OPEN-NUCLEUS THEORY FOR BEEF CATTLE BREEDING SYSTEMS: A REVISITATION

E. Recami *
International Centre for Theoretical Physics, Trieste, Italy

I.U. Packer and M. Tenorio-Vasconcelos
Departamento di Zootechnica,
ESALQ-Escola Superior de Agricultura “Luiz de Queiroz”,
Piracicaba, SP, Brazil.

MIRAMARE - TRIESTE
July 1990

Abstract

A theoretical model for Open-Nucleus Systems is herein described in the case of beef cattle breeding. One of the starting points is the observation that the majority of the standard theoretical models for open-nucleus breeding systems were constructed for the case of discrete generations, i.e. for the cases in which the dam average fertility coefficient is \( f > 2 \) (so that at each generation all parents can be discarded and replaced by males and females selected among their progeny). In the case of cattle herds, when only a fraction of the breeding dams can be replaced (since \( f \approx 0.8 \) and each breeding female as an average has only about 0.4 sons and 0.4 daughters each generation), it is therefore worthwhile to build up anew a rather rigorous theoretical model, with overlapping generations, and check its predictions.

Namely, we apply the new formulæ —explicitly depending on \( \beta_p, \nu_p, \nu_M, K \) and \( R \) — to the system in which all breeding sires are in the Nucleus (and are reared in the nucleus itself), and are mated to both Nucleus and Base dams via artificial insemination. Optimal system design has been looked for by the NAG and MINOS computation programs, operated on Vax computers.

Opening the nucleus in this situation results to be very effective since the (optimum) asymptotic genetic gain per generation for “closed nucleus” systems \( (x = 0) \) results to be, when \( R = P/M \approx 200 \), more than 40% lower than the (optimum) asymptotic genetic gain, \( G^* \), for open nucleus systems. Open nucleus systems confirm themselves, therefore, as very attractive, at least when the initially available breeding sires are not already of exceptional genetic value.

Optimal design corresponds to: (i) having a fraction \( p \approx 16\% \) of the female population in the nucleus; (ii) replacing practically all the (nucleus) breeding sires by the best (nucleus born) males: \( \nu_M = 97 \pm 98\% \); (iii) using for dam replacement all \( (b \approx 100\%) \) the (base and nucleus born) females; (iv) implementing a high upward gene migration \( (x \approx 80\%) \), while all the surplus nucleus-born females are to be used as base replacements. This corresponds to replace, at each generation, also almost all the nucleus dams \( (\nu_F \approx 95 \pm 100\%) \), and the largest possible fraction of base dams \( (b_F \approx 30\%, \text{ a value changing with } p) \).

Analogous conclusions hold of course for sheep breeding systems.
I – INTRODUCTION

The nucleus breeding systems [see, e.g., Bichard, 1971] (let us fix our ideas on cattle herds or sheep flocks) are used when a set of exceptional breeding sires are not available initially, and one wants quickly to produce better and better sires from a Nucleus herd formed with the best dams. In the simplest systems, the remaining dams constitute the Base herd: even if more than 2 levels can exist, and there may be several herds in one level.

Opening the nucleus, by allowing migration into it of the best genes from the base, increases the system genetic gain rate (see, e.g., Jackson & Turner, 1972). However, the majority of the existing theoretical approaches (cf., e.g., Mueller, 1984; Mueller & James, 1984a, 1984b; James, 1982; Hopkins, 1978; Nicholls, 1987; and James, 1977, 1989) to the Open-Nucleus Systems were constructed for the discrete generations case, i.e. by assuming the dam average fertility level to be \( f > 2 \): so that at each generation the parents can be all discarded, and replaced by males and females selected among their progeny. On the contrary, in the case of cattle herds only a fraction of the breeding dams can be replaced at each generation (since \( f \approx 0.8 \) and each breeding female as an average has only \( K \equiv f/2 \approx 0.4 \) sons and \( K \approx 0.4 \) daughters each generation). For cattle herds, therefore, when we need overlapping generations approaches, it seems worthwhile to build up anew a rather rigorous theoretical model, in which the genetic gain does explicitly depend on \( K \), on the ratio breeding dams over breeding sires, and on the percentages of the (nucleus and base) breeding sires and dams replaced at each generation. This is the aim of the present paper. Analogous results will hold —mutatis mutandis— for sheep flocks.

We shall consider an overlapping generations model, in which all the breeding sires stay in the nucleus and are moreover reared in the nucleus only. They are mated to all the nucleus and base dams by the help of AI (Artificial Insemination) [cf. Smith, 1988].

II – METHOD: THEORY REFORMULATION

A) Assumptions

In our overlapping generations model, let us consider a hierarchical population of \( F \) breeding females (e.g., \( F = 1000 \)), with \( M \) breeding sires (e.g., \( M = 5 \)). A fraction \( p \) of the breeding dams is in the elite herd, the Nucleus, and the remainder is in the Base; so that we shall have \( Fp \) dams in the nucleus and \( F(1 - p) \) dams in the base. The breeding sires stay in the nucleus and are reared in the nucleus itself, in the sense that at each generation the best percentage \( a \) of the nucleus-born males are selected as sire replacements. In other words, we are assuming \( v = 0 \); \( w = 1 \), according to the standard terminology (James, 1977). Let us repeat that, however, the \( M \) breeding sires are mated to both the nucleus and base dams, via the help of AI.

At each generation, the best percentage \( b \) of all female progeny are selected as dam replacements: we assume the same value \( b \) both in the nucleus and in the base.

The selection objective is a normally-distributed trait. As commonly done we shall assume, more precisely, that all sub-populations (nucleus dams, base dams, nucleus male progeny, nucleus female progeny, base female progeny) are normally distributed, with the same phenotypic and genetic standard deviations, with respect to such a trait. [The trait heritability \( h^2 \) is also considered – of course – as constant over the various subpopulations].

Our last assumption is that the dam fertility coefficient is constant, and the same for nucleus and base breeding dams: for instance, \( f = 0.8 \).

Since our system is of the open-nucleus type, at each generation a fraction \( x \) of the nucleus dam replacements will come from the base, i.e. will be selected from among the base-born (best) females. The surplus females born in the nucleus are transferred —at least partially— to the base, and will constitute a fraction \( y \) of all the base dam replacements.

The system genetic gain per generation will be a function of the seven variables \( p, a, b, x, y, f \), and \( R \equiv F/M \), which can reduce to six if we adopt the value \( f = 0.8 \) and to five if we choose \( K = 200 \) (e.g., \( F = 1000 \) and \( M = 5 \), as mentioned above).

For future convenience, let us put \( K = f/2 \) (e.g., \( K = 0.4 \)).

B) Male and female Replacement fractions

As functions of the fundamental variables \( p, a, b, x, y, f \), and \( R \equiv F/M \), we can first evaluate the percentages \( \nu_M, \nu_F, \beta_F \) of nucleus breeding sires, nucleus breeding dams and base breeding dams, respectively, that we replace at each generation with their best progeny. Let us notice that the number of animals to be replaced are, correspondingly, \( \nu_M M, \nu_F Fp \) and \( \beta_F F(1 - p) \).

For instance, in the simple case of the sires we have that \( \nu_M M = a Fp K \), wherefrom

\[
\nu_M = a p K \frac{F}{M} \quad [K \equiv f/2]
\]

In the case of the dams, by recalling that the progeny females available at each generation are \( b F K \), we get firstly the condition
\( \nu T F + \beta T F (1 - p) = b K F. \)

Secondly, we can evaluate \( \nu T , \beta T \) from the "conservation relations"

\( (1 - x) \nu T + y (1 - p) \beta T = p b K \)
\( (1 - y) (1 - p) \beta T + x p \nu T = (1 - p) b K \)

and get the expressions

\( \nu T = \frac{b K - p - y}{p - 1 - x - y} \)
\( \beta T = \frac{b K - 1 - x - p}{1 - p - 1 - x - y} \)

which automatically satisfy condition (2a) (since eqs. (2a), (2b), (2c) are linearly dependent: e.g., the former is nothing but the sum of the two remaining equations). Notice, for instance, that we have \( F P \) dams in the nucleus; that at each generation we shall have \( F P K \) nucleus-born males and \( F P K \) nucleus-born females; that the total number of nucleus progeny females selected for dams replacement (partly in the nucleus itself, partly in the base) is \( b P K \); and so on.

C) Basic Theory

The total population consists of two parts: the nucleus, with higher mean breeding value \( (m_N) \), and the base, with lower mean breeding value \( (m_B) \); let it be \( A = m_N - m_B \). Such values change, of course, at each new generation: let us denote by \( C \) the change in mean breeding value from one generation to the next one (change due to the selections applied).

Let us rephrase the theory put forth in James (1977, 1980), by correcting the fundamental equations for the case \( K < 1 \) (as it is for cows and sheep). The average genetic selection differentials \( C \) of animals used in the elite and base herds write, if we assume \( C = \frac{1}{2} [C(M) + C(F)] \), take account of the herds' structure and recall that the nucleus breeding sires are mated also to the base dams:

\( C_N = \frac{1}{2} (1 - \nu T) D_{OFN} + \nu T [(1 - x) D_{NFW} + x D_{BFN}] + (1 - \nu M) D_{OMN} + \nu M D_{MN} \)
\( C_B = \frac{1}{2} (\beta T [(1 - y) D_{DIF} + y D_{DFN}] + (1 - \nu T) D_{DFB} + (1 - \nu M) D_{DMN} + \nu M D_{MN} \)

where the \( D \)s represent the genetic selection differentials. For instance \( D_N F \) and \( D_B F \) refer to the Nucleus-born and Base-born Females, respectively, that are used as nuclear replacements; while \( D_N M \) and \( D_B M \) refer to the Base-born and Nucleus-born Females, respectively, which are used as base replacements; and, finally, \( D_{MN} \) refers to the Nucleus-born Males used as replacements in the Nucleus. We have to introduce here, however, also the differentials referring to the animals that could not be (or were not) replaced: \( D_{OFN} \), \( D_{DFN} \), \( D_{OMN} \) refer to the "old" Females or Males that did remain in the Nucleus or Base, respectively, from the previous generation to the one presently considered. Notice that also the "old" animals did undergo a positive selection, since only the best ones remain in the herd (the worst being discarded in the replacement process).

By correcting for the difference \( A \) between nucleus and base means, we have for the genetic gain per generation:

\( G_N = C_N - \frac{1}{2} \nu T \sigma \)
\( G_B = C_B + \frac{1}{2} (\beta T + \nu M) \sigma \)

Let us look for a steady-state situation in which \( G_N - G_B \) tends to zero, i.e. the rate of progress becomes asymptotically the same in the nucleus and in the base. Actually, it is easy to see that \( G_N \) and \( G_B \) tend to zero when \( A \) assumes the (asymptotic) value

\( A^* = \frac{C_N - C_B}{\nu T + \beta T + \nu M} \)

In the steady-state regime (when \( A = A^* \)), we shall have \( G_N = G_B = G^* \). The asymptotic rate of progress \( G^* \) (genetic gain per generation) of both nucleus and base can be evaluated either from eq (6a) or from eq (6b), which yield -- as it is usually verified -- the same result

\( G^* = \frac{\nu T \sigma C_N + (\beta T + \nu M) C_B}{\nu T + \beta T + \nu M} \)

D) Standardized Genetic Selection Differentials

Our genetic selection differentials can be written as \( D = S \sigma h^2 \), where \( \sigma \) is the phenotypic standard deviation, \( h^2 \) is the heritability of the considered trait and \( S \) is the "standardized selection differential" (also called selection intensity, \( s \)). Since we supposed \( \sigma \) and \( h^2 \) to be constant over the various subpopulations, we can just replace \( D \) by \( S \sigma \); so that our \( D \)s, as well as our equations (5) + (8), will result expressed in units \( h^2 \).

Let us then consider a reduced normal distribution, with a density function \( \varphi = \varphi (u) \) corresponding to mean \( \mu = 0 \) and standard deviation \( \sigma = 1 \). When we select the best fraction \( q \) (\( 0 \leq q \leq 1 \)) out of a normally distributed population, it is well-known that the standardized selection differential [in units \( \sigma h^2 \)]

\( D \equiv \int_{-q}^{q} \varphi (u) du / \int_{-\infty}^{\infty} \varphi (u) du \)

results to be (cf., e.g., James, 1976; Mueller & James, 1983):

\( D = q^2 / \varphi (u) \equiv \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{u^2}{2} \right) \)
where \( Z = \varphi(V) \), quantity \( V \) being the truncation value of the normal curve corresponding to the considered “best fraction” \( q \). Notice that \( V \) as a function of \( q \), is given by the inverse of the so-called error function: \( V = \text{ERF}^{-1}(q) \).

Anyway, once \( q \) is known, the quantity \( D = D(q) \) can be evaluated, e.g. by linearly interpolating the values given by the selection intensity Table (see, e.g., Becker, 1975; and Pearson, 1931). In correspondence with the \( Ds \) entering eqs. (5a), (5b) we get the following values of \( q \):

\[
\begin{align*}
D_{OFN} & \to q \equiv q_{OFN} = 1 - \nu_p \\
D_{NPN} & \to q \equiv q_{NPN} = (1 - z)\nu_p / K \\
D_{BFN} & \to q \equiv q_{BFN} = \frac{z}{K} \\ 
D_{OFN} & \to q \equiv q_{OFN} = 1 - \nu_M \\
D_{NPN} & \to q \equiv q_{NPN} = a \\
D_{BFN} & \to q \equiv q_{BFN} = (1 - y)\beta_p / K \\
D_{NPN} & \to q \equiv q_{NPN} = \frac{y}{K} \cdot \frac{1 - p}{p} \\
D_{OFN} & \to q \equiv q_{OFN} = 1 - \beta_F
\end{align*}
\]

which satisfy the two conditions

\[
\begin{align*}
q_{NPN} + q_{BFN} &= b \\
q_{BFN} + q_{BFN} &= b.
\end{align*}
\]

One should pay attention to the fact that, from among the (best) proportion \( b \) of progeny females (selected for replacement), the worst fraction is chosen as base replacements. So that

\[
\begin{align*}
D_{BFN} &= D_{BFN} - D(b) \\
D_{NPN} &= D_{NPN} - D(b),
\end{align*}
\]

where the notation \( D(b) \) is self-explanatory.

III - CALCULATIONS

We have looked for the maximum of the steady-state genetic gain per generation \( G' \), eq. (8), as a function of the five independent variables \( p, a, b, x, y \), for various values of \( R = F/M \). We kept \( K \) fixed \( (f = 0.8; K = 0.4) \). We initially used the MINOS programs, and later on the NAG-E04 programs, operated on a VAX computer: such programs are usually part of the VAX libraries, but must be adapted – of course – to the present case.

One has to find out the maxima of the function \( G' = G'(p, a, b, x, y) \) in correspondence with a multiply connected region of the five variables: a region bounded by five non-linear constraints (besides by five fixed bounds and two linear constraints). The constraints correspond, essentially, to the fact that the fundamental variables \( p, a, b, x, y \), the quantities \( \nu_M, \nu_p, \beta_p \) (given by eqs. (2), (3), (4)) and the quantities \( q \) appearing in eqs. (10), representing percentages, must all vary between 0 and 1; and this has to be explicitly imposed.

The fixed bounds have been chosen as follows

\[
\begin{align*}
0.01 & \leq p \leq 0.50 \\
0.00 & \leq a \leq 0.25 \\
0.00 & \leq b \leq 1.00 \\
0.00 & \leq x \leq 0.99 \\
0.00 & \leq y \leq 0.99.
\end{align*}
\]

But the more delicate bounds correspond to:

(i) imposing that

\[
\nu_p \leq 1; \beta_p \leq 1; q_{NPN} \leq 1; q_{BFN} \leq 1;
\]

these inequalities yield four non-linear constraints;

(ii) imposing that

\[
\nu_M \leq 1;
\]

this inequality yields the additional (simple) non-linear constraint \( ap \leq 2.5/R \);
(iii) imposing that

\[ \nu_r \leq 0 ; \beta_r \leq 0 ; \]

these two inequalities yield two allowed regions [either \( x + p \leq 0.99 \) and \( y \leq p \), or \( x + p \geq 1.01 \) and \( y \geq p \)] that have been explored separately.

For comparison with the closed-nucleus case, we ran our programs also with \( x = 0 \). Finally, let us mention that the NAG-E04UAF parameter RHO has to be fine-tuned; and that, to use MINOS, i.e. to evaluate the gradient of \( G^* \), we made recourse to the algebraic computer programs MUMATH.

IV – RESULTS

A) Main results and conclusions

We did explore the allowed regions of the parameters looking for the absolute maximum of \( G^* \). Actually, our calculation programs look for the maximum nearest to the starting point: various, reasonable starting points are therefore to be tried. We made our most accurate exploration for \( R = 200 \), which corresponds to having, e.g., 5 breeding sires for a total population of 1000 breeding dams.

Our results state that the maximum \( G^* \) is got when:

(i) \( \beta_r \approx 0.30 \), i.e. when we replace at each generation about 30% of the base breeding dams. This is the main difference with the theoretical model in James (1977). We are going to see that our remaining results don't differ too much from the ones in James (1977), so that they can be said to confirm – mutatis mutandis – the conclusions by James. However, for future developments our equations (5) \( \div \) (8) are to be used, for instance, instead of eqns. (1) \( \div \) (4) in James (1977); the latter failing to depend, as they should in the overlapping generations case,, on \( \beta_r, \nu_r, \nu_M, K \) and \( R \).

Our further results can be stated as follows:

(ii) the elite (nucleus) population has to be about 16% of the total female population \( (p \approx 11 \div 20\%) \);

(iii) at each generation, we have to replace practically all the (nucleus) breeding sires by the best nucleus born males: \( \nu_M \approx 97 \div 99\% \);

(iv) all the (nucleus born and base born) progeny females are to be used as dam replacements: \( b = 100\% \);

(v) we have to replace, as well, almost all the breedings dams of the nucleus by the best (nucleus born and base born) females: \( \nu_r \approx 96 \div 100\% \);

(vi) more precisely, about 80% of the nucleus dams replacements must be base born females: \( x \approx 80\% \);

(vii) all the nucleus born females, not used for replacement in the nucleus itself, are to be used as base replacements.

For instance, the maximum steady-state genetic gain per generation, \( G^* \approx 2.20 \), has been got for

\[ \beta_r \approx 16\% ; \nu_r \approx 7.6\% ; \nu_M \approx 100\% ; x \approx 82\% ; y \approx 14.7\% \]

which correspond to \( \beta_r \approx 29\% ; \nu_r \approx 98.4\% ; \nu_M \approx 97.7\% \).

The value of \( G^* \), near its maximum, depends slightly on \( p \), a little more on \( z \) and \( y \), and even more on \( a \) and \( b \). For high values of \( G^* \), actually, the values of \( a \) and \( y \) do accompany the values of \( p \) and \( z \), in order that our general results (iii) and (vii) above be always satisfied. For instance, by loosing only 1.3% of the genetic gain, the breeders can adopt, instead of relations (18a), the values

\[ p \approx 10.5\% ; a \approx 11.5\% ; b \approx 100\% ; z \approx 84\% ; y \approx 8.4\% \]

corresponding to \( \beta_r \approx 33.4\% ; \nu_r \approx 96.4\% ; \nu_M \approx 97\% \); or, by loosing 2.8% of the genetic gain with respect to values (18a), the breeders can choose

\[ p \approx 20\% ; a \approx 6.1\% ; b \approx 100\% ; z \approx 72\% ; y \approx 13\% \]

Far from the maximum, however, \( G^* \) depends strongly on \( b \) and \( y \), and very strongly on \( a \) and \( z \), so that it is essential choosing high values \( (z > 70\%) \) of \( z \), while \( a \) must always be chosen so that result (iii) is satisfied.
B) Further comments.

The maximal value $G^* = 2.20$ can appear as very high; but we must recall that it is expressed not in terms of standard deviations $\sigma$ but in terms of $h^2/\sigma$. If the heritability of the considered trait is $h^2 = 0.5$, then we shall get $G^* \approx 1.10$, i.e. an asymptotic genetic progress rate $G^* \approx 1.10\sigma$ per generation.

Anyway, the open nucleus systems results to be very efficient (when, let us repeat, $x > 0.70; b > 1.0$ and $p < 0.15$, and $a, y$ satisfy the requirements [iii], (vii) above), at least under the considered assumptions [i.e., when the nucleus breeding sires are mated also to the base dams, via the help of AI, but $R \approx 200$].

To compare the open-nucleus efficiency with the "closed-nucleus" one, we run our programs also for $x = 0$. In the "closed-nucleus" cases, the maximal value for $G^*$ resulted to be $G^* \approx 1.55$. Therefore, one looses more than 40% of the steady-state progress rate per generation when passing from an open-nucleus to a closed-nucleus system.

Incidentally, let us mention that, within our model, the maximal $G^*$ values for the case $x = 0$ have been got for $p \to 100$%. This confirms that, when $x = 0$, it is not worthwhile any longer to split the available herd in two parts. However, quantity $a$ must always take on the value corresponding to complete replacement of the breeding sires by the best males produced by the system. On the contrary, near its maximum $G^*$ depends slightly on $b$, so that it seems sufficient to choose a value in the range $60% < b < 80%$.

At least, let us come to the influence of $R = F/M$ on $G^*$. Our results for the open-nucleus case show that the maximal $G^*$ value (strongly) increases with $R$; so that it goes down to $G^* \approx 2.0h^2\sigma$ for $R \approx 100$, but goes up to $G^* \approx 2.4$ for $R \approx 1000$, to $G^* \approx 3.0$ for $R \approx 4000$ and $G^* \approx 3.2$ for $R \approx 20000$. However, also the associated values for the "closed nucleus" case ($x = 0$) do vary, so that opening the nucleus appears to loose effectiveness when the sire selection becomes very high.

V – ACKNOWLEDGEMENTS

The authors are grateful to J. W. James and C. Smith for stimulating Bitnet correspondence; to the colleagues J. A. Scaramuzzi, J. Zurro and particularly J. M. Martinez for kind assistance during the first phases of the numerical computer calculations; and to V. T. Zanchin for analogous assistance in the algebraic computer calculations. Moreover, they thank A. Insolla, A. M. Liquori, E. C. Lima, L. Lo Monaco, N. Paver, P. Pulino and L. T. Santos for useful discussions. This work was completed while one of the authors (ER) was visiting the I.C.T.P.: he is grateful to Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste, Italy. This work was partly supported by CNQI, CAFES, IBM-60-Brazil and by INPN, Sezione di Catania, HFI, CNR.

REFERENCES


PEARSON, K. 1931. Tables for Statisticians and Biometricians, Part II. Biometrika, London. Table II.