

## Optimising multi-tier open nucleus breeding schemes

R. K. Shepherd<sup>1</sup> and B. P. Kinghorn<sup>2</sup>

<sup>1</sup> Queensland Department of Primary Industries, P.O. Box 6014, Rockhampton Mail Centre, Queensland 4702, Australia

<sup>2</sup> Department of Animal Science, University of New England, Armidale, New South Wales 2351, Australia

Received November 11, 1991; Accepted May 20, 1992

Communicated by L. D. Van Vleck

**Summary.** The constant migration (**CM**) method and the *ebv* migration (**EBVM**) method of optimising the design of multi-tier open nucleus breeding schemes are presented and compared. The equation for the equilibrium rate of genetic gain of a three-tier open nucleus scheme is determined using the **CM** method. The major advantage of the **EBVM** method is the reduction in the number of parameters which have to be varied in order to locate the maximum equilibrium rate of genetic gain. For the **CM** method for the number of variable parameters is 5, 14, 27 and  $(2n+1)(n-1)$  for unrestricted male and female migration in schemes with 2, 3, 4 and  $n$  tiers respectively. The corresponding number of variable parameters for the **EBVM** method is 1, 2, 3 and  $n-1$  respectively. A procedure is given for the **EBVM** method so as to accommodate variance loss due to selection and variance gain due to the mixing of groups with a different mean breeding value.

**Key words:** Open nucleus – Selection – Optimum design – Migration methods

### Introduction

Animal breeding programmes are usually hierarchical in structure and often represented by a pyramid consisting of a number of tiers. Elite nuclei or studs are at the apex of the pyramid. Traditionally, gene flow is all downward in the pyramid and, in particular, the top tier or nucleus remains closed to gene transfer from lower tiers. By allowing the nucleus to import genes from lower tiers, it has been shown theoretically that open nucleus schemes can increase the rate of genetic improvement and can decrease the rate of inbreeding compared with closed nucleus systems (James 1977, 1978).

James (1977) developed the theory for two-tier open nucleus systems based on between-tier migration rates which remain constant over generations. He also used his constant migration (**CM**) theory to determine the population structure and migration rates which maximised the equilibrium rate of genetic gain for two-tier schemes in which female migration was unrestricted while all sires were bred in the nucleus. Mueller and James (1983) developed a **CM** theory in which the assumption of constant genetic variance used by James (1977) was relaxed and, in particular, the loss of genetic variance due to selection was taken into account.

Hopkins and James (1978) gave the theory for optimising two-tier open nucleus systems when generations overlap. This theory was based on migration rates which were allowed to vary for each time period in order that the genetic mean of the next progeny crop was maximised. In essence, the migration rates were a result of truncation selection on the estimated breeding value across nominated parental genetic groups. This method is called the *ebv* migration (**EBVM**) method. Hopkins (1978) used the **EBVM** method to optimise the design of various two-tier open nucleus schemes with overlapping generations. Neither Hopkins and James (1978) nor Hopkins (1978) incorporated the loss of genetic variance due to selection in their method.

Most theory on open nucleus breeding systems relate to the two-tier problem and the same is true for evaluations of optimum designs. As stated by James (1989, p 191), the reason for this has been mainly due to the large number of parameters involved in the optimisation of multi-tier systems. However as shown in the present paper, the number of variable parameters depends on the optimisation procedure chosen. The objective of this paper is to present and compare the **CM** and **EBVM** methods for optimising the design of multi-tier open nucleus

breeding systems. The methods will be illustrated using a three-tier open nucleus example. A procedure which accounts for variance loss due to selection will also be given for the **EBVM** method. Later papers will use these methods to evaluate the improvement in genetic progress and the reduction in inbreeding resulting from optimum multi-tier open nucleus designs.

### Constant migration method

The **CM** method is comprehensively discussed by James (1977) for a two-tier open nucleus breeding system. This section generalises the **CM** method to an  $n$ -tier open nucleus system.

The population size of each tier is assumed to be infinite and generations are assumed to be discrete. The symbols  $a$  and  $b$  are used to represent, respectively, the proportions of all male and female offspring used for breeding. The vector  $p$  consists of the  $n-1$  elements  $p_i$  where each  $p_i$  represents the fraction of the population in tier  $i$ . The proportion of the population in the base or tier  $n$  ( $p_n$ ) is  $1 - \sum_{i=1}^{n-1} p_i$ . With unrestricted migration parents can be born in any tier and mated in any tier. Let the vector  $x$  consist of elements which represent the proportion of females mated in tier  $i$  that were born in tier  $j$  where  $i < j$ . Hence  $x$  is the upward female migration rate vector. The number of elements in  $x$  is  $\frac{1}{2}n(n-1)$  which is the number of elements above the diagonal in a square matrix of order  $n$ . Some of these elements will be set to zero if migration is not allowed between certain tiers. The ordering of the elements of  $x$  is arbitrary. The downward female migration rate vector  $y$  consists of elements which represent the proportion of females mated in tier  $j$  that were born in tier  $i$  where  $i < j$ . Similarly,  $v$  and  $w$  represent the male upward and downward migration rate vectors, respectively.

In the **CM** method, the elements of the migration rate vectors ( $x, y, v, w$ ) are assumed constant from one generation to the next so that equations can be determined for the equilibrium genetic gain and the equilibrium genetic lags between tiers. To optimise the equilibrium rate of genetic gain the optimum equilibrium values of ( $x, y, v, w$ ) are determined and are assumed to operate each generation. In essence the **CM** method considers the distribution of the estimated breeding value within each tier to be truncated at a number of points such that relative to the tier mean the truncation points are unchanged each generation. The best animals in each tier are mated in the top tier and so on. However, before equilibrium the best animals in the population are not necessarily mated in the nucleus, as the truncation points are only optimal at equilibrium.

The **CM** equations are determined as follows. If  $G^t$  is the vector of tier genetic gains from generation  $t-1$  to

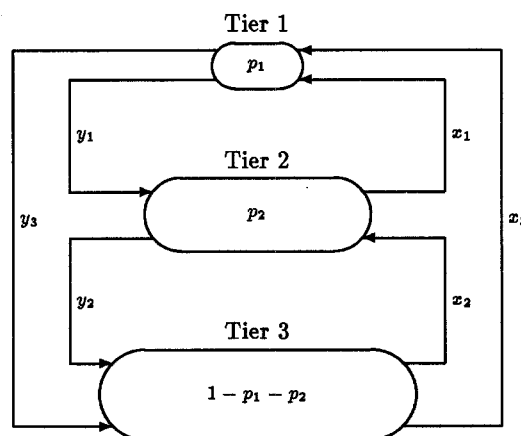


Fig. 1. The pattern of unrestricted female migration in a three-tier open nucleus breeding scheme where  $p_i$  is the proportion of the population in tier  $i$ . The female migration rates upward ( $x$ ) and downward ( $y$ ) indicate the proportion of dams mated in a tier (shown by the arrowhead) that were born in another tier

generation  $t$  and  $\mu^t$  is the vector of tier means at generation  $t$  it follows that

$$G^t = \mu^t - \mu^{t-1}$$

Writing  $\mu^t$  in terms of  $\mu^{t-1}$  and the assumed constant selection differentials, this equation becomes

$$G^t = C + T L^{t-1} \quad (1)$$

where  $C$  is the vector of average genetic selection differentials of animals mated in each tier,  $L^{t-1}$  is the vector of all genetic lags between tiers at generation  $t-1$  [dimension  $\frac{1}{2}n(n-1)$ ], and  $T$ , a matrix of migration rates, is of dimension  $n$  by  $\frac{1}{2}n(n-1)$ . There are  $\frac{1}{2}n(n-1)$  lags but only  $n-1$  are independent. Now the matrix equation 1 can be rewritten using two results. Firstly the change in each genetic lag is the difference between the genetic gains of the two tiers involved (e.g.  $\Delta L_{11}^t = G_{T_1}^t - G_{T_2}^t$ ). Secondly these changes approach zero as the genetic lags approach constant equilibrium values [see Shepherd (1991) for an investigation of the existence of equilibrium]. Using these results, equation 1 becomes

$$T^* L^\infty = C^* \quad (2)$$

where  $T^*$ , a matrix of migration rates, is of dimension  $\frac{1}{2}n(n-1)$  by  $\frac{1}{2}n(n-1)$ ,  $L^\infty$  is the vector of equilibrium genetic lags, and  $C^*$  is a matrix of differences between the elements of  $C$ . Now there are  $(\frac{1}{2}n-1)(n-1)$  constraints on the equilibrium genetic lags (e.g., the lag between tiers 1 and 3 is the sum of the lags between 1 and 2, and 2 and 3). In matrix notation the constraints are of the form  $L^\infty = X L^\infty$ . Using these constraints in conjunction with equation 2 results in a unique solution for  $L^\infty$ . At equilibrium all tiers progress at the same rate, so substituting  $L^\infty$  into equation 1 produces a formula for the equilibrium rate of genetic gain  $G^\infty$  which is of the general functional

form

$$G^\infty = \text{function}(a, b, p, x, y, v, w) \quad (3)$$

For any species denoted by a particular  $(a, b)$  combination, the equilibrium rate of genetic gain  $G^\infty$  is maximised by searching a grid of values in the  $(p, x, y, v, w)$  parameter space for this maximum value.

The remainder of this section illustrates the form of equation 3 for a three-tier open nucleus system in which male and female migration is unrestricted between tiers. Figure 1 shows the pattern of unrestricted female migration;  $x_3$  is the proportion of tier-1 dams which are born in tier-3, while  $y_2$  is the proportion of tier-3 dams which are born in tier-2. A similar figure can be drawn for males by replacing  $x$  with  $v$  and replacing  $y$  with  $w$ .

Applying the CM method to this three-tier scheme [see Shepherd (1991) for details], results in the following equation for the equilibrium rate of genetic gain  $G^\infty$

$$G^\infty = \frac{\lambda_1 C_{T_1} + \lambda_2 C_{T_2} + \lambda_3 C_{T_3}}{\lambda_1 + \lambda_2 + \lambda_3} \quad (4)$$

where  $C_{T_i}$  is the average genetic selection differential of all animals mated in tier  $i$ , and

$$\lambda_1 = u_2 d_3 + d_1 d_2 + d_1 d_3$$

$$\lambda_2 = u_1 d_2 + u_1 d_3 + u_3 d_2$$

$$\lambda_3 = u_1 u_2 + u_2 u_3 + u_3 d_1$$

in which  $u = [\frac{1}{2}(v + x)]$  and  $d = [\frac{1}{2}(w + y)]$  are the average upward and downward migration rate vectors respectively. Appendix I shows that each  $C_{T_i}$  is a function of the elements of vectors  $p, x, y, v$  and  $w$ . So equation 4 is now in the form of equation 3. Hence for any  $(a, b)$  combination, the equilibrium rate of genetic gain can be maximised by searching a grid of values in the  $(p, x, y, v, w)$  parameter space. In this example the dimension of the parameter space is 14 as there are two  $p$ s and three elements in each of the four migration rate vectors.

Two special cases of equation 4 warrant further examination. If the third tier doesn't supply any breeding animals upward ( $u_2 = u_3 = 0$ ) then, after cancelling  $d_2 + d_3$  from the numerator and the denominator,

$$G^\infty = \frac{d_1 C_{T_1} + u_1 C_{T_2}}{d_1 + u_1}.$$

This is the equation for the equilibrium rate of genetic gain of a two-tier open nucleus system as the third tier no longer influences the asymptotic genetic gain. The three-tier equations given by James (1977) can be formed from the above equations by setting  $w_1 = w_2 = 1$ ,  $v_i = 0$  for all  $i$ , and  $u_3 = d_3 = 0$ . That is, female migration is allowed in both directions but only between adjacent tiers, while each tier's sires are bred in the tier above. More generally, if male and female migration is only allowed between adjacent tiers ( $u_3 = d_3 = 0$ ), but in both directions, then  $G^\infty = \frac{d_1 d_2 C_{T_1} + u_1 d_2 C_{T_2} + u_1 u_2 C_{T_3}}{d_1 d_2 + u_1 d_2 + u_1 u_2}$ .

Although the CM method can be used to derive the equation for the equilibrium rate of genetic gain for any open nucleus system, the amount of algebra increases exponentially with the number of tiers. Shepherd (1991) derived the equation for the equilibrium rate of genetic gain for a four-tier open nucleus system in which both male and female migration was unrestricted between tiers.

### Ebv migration method

In the EBVM method the elements of the migration rate vectors  $(x, y, v, w)$  are allowed to vary each generation according to the  $n$ -tier generalisation of the two-tier principle discussed by Hopkins and James (1978). With unrestricted migration the principle is as follows. For each generation the best animals, irrespective of tier of birth, are randomly mated in tier 1. The next best animals are randomly mated in tier 2 and so on down to tier  $n$ . In other words the estimated breeding value (ebv) of any dam in tier  $i$  is not only greater than the ebv of any dam in tier  $j$  if  $i < j$  but also less than the ebv of any dam in tier  $k$  if  $i > k$ . A similar statement applies to sires.

The following algorithm describes the EBVM method for a three-tier system allowing unrestricted migration between tiers. In each generation proceed as follows.

*Step 1.* Using known genetic means and variances, line up the assumed normal distributions of progeny ebv in each tier using the same abscissa scale.

*Step 2.* Select as dams for tier 1 all female progeny whose ebv is larger than a common truncation point  $t_1$  which is chosen such that the best  $b p_1$  of all female progeny become tier-1 dams. In mathematical terms, the proportion of female progeny in each tier selected as dams for tier 1 ( $q_{i1}$ ) must satisfy the equation  $\sum_i p_i q_{i1} = b p_1$  while the

standardised ebv truncation point in tier  $i$  ( $X_{i1}$ ) must satisfy  $t_1 = \mu_i + X_{i1} \sigma_I$  where  $\mu_i, \sigma_I$  and the subscripts of  $q$  are defined in appendix I.

*Step 3.* Select as dams for tier  $j$  ( $j = 2, 3$ ) all female progeny whose ebv is not only smaller than the previous common truncation point  $t_{j-1}$  but also greater than another common truncation point  $t_j$  which is chosen such that the best  $b P_j$  of all female progeny become dams in tiers 1 to  $j$  with the best  $b P_{j-1}$  in tiers 1 to  $j-1$ , where  $P_j = \sum_{k=1}^j p_k$ .

In mathematical terms, determine the proportion of female progeny in each tier selected as dams for tier  $j$  ( $q_{ij}$ ) using  $q_{ij} = Q_{ij} - Q_{i,j-1}$  where  $Q_{i1} = q_{i1}$  and the other proportions  $Q_{ij}$  satisfy  $\sum_i p_i Q_{ij} = b P_j$  and the standardised

ebv truncation point in tier  $i$  ( $X_{ij}$ ) must satisfy  $t_j = \mu_i + X_{ij} \sigma_I$  for  $i = 1, 2, 3$ .

*Step 4.* In an analogous manner repeat steps 2 and 3 for males.

*Step 5.* The migration rates ( $x, y, v, w$ ) are calculated as follows:

$$\begin{aligned} x_1 &= \frac{p_2 q_{21}}{b p_1}, x_2 = \frac{p_3 q_{32}}{b p_2}, x_3 = \frac{p_1 q_{31}}{b p_1} \\ y_1 &= \frac{p_1 q_{12}}{b p_2}, y_2 = \frac{p_2 q_{23}}{b p_3}, y_3 = \frac{p_1 q_{13}}{b p_3} \\ v_1 &= \frac{p_2 q_{21}}{a p_1}, v_2 = \frac{p_3 q_{32}}{a p_2}, v_3 = \frac{p_1 q_{31}}{a p_1} \\ w_1 &= \frac{p_1 q_{12}}{a p_2}, w_2 = \frac{p_2 q_{23}}{a p_3}, w_3 = \frac{p_1 q_{13}}{a p_3} \end{aligned}$$

These equations are simply rearrangements of the formulae in appendix I for the female proportions selected and the corresponding formulae for males. Unlike the assumed constant migration rates in the **CM** method these migration rates are allowed to vary each generation.

*Step 6.* The mean progeny breeding value for each tier is calculated using the parents chosen for each tier. To do this the genetic selection differentials ( $D$ ) of the 18 types of parents need to be determined (two sexes by three tiers of birth by three tiers of mating). The genetic selection differentials are given by  $i\sigma_I$  where  $i$  is the standardised selection differential for the type of parent considered and  $\sigma_I$  is the standard deviation of the selection index  $I$ . The expressions for the selection differentials  $i$  are the same as those for the **CM** method in appendix I. Using these expressions, the mean progeny breeding value of tier  $j$  in generation  $t+1$  is

$$\mu_j^{t+1} = \frac{1}{2} \sum_i [a_{ij}(\mu_i^t + D_{ij}) + a_{ij}^*(\mu_i^t + D_{ij}^*)]$$

where  $D_{ij} = i_{ij}\sigma_I$  and  $D_{ij}^* = i_{ij}^*\sigma_I$ . The  $a_{ij}$  and  $a_{ij}^*$  are defined in appendix I.

*Step 7.* Repeat steps 1 to 6 for the next generation.

Eventually all tiers make progress at the same rate. Hence, although a closed-form solution like equation 4 cannot be found for the equilibrium rate of genetic gain  $G^\infty$ , we have implicitly that

$$G^\infty = \text{function}(a, b, p) \quad (5)$$

irrespective of the initial genetic means of each tier. For any species denoted by a particular ( $a, b$ ) combination, the equilibrium rate of genetic gain  $G^\infty$  is maximised by searching a grid of values in the  $p$  parameter space for this maximum value. For the three-tier example the maximum  $G^\infty$  is found by searching a two-dimensional grid of values in the ( $p_1, p_2$ ) parameter space.

To numerically implement the **EBVM** algorithm requires the following: a procedure for evaluating functions of a single normal distribution and a procedure for determining a common truncation point across a number of normal distributions. The NAG or IMSL computer sub-routine libraries can be used to evaluate the truncation point and the selection intensity for a standard normal

distribution. A bisection-search algorithm can be used to locate a common truncation point across a group of normal distributions.

## Discussion and comparison of the methods

The main properties of the **CM** and **EBVM** methods are summarised in Table 1. The major advantage of the **EBVM** method is the reduction in the number of parameters which have to be varied in order to locate the maximum equilibrium rate of genetic gain. For the **CM** method the number of variable parameters is 5, 14, 27 and  $(2n+1)(n-1)$  for unrestricted male and female migration in schemes with 2, 3, 4 and  $n$  tiers respectively. The corresponding number of variable parameters for the **EBVM** method is 1, 2, 3 and  $n-1$  respectively. The basic feature of the **EBVM** method is better described in terms of geometry of the  $G^\infty$  surface. For each  $p$  vector there is a particular set of ( $x, y, v, w$ ), denoted ( $x, y, v, w$ )<sub>opt</sub>, which maximises  $G^\infty$ . For example, Table 2 shows the ( $x, y, v, w$ )<sub>opt</sub> values which maximise  $G^\infty$  for each  $p$  in a particular two-tier system. Any other set of ( $x, y, v, w$ ) will result in a lower genetic gain at equilibrium for each  $p$ . Basically, with the **EBVM** method the search is only over the  $n-1$  dimensional plateau of ( $x, y, v, w$ )<sub>opt</sub> values in the  $(2n+1)(n-1)$  dimensional space. For two tiers the search is only over the one dimensional  $p$  ridge in the five dimensional ( $p, x, y, v, w$ ) space as shown in Table 2.

Another feature of the **CM** method is that the number of dimensions of the search space depends on the pattern of migration allowed between tiers. For example, in an  $n$ -tier system the search space is of dimension  $(2n+1)(n-1)$  for unrestricted male and female migration, and of dimension  $3(n-1)$  if female migration is allowed in both directions but only between adjacent tiers while each tier's sires are bred in the tier above. The number of dimensions of the search space for the **EBVM** method is independent of the pattern of migration allowed between tiers.

Table 2 shows the equilibrium genetic variance between the nucleus and base tiers  $V_{Ab}^\infty$  for each  $p$  value in a two-tier system. Hence it is easy to show with the **EBVM** method that the  $p$  value (0.07) which maximizes  $G^\infty$  is much smaller than the nucleus size (0.25) which produces the greatest between-population variance. The discrepancy in optimum nucleus size is because  $G^\infty$  is no longer simply related to the population genetic variance by the equation  $G^\infty = ir\sigma_A^2$ , where  $r$  is the accuracy of selection, as in a normally distributed population. The genetic lag between the nucleus and base has produced substantial non-normality in the population-distribution of breeding values. The **CM** method can also be used to show this result but optimisations with respect to the four migration rates have to be performed for each  $p$ .

**Table 1.** The main properties of the **CM** and **EBVM** methods

CM method	EBVM method
Equation for equilibrium genetic gain	Algorithm for equilibrium genetic gain
Equation for equilibrium genetic lags	Algorithm for equilibrium genetic lags
Non-optimum migration until equilibrium	Optimum migration at each generation
Dimension depends on migration	Dimension independent of migration
Unrestricted migration has a $(2n+1)(n-1)$ dimensional search space	Unrestricted migration has an $(n-1)$ dimensional search space

**Table 2.** The values of  $(x, y, v, w)$  which maximise the equilibrium genetic gain  $G_{\text{open}}^{\infty}$  for various nucleus sizes  $p$  of a two-tier open nucleus system.  $R^{\infty} (= G_{\text{open}}^{\infty}/G_{\text{closed}}^{\infty})$  is the relative response rate with respect to a closed nucleus.  $V_{Ab}^{\infty} [= p(1-p)(\mu_N^{\infty} - \mu_B^{\infty})^2]$  is the additive genetic variance between nucleus and base tiers at equilibrium in units of  $\sigma_I$ . The male and female replacement rates are 0.01 and 0.5 respectively. Constant variances are assumed

$p$	$x_{\text{opt}}$	$y_{\text{opt}}$	$v_{\text{opt}}$	$w_{\text{opt}}$	$R^{\infty}$	$V_{Ab}^{\infty}$
0.01	0.47	0.01	0.03	0.38	1.085	0.048
0.04	0.51	0.06	0.09	0.59	1.110	0.099
0.07	0.51	0.10	0.11	0.66	1.114	0.126
0.10	0.49	0.14	0.12	0.69	1.113	0.143
0.13	0.48	0.18	0.13	0.72	1.111	0.154
0.16	0.46	0.21	0.13	0.74	1.108	0.161
0.19	0.44	0.25	0.13	0.76	1.104	0.166
0.22	0.43	0.28	0.13	0.77	1.100	0.168
0.25	0.41	0.31	0.13	0.79	1.096	0.169
0.28	0.39	0.35	0.13	0.80	1.092	0.168
0.31	0.38	0.38	0.13	0.81	1.088	0.166

A feature of the **EBVM** method is that it gives the optimal migration rates at each generation for maximising progeny merit (Hopkins and James 1978), whereas the **CM** method uses constant migration rates at each generation. Shepherd (1991) showed that migration on estimated breeding value is required at each generation when maximising the genetic mean of the population over any time horizon irrespective of whether or not the  $p$  vector varies from generation to generation. Hence the **EBVM** method will reach equilibrium quicker than any other method with a fixed  $p$  as its genetic mean and will always be greater than, or equal to, the genetic mean of any other method. To obtain this advantage the **EBVM** method assumes and uses knowledge of the genetic lags between tiers in the foundation generation. This knowledge is rarely known accurately in practice. However, as genetic links are created between tiers by migration, the genetic lags can be estimated on a pooled within- and

between-tier basis by eliminating the environmental effects associated with between-tier comparisons.

The assumption of constant variance can be relaxed in either method using the  $n$ -tier generalisation of the two-tier theory presented by Mueller and James (1983). Appendix II shows the modifications required to the **EBVM** algorithm so as to relax this assumption in the three-tier open nucleus example. In most industry schemes the accuracy of selection is higher in males and also in tiers with greater expected genetic means. Although not explicitly shown in this paper, both methods can take account of variance loss due to selection irrespective of whether the accuracy of selection differs between tiers and sexes.

The multi-tier open nucleus theory given in this paper was developed for the discrete generation model. Hopkins and James (1978) gave the theory for two-tier open nucleus systems with overlapping generations when *parent selection* and *progeny selection* strategies are used. *Progeny selection* is when breeding animals, once selected, are guaranteed a place in the breeding unit until they die or are culled for age. *Parent selection* is when previously selected animals must compete each year to remain in the breeding unit. Meuwissen (1989) described an **EBVM** method for optimising the design of a dairy cattle breeding programme using BLUP selection. Although he stated that there were three tiers, he really only used a two-tier structure in the design of the open nucleus system because he defined the genetic groups (nucleus born males, nucleus born females, and base born females) as his three tiers. Basically he used a *parent selection* strategy in a two-tier open nucleus structure which incorporated progeny testing. The theory developed by Hopkins and James (1978), and Meuwissen (1989) for two-tier systems with overlapping generations can be easily extended in principle to multi-tier open nucleus systems. However the notation and methods become cumbersome due to the large number of genetic groups determined by the number of age classes, the place of birth, and the pattern of migration, allowed between tiers.

Finally, the ease of optimisation of multi-tier systems with the **EBVM** method opens up an area of quantitative genetics which, as discussed by Smith and Hammond (1987), has not been modelled accurately in the past. The problem is the prediction of response to selection within a population when mating is assortative rather than random. The results of work in this area will be reported in later papers.

**Acknowledgements.** This work was carried out while R.K.S. held an Australian Wool Corporation Postgraduate Scholarship and was on study leave at the University of New England. A referee is thanked for suggesting improvements to the notation.

## Appendix I

Genetic selection differential ( $C_{T_j}$ ) as a function of  $p$ ,  $x$ ,  $y$ ,  $v$ , and  $w$

Let  $a_{ij}$  be the proportion of tier  $j$  dams which are born in tier  $i$ . Then the matrix  $A$  is as follows

$$A = \begin{bmatrix} 1-x_1-x_3 & y_1 & y_3 \\ x_1 & 1-y_1-x_2 & y_2 \\ x_3 & x_2 & 1-y_2-y_3 \end{bmatrix}$$

The corresponding proportions for males are denoted  $a_{ij}^*$  with matrix  $A^*$ . Using these definitions the average genetic selection differential of all animals mated in tier  $j$  ( $C_{T_j}$ ) is as follows

$$C_{T_j} = \frac{1}{2} \sum_i (a_{ij} D_{ij} + a_{ij}^* D_{ij}^*)$$

where  $D_{ij}^*$  and  $D_{ij}$  are the male and female genetic selection differentials respectively. The genetic selection differentials are defined as follows. At generation  $t$  is the mean breeding value of tier 1-born males mated in tier 2 is  $\mu_1' + D_{12}^*$  where  $\mu_1'$  is the mean breeding value of tier 1 at generation  $t$  and  $D_{12}^*$  is the genetic selection differential of tier 1-born males mated in tier 2. Similarly,  $D_{31}$  is the genetic selection differential of tier 3-born females mated in tier 1.

Parents are chosen by truncation selection on the basis of an index  $I$  which is normally distributed. The female (male) genetic selection differentials  $D(D^*)$  are given by  $i\sigma_I$  ( $i^*\sigma_I$ ) where  $i(i^*)$  is the standardised selection differential for this group of female (male) parents and  $\sigma_I$  is the standard deviation of the selection index  $I$ . The standardised selection differentials are readily calculated because the index is normally-distributed. Let  $S(q)$  denote the standardised selection differential achieved by truncation selection of the best proportion  $q$  of animals. Define  $q_{31}$  as the proportion of tier 3-born females mated in tier 1 and so on. Then the nine female proportions selected are as follows

$$\begin{aligned} q_{11} &= b(1-x_1-x_3), \quad q_{12} = \frac{by_1p_2}{p_1}, \quad q_{13} = \frac{by_3p_3}{p_1} \\ q_{21} &= \frac{bx_1p_1}{p_2}, \quad q_{22} = b(1-x_2-y_1), \quad q_{23} = \frac{by_2p_3}{p_2} \\ q_{31} &= \frac{bx_3p_1}{p_3}, \quad q_{32} = \frac{bx_2p_2}{p_3}, \quad q_{33} = b(1-y_2-y_3) \end{aligned}$$

The standardised selection differentials of tier  $k$ -born females mated in tier  $j$  ( $i_{kj}$ ) are given by

$$i_{k1} = S(q_{k1})$$

and for  $j > 1$  by

$$i_{kj} = [Q_{kj} S(Q_{kj}) - Q_{k,j-1} S(Q_{k,j-1})] / q_{kj}$$

where  $Q_{kj} = \sum_{l=1}^j q_{kl}$ .

Then  $D_{kj} = i_{kj} \sigma_I$ . The male genetic selection differentials ( $D_{kj}^*$ ) are obtained in a similar manner. Hence we have shown that each  $C_{T_j}$  is a function of elements of the vectors  $p$ ,  $x$ ,  $y$ ,  $v$ , and  $w$ .

## Appendix II

Accommodating variance loss and gain in the **EBVM** algorithm

This appendix outlines the alterations necessary to the **EBVM** algorithm of an unrestricted three-tier open nucleus system in order to allow for different losses of genetic variance from the selection of males and females from each tier, and to allow for variance increase from the mixing of groups with different mean breeding values. The methodology used to calculate the variance of progeny breeding values at each generation is an extension of that given by Mueller and James (1983) for constant migration in two-tier systems.

Consider a single trait with heritability  $h_1^2$ ,  $h_2^2$  and  $h_3^2$  and additive genetic variance  $V_{A1}$ ,  $V_{A2}$  and  $V_{A3}$  in tier 1, tier 2 and tier 3 respectively. The logic described in steps 1 to 4 is still correct, but account has to be taken of the different genetic variances in each tier. Hence the identities for the proportions selected remain correct in steps 2 to 4 but the standardised truncation points  $X$  for females must now satisfy:

$$\mu_1 + X_{11} h_1 \sigma_{A1} = \mu_2 + X_{21} h_2 \sigma_{A2} = \mu_3 + X_{31} h_3 \sigma_{A3}$$

$$\mu_1 + X_{12} h_1 \sigma_{A1} = \mu_2 + X_{22} h_2 \sigma_{A2} = \mu_3 + X_{32} h_3 \sigma_{A3}$$

$$\mu_1 + X_{13} h_1 \sigma_{A1} = \mu_2 + X_{23} h_2 \sigma_{A2} = \mu_3 + X_{33} h_3 \sigma_{A3}$$

where  $\sigma_{Ai} = \sqrt{V_{Ai}}$ . Similar identities can be given for males.

The calculation of the migration rates in step 5 remains unchanged as does the calculation of the standardised selection differentials  $i$  and  $i^*$  in step 6. However the genetic selectional differentials  $D_{ij}$  and  $D_{ij}^*$  are now equal to  $i_{ij} h_i \sigma_{Ai}$  and  $i_{ij}^* h_i \sigma_{Ai}$  respectively. A new step is required between steps 6 and 7 to calculate the additive genetic variance of each tier's new progeny crop as genetic variance is lost due to the selection of parents. Only the female equations will be described as the male equations are obtained in a similar manner.

*Step 6a.* Calculate the mean breeding values of the selected female groups which are born in tier  $i$  and either mated in tier  $j$  ( $m_{ij}$ ) or in tiers 1 to  $j$  ( $M_{ij}$ ) using the following equations.

$$m_{ij} = \mu_i + i_{ij} h_i \sigma_{Ai}$$

$$M_{ij} = \mu_i + i_{Q_{ij}} h_i \sigma_{Ai}$$

where  $i_{Q_{ij}} = S(Q_{ij})$ . Now the breeding value variance ( $V_{Aij}$ ) of dams born in tier  $i$  which are mated in tier  $j$  is

given by

$$V_{A_{i1}} = V_{A_i} K_{i1}$$

and for  $j > 1$  by

$$V_{A_{ij}} = V_{A_i} \frac{Q_{ij} K_{ij} - Q_{i,j-1} K_{i,j-1}}{q_{ij}} - \frac{Q_{i,j-1} (M_{i,j-1} - m_{ij})^2}{Q_{ij}}$$

where  $K_{ij} = 1 - i_{Q_{ij}} (i_{Q_{ij}} - X_{Q_{ij}}) h_i^2$  and  $Q_{i1} = q_{i1}$  in an obvious notation.

Now the mean breeding value ( $m_j$ ) of all dams used in tier  $j$  is  $m_j = \sum_i a_{ij} m_{ij}$ . Using this we can calculate the pooled variance of all dams used in tier  $j$  ( $V_{FA_j}$ ) taking into account the extra variation due to difference in mean breeding value of the selected female groups.

$$V_{FA_j} = \sum_i a_{ij} [V_{A_{ij}} + (m_{ij} - m_j)^2]$$

Now the variance of progeny breeding value in tier  $j$  at generation  $t+1$  ( $V_{A_j}^{t+1}$ ) is derived from the parental variances at generation  $t$  according to the following equation.

$$V_{A_j}^{t+1} = \frac{1}{4} V_{FA_j}^t + \frac{1}{4} V_{MA_j}^t + \frac{1}{2} V_{S_j}^t$$

where  $V_{MA_j}^t$  is the pooled variance of sires mated in tier  $j$  at generation  $t$  and  $V_{S_j}^t$  is the segregation variance in tier  $j$  at generation  $t$ . If the segregation variances vary between tiers then they change over generations according to the following equations.

$$V_{S_j}^{t+1} = \sum_i (a_{ij} + a_{ij}^*) V_{S_i}^t$$

for  $j = 1, 2, 3$ . This completes step 6a.

Now steps 1 to 6a can be performed again for the next generation using the new variance of breeding values and the new heritability in each tier.

## References

- Hopkins IR (1978) Some optimum age structures and selection methods in open nucleus breeding schemes with overlapping generations. *Anim Prod* 26:267-276
- Hopkins IR, James JW (1978) Theory of nucleus breeding systems with overlapping generations. *Theor Appl Genet* 53:17-24
- James JW (1977) Open nucleus breeding systems. *Anim Prod* 24:287-305
- James JW (1978) Effective population size in open nucleus breeding schemes. *Acta Agric Scand* 28:387-392
- James JW (1989) Design of animal breeding programmes. In: Hill WG, Mackay TFC (eds) *Evolution and animal breeding. Reviews on molecular and quantitative approaches in honour of Alan Robertson*. CAB International, Wallingford, UK
- Meuwissen THE (1989) A deterministic model for the optimization of dairy cattle breeding based on BLUP breeding value estimates. *Anim Prod* 49:193-202
- Mueller JP, James JW (1983) Effects of reduced variance due to selection in open nucleus breeding systems. *Aust J Agric Res* 34:53-62
- Shepherd RK (1991) Multi-tier open nucleus breeding schemes. PhD thesis, University of New England, Armidale
- Smith SP, Hammond K (1987) Assortative mating and artificial selection: a second appraisal. *Genet Sel Evol* 19:181-196