Theor Appl Genet (1993) 86:234–236

Estimation of components of genetic variance and heritability for flowering time and yield in gerbera using Derivative-Free Restricted Maximum Likelihood (DFRML)

Y. Yu [1], **J. Harding** [1], **T. Byrne** [1], **and T. Famula** [2]

[1] Department of Environmental Horticulture, University of California, Davis, CA 95616 USA
[2] Department of Animal Science, University of California, Davis, CA 95616, USA

Received December 15, 1991; Accepted October 7, 1992
Communicated by A. L. Kahler

Summary. Additive genetic components of variance and narrow-sense heritabilities were estimated for flowering time (FT) and cut-flower yield (Y) for six generations of the Davis Population of gerbera using Derivative-Free Restricted Maximum Likelihood (DFRML). Additive genetic variance accounted for 54% of the total variability for FT and 30% of the total variability for Y. The heritability of FT (0.54) agreed with previous ANOVA-based estimates. However, the heritability of Y (0.30) was substantially lower than estimates using ANOVA. The advantages of DFRML and its applications in the estimation of components of genetic variance and heritabilities of plant populations are discussed.

Key words: Additive variance – Maximum likelihood – General Linear Model – Selection – Gaussian elimination – Plant pedigree matrix

Introduction

Historically, components of genetic variance and heritability for plant populations have been estimated from least squares or ANOVA-based methods (Falconer 1989) on the basis of particular experimental designs, i.e., NCII (Comstock and Robinson 1948). For unbalanced data, components of variance have been estimated from quadratics that mimic those of a balanced ANOVA. While this approach offers unbiased estimates and computation is relatively simple, it offers few other desirable statistical properties, and basic assumptions are often violated when fitting a genetic model. For example, when estimating the components of genetic variance from ANOVA, individuals in the population are often assumed to be independent and unselected (Becker 1981). However, it is common in breeding programs to select parents to improve traits rather than choose them at random. In addition, some inbreeding will result in any finite population.

Maximum likelihood (ML) has been used extensively in animal breeding and human genetics (e.g., Meyer 1983; Henderson 1984; Lange et al. 1976). The use of ML for the estimation of variance components in quantitative genetics was introduced by Hartley and Rao (1967) and later modified by Patterson and Thompson (1971) to *restricted* maximum likelihood (RML). Although ML is not uniformly ideal, it is an intuitively appealing strategy for point estimation. Sorenson and Kennedy (1984) suggested that RML could eliminate the selection bias inherent in ANOVA-based estimates. Since RML utilizes information from relatives, it is not constrained by experimental design, provided pedigrees are known. However, computation has been time-consuming, and RML estimates have not been practical for large unbalanced data sets. Recently, Grasser et al. (1987) presented DFRML, a derivative-free approach, which simplifies computation considerably. Computer programs (e.g., Meyer 1988) make the computation of DFRML practical, even for a large data set.

This study was designed to: (1) estimate the components of genetic variance and heritability of flowering time and cut-flower yield for the Davis Population of gerbera, using DFRML, and (2) compare the results from DFRML to those obtained from ANOVA-based methods.

Materials and methods

Population and traits

Samples were taken from generations 8 through 13 of the Davis Population of gerbera (*Gerbera hybrida*, Compositae). Individu-

---

Correspondence to: Y. Yu

al plants were grown at random in a controlled-temperature greenhouse in a uniform sand/organic substrate and irrigated with liquid fertilizer. Parents were selected each year on the basis of a Smith-Hazel index (Smith 1936; Hazel and Lush 1942) and crossed in an NCII design (except for generation 9), producing one breeding cycle per year. Descriptive parameters are shown in Table 1 (for details, see Harding et al. 1981, 1985; Yu 1991). A complete pedigree was maintained for these generations.

The two traits analyzed in this study are flowering time (FT) and cut-flower yield (Y). FT is defined as the number of days from seeding to harvest stage; Y is the number of flowers harvested over a 24-week period beginning with the first week in September for each generation (Yu et al. 1991). In the generations of this study, *Y* was always subjected to selection, but FT was not selected directly.

Statistical analysis

RML estimates of variance components requires the generalized inverse of the coefficient matrix of the mixed model equation (MME) (Henderson 1984). This is a daunting task, even for small sets of data. A more efficient algorithm (DFRML) makes use of Gaussian elimination at one-third the computational effort of inversion. This saving makes application to large data sets tractable. A complete description of the DFRML algorithm is presented by Graser et al. (1987). A brief review of this approach is informative. It follows the additive genetic model:

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} + \mathbf{Z}\,\mathbf{a} + \mathbf{e} \tag{1}$$

where **y** is a n × 1 vector of observations, **b** is a p × 1 vector of unknown fixed effects (e.g., treatment), **a** is a q × 1 vector of unobservable additive genetic effects, and **e** is an n × 1 vector of residuals. **X** and **Z** are known incidence matrices relating observations in **y** to effects in **b** and **a**, where the order of **a** (q) can exceed the number of observations (n). This model assumes **a** and **e** are multivariate normally distributed random variables with null means and

$$\mathrm{Var}\begin{vmatrix}\mathbf{a}\\ \mathbf{e}\\ \mathbf{y}\end{vmatrix} = \begin{vmatrix}\mathbf{A}\sigma_a^2 & \mathbf{O} & \mathbf{A}\mathbf{Z}'\sigma_a^2\\ \mathbf{O} & \mathbf{R} & \mathbf{R}\\ \mathbf{Z}\mathbf{A}\sigma_a^2 & \mathbf{R} & \mathbf{V}\sigma_e^2\end{vmatrix} \tag{2}$$

where **A** is the matrix of numerator relationships, σ_a^2 is the additive genetic variance, σ_e^2 is the residual variance, **R** is the matrix of environmental covariance (**R** = **I**), and

$$\mathbf{V} = (\mathbf{I} + \mathbf{Z}\,\mathbf{A}\,\mathbf{Z}'\,\mathbf{r})$$

for $\mathbf{r} = \sigma_a^2/\sigma_e^2$, and $h^2 = \mathbf{r}/(1+\mathbf{r})$. The objective of DFRML is to find those values of σ_a^2 and σ_e^2 (and thus **r**) that maximize the multivariate normal likelihood function. This objective is no different than RML estimation under any other algorithm, i.e., for a fixed prior value of **r**,

$$\sigma_e^2 = \frac{\mathbf{y}'\,\mathbf{P}\mathbf{y}}{\mathrm{N} - \mathrm{rank}(\mathbf{X})} \tag{3}$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{V}^{-1}$; therefore, $\sigma_a^2 = \mathbf{r}\,\sigma_e^2$.

The simplicity of DFRML is finding a means of computing $\mathbf{y}'\,\mathbf{P}\mathbf{y}$ that avoids evaluating $\mathbf{V}^{-1}$. To do so, consider the partitioned matrix

$$\begin{vmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A} - \frac{\mathbf{1}}{\mathbf{r}} & \mathbf{Z}'\mathbf{y}\\ \mathbf{y}'\mathbf{X} & \mathbf{y}'\mathbf{Z} & \mathbf{y}'\mathbf{y}\end{vmatrix} = \begin{vmatrix}\mathbf{C} & \mathbf{f}\\ \mathbf{f}' & \mathbf{y}'\mathbf{y}\end{vmatrix} \tag{4}$$

Table 1. Descriptive parameters of generations 8 through 13 of the Davis population of gerbera

Generation	Year grown	The number of			
		Parent plants	Crossing blocks	Full-sib families	Population size
8	1982–83	80	10	142	397
9	1983–84	50	1	100	399
10	1984–85	20	2	49	391
	1985–86	No plants grown; seeds stored for 1 year			
11	1986–87	40	4	71	552
12	1987–88	40	4	62	552
13	1988–89	40	4	100	735

Table 2. Components of variance and heritabilities of flowering (FT) and cut-flower yield (Y) for the Davis Population of gerbera

Traits	N	σ_p^2	σ_a^2	σ_e^2	h^2
FT	3,021	149.07	79.91	69.16	0.53
Y	3,029	84.84	25.45	59.39	0.30

where **C** and **f** are the left-hand and right-hand sides of the mixed model equations, respectively. Algebraically, if **C** is absorbed into $\mathbf{y}'\mathbf{y}$, the total sum of squares is replaced with $\mathbf{y}'\mathbf{P}\mathbf{y}$. Smith and Grasser (1986) show that the absorption of **C** does not require inversion, but can be accomplished through Gaussian elimination. Thus, for a fixed value of **r**, we build an array as in Eq. 4, absorb C row by row to compute $\mathbf{y}'\mathbf{P}\mathbf{y}$, and from that estimate σ_a^2 and σ_e^2. The next step is to evaluate that part of the log likelihood which depends upon **r**, i.e.,

$$\mathrm{L} = -\frac{1}{2}\left[\{\mathrm{N} - \mathrm{rank}(\mathbf{X}) - \mathrm{q}\}\,\log\sigma_e^2 + \log|\mathbf{C}| + \mathrm{q}\log\sigma_a^2 + \frac{\mathbf{y}'\mathbf{P}\mathbf{y}}{\sigma_e^2}\right] \tag{5}$$

The value of **r** that maximizes Eq. 5 can be found by iteration. Only log |**C**| is not immediately available. However, during the row by row absorption, log|**C**| can be found as the sum of the logs of the diagonal pivot elements of **C**. Hence, for a fixed **r**, **C** is absorbed while computing log|**C**|, $\mathbf{y}'\mathbf{P}\mathbf{y}$ is used to estimate σ_e^2, and Eq. 5 can be evaluated. Iteration is continued through possible values of **r**, computing σ_e^2, σ_a^2, and Eq. 5 until a maximum value of **r** is found. This provides RML estimates of σ_a^2, σ_e^2 and h^2.

The DFRML calculation was carried out separately for traits FT and Y using a VAX computer wtih a program by Meyer (1988). In each case, the analysis was based on a complete data matrix of all individuals (> 3,000) from generations 8 through 13.

Results

The results of DFRML analyses are shown in Table 2. The estimate of the additive genetic component for FT was 79.91, which is 54% of the total phenotypic variance; the remainder of 46% is various non-additive and environmental variances. The estimate for Y was 84.84, which is 30% of the total variance; the non-additive and environmental effects accounted for 70%.

Discussion

The estimate of heritability for FT was 0.54 using DFRML for the composite data matrix from generations 8 through 13. Estimates of heritability for this trait for the same generations have been reported (Yu et al. 1991) using the method of least squares, GLM (SAS 1988). Individual estimates obtained from each generation using the NCII design ranged from 0.27 to 0.67, with a mean of 0.50; there was little difference between estimates made using DFRML or GLM/NCII. The estimate of heritability for Y was 0.30 using DFRML. The mean of estimates of heritability for this trait was 0.46 over the same generations using GLM/NCII (Yu 1991). Thus, there is an appreciable difference in heritabilities for Y between the methods of DFRML and GLM/NCII.

Since Y was included in the selection indices, a selection differential was determined from indices over generations 8 through 13 (from Yu 1991), and realized heritability was estimated to be 0.26. Therefore, it appears that the DFRML estimate (0.30) more accurately predicts response to selection for Y (0.26) than does the GLM/NCII estimate (0.46). This may result from the fact that an assumption of no selection is required by the GLM/NCII method, while estimates from DFRML are free of this assumption (Sorenson and Kennedy 1984; Gianola and Fernando 1986).

Estimates may also differ because DFRML accounts for all of the genetic relationships in the entire data matrix **A**, and GLM/NCII only accounts for full- and half-sib relationships within sub-matrices from each generation. However, genetic information beyond full- and half-sibs may add little precision in this study because more than 3,000 individuals were included. Estimates of heritability for FT, which was not included in the selection indices, were nearly identical when calculated using either DFRML or GLM/NCII. Estimates of heritability for Y, which *was* indexed in the selection indices, were much higher when calculated with GLM/NCII. This suggests that the assumption of no selection in the GLM/NCII procedure is important. For this reason, estimates of heritability for the selected trait Y reported by Harding et al. (1985) may be biased. The DFRML estimate of 0.30 is supported by the reported realized heritability of 0.26. Both DFRML and GLM/NCII provided similar estimates of heritability for flowering time, a trait that was not included in the selection indices.

The results of this study lead us to recommend that plant breeders use DFRML when studying a trait that has been subjected to selection.

Acknowledgements. The authors wish to thank Karin Meyer for the use of her DFRML programs and Curt Finley for his assistance in running the program.

References

Becker WA (1981) Manual of quantitative genetics, 4th edn. Academic Enterprises, Pullman, Wash.

Comstock RE, Robinson HF (1948) The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4:254–266

Falconer DS (1989) Introduction to quantitative genetics. Longman Scientific and Technical and John Wiley and Sons, New York

Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. J Anim Sci 63:217

Graser H-U, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. J Anim Sci 64:1362

Harding J, Byrne TG, Nelson RL (1981) Estimation of heritability and response to selection for cut-flower yield in Gerbera. Euphytica 30:313–322

Harding J, Drennan D, Byrne TG (1985) Components of genetic variation of cut-flower yield in the Davis population of Gerbera. Euphytica 34:759–767

Hartley HO, Rao JNK (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika 54:93

Hazel LN, Lush JL (1942) The efficiency of three methods of selection. J Hered 33:393

Henderson CR (1984) ANOVA, MIVQUE, RML and ML algorithms for estimation of variances and covariances. In: David HA, David HT (eds) Statistics: an appraisal. Proc 50th Anniversary Conf Iowa State Stat Lab. The Iowa State University Press, Ames, Iowa, pp 257–286

Lange K, Westlake J, Spence MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 39:486–491

Meyer K (1983) Maximum likelihood procedures for estimating genetic parameters for later lactions of dairy cattle. J Dairy Sci 66:1988–1997

Meyer K (1988) DFRML – A set of programs to estimate variance components under an individual animal model. J Dairy Sci 71 [Suppl 2]:33 (Abstr)

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

SAS Institute (1988) SAS/STAT user's guide, version 6.03 edn. SAS Institute Inc, Cary, N.C.

Smith HF (1936) A discriminant function for plant selection. Ann Eugenics 7:240–250

Smith SP, Graser H-U (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood. J Dairy Sci 69:1156

Sorenson DA, Kennedy BW (1984) Estimation of genetic variances from unselected and selected populations. J Anim Sci 59:1213

Yu Y (1991) Genetics of flowering time in gerbera (*Gerbera hybrida,* Compositae). PhD thesis, University of California, Davis

Yu Y, Byrne TG, Harding J (1991) Quantitative analysis of flowering time in the Davis population of Gerbera. I. Components of genetic variances and heritability. Euphytica 53:19–23