Proc. Assoc. Advmt. Anim. Breed. Genet. 19:15-22

AUTOMATED MATE SELECTION ANALYSES

B.P. Kinghorn

University of New England, Armidale, NSW 2351

SUMMARY

This paper outlines methods used to help ensure robust outcomes from mate selection analyses. In particular, the balance between genetic gain and genetic diversity is maintained appropriately despite changing emphasis on other factors in the objective function, such as progeny inbreeding and trait distribution management. This is needed to provide automated analyses with minimal human intervention and routine delivery of mating lists to accompany EBV results from genetic evaluation services.

INTRODUCTION

The classic approach to implementing an animal breeding program is *strategic* in nature. Rules are developed that cover a range of issues, for example:

- Rank on index and choose the best as parents
- Avoid extensive use of any one sire or any one sire family
- Avoid full-sib mating
- Do not use carriers of genetic defects, or, at least, do not mate carriers
- Use Breed X sires to mate cows and Breed Y sires to mate heifers
- Perform corrective mating for trait Z.

These rules are then followed as closely as possible. However, given that some rules are antagonistic or competitive with others, compromise must be made to achieve optimal outcomes.

One key component in this list is effectively *tactical* in nature: ranking on index depends on prevailing animals. A valuable extension is to use optimal contributions (Meuwissen, 1997; Meuwissen and Sonneson, 1998; Grundy *et al.* 1998) such that higher indexing males (and/or lowly related males) are generally assigned more matings than lower indexing selected males, but in a manner that simultaneously manages genetic diversity.

This can be further extended to give a full mating list that also accommodates other issues, such as progeny inbreeding, genetic defects and trait distributions. This is a fully tactical implementation, whereby the analysis uses all prevailing information to derive a balanced outcome across issues, guided by breeder experience and attitude.

Such tactical implementation systems have been used in various species for over 10 years, including Total Genetic Resource Management (TGRM) in Australia (Kinghorn and Shepherd, 1999, Upton *et al.* 2001). An impediment to widespread use has been the need for a custom analysis for each mating list, with user-guidance to ensure desired outcomes. The associated high cost of service also inhibits uptake. Moreover, speed was very slow for breeding structures involving several groups for each sex. This was a notable problem in pigs, with many breeding lines to be mated each week.

These problems were solved with steps taken to remove the need for continual user guidance and to increase speed substantially (Newman *et al.* 2009; Kinghorn, 2011). Weekly automated analyses have been carried out since 2007 in 32 breeding programs covering 17 lines of pigs in 6 countries (Scott Newman, Pig Improvement Company, *pers com*). The full breeding information system can be essentially automatic, with minimal user intervention required between performance recording and delivery of mating lists (Newman *et al.* 2009).

Experience with these developments in pigs suggests a new mode of implementation of mate selection in the extensive industries. This paper will present approaches to be used for automated mate selection analyses, aiming at widespread delivery of recommended mating lists carried out cheaply due to lack of need for human intervention.

METHODS

Key components for automation include (1) launching analyses, which can be triggered by completion of genetic evaluation analyses; (2) stopping analyses through diagnosis of convergence (Kinghorn, 2008); and (3) taking steps to give a robust pattern of outcomes, which is addressed in this paper. The two key outcomes are the levels of genetic gain and genetic diversity, as indicated by predicted progeny mean index and parental coancestry respectively (see Figure 1).

In TGRM, these two outcomes are balanced by use of a weighting (λ) in the objective function (*OF*):

$$OF = x'G/2M + (x'Ax/(4M\uparrow 2)$$

… where $x'G/2M$ is predicted mean progeny index; $x'Ax/4M^2$ is mean parental coancestry; x is the vector of contributions from male and female candidates, expressed as number of matings allocated to each, such that x sums to $2M$, where M is the total number of matings to be made; A is the numerator relationship matrix (or potentially a genomic relationship matrix); G is the vector of candidate index values, typically multi-trait EBVs calculated from pedigree or markers or both..

λ is chosen to give the desired balance between these two key issues. However, when other issues are added, this balance is disrupted, such that user intervention is required to restore desired balance.

A mate selection analysis covering multiple issues is analogous to a selection index analysis covering multiple traits: If we have a 2-trait index giving a certain proportionality of predicted response between these two traits, then adding a third trait without changing the relative index weighting between the first two traits will generally change the pattern of response for these two traits. For the mate selection case this was solved by moving away from the weighted score paradigm, using the following objective function, which in this case includes emphasis on progeny inbreeding (Kinghorn, 2011):

If

$$aTan\left[\frac{(x_0'Ax_0 - x'Ax)\Big/(x_0'Ax_0 - x_{90}'Ax_{90})}{(x'G - x_{90}'G)\Big/(x_0'G - x_{90}'G)}\right] < TD$$

then

$$OF = \frac{\frac{x'G}{2M} - \frac{x_{90}'G}{2M}}{Cos(TD) * \left(\frac{x_0'G}{2M} - \frac{x_{90}'G}{2M}\right)} - 1.F$$

where

$$F = \frac{\Sigma_{i=1}^{M} 0.5\ a_{Male_i.Female_i}}{M}$$

or else $$OF = -10^{20} - \frac{x'Ax}{8M^2}$$

… where x_0 is the vector of optimal contributions that maximize the progeny index and x_{90} is the vector of optimal contributions that minimize parental coancestry (x_0 and x_{90} having been determined by this stage, *0* and *90* relate to degrees in Figure 1); *a* is an element from the numerator relationship matrix *A*; *G* is the vector of candidate index or EBV values; *F* is the mean inbreeding coefficient in progeny that would result from the current mate selection solution, as defined by the parents ($\llbracket Male \rrbracket \downarrow i, \llbracket\ Female \rrbracket \downarrow i)$ of the i^{th} mating; TD = TargetDegrees is the degree line, set to 25 degrees in Figure 1, below which value a solution is taken to be illegal. The latter is effected with an objective function value of -10^{20} but with an additional penalty on high coancestry to help approach legality in the case that all solutions are illegal. This is Balance Strategy 3 of Kinghorn (2010). Other strategies include, for example, the setting of a maximum value for parental coancestry.

Notice that for legal solutions, the mean predicted progeny merit $\left(\mathbf{x}'\mathbf{G}/_{2\mathbf{M}}\right)$ is expressed as a deviation from the minimum merit previously found with full emphasis on reduced coancestry $\left(\mathbf{x}'_{90}\mathbf{G}/_{2\mathbf{M}}\right)$, and then scaled by a denominator that uses a trigonometric function to give an expected range from 0 to 1, assuming a circular shape for the frontier in Figure 1.

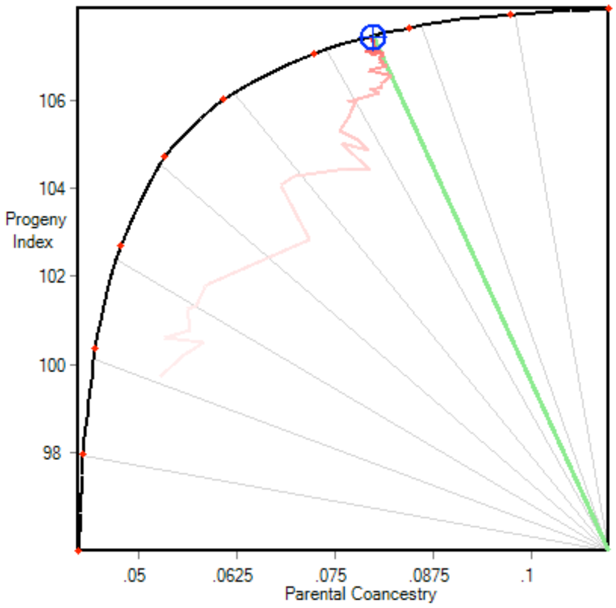


Figure 1. A response frontier. The curve is the frontier of optimal contributions, where each point on the frontier represents an optimal mating list for the corresponding relative emphasis on progeny index and parental coancestry. The top-right of the frontier is 0 degrees, with full

emphasis on progeny index, and the bottom-left is 90 degrees, with full emphasis on lowered parental coancestry. The solution has settled on the frontier at the 25 degree 'target degree' line.

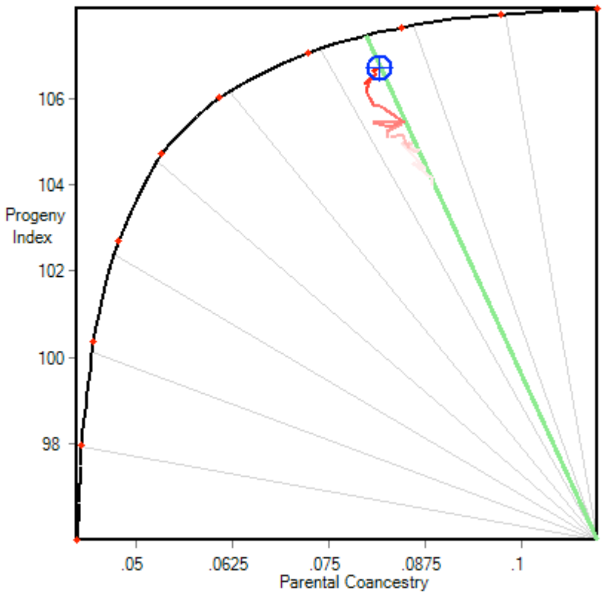


Figure 2. The solution following a strong weighting against mean progeny inbreeding. The frontier is not reached because of emphasis on this third issue, but the target degree line is adhered to.

Thus, given appropriate constraint, such as the target degree line in Figure 2, the combined results for gain and diversity lie on a single scale from 0 to 1. By aiming to give the same or similar scale range for other issues (easy for progeny mean inbreeding which is already on this scale), there is a more consistent impact of chosen weighting values on all components in the objective function. Adoption of such a strategy in developing components of the objective function gives more consistent outcomes across different runs, for example for different farms, or for the same population at different mating cycles. This is an important step towards automated analyses that require little or no human intervention.

The best position to aim at on the frontier in Figure 1 or 2 depends on the shape of that frontier. This would normally require user intervention to inspect that shape, but a different approach can be taken in the interests of automation, as illustrated in Figure 3. To generate this figure, balance strategy 5 "Project to Target Degrees line" of Kinghorn (2010) was adopted, and the point arrived at on the frontier is that which maximises the distance from the origin to its projection on the 25 degree target degree line. This essentially treats that line as an index to be maximised, and recognises that the best outcome is not necessarily the point where that line crosses the frontier. In this case, the shape of the frontier is such that moving from the targeted degrees to the optimal result gives a large reduction in parental coancestry in exchange for a small compromise in progeny index. This is typical where there are many lowly related male candidates of similar index value.

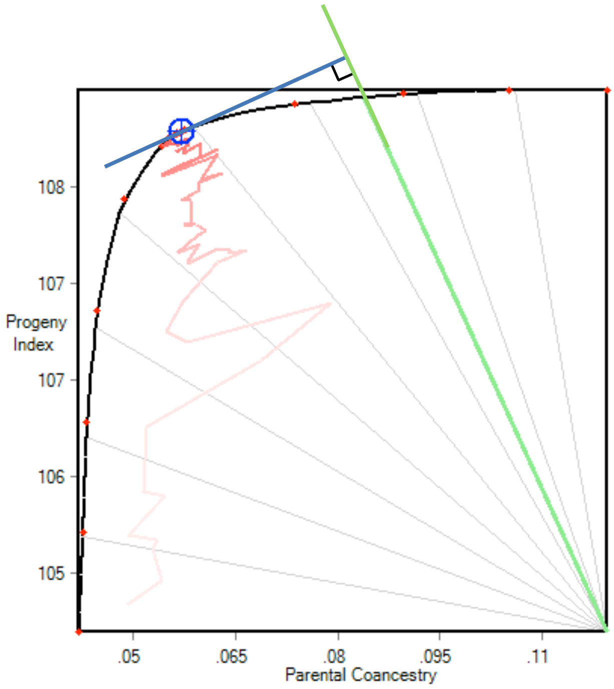


Figure 3. Accounting for the prevailing shape of the frontier by maximising the projection to the target degree line. The optimal solution is the point which, when projected to the target degree line, gives the biggest deviation from the origin. In this case, a small compromise in progeny index gives a big reduction in parental coancestry.

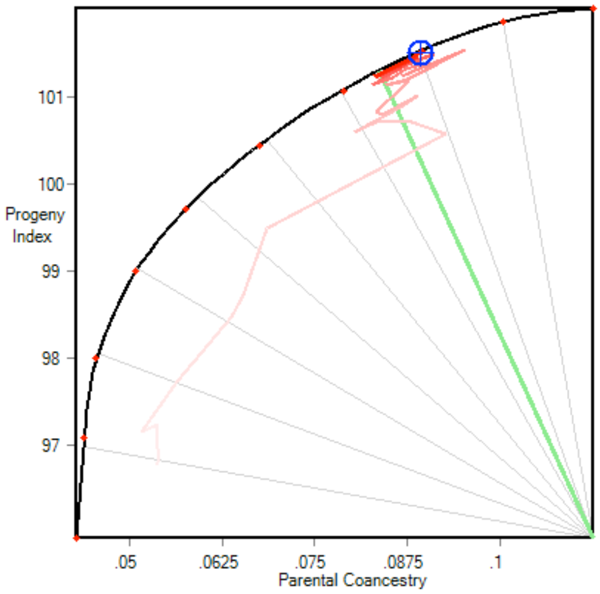


Figure 4. A case where projection gives achieved degrees that are lower than the target degrees.

Figure 4 shows a case where there are only a few candidates of notably high index value, and spreading contributions to give lower coancestry results in rapid loss in mean progeny index. In this rather extreme case, balance strategy 5 gives an optimal result on the other side of the target degree line compared to Figure 3. For most frontiers, balance strategy 5 results in higher realised degrees, such that the declared target should typically be lowered for routine use. It is possible to aim for a result that is intermediate, penalising the outcome for deviation from the declared target degrees, i.e., balance strategy 6 of Kinghorn (2010).

DISCUSSION

The steps taken to help ensure a good balance between progeny index and parental coancestry give some confidence in running mate selection analyses that are unattended by a human operator.

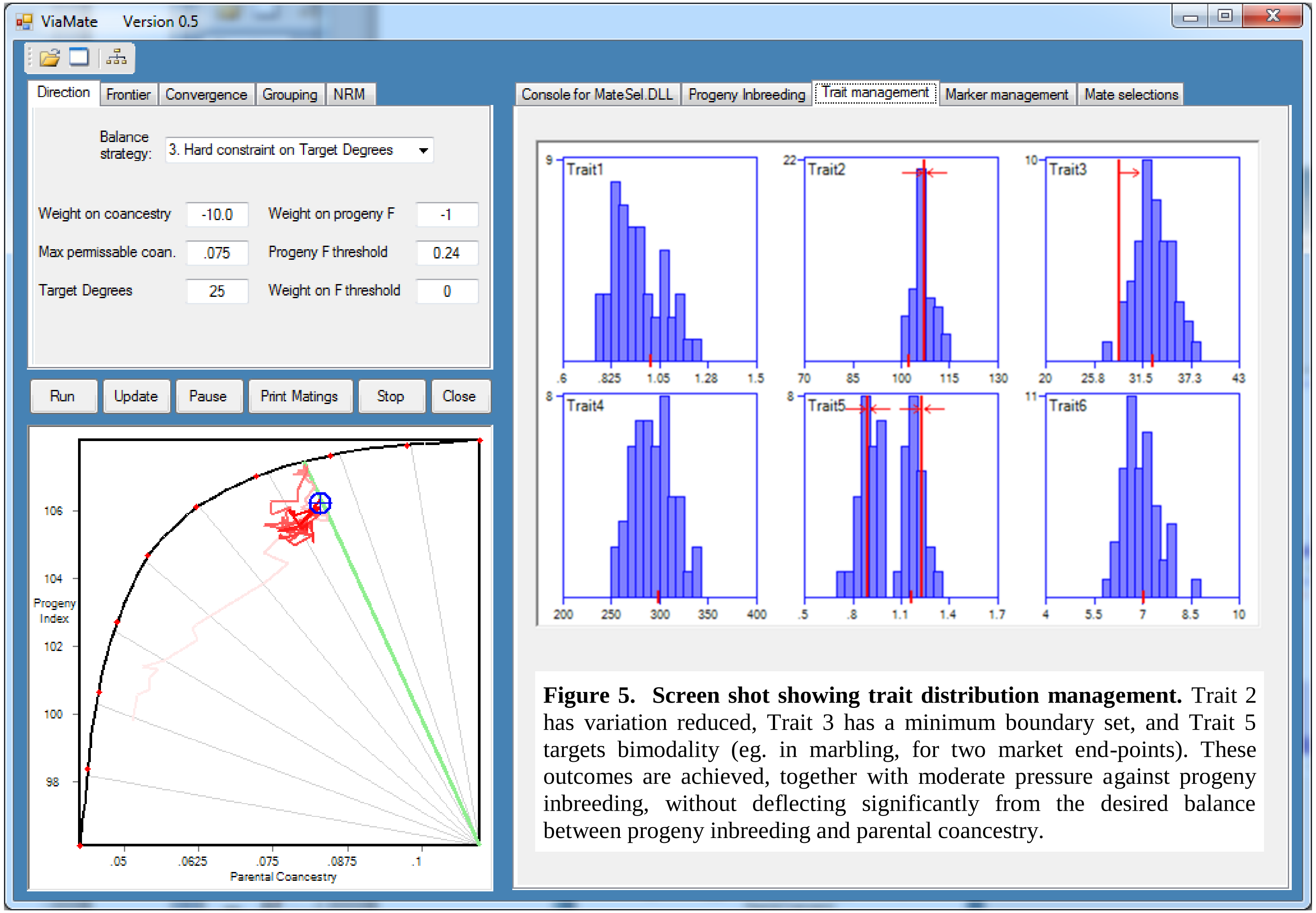
However, before such runs are made, work is required to find settings (weightings, modes of constraint, etc) that generally lead to desired outcomes. To facilitate this, analyses can be monitored graphically, and settings changed dynamically to discover the range of outcomes possible. This gives a flexible basis to find settings that give desired outcomes.

Policies reflected in these settings can be set separately for different breeding populations, for example: aiming for higher genetic diversity in breeds that are threatened, or in highly elite herds that have little or no immigration from outside; targeting elimination of a recessive genetic defect over a given period; or increasing genetic variance for a specified trait, as a prelude to new line development.

Runs that are fully automated may have some such pre-set emphasis on trait distributions, genotype and/or allele frequencies for genetic defects, and a range of other issues. However, for analyses involving many issues it will be preferable to use a graphical user interface for each individual analysis, as in TGRM and the prototype program shown in Figure 5, to explore the range of possible outcomes and settle on the most suitable mating list.

ACKNOWLEDGMENTS

The project to automate mate selection analyses in Australia involves the Agricultural Business Research Institute, Angus Australia, Meat and Livestock Australia, Pfizer Animal Genetics and the University of New England. Scott Newman and Valentin Kremer are thanked for collaborations in this area.



Balance strategy: 3. Hard constraint on Target Degrees

Weight on coancestry -10.0

Weight on progeny F -1

Max permissable coan. .075

Progeny F threshold 0.24

Target Degrees 25

Weight on F threshold 0

Figure 5. Screen shot showing trait distribution management. Trait 2 has variation reduced, Trait 3 has a minimum boundary set, and Trait 5 targets bimodality (eg. in marbling, for two market end-points). These outcomes are achieved, together with moderate pressure against progeny inbreeding, without deflecting significantly from the desired balance between progeny inbreeding and parental coancestry.

REFERENCES

Kinghorn B.P. (2008) Chapter 10 in: Gondro C, Kinghorn BP: Application of Evolutionary Algorithms to solve complex problems in Quantitative Genetics and Bioinformatics. Centre for Genetic Improvement of Livestock University of Guelph. 4 to 8 August 2008. http://www-personal.une.edu.au/~bkinghor/Evolutionary_Algorithms_CGIL2008.pdf

Kinghorn B.P. (2010) http://www-personal.une.edu.au/~bkinghor/pedigree.htm

Kinghorn B.P. (2011) *Genetics Selection Evolution*. 43:4.

Kinghorn B.P. and Shepherd R.K. (1999) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **13**:130.

Meuwissen T.H.E. (1997) *J. Anim. Sci*. **75**: 934.

Meuwissen T.H.E. and Sonesson A.K. (1998) *J. Anim. Sci*. **76**: 2575.

Grundy B., Villanueva, B., Woolliams J.A. (1998) *Genet. Res*. **72**:159.

Newman S., Kremer V.D. and Kinghorn B.P. (2009) *58th Annual National Breeders Roundtable*. St Louis Missouri, 7-8 May 2009

Upton W.H., Meszaros S.A., Barwick S.A. and Kinghorn B.P. (2001) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **14**: 457.