

## Estimation of Genetic Parameters and Breeding Values in Honey Bees

*E.W. Brascamp\**, *R.F. Veerkamp<sup>†</sup>* and *P. Bijma\**

\*Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

<sup>†</sup>Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Wageningen, The Netherlands

**ABSTRACT:** Estimation of breeding values and variance components in honey bees is complex due to bees' reproduction system. It complicates calculation of the numerator relationships matrix (NRM). Mixed model methodology to estimate breeding values in honey bees was developed by Bienefeld (2007) (BER). To invert NRM they used a diagonal matrix of Mendelian sampling terms as an approximation of a more realistic block diagonal matrix. Brascamp and Bijma (2014) developed the necessary algebra for elements in NRM including these blocks and compared breeding value estimation using both methods with the same genetic parameters for both. In this paper we estimate these genetic parameters with ASReml. It appears that BER in particular underestimates the additive genetic variance for worker effect and also estimates the genetic correlation (true value -0.50) closer to zero. For the ranking of breeding animals differences are small.

**Key words:** honey bees; estimation genetic parameters; estimation breeding values.

### INTRODUCTION

To estimate genetic parameters and breeding values in farm animals preference is for the animal model and associated mixed model methodology. Computationally the matrix of Mendelian sampling terms **D** (Henderson (1976) and Quaas (1976)) plays a central role because it is a diagonal matrix which supports simple derivation of the inverse of the numerator relationships matrix (**A**) in large data sets. The relation between the inverse of **A** and **D** is given by  $\mathbf{A}^{-1} = (\mathbf{I} - \mathbf{M})'\mathbf{D}^{-1}(\mathbf{I} - \mathbf{M})$ , where **I** is the identity matrix and **M** relates animals to their parents. This approach is not directly applicable in honey bees because of the way honey bees reproduce (Bienefeld, Ehrhardt and Reinhardt (2007)) which causes **D** to be block diagonal. **D** is block diagonal because groups of full sibs within a colony can derive from the same drone while paternal half sibs cannot, such that the paternal part of the additive genetic relationship between full sibs and that between half sibs are not equal.

For the background it may be useful to describe some aspects of the reproduction of honey bees. Queens are fertilized by a various drones when flying in open air. A drone mates only once because it dies as a result of mating. After fertilization the queen returns to its colony. The pedigree of drones is controlled by allowing only certain drone producing queens and their colonies on mating stations, for example islands. Artificial insemination also is a possibility. Drones are haploid and produce a multitude of identical sperm cells. A colony therefore consists of groups of workers each descending from a different drone.

Workers within a group have an additive genetic relationship of  $\frac{3}{4}$ . The additive genetic relationship between workers in different groups depends upon whether both paternal drones are from the same drone producing queen or not, and in the latter case upon the additive genetic relationship between both drone producing queens. Drones therefore can be considered as flying gametes producing identical sperm cells, and in that way the drone producing queen can be considered as the sire. Because several drone producing queens deliver the drones mating to one queen we define as a sire the group of those drone producing queens. In a pedigree, a group of full sibs, i.e. deriving from the same queen and sire, consists of a colony and any queen and sire reared from that colony. Selection candidates are unfertilized queens in a colony and their estimated breeding value is that of the group of workers that constitutes the colony.

In the approximation of Bienefeld, Ehrhardt and Reinhardt (2007), earlier described by Bienefeld and Pirchner (1990) (we abbreviate this method as BER) the paternal part of the additive genetic relationship between full sibs and that between half sibs are made equal to their average and **D** is assumed diagonal. The weighing of information from full sibs vs half sibs is clearly not optimal in that way. Bienefeld and Pirchner (1990) and Ehrhardt et al. (2010) presented estimates of genetic parameters using Restricted Maximum Likelihood with BER.

Brascamp and Bijma (2014) (method BB) derived the necessary algebra to calculate **D** accounting for the role of drones in the reproductive system as well as the fact that drones derive from a limited number of drone producing queens. They also investigated the consequences for breeding value estimation.

In this study we estimate genetic parameters using BB and BER, in particular the heritabilities for worker effect and queen effect (both contributing to phenotypes of colonies) and the genetic correlation between these effects. Analyses are done using a simulated dataset. We also explore whether differences in these estimates –if any– affect the relative accuracy of breeding value estimation using both methods.

### MATERIALS AND METHODS

#### Model and components of the D-matrix.

Observations (the phenotype) in honey bees are made on the level of a colony, consisting of a queen and ten thousands of workers. Following Brascamp and Bijma (2014) we define the phenotypic value of a colony as:

$$P_c = \bar{A}_w^W + A_d^Q + E_c \quad (1)$$

where  $P_c$  is the phenotypic value of a colony,  $\bar{A}_w^W$  the average breeding value for worker effect of the group of

**Table 1.** Diagonals (queen, sire and colony) and off-diagonals for full sibs (from the same queen and sire) of the matrix of Mendelian sampling terms  $D$  (multiplied by  $\sigma_A^2$ ) and the additive genetic relationship between full sibs

source	Element in $D$
Queen	$\frac{1}{4}\sigma_A^2(1 - F_d) + \frac{1}{4}\sigma_A^2(1 - F_s) + \frac{1}{4}\sigma_A^2\frac{(S-1)}{S}(1 + F_s - a_{FS})$
Sire	$\frac{var(\delta_i)}{S} + \frac{S-1}{S} cov(\delta_{FS})$
Colony	$cov(\delta_{FS})$
Off-diagonal	$cov(\delta_{FS}) = \frac{2-a_{FS}}{4D} \sigma_A^2$
$a_{FS}$	$\frac{1+F_d}{4} + \frac{1}{2D} + \frac{1+F_s}{4S} + \frac{(DS-D-S)a_{FS_{i-1}}}{4DS} + \frac{a_{sd}}{2}$

The diagonal elements for queens are a function of the inbreeding coefficients of dams ( $d$ ) and sires ( $s$ ), the number of drone producing queens ( $S$ ) and the additive genetic relationship between drone producing queens. The diagonal elements for sires additionally are a function of the covariance between Mendelian sampling terms of full sibs, which happens to equal the diagonal elements for colonies and also the off-diagonals for full sibs. The equation for off-diagonals and also for  $a_{FS}$  depend upon the numbers and the distributions of number of offspring per drone and number of drones per drone producing queen, which are Poisson in Table 1. Note that  $a_{FS}$  is recursive in that the relationship is a function of the same relationship in the sires (among drone producing queens).

workers that constitutes the colony,  $A_d^Q$  the queen effect of the queen in the colony (the dam of the workers) and  $E_c$  is the effect of the environment. Furthermore, ignoring superscripts  $W$  and  $Q$ ,

$$\bar{A}_i = \frac{1}{2}A_d + \frac{1}{2}\bar{A}_s + \bar{\delta}_i \quad (2)$$

describes the average breeding value of a group of workers as the result of the breeding value of the dam of the workers  $A_d$  and the average breeding value of the queens producing the drones (jointly being the sire) mated to the queen ( $\bar{A}_s$ ) and an averaged Mendelian sampling term  $\bar{\delta}_i$ . Note that a similar equation holds for the breeding value of the sire,  $\bar{A}_s$ , being the average breeding value of drone producing queens. The breeding value of a queen can be described as

$$A_i = \frac{1}{2}A_d + \frac{1}{2}\bar{A}_s + \delta_i \quad (3)$$

Brascamp and Bijma (2014) derived elements of the  $D$ -matrix for various assumptions on the numbers of offspring per drone in a colony and the numbers of drones per drone-producing queens. These elements are given in Table 1 assuming that these parameters follow a Poisson distribution.

In Bienefeld et al. (2007) the breeding values of queens, sires and colonies is described by

$$A_i = \frac{1}{2}A_d + qA_s + \delta_i \quad (4)$$

where  $q = \sqrt{\frac{a_{FS} - \frac{1}{4} + a_{pHS}}{2}}$ , with  $a$  the additive genetic relationships for full sibs and paternal half sibs; In this equation  $a_{FS}$  and  $a_{pHS}$  are equilibrium values with zero inbreeding.

**Simulations.** Using equations (1)-(3) we simulated replicates of a dataset with 10 years of data each. For each replicate another pedigree was simulated. The first three years consist of base sires and years 2 and 3 of base queens with colonies. Consecutive years consist of sires, queens and colonies deriving from the base animals. Simulations were done with 10 sires and 500 queens with colonies

annually. Each queen produced 10 full sib groups consisting of 5 full sibs each. Mates (sires) were assigned to queens randomly with equal numbers of matings for each sire. To estimate genetic parameters data were simulated without selection taking place. For further details see Brascamp and Bijma (2014). We looked into two scenarios with 100 replicates of simulated data each. In the first scenario each queen was fertilized by 12 drones from 8 drone producing queens constituting a sire. In the second, more realistic scenario, one third of the queens was inseminated by AI with 6 drones from 6 drone producing queens, one third was considered to be mated on large mating stations (12 drones and 30 drone producing queens) and one third on small mating stations (12 and 8). For BER only the latter numbers were used in accordance with Bienefeld et al. (2007) as this method cannot accommodate varying numbers. Note that in these simulations we assume that drone producing queens are full sibs, which for example is the case in the German breeding programme which uses breeding values estimated on the basis of BER.

Variances for environment, additive genetic worker effect and additive genetic queen effect have 2, 1, 0.5 as input, while the genetic correlation between worker and queen effect is set to -0.50.

To look into the effect of resulting genetic parameters on the accuracy of breeding value estimation we ran 100 replicates of simulated data with the same population structure as before and estimated breeding values for BB and BER. Because we found that BB produces unbiased genetic population parameters (see results) while BER does not, for BER we looked at two variants using either the population genetic parameters or those estimated with BER itself.

**Analyses.** We analysed the data using ASReml 3.0 (Gilmour et al. (2009)), to produce estimates of the (co)variance components for the breeding values for worker and queen effect and the residual effect for each replicate. The model contained only one fixed effect ( $\mu$ ) and the random worker and queen effects and residual. Inputs were the simulated observations on colonies with colony and queen identifications and furthermore the inverse of the NRM for either BB or BER. Accuracies of breeding value estimation were calculated as the correlations between estimated and true breeding values in year 10 of young unfertilized queens for each replicate. We present averages and standard deviations of the means (standard errors) to evaluate results.

## RESULTS AND DISCUSSION

**Genetic parameters.** Table 2 summarizes the estimated genetic parameters for scenario 1 where all matings were simulated with 12 drones from 8 drone producing queens. It appears that BB estimates genetic parameters very closely to input and may be considered unbiased. BER in particular underestimates the additive genetic variance for worker effect. This probably is due to the fact that using BER the breeding value of a colony is modelled relating to a single animal, while it is the mean of a group of workers. In both BB and BER the breeding value of a queen is modelled as relating to an individual animal and both methods estimate the associated variance

**Table 2.** Means (m) and standard errors (s) of genetic parameters of 100 replicates of scenario 1.

method	(co)variance				$h_W^2$	$h_Q^2$	$r_g$
	$A^W$	$A^Q$	$cov$	$E$			
true	1.00	0.50	-.35	2.00	0.38	0.19	-.50
m BB	1.03	0.51	-.37	2.00	0.39	0.19	-.50
m BER	0.63	0.43	-.20	1.68	0.25	0.16	-.37
s BB	0.03	0.02	0.02	0.01	0.01	0.01	0.02
s BER	0.02	0.01	0.01	0.01	0.01	0.01	0.02

W and Q refer to worker and queen effect. In scenario 1 the numbers of drones and drone producing queens are constant (12 and 8). The values for heritabilities and  $r_g$  are derived from those of the (co)variance components.

**Table 3.** Means (m) and standard errors (s) of genetic parameters of 100 replicates of scenario 2.

method	(co)variance				$h_W^2$	$h_Q^2$	$r_g$
	$A^W$	$A^Q$	$cov$	$E$			
true	1.00	0.50	-.35	2.00	0.38	0.19	-.50
m BB	0.98	0.52	-.35	1.99	0.37	0.20	-.49
m BER	0.60	0.41	-.17	1.70	0.20	0.16	-.33
s BB	0.03	0.02	0.02	0.01	0.01	0.01	0.02
s BER	0.02	0.01	0.01	0.01	0.01	0.01	0.02

W and Q refer to worker and queen effect. In scenario 2 the numbers of drones and drone producing queens vary. For one thirds of the matings the numbers are 6 and 6, 12 and 30 and 12 and 8, respectively. The values for heritabilities and  $r_g$  are derived from those of the (co)variance components.

component similarly. The genetic correlation between worker and queen effects is estimated by BER more closely to zero (less negative) than by BB. Also in scenario 2, where there are three different systems of mating (AI, large and small mating stations) BB estimates the genetic parameters close to input (Table 3). The differences between BB and BER are similar to scenario 1, with the exception of the genetic correlation which BER seems to estimate still closer to zero. We have no explanation for the different estimates of the genetic correlation using BB or BER. Ironically, we initiated this research because of very strong negative estimates for genetic correlations using BER by Bienefeld and Pirchner (1990) and we hypothesised that it was due to the model used. Their estimates were around -0.90 for several traits. Later estimates by Ehrhardt et al. 2010, were moderately negative with -0.63 for Varroa population growth and -0.13 for hygienic behaviour. Our analysis shows, however, that BER produces estimates biased towards zero.

**Table 4.** Regressions of TBV on EBV and correlations between TBV and EBV in year 10 of 100 replicates of scenario 1. Standard errors between brackets.

method	genetic parameters	regression	correlation
		BB	0.994 (.019)
BER	BB	0.690 (.013)	0.319 (.006)
BER	BER	0.870 (.016)	0.331 (.007)

TBV and EBV relates to true and estimated breeding values, respectively

**Estimation of breeding values.** Table 4 shows regression coefficients of true (simulated) breeding values (TBV) on estimated breeding values (EBV). For BB these values don't differ significantly from unity which is according to theory. For BER the regression coefficients are smaller than 1. Nevertheless the correlations between EBV's and TBV's are similar though higher for BB. Values are lowest for BER using population parameters. When the BER estimated genetic parameters are used the correlations between EBV and TBV are intermediate.

## CONCLUSION

The methods to derive the NRM in honey bees presented here yields unbiased estimates of genetic parameters and breeding values. REML estimates of genetic parameters using BER appear in particular to underestimate the variance component for the additive genetic variance for worker effect. In our analyses, with a true genetic correlation of -0.50 the BER-estimate of the genetic correlation shows a bias towards zero. We estimated breeding values using BB and BER. It appears that the ranking of animals using both methods is similar but favourable for BB.

## ACKNOWLEDGMENTS

EWB is grateful to Hendrik-Jan Megens and Alex Hulzebosch showing him the way on the Breed4Food (B4F) cluster, the joint High Performance Compute (HPC) infrastructure of the Animal Breeding and Genomics Centre. PB acknowledges financial support of the Netherlands Organization for Scientific Research (NWO).

## LITERATURE CITED

- Bienefeld, K., Ehrhardt, K. and Reinhardt, F. (2007). *Apidologie* 38:77-85.
- Bienefeld, K. and Pirchner, F. (1990). *Apidologie* 21: 175-183.
- Brascamp, E.W. and Bijma P. (2014). Submitted to *Gen. Sel. Evol.*
- Ehrhardt, K., Büchler, R. and Bienefeld, K. (2010) *Proc 9th WCGALP*, P0565\_PP1-95.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R. (2009). *ASReml* 3.0.
- Henderson, C.R. (1976). *Biometrics* 32:69-83.
- Quaas, R.L. (1976). *Biometrics* 32:949-953.