsky [4,5] proposed a modified augmented design (MAD) with two subtypes. The type 1 MAD is used for square plots [4] and the type 2 MAD (MAD2) for rectangular plots [5]. This modified design is superior to the general augmented design in systematic placement of control and test genotypes within a block to enhance adjustment for soil heterogeneity [4].

MAD2 is used largely for early evaluation of breeding lines in crops such as wheat [6,7], potato [8], soybean [9], barley [10,11], sugarcane [12,13], and maize [14]. It is also used in flax breeding programs in Canada for field evaluation of flax yield, seed oil component, disease resistance, and other traits of agronomic and economic importance and for purposes of QTL identification, association mapping, and genomic selection [15–18]. In genetic experiments, individuals may have adequate amounts of seed for replicated trials, but it may be impractical to accommodate hundreds of genotypes in one homogeneous block of a field, owing to soil heterogeneity. Our earlier study [19] indicated that soil heterogeneity can be sufficiently adjusted for traits in MAD2 trials, suggesting that genetic variance of traits can be determined using a MAD2 approach.

Heritability and genetic correlation are crucial genetic parameters for quantitative traits because they can be used to predict the response to selection in plant breeding. Because the theoretical statistical distributions of these genetic parameter estimators are unknown, approximate tests of significance can be performed only on the basis of sampling errors. Methods for estimating sampling variances of the genetic correlation coefficient and heritability in some replicated experimental designs have been reported [20–24].

We have improved upon previous methods of MAD2 statistical analysis in adjusting for soil heterogeneity [19]. Owing to the lack of replication of test genotypes in the design, however, the total variance for test genotypes cannot be partitioned into its genetic and error components, and for this reason the method is unable to estimate genetic parameters. Here we present a method for estimating broad-sense heritability (H^2) and genetic correlation coefficients (r_g) of quantitative traits in the MAD2. We also derive the statistical formulas for estimating their sampling variances. We used computer simulations to evaluate the reliability of the proposed methods. As a case study using flax, we estimated the genetic parameters of three quantitative traits in a biparental recombinant inbred line (RIL) population of 243 lines.

2. Methods

2.1. Experimental design and statistical analysis

A typical MAD2 has $r * c$ whole plots structured as a grid of r rows and c columns. Each whole plot is split into k (an odd number, usually five or seven) parallel rectangular subplots. The whole experiment has a total of $r * c * k$ subplots. A control genotype is assigned to the central subplot of each whole plot (plot control). Two additional control genotypes serve as subplot controls randomly assigned to subplots in randomly selected whole plots with n replicates. Thus, the entire trial accommodates $rck - rc - 2n$ test genotypes that are randomly allocated to the remaining subplots (see Fig. 1 in [19] for the field layout).

Control plots are used to estimate row (R), column (C) and R × C interaction effects and to test for additive soil variation in the row and/or column directions. The two subplot controls plus one plot control are used to estimate the subplot error and test for non-additive soil variation in multiple directions across the field [9,19]. The test results are used to determine whether data adjustment is needed and which method of adjustment should be used. Three methods have been proposed to adjust test genotypes to reduce or remove effects due to soil heterogeneity [4,5,9]. For MAD2, method 1 is used if the row or column effects or both are significant, method 3 is used if the R × C interaction is significant [5,9,25] and a combined methods 1 and 3 approach is suggested in most cases [19]. A detailed statistical analysis for MAD2 trials has been described [19].

2.2. Case study

An RIL population with 243 lines derived from a cross between "CDC Bethune" and "Macbeth" (BM) was used to evaluate genetic variation. The single MAD2 trial consisted of 49 whole plots (7 × 7 grids), each splits into seven parallel subplots (1.5 m × 2.0 m with a 20-cm row spacing). CDC Bethune with 49 replicates was used as the plot control, and 7 replicates of both Hanley and Macbeth served as subplot controls. Field trials with the same design were conducted at two locations in Canada (Morden, Manitoba and Kernen Farm near Saskatoon, Saskatchewan) from 2009 to 2012 [18]. Genetic parameters and their sampling variances were estimated for three traits: oil content (OIL), iodine value (IOD), and linolenic acid content (LIN). The raw phenotypic data are presented in Table S1.

2.3. Estimation of genetic parameters

Observations of test genotypes and control genotypes after statistical adjustment [19] are expected to exclude the effect of soil heterogeneity; thus, the variation among replications of each control genotype should be caused only by random errors. The adjusted dataset in the trials corresponds to that obtained from a completely random design. Because each test genotype has a single adjusted observation, the total variance among test genotypes cannot be partitioned into genetic and error variances. However, the total variance within each control genotype, which is caused by random error, can be treated as the error variance of the test genotypes because it is reasonable to assume that any

Table 1 – Analyses of variance and covariance for model 1.

Source	*df*	MS	EMS	COV	ECOV
Genotype variance and covariance analyses					
Genotype (G)	$g - 1$	A_{ii}	$\sigma_e^2 + \sigma_G^2$	A_{ij}	$COV_e + COV_G$
Error variance and covariance analyses					
Control (C)	$t - 1$	B_{ii}	$\sigma_e^2 + n\kappa_C^2$	B_{ij}	$COV_e + nCOV_C$
Error	$rc + 2m - t$	C_{ii}	σ_e^2	C_{ij}	COV_e

DF: degrees of freedom; MS: mean square; EMS: expected mean square; COV: covariance; ECOV: expected covariance; *g*: number of genotypes; *t*: number of control genotypes; *n*: average number of replicates for each control genotype (see Formula (7) in text); *r* and *c* are the number of rows and columns, respectively; and *m* is the number of replicates for two subplot controls.

error effect of test genotypes or control genotypes follows the same normal distribution with $N(0, \sigma_e^2)$, where σ_e^2 is the error variance. Accordingly, the genetic variance can be estimated by subtraction of the error variance from the total variance of the test genotypes.

Thus, the genetic correlation coefficient ($\hat{r}_g$), error correlation coefficient ($\hat{r}_e$), phenotypic correlation coefficient ($\hat{r}_p$) between two traits *i* and *j* (i, *j* = 1, 2), and the broad-sense heritability ($\hat{H}_i^2$) of any single trait can be defined as.

$$\hat{r}_g = \widehat{COV}_{Gij} / \left(\widehat{COV}_{Gii}\widehat{COV}_{Gjj}\right)^{(1/2)} \tag{1}$$

$$\hat{r}_e = \widehat{COV}_{Eij} / \left(\widehat{COV}_{Eii}\widehat{COV}_{Ejj}\right)^{(1/2)} \tag{2}$$

$$\hat{r}_p = \widehat{COV}_{Pij} / \left(\widehat{COV}_{Pii}\widehat{COV}_{Gjj}\right)^{(1/2)} \tag{3}$$

$$\hat{H}_i^2 = \widehat{COV}_{Gii} / \widehat{COV}_{Pii}, \tag{4}$$

where $\widehat{COV}_P$, $\widehat{COV}_G$, and $\widehat{COV}_E$ represent the phenotypic, genetic and error variances of single traits (i = *j*) or covariances of two traits (i ≠ *j*), respectively. Estimation of these variances and covariances is dependent on statistical models.

2.3.1. Model 1: Single trial

For a single trial with *g* test genotypes and *t* control genotypes (including main plot controls and subplot controls), the adjusted observation of any test genotype with no replication can be expressed as.

$$y_i = \mu + G_i + \varepsilon_i (i = 1, 2, \ldots, g), \tag{5}$$

where $y_i \sim N(\mu, \sigma_P^2)$, $G_i \sim N(0, \sigma_G^2)$ and $\varepsilon_i \sim N(0, \sigma_e^2)$. σ_P^2, σ_G^2, and σ_e^2 are phenotypic, genetic and error variances, respectively. The error variance σ_e^2 is estimated based on *t* replicated control genotypes. For a given trait *i* (*i* = 1, 2), the analyses of variance and covariance are shown in Table 1.

For the two traits i and *j* (i, *j* = 1, 2), the error, genetic and phenotypic variances and covariances can be estimated as $\widehat{COV}_{Eij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Pij}$ as follows:

$$\begin{cases} \widehat{COV}_{Eij} = C_{ij} \\ \widehat{COV}_{Gij} = A_{ij} - C_{ij} \\ \widehat{COV}_{Pij} = \widehat{COV}_{Gij} + \widehat{COV}_{Eij} - C_{ij} (\text{on a plot basis}) \\ \widehat{COV}_{Pij} = \widehat{COV}_{Gij} + \dfrac{\widehat{COV}_{Eij}}{n} = \dfrac{[nA_{ij} + (1-n)C_{ij}]}{n} (\text{on a genotype mean basis}), \end{cases} \tag{6}$$

Table 2 – Analyses of variance and covariance for model 2.

Source	DF	MS	EMS	COV	ECOV
Genotype variance and covariance analyses					
Genotype (G)	$g - 1$	A_{ii}	$\sigma_e^2 + \sigma_{GE}^2 + e\sigma_G^2$	A_{ij}	$COV_e + COV_{GE} + eCOV_G$
Environment (E)	$e - 1$	B_{ii}	$\sigma_e^2 + \sigma_{GE}^2 + g\sigma_E^2$	B_{ij}	$COV_e + COV_{GE} + gCOV_E$
G × E	$(g - 1)(e - 1)$	C_{ii}	$\sigma_e^2 + \sigma_{GE}^2$	C_{ij}	$COV_e + COV_{GE}$
Error variance and covariance analyses					
Control (C)	$t - 1$	D_{ii}	$\sigma_e^2 + en\kappa_C^2$	D_{ij}	$COV_e + enCOV_C$
Environment (E)	$e - 1$	E_{ii}	$\sigma_e^2 + tn\sigma_E^2$	E_{ij}	$COV_e + tnCOV_E$
C × E	$(t - 1)(e - 1)$	F_{ii}	$\sigma_e^2 + n\sigma_{CE}^2$	F_{ij}	$COV_e + nCOV_{CE}$
Error	$e(rc + 2\ m - t)$	G_{ii}	σ_e^2	G_{ij}	COV_e

e: number of environments. See Table 1 for other notes.

where n is the number of replicates and C_{ij} and A_{ij} are the error and genotype covariance for trait i and j in Table 1, respectively. Because the number of replicates per control genotype differs in the MAD2 design, the number of replicates used for phenotypic variance estimation as described above is estimated [26,27] as

$$n=\left(\left(\sum n_k\right)^2-\sum n_k^2\right)/\left(\left(\sum n_k\right)(t-1)\right), \tag{7}$$

where n_k is the number of replicates for the kth control genotype and t is the number of control genotypes used, usually 3 in MAD2.

2.3.2. Model 2: Trials in multiple environments

For the joint analysis of data in multiple environments or trials with the same design (each trial from different years and sites treated as environments), the adjusted observation of any test genotype with e environments and without replication can be expressed as

$$y_{ij}=\mu+G_i+E_j+(GE)_{ij}+\varepsilon_{ij,}(i=1,2,\dots,g;j=1,2,\dots,e), \tag{8}$$

where $y_{ij}\sim N(\mu,\ \sigma_P^2)$, $G_i\sim N(0,\ \sigma_G^2)$, $E_j\sim N(0,\ \sigma_E^2)$, $(GE)_{ij}\sim N(0,\ \sigma_{GE}^2)$, and $\varepsilon_{ij}\sim N(0,\ \sigma_e^2)$. σ_P^2, σ_G^2, σ_E^2, σ_{GE}^2, and σ_e^2 are the phenotypic, genetic, environmental, genotype-by-environment interaction (G × E), and error variances, respectively. σ_e^2 is jointly estimated based on e trials with t replicated control genotypes in each trial. For a given trait i (i = 1, 2), the analyses of variance and covariance are shown in Table 2.

For the two traits i and j (i, j = 1, 2), the error, genetic, G × E, and phenotypic variance and covariance can be estimated as $\widehat{COV}_{Eij}$, $\widehat{COV}_{Gij}$, $\widehat{COV}_{(GE)ij}$, and $\widehat{COV}_{Pij}$ as follows:

$$\begin{cases} \widehat{COV}_{Eij}=G_{ij} \\ \widehat{COV}_{Gij}=\dfrac{1}{e}\left(A_{ij}-C_{ij}\right) \\ \widehat{COV}_{(GE)ij}=C_{ij}-G_{ij} \\ \widehat{COV}_{Pij}=\widehat{COV}_{Gij}+\widehat{COV}_{(GE)ij}+\widehat{COV}_{Eij}=\dfrac{1}{e}\left[A_{ij}+(e-1)C_{ij}\right]\ (on\ a\ plot\ basis) \\ \widehat{COV}_{Pij}=\widehat{COV}_{Gij}+\dfrac{\widehat{COV}_{(GE)ij}}{e}+\dfrac{\widehat{COV}_{Eij}}{en}=\dfrac{1}{en}\left[nA_{ij}+(1-n)G_{ij}\right](on\ a\ genotype\ mean\ basis). \end{cases} \tag{9}$$

where G_{ij}, C_{ij}, and A_{ij} are the covariances for error, G × E, and genotype for trait i and j in Table 2, respectively. Genetic parameters can be estimated using Formulas (1)–(4) and (9).

2.3.3. Model 3: Trials in multiple years and sites

Specifically for the joint analysis of data in multiple years and sites, the adjusted observation of any test genotype during y years at s sites with no replication can be expressed as

$$\begin{aligned} &y_{ijk}=\mu+G_i+Y_j+(GY)_{ij}+S_k+(GS)_{ik}+(YS)_{jk}+(GYS)_{ijk}+\varepsilon_{ijk} \\ &(i=1,2,\dots,g;j=1,2,\dots,y;k=1,2,\dots,s) \end{aligned} \tag{10}$$

where $y_{ijk}\sim N(\mu,\ \sigma_P^2)$, $G_i\sim N(0,\ \sigma_G^2)$, $Y_j\sim N(0,\ \sigma_Y^2)$, $(GY)_{ij}\sim N(0,\ \sigma_{GY}^2)$, $S_k\sim N(0,\ \sigma_S^2)$, $(GS)_{ik}\sim N(0,\ \sigma_{GS}^2)$, $(YS)_{jk}\sim N(0,\ \sigma_{YS}^2)$, $(GYS)_{ijk}\sim N(0,\ \sigma_{GYS}^2)$, and $\varepsilon_{ijk}\sim N(0,\ \sigma_e^2)$. σ_P^2, σ_G^2, σ_Y^2,σ_{GY}^2, σ_S^2 σ_{GS}^2, σ_{YS}^2, σ_{GYS}^2, and σ_e^2 are the variances for phenotype, genotype (G), year (Y), G × Y, site (S), G × S, Y × S, G × Y × S, and error, respectively. σ_e^2 is jointly estimated based on t replicated control genotypes during y years at s sites. For a given trait i (i = 1, 2), the analyses of variance and covariance are shown in Table 3.

Table 3 – Analyses of variance and covariance for model 3.

Source	DF	MS	EMS	COV	ECOV
Genotype variance and covariance analyses					
Genotype (G)	$g-1$	A_{ii}	$\sigma_e^2+\sigma_{GYS}^2+s\sigma_{GY}^2+y\sigma_{GS}^2+ys\sigma_G^2$	A_{ij}	$COV_e+COV_{GYS}+sCOV_{GY}+yCOV_{GS}+ysCOV_G$
Year (Y)	$y-1$	B_{ii}	$\sigma_e^2+\sigma_{GYS}^2+s\sigma_{GY}^2+g\sigma_{YS}^2+gs\sigma_Y^2$	B_{ij}	$COV_e+COV_{GYS}+sCOV_{GY}+gCOV_{YS}+gsCOV_Y$
Site (S)	$s-1$	C_{ii}	$\sigma_e^2+\sigma_{GYS}^2+y\sigma_{GS}^2+g\sigma_{YS}^2+gy\sigma_S^2$	C_{ij}	$COV_e+COV_{GYS}+yCOV_{GS}+gCOV_{YS}+gyCOV_S$
G × Y	$(g-1)(y-1)$	D_{ii}	$\sigma_e^2+\sigma_{GYS}^2+s\sigma_{GY}^2$	D_{ij}	$COV_e+COV_{GYS}+sCOV_{GY}$
G × S	$(g-1)(s-1)$	E_{ii}	$\sigma_e^2+\sigma_{GYS}^2+y\sigma_{GS}^2$	E_{ij}	$COV_e+COV_{GYS}+yCOV_{GS}$
Y × S	$(y-1)(s-1)$	F_{ii}	$\sigma_e^2+\sigma_{GYS}^2+g\sigma_{YS}^2$	F_{ij}	$COV_e+COV_{GYS}+gCOV_{YS}$
G × Y × S	$(g-1)(y-1)(s-1)$	G_{ii}	$\sigma_e^2+\sigma_{GYS}^2$	G_{ij}	COV_e+COV_{GYS}
Error variance and covariance analyses					
Control (C)	$t-1$	H_{ii}	$\sigma_e^2+ysn\kappa_C^2$	H_{ij}	$COV_e+ysn_cCOV_C$
Year (Y)	$y-1$	I_{ii}	$\sigma_e^2+n\sigma_{CYS}^2+sn\sigma_{CY}^2+gn\sigma_{YS}^2+gsn\sigma_Y^2$	I_{ij}	$COV_e+nCOV_{CYS}+snCOV_{CY}+gnCOV_{YS}+gsnCOV_Y$
Site (S)	$s-1$	J_{ii}	$\sigma_e^2+n\sigma_{CYS}^2+yn\sigma_{CS}^2+gn\sigma_{YS}^2+gyn\sigma_S^2$	J_{ij}	$COV_e+nCOV_{CYS}+ynCOV_{CS}+gnCOV_{YS}+gynCOV_S$
C × Y	$(t-1)(y-1)$	K_{ii}	$\sigma_e^2+n\sigma_{CYS}^2+sn\sigma_{CY}^2$	K_{ij}	$COV_e+nCOV_{CYS}+snCOV_{CY}$
C × S	$(t-1)(s-1)$	L_{ii}	$\sigma_e^2+n\sigma_{CYS}^2+yn\sigma_{CS}^2$	L_{ij}	$COV_e+nCOV_{CYS}+ynCOV_{CS}$
Y × S	$(y-1)(s-1)$	M_{ii}	$\sigma_e^2+n\sigma_{CYS}^2+tn\sigma_{YS}^2$	M_{ij}	$COV_e+nCOV_{CYS}+tnCOV_{YS}$
C × Y × S	$(t-1)(y-1)(s-1)$	N_{ii}	$\sigma_e^2+n\sigma_{CYS}^2$	N_{ij}	COV_e+nCOV_{CYS}
Error	$ys(rc+2\ m-t)$	O_{ii}	σ_e^2	O_{ij}	COV_e

y: number of years; s: number of sites. See Tables 1 and 2 for other notes.

Table 4 – Correction coefficients in Formula (17) for sampling variance estimation for model 1.

$\hat{\theta}$	$\widehat{COV}_{Pst}$		$\widehat{COV}_{Gst}$		$\widehat{COV}_{Est}$	
	C_1	C_2	C_1	C_2	C_1	C_2
$\widehat{COV}_{Pqr}$ [a]	n^2	0	n^2	0	0	0
$\widehat{COV}_{Pqr}$ [b]	n^2	$(1-n)^2$	n^2	$-n(1-n)$	0	$n(1-n)$
$\widehat{COV}_{Gqr}$			n^2	n^2	0	$-n^2$
$\widehat{COV}_{Eqr}$					0	n^2

[a] On a plot basis.
[b] On an entry-mean basis.

For the two traits i and j ($i, j = 1, 2$), the variances and covariances for error, G, Y, G × Y, G × S, and G × Y × S can be estimated separately as $\widehat{COV}_{Eij}$, $\widehat{COV}_{Gij}$, $\widehat{COV}_{Yij}$, $\widehat{COV}_{(GY)ij}$, $\widehat{COV}_{(GS)ij}$, and $\widehat{COV}_{(GYS)ij}$, respectively:

$$\begin{cases} \widehat{COV}_{Eij} = O_{ij} \\ \widehat{COV}_{Gij} = \dfrac{1}{ys}\left(A_{ij} - D_{ij} - E_{ij} + G_{ij}\right) \\ \widehat{COV}_{(GY)ij} = \dfrac{1}{s}\left(D_{ij} - G_{ij}\right) \\ \widehat{COV}_{(GS)ij} = \dfrac{1}{y}\left(E_{ij} - G_{ij}\right) \\ \widehat{COV}_{(GYS)ij} = G_{ij} - O_{ij} \\ \widehat{COV}_{Pij} = \widehat{COV}_{Gij} + \widehat{COV}_{(GY)ij} + \widehat{COV}_{(GS)ij} + \widehat{COV}_{(GYS)ij} + \widehat{COV}_{Eij} \\ = \dfrac{1}{ys}\left[A_{ij} + (y-1)D_{ij} + (s-1)E_{ij} + (1-y)(1-s)G_{ij}\right] (\text{on a plot basis}) \\ \widehat{COV}_{Pij} = \widehat{COV}_{Gij} + \dfrac{\widehat{COV}_{(GY)ij}}{y} + \dfrac{\widehat{COV}_{(GS)ij}}{s} + \dfrac{\widehat{COV}_{(GYS)ij}}{ys} + \dfrac{\widehat{COV}_{Eij}}{ysn} \\ = \dfrac{1}{ysn}\left(nA_{ij} + (1-n)O_{ij}\right) \text{ (on a genotype mean basis).} \end{cases} \quad (11)$$

where O_{ij}, A_{ij}, D_{ij}, E_{ij}, and G_{ij} are the covariance for error, G, G × Y, G × S, and G × Y × S for traits i and j in Table 3, respectively. Similarly, several genetic parameters can be estimated by applying Formula (11) to Formulas (1)–(4).

2.4. Estimation of sampling variances

The Delta method [28,29] was used to derive the formulas for sampling errors for several genetic parameters. General formulas for sampling errors of several genetic parameters are available [22,24,30]:

$$\begin{cases} V(\hat{r}_p) = \hat{r}_p^2\left[\dfrac{V\left(\widehat{COV}_{Pij}\right)}{\widehat{COV}_{Pij}^2} + \dfrac{V\left(\widehat{COV}_{Pii}\right)}{4\widehat{COV}_{Pii}^2} + \dfrac{V\left(\widehat{COV}_{Pjj}\right)}{4\widehat{COV}_{Pjj}^2} - \dfrac{COV\left(\widehat{COV}_{Pij}, \widehat{COV}_{Pii}\right)}{\widehat{COV}_{Pij}\widehat{COV}_{Pii}} - \dfrac{COV\left(\widehat{COV}_{Pij}, \widehat{COV}_{Pjj}\right)}{\widehat{COV}_{Pij}\widehat{COV}_{Pjj}} + \dfrac{COV\left(\widehat{COV}_{Pii}, \widehat{COV}_{Pjj}\right)}{2\widehat{COV}_{Pii}\widehat{COV}_{Pjj}}\right] \\ V(\hat{r}_g) = \hat{r}_g^2\left[\dfrac{V\left(\widehat{COV}_{Gij}\right)}{\widehat{COV}_{Gij}^2} + \dfrac{V\left(\widehat{COV}_{Gii}\right)}{4\widehat{COV}_{Gii}^2} + \dfrac{V\left(\widehat{COV}_{Gjj}\right)}{4\widehat{COV}_{Gjj}^2} - \dfrac{COV\left(\widehat{COV}_{Gij}, \widehat{COV}_{Gii}\right)}{\widehat{COV}_{Gij}\widehat{COV}_{Gii}} - \dfrac{COV\left(\widehat{COV}_{Gij}, \widehat{COV}_{Gjj}\right)}{\widehat{COV}_{Gij}\widehat{COV}_{Gjj}} + \dfrac{COV\left(\widehat{COV}_{Gii}, \widehat{COV}_{Gjj}\right)}{2\widehat{COV}_{Gii}\widehat{COV}_{Gjj}}\right] \\ V(\hat{r}_e) = \hat{r}_e^2\left[\dfrac{V\left(\widehat{COV}_{Eij}\right)}{\widehat{COV}_{Eij}^2} + \dfrac{V\left(\widehat{COV}_{Eii}\right)}{4\widehat{COV}_{Eii}^2} + \dfrac{V\left(\widehat{COV}_{Ejj}\right)}{4\widehat{COV}_{Ejj}^2} - \dfrac{COV\left(\widehat{COV}_{Eij}, \widehat{COV}_{Eii}\right)}{\widehat{COV}_{Eij}\widehat{COV}_{Eii}} - \dfrac{COV\left(\widehat{COV}_{Eij}, \widehat{COV}_{Ejj}\right)}{\widehat{COV}_{Eij}\widehat{COV}_{Ejj}} + \dfrac{COV\left(\widehat{COV}_{Eii}, \widehat{COV}_{Ejj}\right)}{2\widehat{COV}_{Eii}\widehat{COV}_{Ejj}}\right] \\ V\left(\hat{H}_i^2\right) = \hat{H}_i^4\left[\dfrac{V\left(\widehat{COV}_{Gii}\right)}{\widehat{COV}_{Gii}^2} + \dfrac{V\left(\widehat{COV}_{Pii}\right)}{\widehat{COV}_{Pii}^2} - \dfrac{2COV\left(\widehat{COV}_{Gii}, \widehat{COV}_{Pii}\right)}{\widehat{COV}_{Gii}\widehat{COV}_{Pii}}\right] \end{cases} \quad (12)$$

Table 5 – Correction coefficients in Formula (18) for sampling variance estimation for model 2.

$\hat{\theta}$	$\widehat{COV}_{Pst}$			$\widehat{COV}_{Gst}$			$\widehat{COV}_{Est}$		
	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
$\widehat{COV}_{Pqr}$ [a]	n^2	$n^2(e-1)^2$	0	n^2	$-n^2(e-1)$	0	0	0	0
$\widehat{COV}_{Pqr}$ [b]	n^2	0	$(1-n)^2$	n^2	0	0	0	0	$en(1-n)$
$\widehat{COV}_{Gqr}$				n^2	n^2	0	0	0	0
$\widehat{COV}_{Eqr}$							0	0	$(en)^2$

[a] On a plot basis.
[b] On an entry-mean basis.

Table 6 – Correction coefficients in Formula (19) for sampling variance estimation for model 3.

$\hat{\theta}$	$\widehat{COV}_{Pst}$					$\widehat{COV}_{Gst}$					$\widehat{COV}_{Est}$				
	C_1	C_2	C_3	C_4	C_5	C_1	C_2	C_3	C_4	C_5	C_1	C_2	C_3	C_4	C_5
$\widehat{COV}_{Pqr}$ [a]	n^2	$n^2(y-1)^2$	$n^2(s-1)^2$	$n^2(y-1)^2(s-1)^2$	0	n^2	$-n^2(y-1)$	$-n^2(s-1)$	$n^2(y-1)(s-1)$	0	0	0	0	0	0
$\widehat{COV}_{Pqr}$ [b]	n^2	0	0	$4n^2$	$(1-n)^2$	n^2	0	0	$-2n^2$	0	0	0	0	0	0
$\widehat{COV}_{Gqr}$						n^2	n^2	n^2	n^2	n^2	0	0	0	0	0
$\widehat{COV}_{Eqr}$											0	0	0	0	$(nys)^2$

[a] On a plot basis.
[b] On an entry-mean basis.

We noticed that $\widehat{COV}_{Pij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Eij}$ in Formulas (6), (9), and (11) are linear functions of moments, θ $(m_1, m_2, \ldots, m_k)$:

$$\theta(m_1, m_2, \ldots, m_k) = \sum\nolimits_{(i=1)}^{k} a_i m_i, \tag{13}$$

where m_i corresponds to the mean square of a variation source in Tables 1, 2, and 3. Then the variance of θ in Formula (14) can be estimated [31]:

$$V(\theta) = \sum\nolimits_{(i=1)}^{k} a_i^2 V(m_i) + \sum\nolimits_{(i \neq j)}^{k} a_i a_j COV(m_i, m_j)\cdot \tag{14}$$

Similarly, the approximate covariance between two functions of moments $\theta_l(m_1, \ldots, m_k)$ $(l = 1, 2)$ is given by [31]:

$$COV(\theta_1, \theta_2) = \sum\nolimits_{(i,j=1)}^{k} a_i a_j COV(m_i, m_j)\cdot \tag{15}$$

$V(m_i)$ and $COV(m_i, m_j)$ in Formulas (15) and (16) can be calculated using the following formulas [32,33]:

$$\begin{cases} V(m_{qr}) = \dfrac{1}{df+2}\left(m_{qq}m_{rr} + m_{qr}^2\right) \\ COV(m_{qr}, m_{st}) = \dfrac{1}{df+2}\left(m_{qs}m_{qt} + m_{qt}m_{rs}\right), \end{cases} \tag{16}$$

where q, r, s, t = 1, 2 and df are the degrees of freedom. The denominator value $df + 2$ has been suggested [34] to yield unbiased estimates.

Suppose that genotype and environment are independent. By applying Formulas (14)–(16) to Formulas (6), (9), and (11), we can calculate the variances of $\widehat{COV}_{Pij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Eij}$ (i, j = 1, 2; $i = j$ or $i \neq j$), and covariances of any two of them, which are finally used to estimate the variances of correlation coefficients ($\hat{r}_p$, $\hat{r}_g$, $\hat{r}_e$), and $\hat{H}_i^2$.

For model l, we derived a general formula to calculate the variances of $\widehat{COV}_{Pij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Eij}$ (i, j = 1, 2; $i = j$ or $i \neq j$) and the covariances between any two of them:

$$COV(\hat{\theta}_{qr}, \hat{\theta}_{st}) = \frac{1}{n^2}\left[C_1 \frac{A_{qs}A_{rt} + A_{qt}A_{rs}}{d_A} + C_2 \frac{C_{qs}C_{rt} + C_{qt}C_{rs})}{d_C}\right], \tag{17}$$

where $\hat{\theta}$ represents $\widehat{COV}_P$, $\widehat{COV}_G$, $\widehat{COV}_E$; q, r, s, t = 1, 2; n is the number of replicates of control genotypes estimated from Formula (7); d_A = $(g - 1) + 2$ and d_C = $(rc + 2\,m - t) + 2$ from Table 1; and C_1 and C_2 are the correction coefficients listed in Table 4 for calculation of different variances or covariances.

For model 2, a similar general formula was derived to calculate variances of $\widehat{COV}_{Pij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Eij}$ and covariances of any two of them:

$$COV(\hat{\theta}_{qr}, \hat{\theta}_{st}) = \frac{1}{(en)^2)}\left(C_1 \frac{(A_{qs}A_{rt} + A_{qt}A_{rs})}{d_A} + C_2 \frac{C_{qs}C_{rt} + C_{qt}C_{rs})}{d_C} + C_3 \frac{G_{qs}G_{rt} + G_{qt}G_{rs})}{d_G}\right), \tag{18}$$

where e is the number of environments; n is the number of replicates estimated with Formula (7); d_A = $(g - 1) + 2$, d_C = $(g - 1)(e - 1) + 2$ and d_G = $e(rc + 2\,m - t) + 2$ in Table 2; and C_1, C_2, and C_3 are the correction coefficients listed in Table 5 for calculation of different variances or covariances.

For model 3, we derived the following general formula to calculate variances of $\widehat{COV}_{Pij}$, $\widehat{COV}_{Gij}$, and $\widehat{COV}_{Eij}$ and covariances of any two of them:

$$COV(\hat{\theta}_{qr}, \hat{\theta}_{st}) = \frac{1}{(ysn)^2}(C_1 \frac{A_{qs}A_{rt} + A_{qt}A_{rs}}{d_A} + C_2 \frac{D_{qs}D_{rt} + D_{qt}D_{rs}}{d_D} + C_3 \frac{E_{qs}E_{rt} + E_{qt}E_{rs}}{d_E} C_4 \frac{G_{qs}G_{rt} + G_{qt}G_{rs}}{d_G} + C_5 \frac{O_{qs}O_{rt} + O_{qt}O_{rs}}{d_O}), \tag{19}$$

where y is the number of years; s is the number of sites; n is the number of control replicates estimated with Formula (7);

$d_A = (g-1)+2$, $d_D = (g-1)(y-1)+2$, $d_E = (g-1)(s-1)+2$, $d_G = (g-1)(y-1)(s-1)+2$ and $d_O = ys(rc+2\,m-t)+2$ from Table 3; and C_1, C_2, C_3, C_4, and C_5 are the correction coefficients listed in Table 6 for calculation of different variances or covariances.

2.5. *Computer simulations*

Based on the MAD2 design scheme, we simulated single MAD2 trials for estimation of two genetic parameters: H^2 of a trait and r_g between two traits. The purposes of the simulations were to (1) validate the proposed method for estimating genetic parameters in the MAD2 trials and (2) assess the accuracy of the derived theoretical formulas of the sampling variances for the two genetic parameters. We compared $H^2$and r_g values with the simulated $\hat{H}^2$ and $\hat{r}_g$ to determine whether these parameters were accurately estimated.

A single MAD2 trial with 10 × 10 whole plots and five subplots in each whole plot was simulated. The dataset of 390 test genotypes with one observation, one main plot control with 100 replicates, and two subplot controls with five replicates each were generated based on assumptions in Formula (5) and given values for heritability and genetic correlation of the test genotype population. All simulations were performed using R software (https://www.r-project.org/), and the R code is available upon request.

2.5.1. *Broad-sense heritability* (H^2)

Given the σ_G^2 and H^2 of a trait, we can calculate the error variances as $\sigma_e^2 = \sigma_G^2\,(1\text{-}H^2)/H^2$ on a plot basis. Thus, we can simulate the effect of different error variances on the estimation of H^2 in MAD2. Data generation was performed as follows: (1) given the μ and σ_G^2 of a trait, we generated a set of normal random numbers for 390 test genotypes plus three control genotypes following $N(\mu, \sigma_G^2)$, corresponding to the genetic values of test and control genotypes; (2) given H^2, we calculated σ_e^2 and generated 100 sets of normal random numbers with $N(0, \sigma_e^2)$, corresponding to the error effect of 100 replicates; and (3) we merged genetic values and error effects to generate phenotypic values of test and control genotypes of 100 replicates, creating a matrix of 393 rows and 100 columns, following $N(\mu, \sigma_P^2)$ and representing phenotypic values of the single MAD2 trial; (4) we randomly chose 390 rows with one column to simulate test genotypes without replication, one row with all 100 columns to simulate the plot control with 100 replicates, and two rows with five columns to simulate two subplot controls with five replicates. For each given H^2 value from 0.1 to 0.9 with an interval of 0.1, a total of 1000 simulations were performed. For each, the data were analyzed using model 1 (Table 1) and $\hat{H}^2$ and its sampling error were estimated using Formulas (4) and (12). The standard deviation of $\hat{H}^2$ in 1000 samples was calculated to represent an actual sampling error (henceforth termed "simulated" sampling error) for comparison with those calculated based on Formula (12).

2.5.2. *Genetic correlation* (r_g)

Given two traits (1 and 2) following $N(\mu_1, \sigma_{G1}^2)$ and $N(\mu_2, \sigma_{G2}^2)$ with r_g, we generated two sets of correlated random numbers to simulate genetic values of traits as follows: (1) we generated two sequences of uncorrelated standard normal distributed random numbers X_1 and X_2; (2) we defined a new variable $Y = r_g X_1 + \sqrt{1-r_g^2}\,X_2$ that had a genetic correlation of r_g with X_1; and (3) we transformed X_1 and Y into two new variables following the given normal distribution: $X_1' = X_1\sigma_{G1} + \mu_1$ and $X_2' = Y\sigma_{G2} + \mu_2$. To simplify the simulation, we set the error correlation r_e between the two traits to zero. We then generated two sets of independent random numbers for the error effects of the two traits. All other procedures followed the principles described above.

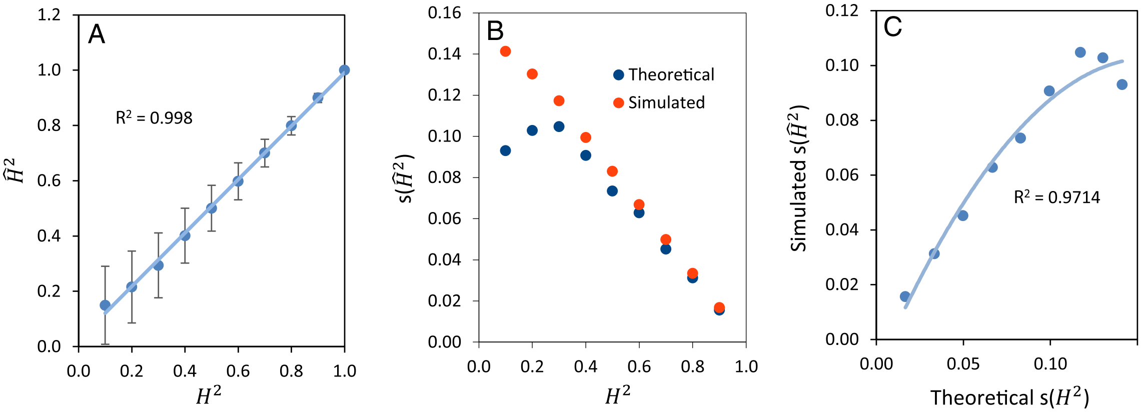


Fig. 1 – Simulation of broad-sense heritability ($\hat{H}^2$). (A) Simulation-based estimated $\hat{H}^2$ and its sampling error $s(\hat{H}^2)$ in relation to H^2. (B) Simulated and theoretical sampling errors ($s(\hat{H}^2)$) in relation to H^2. (C) Relationship between simulated $s(\hat{H}^2)$ and theoretical $s(\hat{H}^2)$.

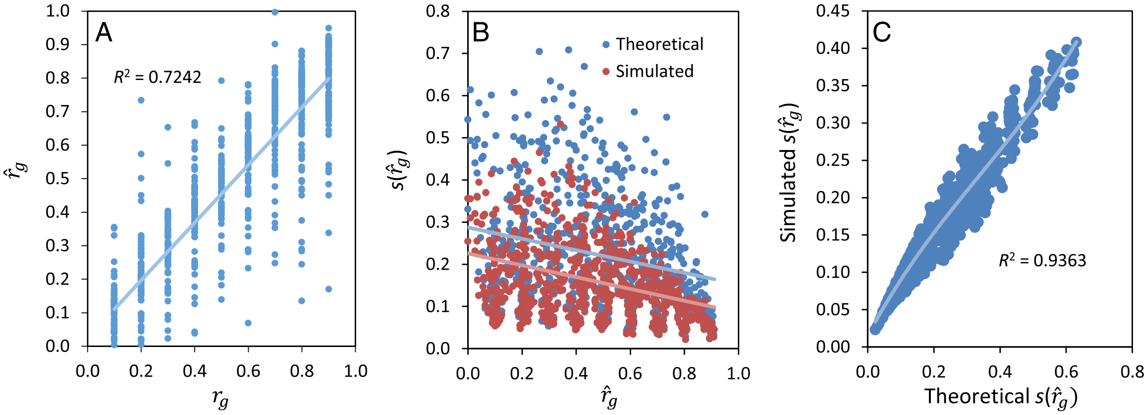


Fig. 2 – Simulation of genetic correlation coefficient ($\hat{r}_g$). (A) Estimated $\hat{r}_g$ based on simulation data in relation to given r_g. (B) Simulated and theoretical s($\hat{r}_g$) in relation to $\hat{r}_g$. (C) Relationship between simulated s($\hat{r}_g$) and theoretical s($\hat{r}_g$). The dots in plots represent averages of estimates from 1000 simulations.

2.5.3. Simulation of trial data from multiple years and sites

When trial data from multiple years and sites are available, both models 2 and 3 can be used for genetic parameter estimation. Model 1 can also be applied for analysis of single trials. To compare these three statistical models, we simulated trial data from four years and two sites per year that were similar to those of the case study. The same trial design and simulation procedure as the single trial were used but several major effects for years and sites, and some interaction effects, were added to the linear model (Formula (10) and Table 3). A total of eight trials were produced for a given H^2. All three models were used to estimate H^2.

2.6. Pipeline programs

The ANOVA and covariance analyses in Tables 1, 2, and 3 were implemented using SAS software (SAS Institute Inc., Cary, USA). The results from SAS served as input to a Perl program and were further analyzed to estimate several genetic parameters and their sampling variances. A new module including a SAS and a Perl program was appended to the MAD pipeline [19].

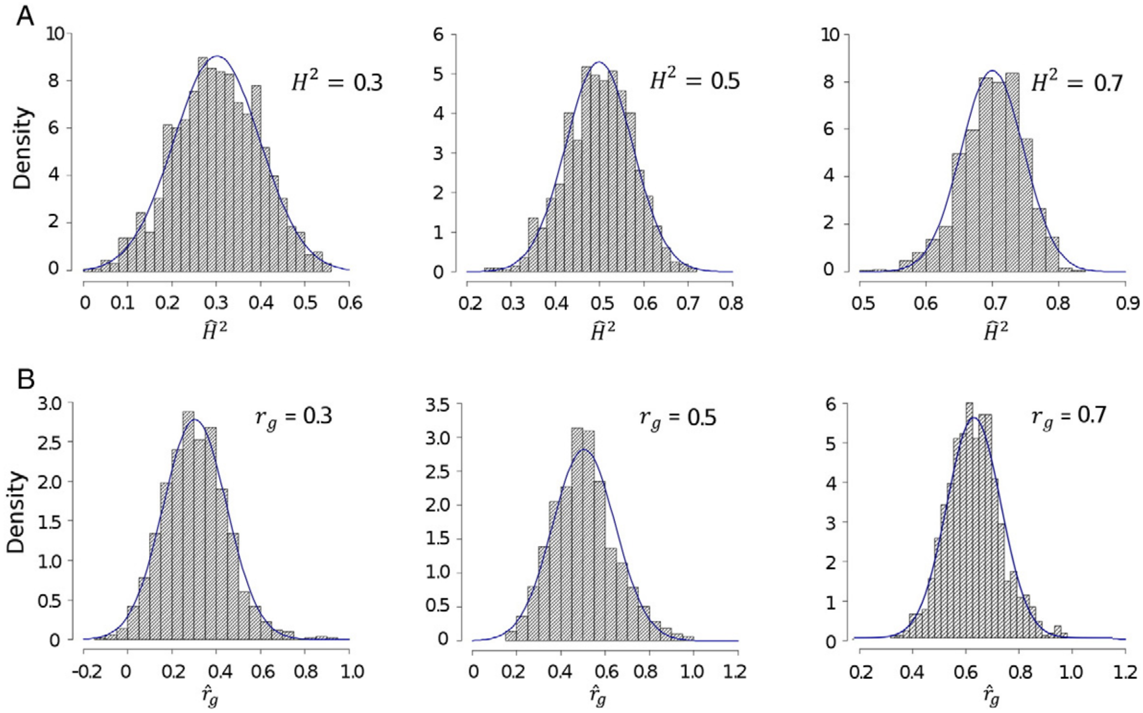


Fig. 3 – Sampling distributions of broad-sense heritability ($\hat{H}^2$) (A), and genetic correlation coefficient ($\hat{r}_g$) (B) at several parameter values.

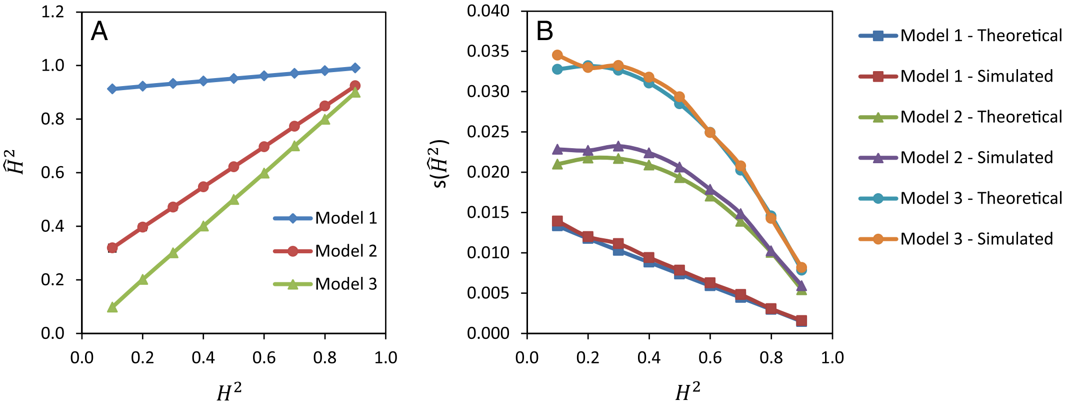


Fig. 4 – Simulation of broad-sense heritability ($\hat{H}^2$) under different statistical models, assuming significant genotype-by-environment (year and site) interaction effects. (A) Estimated $\hat{H}^2$ in relation to H^2. (B) Simulated and theoretical sampling errors ($s(\hat{H}^2)$) in relation to H^2.

3. Results

3.1. Computer simulations

3.1.1. Estimation of genetic parameters and their sampling errors

Given the different H^2 values from 0.1 to 1.0, the average $\hat{H}^2$ estimates of 1000 simulated datasets were highly correlated ($R^2 = 0.998$) with H^2 (Fig. 1A); both theoretical and simulated sampling errors ($s(\hat{H}^2)$) decreased with increasing H^2 (Fig. 1B); and the simulated $s(\hat{H}^2)$ was highly correlated with the theoretical $s(\hat{H}^2)$ (Fig. 1C). The $s(\hat{H}^2)$ values estimated from the two methods were consistent except when H^2 was less than 0.3. These results indicate that estimation of H^2 and its $s(\hat{H}^2)$ using the derived theoretical formula is reliable and that the reliability of the estimates increases with H^2.

Similarly, we simulated trial data for estimation of r_g for values ranging from 0.1 to 0.9. r_g was calculated based on the genetic covariance and variance of two traits. Considering that two traits may have different heritabilities, we generated data for 729 parameter combinations of different r_g (0.1–0.9), H_1^2 (0.1–0.9) and H_2^2 (0.1–0.9) each with 1000 simulations. A significant correlation between r_g and $\hat{r}_g$ ($R^2 = 0.7242$) was observed (Fig. 2A), but this relationship was more complex than that between $\hat{H}^2$ and H^2 (Fig. 1A). Large sampling errors were observed for any given r_g, which may result from the bias caused by the correlated errors of two traits (r_e). We also noticed that the theoretical $s(\hat{r}_g)$ was slightly higher than the simulated $s(\hat{r}_g)$ (Fig. 2B), though the theoretical $s(\hat{r}_g)$was also highly correlated with the simulated $s(\hat{r}_g)$ (Fig. 2C).

3.1.2. Sampling distribution of genetic parameters

Using 1000 simulations (or samples) for each given parameter or a combination of parameters, we can calculate the sampling error for each simulation and assess the sampling distribution of the parameter. Most samples appeared to be near- or normally distributed for $\hat{H}^2$ and $\hat{r}_g$. Fig. 3A and B shows several typical examples of the sampling distributions for $\hat{H}^2$ and $\hat{r}_g$, respectively. Based on all the simulated samples in two simulation experiments, 97% and 92% of the samples for $\hat{H}^2$ and $\hat{r}_g$, respectively, were normally distributed ($P > 0.05$) and the remainders followed an approximate normal distribution, suggesting that the theoretically estimated sampling error of a parameter estimate can be used to derive an approximate assessment of the significance of an estimate different from zero with a Z test.

3.1.3. Comparison of statistical models

For the joint data analysis of trials from multiple years and sites, two statistical models, model 2 (Table 2) and model 3 (Table 3), are suitable. Technically, model 1 (Table 1) can also be used for a single-trial analysis. The question was whether all

Table 7 – $\hat{H}^2$ and $s(\hat{H}^2)$ for three traits (OIL, IOD and LIN) in the BM population.

Model [a]	Unit [b]	OIL	IOD	LIN
Model 3	Genotype mean	0.916 ± 0.063 **	0.957 ± 0.025 **	0.954 ± 0.025 **
Model 2	Genotype mean	0.888 ± 0.011 **	0.963 ± 0.004 **	0.964 ± 0.004 **
Model 1	Genotype mean	0.996 ± 0.001 **	0.997 ± 0.001 **	0.997 ± 0.001 **
Model 3	Plot	0.490 ± 0.084 **	0.748 ± 0.059 **	0.748 ± 0.060 **
Model 2	Plot	0.400 ± 0.025 **	0.676 ± 0.021 **	0.675 ± 0.021 **
Model 1	Plot	0.905 ± 0.017 **	0.919 ± 0.014 **	0.925 ± 0.013 **

[a] Model 3: joint analysis of 4 years × 2 sites; Model 2: joint analysis using eight environments (each site/year as an environment); and Model 1: one single trial (2012 at Morden) is shown as an example.
[b] Genotype mean: on an entry-mean basis; Plot: on a plot basis.
** Represents statistical significance at the 0.01 probability level.

Table 8 – $\hat{r}_g$ and $s(\hat{r}_g)$ between three traits (OIL, IOD and LIN) in the BM population.

Model	OIL vs. IOD	OIL vs. LIN	IOD vs. LIN
Model 3	−0.277 ± 0.187	−0.261 ± 0.188	0.963 ± 0.015**
Model 2	−0.268 ± 0.065**	−0.259 ± 0.065**	0.961 ± 0.005**
Model 1	−0.336 ± 0.065**	−0.330 ± 0.064**	0.957 ± 0.006**

See Table 7 for the same notes.

three models could accurately estimate the genetic parameters when significant genotype-by-environment interaction effects were present. To compare the three statistical models for the same sets of data, we simulated trial data from 4 years and two sites (similarly to the case study). The results showed that only model 3 produced accurate H^2 estimates, whereas models 2 and 1 overestimated H^2, especially at low H^2 values (Fig. 4A). The theoretically estimated sampling errors of $\hat{H}^2$ fitted the simulated ones well in all three models (Fig. 4B). The sampling errors of $\hat{H}^2$ in model 3 were higher than those in models 2 and 1. Although $\hat{H}^2$ in model 1 had the lowest sampling errors, they deviated greatly from the correct values.

3.2. Case study

OIL, IOD and LIN are three phenotypic traits important in flax breeding for flaxseed or linseed. For the trial data of the BM population from 4 years at two sites, we first performed data adjustment using the MAD pipeline [19]. Then, using the adjusted observations, we also calculated the $\hat{H}^2$ (Table 7) and $\hat{r}_g$ (Table 8) for the three traits and their sampling errors on a single-plot and a genotype mean basis. Two statistical models (models 2 and 3) were applied to the same dataset. We also estimated the genetic parameters using model 1 independently for each of eight trials. Similar estimates for all two parameters were obtained using both model 2 and model 3 to account for the possibility of their high heritability. As expected, higher estimates of $\hat{H}^2$ and $\hat{r}_g$ of the three traits were obtained from model 1 (Tables 7 and 8). The sampling error estimates from model 3 were consistently higher than those from models 2 and 1 (Tables 7 and 8), in accordance with the simulation results (Fig. 4). Because the two genetic parameters follow a normal sampling distribution (Fig. 3), we could perform an approximate Z test to determine whether the estimates of the parameters were significantly different from zero. All three traits had high and statistically significant ($P < 0.01$) heritability estimates. For $\hat{r}_g$, the $\hat{r}_g$ estimates of all possible trait pairs were significant in model 2 and model 1, but the estimates of some trait pairs in model 3 were not significant because of their higher sampling errors. In addition, the estimates of $\hat{H}^2$ based on the genotype mean were larger than those based on single plots because the estimation of phenotypic variances differed (Formulas (2), (9), and (11)).

4. Discussion

An augmented design is usually applied by breeders to a large number of lines that are to be planted in a field of limited size. Error variance and genetic parameters may be estimated from replicated controls in unreplicated experimental designs such as MAD2. In the present study, genetic variance (covariance) was calculated based on total phenotypic variance (covariance) estimated from the test genotypes minus error variance (covariance) estimated from the control genotypes. This separate analysis approach provides approximate estimates of genetic parameters based on the MAD2 design, although it is not optimal for some cases. Our simulation results suggest that the method we propose is highly accurate for estimating H^2 with the reliability of the estimates increasing with trait heritability. Estimates of r_g had larger sampling errors than those of H^2, indicating that the latter is less subject to environmental effects.

We derived approximate theoretical sampling error formulas for the two genetic parameters using the Delta method [28,29]. We found that the theoretical sampling errors of all two genetic parameters were highly consistent with the simulated sampling errors, except for a few cases at very low

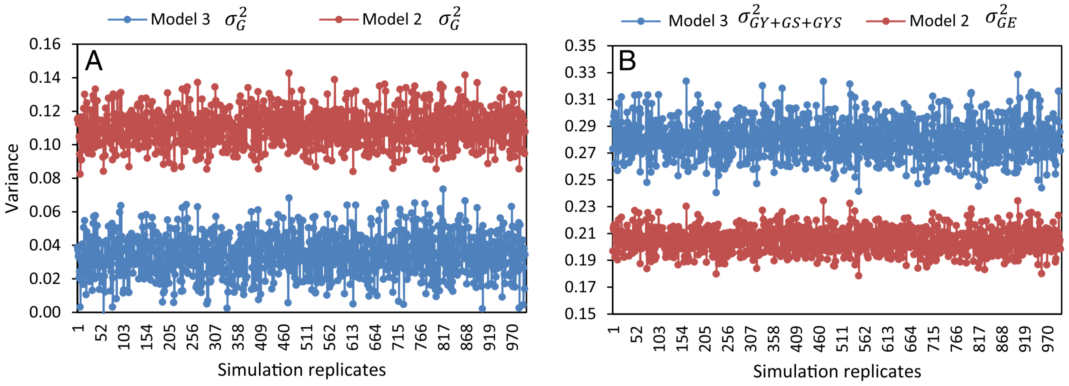


Fig. 5 – Partition of genetic variance (A) and genotype-by-environment variance (B) in 1000 simulation replicates at H^2 = 0.1 for models 2 and 3.

heritability (Figs. 1, 2, and 3) suggesting that estimation of the sampling errors for two genetic parameters in MAD2 is reliable and that it can be used to test whether the estimated genetic parameters are significantly different from zero.

Theoretically, the total variance of the test genotypes (the mean square A_{ii} in Table 1) will be greater than the error variance in a single trial. Accordingly, we were able to obtain genetic variance as total variance minus the error variance. However, because a limited number of control genotypes (three in our case) were used to estimate the error variance, the latter estimate is occasionally greater than the total variance of the test genotypes as a consequence of sampling bias. This results in negative genetic variance estimates and failure to estimate genetic parameters. In our simulation, when H^2=0.1,22.5% of simulation data sets failed to yield estimates of genetic parameters, but when H^2=0.3, only 0.6% of simulation data sets failed; and when H^2>0.3, none failed. When the heritability of a trait is very low (e.g. <0.1), the method proposed in this paper is sometimes unable to estimate genetic parameters precisely. In addition, there is some risk of misadjustment in this design if control genotypes show a different error variance or perform differently from the unreplicated entries [35]. Some alternatives have been proposed to reduce this risk, such as partially replicated (p–rep) designs, where a proportion of the test entries are replicated at each location [36–38].

There are two units used to measure phenotypic variances: one based on the single plot and the other based on the genotype mean. The two measurement units will generate different estimates of H^2; however, the estimation of r_g is not affected because the numerator and denominator of Formula (1) for calculating $\hat{r}_g$ involve only genetic components. The estimates of phenotypic variance based on the genotype mean were always larger than those based on the plot (Tables 7 and 8) because the error and interaction variance components were divided by the corresponding number of observations in the measurement unit on a genotype mean basis (Formulas (6), (9), and (11)). Because MAD2 is an unreplicated unbalanced design, each adjusted observation comes from single plots only, and estimates based on the plot may be more reasonable estimates of genetic variation.

Three statistical models were considered. Because model 1 deals only with single-trial data, the genetic variance contains an undecomposable genotype-by-environment interaction and consequently H^2 and r_g are always overestimated (Fig. 4A, Tables 7 and 8). For this reason, we suggest a joint analysis of trials from multiple environments (different years and/or sites) with model 2 or model 3. However, in the presence of significant genotype-by-environment effect, H^2 is generally overestimated in both models 1 and 2 (Fig. 4A). Theoretically, in model 2, the total variation of the test genotypes is partitioned into three components: G, E, and G × E (Table 2), whereas in model 3, E is further partitioned into Y, S, and Y × S, and G × E into G × Y, G × S, and G × Y × S. Hence, ANOVA of the same dataset had identical sum of squares (SS) of G in models 2 and 3; the SS of E was equal to the summation of the SS of Y, S and Y × S; and the SS of G × E was equal to the summation of the SS of G × Y, G × S, and G × Y × S. Both models also yielded the same error variances. The two models applied different formulas (Formulas (9) and (11)) to estimate σ^2_{G}, σ^2_{GE} or σ^2_{GY}, σ^2_{GS}, and σ^2_{GYS} that resulted in higher σ^2_{G} and lower σ^2_{GE} in model 2 than in 3 (Fig. 5) and consequently in the overestimation of genetic parameters in model 2. However, because more partitioned variance components in model 3 are indirectly estimated, higher sampling errors usually ensue—the major reason for the higher sampling variance of the genetic parameters estimated from model 3. Model 2 yields reasonable estimation accuracy and low sampling variance. Because model 2 treats all years, sites or their combinations as environments, it can be applied when complete data missing for a year or a site occurs, or data from only years or sites are available. Thus, model 2 is a more practical and flexible statistical model for genetic parameter estimation using datasets from multiple years and sites.

5. Conclusions

We have proposed an approximation method to estimate H^2 and r_g and their respective sampling variances for MAD2 trials. The simulation results suggest that H^2 can be reliably estimated in the MAD2 trial. The sampling error estimates based on the derived theoretical formulas coincide with the simulated values and can be applied to statistical tests of estimated genetic parameters.

Acknowledgments

This work was partly supported by an A-base project funded by Agriculture and Agri-Food Canada, the TUFGEN project funded by Genome Canada and other stakeholders, and funds from the Western Grains Research Foundation. The authors thank Andrzej Walichnowski for manuscript editing.

Appendix A. Supplementary data

Supplementary material for this article can be found online at http://dx.doi.org/10.1016/j.cj.2016.01.003.

REFERENCES

[1] W.T. Federer, Augmented (or hoonuiaku) designs, Hawaiian Planter's Record 55 (1956) 191–208.

[2] W.T. Federer, R.C. Nair, D. Raghavarao, Some augmented row-column designs, Biometrics 31 (1975) 361–374.

[3] W.T. Federer, D. Raghavarao, On augmented designs, Biometrics 31 (1975) 29–35.

[4] C.S. Lin, G. Poushinsky, A modified augmented design for an early stage of plant selection involving a large number of test lines without replication, Biometrics 39 (1983) 553–561.

[5] C.S. Lin, G. Poushinsky, A modified augmented design (type 2) for rectangular plots, Can. J. Plant Sci. 65 (1985) 743–749.

[6] A.R. Golparvar, M.M. Gheisari, D. Naderi, A.M. Mehrabi, A. Hadipanah, S. Salehi, Determination of the best indirect selection criteria in Iranian durum wheat (*Triticum aestivum* L.) genotypes under irrigated and drought stress conditions, Genetika 47 (2015) 549–558.

[7] C.H.A. Snijders, Field evaluation of type 2 modified augmented designs for non-replicated yield trials in the early stages of a wheat breeding program, in: P.

Ruckenbauer, F. Raab, R. Kern, K. Buchgraber, A. Schaumberger (Eds.),Bericht uber die Arbeitstagung 2002 der Vereinigung der Pflanzenzuchter und Saatgutkaufleute Osterreichs gehalten vom 26. bis 28 November 2002 in Gumpenstein, 2003.

[8] G.B. Schaalje, D.R. Lynch, G.C. Kozub, Field evaluation of a modified augmented design for early stage selection involving a large number of test lines without replication, Potato Res. 30 (1987) 35–45.

[9] C.S. Lin, H.D. Voldeng, Efficiency of type 2 modified augmented designs in soybean variety trials, Agron. J. 81 (1989) 512–517.

[10] K.W. May, G.C. Kozub, Success of a selection program for increasing grain yield of two-row barley lines and evaluation of the modified augmented design (type 2), Can. J. Plant Sci. 75 (1995) 795–799.

[11] K.W. May, G.C. Kozub, G.B. Schaalje, Field evaluation of a modified augmented design (type 2) for screening barley lines, Can. J. Plant Sci. 69 (1989) 9–15.

[12] K.V. Bhagyalakhsmi, K.G. Somarajan, A modified augmented design for early selection stages in sugarcane and its limitation, Sugar Tech. 1 (1999) 63–66.

[13] S.B. Milligan, L.M. McDonal, Use of a modified augmented design in the unreplicated stages of sugarcane selection, Report of Projects Louisiana Agricultural Experiment Station, Department of Agronomy 1990, pp. 132–135.

[14] C.G. Afolabi, P.S. Ojiambo, E.J.A. Ekpo, A. Menkir, R. Bandyopadhyay, Evaluation of maize inbred lines for resistance to fusarium ear rot and fumonisin accumulation in grain in tropical Africa, Plant Dis. 91 (2007) 279–286.

[15] B.J. Soto-Cerda, S. Duguid, H. Booker, G. Rowland, A. Diederichsen, S. Cloutier, Association mapping of seed quality traits using the Canadian flax (*Linum usitatissimum* L.) core collection, Theor. Appl. Genet. 127 (2014) 881–896.

[16] B.J. Soto-Cerda, S. Duguid, H. Booker, G. Rowland, A. Diederichsen, S. Cloutier, Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping, J. Integr. Plant Biol. 56 (2014) 75–87.

[17] D. Thambugala, S. Cloutier, Fatty acid composition and desaturase gene expression in flax (*Linum usitatissimum* L.), J. Appl. Genet. 55 (2014) 423–432.

[18] S. Kumar, F.M. You, S. Duguid, H. Booker, G. Rowland, S. Cloutier, QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.), Theor. Appl. Genet. 128 (2015) 965–984.

[19] F.M. You, S.D. Duguid, D. Thambugala, S. Cloutier, Statistical analysis and field evaluation of the type 2 modified augmented design (MAD) in phenotyping of flax (*Linum usitatissimum*) germplasms in multiple environments, Aust. J. Crop Sci. 7 (2013) 1789–1800.

[20] E.C.R. Reeve, The variance of the genetic correlation coefficient, Biometrics 11 (1955) 357–374.

[21] A. Robertson, The sampling variance of the genetic correlation coefficient, Biometrics 15 (1959) 469–485.

[22] C.J. Mode, H.F. Robinson, Pleotropism and the genetic variance and covariance, Biometrics 15 (1959) 518–537.

[23] E. Scheinberg, The sampling variance of the correlation coefficients estimated in genetic experiments, Biometrics 22 (1966) 187–191.

[24] J.B. Liu, D. Yang, M.A. You, The sampling variances of co-heritability and genetic correlation coefficients between traits in genetic design of single factor, J. Nanjing Agric. Univ. 17 (1994) 13–20 (in Chinese with English abstract).

[25] C.S. Lin, G. Poushinsky, P.Y. Jui, Simulation study of three adjustment methods for the modified augmented design and comparison with the balanced lattice square design soil variation, statistical models, J. Agric. Sci. 100 (1983) 527–534.

[26] D. Yang, J.Y. Gai, Y.H. Ma, Analysis of genetic parameters of agronomic and seed quality traits of soybean landraces in southern China soybean, Soybean Sci. 9 (1990) 9–18 (in Chinese with English abstract).

[27] M.G. Kendall, A. Stuart, The Advanced Theory of Statistics, Vol. 2, Inference and Relationship, Griffin, London, 1979.

[28] G.W. Oehlert, A note on the Delta method, Am. Stat. 46 (1992) 27–29.

[29] A.C. Davison, Statistical Models, Cambridge University Press, Cambridge, 2003.

[30] M.D. Li, Estimation of the sampling variance of co-heritability, J. Genet. Genomics 20 (1993) 504–513 (in Chinese wth English abstract).

[31] M.G. Kendall, A. Stuart, The Advanced Theory of Statistics, Vol 1, Distribution Theory, Griffin, London, 1977.

[32] G.M. Tallis, Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance, Aust. J. Stat. 1 (1959) 35–43.

[33] M.G. Bulmer, The Mathematical Theory of Quantitative Genetics, Clarendon Press, Oxford, 1980.

[34] O. Kempthorne, An Introduction to Quantitative Genetics, Longman, New York and London, 1957.

[35] R.A. Kempton, The design and analysis of unreplicated field trials, Vortrage Fuer Pflanzenzuchtung 7 (1984) 219–242.

[36] E.R. Williams, J.A. John, D. Whitaker, Construction of more flexible and efficient p-rep designs, Aust. Nz. J. Stat. 56 (2014) 89–96.

[37] E.R. Williams, H.P. Piepho, D. Whitaker, Augmented p-rep designs, Biom. J. 53 (2011) 19–27.

[38] B. Cullis, A. Smith, N. Coombes, On the design of early generation variety trials with correlated data, J. Agric. Biol. Environ. Stat. 11 (2006) 381–393.