

An Iterative Algorithm for Optimum Contribution Selection in Large Scale Breeding Programs

Binyam S. Dagnachew¹ and T.H.E. Meuwissen¹

¹Department of animal and Aquacultural Sciences, Norwegian University of Life Sciences, N-1432 Ås, Norway

ABSTRACT: A novel iterative algorithm, Gencont2, for calculating optimum genetic contributions was developed. It was validated by comparing it with a previous program, Gencont, on three datasets obtained from practical breeding programs of three species (cattle, pig and sheep). The numbers of selection candidates were 2,929, 3,907 and 6,875 for the pig, cattle and sheep datasets respectively. In most cases, both algorithms select the same candidates and gave very similar results in genetic gain. In cases when there were differences in number of animals to select, the extra selected candidates had contributions within the range of 0.006–0.08%. The correlations between assigned contributions were very close to 1; however, Gencont2 considerably decreased the computation time by 90% to 95% (13 to 22 times faster) compared to Gencont. This fast iterative algorithm makes the practical implementation of OC selection feasible in large scale breeding programs.

Keywords: Inbreeding; Optimum genetic contribution; Genetic gain

Introduction

Modern breeding schemes have substantially increased response to selection due to the continuous advancement in selection methods and biological (reproductive) techniques. Selection programs are usually designed to optimize genetic gain with no or an implicit limitation of rates of inbreeding. Even though the main goal in breeding programs is to maximize genetic gain, management of inbreeding is also vital because it leads to lower performance in the long term known as inbreeding depression (Falconer, 1989).

The optimal balance between rate of inbreeding and genetic gain is a core problem in practical animal breeding. In the last decade, a dynamic selection method, known as optimum contribution (OC) selection, has been proposed to deal with this optimization problem (Meuwissen, 1997, Grundy et al., 1998, Pong-Wong and Woolliams, 2007). OC selection seeks to maximize genetic response while putting a restriction on the rate of inbreeding.

Simulation results showed that OC selection could achieve up to 60% more genetic gain compared with truncation selection at the same rate of inbreeding (Meuwissen, 1997, Grundy et al., 1998). The potential application of OC selection method in practical breeding programs has also been presented, for instance, in dairy cattle population (e.g. Koenig and Simianer, 2006), in British sheep and beef cattle populations (Avendaño et al., 2003) and in salmon breeding (Hinrichs et al., 2006). These studies reported higher genetic gain under OC selection at the same level of inbreeding rate compared to traditional truncation selection.

Meuwissen (1997, 2002) presented an algorithm, Gencont, to calculate optimized genetic contributions of selection candidates constrained on a predefined rate of inbreeding. The algorithm requires an inverse of the relationship matrix of the selection, which needs to be re-calculated as some candidates proof unfit for selection. However, there are challenges for implementing this in practical breeding schemes which have large numbers of selection candidates, due to computational limitations to the repeated calculation of the inverses of large matrices. An alternative approach considered here is to obtain optimum contributions iteratively. Therefore, this paper presents an iterative algorithm, Gencont2, for the calculation of optimized genetic contributions.

Materials and Methods

Theory: Let G be the genetic level of the parents and thus of the next generation:

$$G = c^T EBV$$

where EBV is a vector of estimated breeding values of the selected parents and c is a vector of genetic contributions of the selected parents to the next generation. The problem is to find optimum genetic contributions, c , so that G is maximized. Therefore, $c_i \geq 0$ for a candidate that is selected to be a parent, and with the total contributions summing to 1 (i.e. $\sum_{i=1}^n c_i = 1$). In diploid species each sex contributes half the genes to the gene pool (i.e. $\sum c_i = 0.5$ where the sum is over all individuals of a single sex). Then the restriction on contribution per sex is:

$$c^T Q = r \quad [1]$$

where Q is an incidence matrix for sex of the candidates, and r is a vector of 0.5s. Controlling of inbreeding can be done through constraining the group co-ancestry of selection candidates. For a set of genetic contributions of selection candidates, the constraint on the group co-ancestry is:

$$K = 1/2 (c^T A c) \quad [2]$$

where A is additive relationship matrix of the selection candidates and K is a constraint. The constraint, K , was set as $K = C_p + \Delta F(2 - C_p)$, where C_p is the average co-ancestry of the current population and ΔF is the targeted rate of inbreeding.

The optimal c that maximizes G under the above constraints (i.e. Eq. [1] and [2]) is then obtained by maximizing the following objective function:

$$H = c^T EBV - \lambda_0 (2K - c^T A c) - \lambda (c^T Q - r)$$

where λ_0 and λ are Lagrangian multipliers (λ is a vector of length 2). This objective function yields:

$$c = A^{-1} (EBV - Q\lambda) / 2\lambda_0 \quad [3]$$

From the constraint in Eq. [2] follows an equation for λ_0 :

$$\lambda_0^2 = \frac{EBV^T(A^{-1}-A^{-1}Q(Q^T A^{-1}Q)^{-1}Q^T A^{-1})EBV}{8K-1^T(Q^T A^{-1}Q)^{-1}1} \quad [4]$$

For a more detailed derivation of Eq. [3] and [4] see Meuwissen (1997). In Eq. [4] A^{-1} is assumed as an identity matrix and in Eq. [3] it was obtained iteratively. The procedure to find optimal genetic contributions iteratively is described in five steps.

Step 1: Calculate λ_0 by solving Eq. [4] and initiate values for λ (for instance zeros). Calculate optimal contribution, c , by solving Eq. [3] for c .

Step 2: Update λ . Taking the gradient of Eq. [3] with respect to λ yields:

$$Q^T \partial c = -Q^T A^{-1} Q \partial \lambda / 2 \lambda_0$$

From the constraint in Eq. [1], for a new set of c , $\Delta r = r - Q^T c$ and rearranging yields:

$$\partial \lambda = -2 * (Q^T A^{-1} Q)^{-1} * \Delta r * \lambda_0$$

$$\lambda_{new} = \lambda + \partial \lambda$$

Step 3: Update λ_0 . Taking the gradient of Eq. [3] with respect to λ_0 yields:

$$\partial c_t = -A^{-1}(EBV - Q\lambda) \partial \lambda_0 / 2 \lambda_0^2$$

From the constraint in Eq. [2], for a set of new c , $\Delta K = 2K - c^T A c$ and rearranging yields:

$$\partial \lambda_0 = -\frac{\Delta K \lambda_0^2}{c^T (EBV - Q\lambda)}$$

$$\lambda_{0_{new}} = \lambda_0 + \partial \lambda_0$$

Step 4: Re-calculate optimal genetic contributions for the updated values of $\lambda_{0_{new}}$ and λ_{new} by solving Eq. [3] for c .

Step 5: Check for convergence. Convergence was monitored after each round of iteration. If it is not converged, the algorithm will return back to step 2.

In step 4, some of the contributions could be negative for some candidates with poor EBVs. These candidates were kept in the optimization process for the first 500 iterations. If they remained with negative solutions, then their contributions were constrained to zero and they were removed from the optimization process. This iterative algorithm, referred to as Gencont2, was programmed in FORTRAN95. Performance of the program was tested on real datasets and results were compared with the original algorithm, Gencont (Meuwissen, 2002).

Description of datasets: the performance of the iterative was validated using three datasets from practical breeding programs of three species (cattle, pig and sheep). These datasets were provided by Geno (the breeding organization for Norwegian Red), Norsvin (the Norwegian swine breeding organization) and NSG (the Norwegian association of sheep and goat farmers). *Cattle dataset:* the number of selection candidates was 3,907 and the pedigree file contains 23,224 animals. *Pig dataset:* the number of selection candidates was 2,929 and the pedigree file contains 11,945 animals. *Sheep dataset:* the number of selection candidates was 6,875 and the pedigree file contains 82,225 animals. For the cattle dataset the EBVs are based on BLUP estimated breeding values, whereas in the pig and sheep datasets they are based on selection index scores.

Results and Discussion

This paper presents a novel iterative algorithm, Gencont2, for calculating optimized genetic contributions with predefined rate of inbreeding. Figure 1 shows the association between EBV and optimized genetic contributions of selected bulls for the cattle dataset with a targeted rate of inbreeding of 0.01 (left) and 0.05 (right). The algorithm was successful in constraining ΔF to the predefined levels. The number of bulls to be selected (to have nonzero contributions) increased when more severe constraints were placed on ΔF (Figure 1). For example, the number of animals selected increased from 15 to 75 as the predefined rate of inbreeding reduced from 0.05 to 0.01. Consequently, the expected genetic gain was also decreased as the constraint becomes more stringent (Figure 1). These results are expected because as more severe restrictions placed on future inbreeding the contributions from best animals will be reduced and more animals are selected in order to minimize the average relationship. For instance, the maximum percentage of progeny per individual was well under 5% when ΔF was 0.01 compared with 10% when ΔF was 0.05 (Figure 1). Similar results have been reported in other OC selection studies (e.g. Hinrichs et al., 2006).

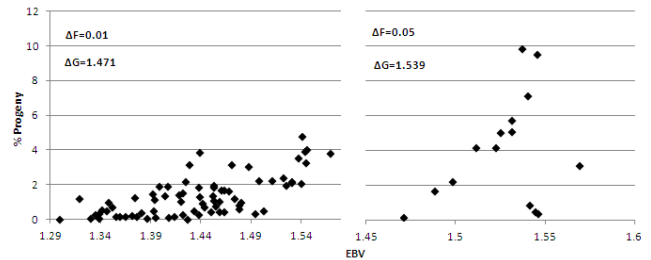


Figure 1: Association between EBV and optimized genetic contribution for cattle dataset with 3907 male selection candidates for two levels of constraints on rate of inbreeding (ΔF). ΔG = genetic gain.

Results of the current algorithm were compared with Gencont (Meuwissen, 2002) on the basis of the optimal number of animals to select, genetic gain and computer time. Analyses of the cattle dataset using both algorithms at different rates of inbreeding are summarized in Table 1. Both algorithms suggested selecting equal optimal number of candidates at 0.05 and 0.01 rates of inbreeding and gave similar results with respect to genetic gain. The selected candidates by the two algorithms were the same and the assigned contributions have correlations very close to 1 (Table 1). However, Gencont2 considerably decreased the computer time by around 92% compared to Gencont (Table 1). For the 0.005 and 0.001 rates of inbreeding, the new method selects fewer numbers of candidates and gives slightly higher genetic gain. The extra selected candidates have contributions between 0.006-0.03%. Moreover, both algorithms successfully met the constraints.

Table 1: Analysis of cattle dataset using Gencont and Gencont2. Different level of rate of inbreeding (ΔF) with respect to genetic gain (ΔG), number of selected individuals and computation time.

ΔF	ΔG^1	# Selected ¹	Time ²	R ³
0.05	1.539	15	8.1	0.999
0.01	1.471	75	7.9	0.999
0.005	1.449 (1.448)	104 (106)	7.9	0.998
0.001	1.424 (1.423)	127 (128)	7.9	0.996

¹if there was difference between the two algorithms, the result in parentheses belong to Gencont

²the amount of computation time used by Gencont2 compared with Gencont in percentage

³Correlation between assigned contributions

The optimal number of boars selected, expected genetic gain and relative computer times at different inbreeding rates (ΔF) for the pig dataset is presented in Table 2. Both algorithms gave very similar results with respect to the optimal number of boars to select and genetic gains when the predefined rate inbreeding were 0.05 and 0.001 (Table 2). The iterative algorithm suggests fewer boars to select when constraints on inbreeding rate were 0.01 and 0.005. The differently selected candidates have contributions within the range of 0.02-0.08% and the correlations between assigned contributions by the two algorithms were > 0.98 (Table 2). For all these analyses both algorithms satisfied the imposed constraints. The iterative algorithm used only from 6.5% up to 9.3% of the computer time to obtain optimal solutions compared to Gencont (Table 2).

Table 2: Analysis of pig dataset using Gencont and Gencont2. Different level of rate of inbreeding (ΔF) with respect to genetic gain (ΔG), number of selected individuals and computation time

ΔF	ΔG^1	# Selected ¹	Time ²	R ³
0.05	129.55	28	6.8	0.999
0.01	125.43(125.39)	73 (77)	6.5	0.986
0.005	124.42 (124.37)	84 (90)	8.2	0.986
0.001	123.44 (123.40)	103	9.3	0.990

¹if there was difference between the two algorithms, the result in parentheses belong to Gencont

²the amount of computation time used by Gencont2 compared with Gencont in percentage

³Correlation between assigned contributions

For the sheep dataset with 6,895 selection candidates, Gencont could not be run because it could not handle such a large dataset. This is because the algorithm requires repeated calculations of inverses of relationship matrices of the selection candidates (Meuwissen, 1997). However, the iterative algorithm was successful in obtaining optimized solutions and achieving predefined constraints (results are not presented).

Inbreeding is a growing concern in animal breeding programs. Advancements in methods for genetic evaluation, such as BLUP, have increased the accuracy of breeding value estimation. This increase in accuracy comes with a

cost of increasing the probability of co-selection of related individuals, which in turn increase inbreeding (Quinton et al., 1992). In the last decades, genetic contributions optimization tools have been developed to manage rate of inbreeding in breeding programs (Meuwissen, 1997, Grundy et al., 1998). The use of OC selection provides a useful tool to control the rate of at which inbreeding accumulates in a population.

One practical challenge in the use of presently existing OC selection algorithms, specifically Gencont, is the heavy computing requirements arising from inverting the relationship matrix for large selection candidates repeatedly. An alternative approach, presented in the current study, is to obtain solutions iteratively and avoid inverting of the relationship matrix. It may be noted that the objective function H could be extended to constrain several (genomic) relationship matrices, which may be relevant if inbreeding is to be constrained at several positions in the genome. This would require updating several λ_0 values in step 3 of the algorithm. As indicated by Sonesson et al. (2012), when genomic selection is used, the pedigree relationship matrix A should be replaced by the genomic relationship matrix.

Conclusion

Optimal contribution selection is a useful tool in controlling rate of inbreeding whilst improving genetic gain. The presented iterative algorithm is many times faster and can handle datasets with large number of selection candidates. This makes practical implementation of OC selection feasible in a large scale to manage the rate of inbreeding in practical breeding programs.

Literature Cited

- Avendaño, S., Villanueva, B., Woolliams, J.A. (2003). J. Anim. Sci. 81:2964-2975.
- Falconer, D. S. (1989). Introduction to Quantitative Genetics. 3rd ed.
- Grundy, B., Villanueva, B., and Woolloams, J.A. (1998). Genet. Res. 72: 159-168
- Hinrichs, D., Wetten, M., and Meuwissen, T.H. (2006). J. Anim. Sci. 84:3212-3218.
- Koenig, S., and Simianer, H. (2006). Lives. Sci. 103:40-53.
- Meuwissen, T.H.E. (1997). J. Anim. Sci. 75:934-940.
- Meuwissen, T.H.E. (2002). Proc. 7th WCGALP, volume 33:769-770.
- Pong-Wong, R. and Woolliams, J.A. (2007). Genet. Sel. Evo. 39:3-25.
- Sonesson, A. K., Woolliams, J.A, and Meuwissen, T.H. (2012). Genet. Sel. Evo. 44:27
- Quinton, M., Smith, C., and Goddard, M.E. (1992). J.Anim. Sci. 70:1060-1067.