

MANAGING COST OF PHENOTYPING

Cécile Massault¹, Brian Kinghorn¹ and Julius van der Werf^{1,2}

¹ School of Environmental and rural science, University of New England, Armidale, NSW 2351

² Australian Cooperative Research Centre for Sheep Industry Innovation, University of New England, Armidale, NSW 2351

SUMMARY

We investigate a way to reduce phenotyping cost with an approach that uses a differential evolution algorithm to optimize which sets of animals to phenotype. The “fitness function” to optimize was the average accuracy of selection candidates (prediction error variance covariance function) when phenotyping 15 animals of four small illustrative pedigrees (different family structure, balanced and unbalanced family sizes). We compared these results to three other strategies (random phenotyping, 15 selection candidates and sires and 12 selection candidates) for different heritability. The tactical approach was either the best strategy or shared highest position on the podium with an equally performing strategy. Phenotyping patterns are impacted by heritability, family structure and family size. This tactical approach to phenotyping is a first step towards a tool that can more generally help decide which animals should be phenotyped or genotyped to optimize breeding programs.

INTRODUCTION

Some traits which impact on profit are not routinely recorded for genetic evaluation or evaluated through a correlated trait, such as traits that are difficult to record (e.g. carcass traits) and/or expensive to measure (e.g. methane emission of cattle). A strategy that reduces costs of phenotyping would allow the integration of valuable novel traits and reduces the cost of implementing genomic selection for some species (i.e. sheep, beef cattle).

We investigate in this paper an approach to manage total cost of phenotyping. The objective of our approach is to find the best set of animals to phenotype for a given number of phenotypes that maximises the EBV accuracy for selection candidates. The total number of phenotypes to be measured will be dictated by the financial resources of the breeder or farmer. We used a differential algorithm to find the set of animals to phenotype that incur the highest average accuracy. The fitness of the solution was evaluated by an objective function that calculated the average accuracy of selection candidates based the prediction error variance covariance (PEVC) given by the traditional best linear unbiased prediction method (BLUP, Henderson 1984). Using four simple pedigrees, we investigate the phenotyping patterns of optimal average accuracy of selection candidates that results from our tactical approach (depending on family structure and family size) when phenotyping 15 animals. We also compared our approach to three others phenotyping strategies (phenotyping random animals, only selection candidates or sires and selection candidates).

MATERIAL AND METHODS

Phenotyping approach. To find which animals should be phenotyped given a maximum number, we used a differential evolution algorithm (DE, Price and Storn 1997), that judges on the fitness of solutions by maximizing accuracy obtained through the fitness function described below. The DE is set up with a population of 10 solutions that evolved for a maximum of 100,000 generations. Two other criteria can stop the evolution prematurely when they have been all met: the DE run for at least 1000 generations without any improvement and solution has to exceed 99.5% of current predicted asymptotic maximum solution (see Kinghorn 2008 for full description).

Fitness function. The fitness function establishes the average accuracy of selection candidates given a set of animals to phenotype for the simple case of a single trait with no fixed effects. The prediction error variance-covariance corresponds to the inverse of the left hand side of the mixed model equations used in traditional BLUP genetic evaluation:

$$PEVC_i = diagonal(Z'Z + \lambda A^{-1})^{-1} \times MSE$$

where Z is the matrix that contains information on which animals is phenotyped, A is the relationship matrix and $\lambda = \frac{1-h^2}{h^2}$. MSE represents the mean squared error, assumed to be 1. Individual accuracy was then calculated as:

$$r_{IH_i} = \sqrt{1 - \lambda PEVC_i}$$

The value maximized by the fitness function was the average accuracy of selection candidates.

Pedigrees. We designed 4 representative pedigrees with specific characteristics to evaluate our approach for common population structures found in animal breeding. The pedigrees are described in **Table 1**. They all comprise 3 generations and the last generation (offspring) represents selection candidates. We investigated the impact of different family structure (half-sib and full-sib), of balanced and unbalanced family sizes. Family sizes in the unbalanced pedigrees are 2 (small family), 10 (medium family), and 18 (large family) offspring. We looked at the particular cases of phenotyping 15 animals which corresponds to half of the selection candidates and for three heritabilities of 0.1, 0.5 and 0.8 (TACT strategy). We also examined the accuracy when phenotyping 15 animals randomly across generations (RAND strategy), phenotyping 15 selection candidates randomly (OFFS strategy) and phenotyping the sire of each family and 12 random offspring (SIRE strategy) for each pedigree and each heritability (0.1, 0.5 and 0.8).

Table 1. Illustrative pedigrees family structures

	Structure	# GP	# sires	# dams	# offspring	Family size		
PED1	HS	66	3	30	30	10	10	10
PED2	FS	12	3	3	30	10	10	10
PED3	HS	66	3	30	30	2	10	18
PED4	FS	12	3	3	30	2	10	18

RESULTS AND DISCUSSION

Table 2 reports the percentage of maximum accuracy achieved by the four different phenotyping strategies. As expected, RAND was the least efficient strategy. Out of the 12 scenarios, TACT always performs best. In 2 cases, TACT was equivalent to OFFS (PED2 heritability of 0.5 and 0.8) and in 2 cases equivalent to SIRE (PED1, heritability 0.1 and 0.5). This indicates that our tactical approach found the best strategy, which also happened to be one of other strategy. Figure 1 shows the distribution of phenotypes for the families when phenotyping 15 animals in PED1 to PED4 for heritability of 0.1, 0.5 and 0.8 using the TACT approach. We are considering three factors that impact on the phenotyping pattern given a fixed number of individuals to phenotype and a fixed number of selection candidates: heritability, family structure and balanced/unbalanced family size.

Table 2. Percentage of maximum accuracy (when all animals are phenotyped) captured when phenotyping 15 animals for PED1 to PED4 for heritability of 0.1, 0.5 and 0.8 with TACT, RAND, OFFS and SIRE phenotyping strategies

	Heritability	TACT	RAND	OFFS	SIRE
PED1	0.1	63	33	51	63
	0.5	69	38	60	69
	0.8	72	40	62	69
PED2	0.1	73	63	71	69
	0.5	84	75	84	80
	0.8	84	76	84	80
PED3	0.1	66	32	61	64
	0.5	71	40	69	69
	0.8	73	42	70	71
PED4	0.1	77	67	75	73
	0.5	85	77	84	82
	0.8	85	77	82	81

Optimal phenotyping patterns vary with heritability except for the case of PED2. We can also observe difference in optimal phenotyping pattern when the structure is different. For example, for the same heritability of 0.1, TACT recommends to phenotype 1 sire and 4 offspring per family with a half-sib structure (PED1), while it recommends to phenotype 5 offspring per family with a full-sib structure (PED2). Finally, we can note that phenotyping patterns are different between balanced and unbalanced family sizes (e.g. PED2 and PED4). Phenotyping patterns are clearly impacted by family structure, heritability and family size.

We also use the tactical approach with a pedigree of 50 half-sib families of various sizes (3-16) for 3 heritabilities and phenotyped patterns observed were similar to the ones described in Figure 1 for PED3. The phenotyping approach is a useful tool to find the best set of animals to phenotype for a given number of phenotypes. The advantage of such an approach is that accuracy is only slightly lower, while cost of phenotyping can be significantly reduced. This could permit the inclusion in the breeding goal of new traits that are expensive to measure, including genotyping. In this study, we maximized accuracy of the youngest cohort of selection candidates, as it affects short term response. Genetic gain is impacted by accuracy, as well as generation interval. Increasing accuracy of younger animals tends to reduce the generation interval and therefore increases genetic gain. Optimal response in the longer term is obtained when genetic diversity is also considered. Phenotyped animals are more informative and are more likely to be selected for the next generation and phenotyping many related animals is more likely to increase inbreeding. Constraints on inbreeding as used in optimum contributions selection (Sonesson and Meuwissen 2000) can also be added to the tactical phenotyping approach described here. A further step is to extend the tactical approach to phenotyping to a multi-trait case, and subsequently to include genotyping for the case of genomic selection. Further work will also address the need to reduce the computational time e.g. by using approximate accuracies (Meyer 1989) for larger pedigrees.

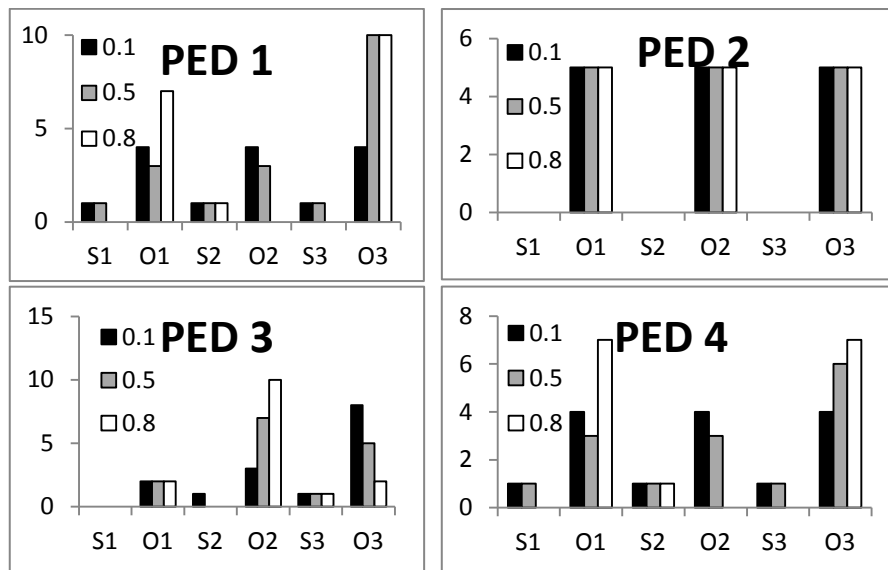


Figure 1. Optimal phenotyping patterns for PED1, PED2, PED3 and PED4 when phenotyping 15 animals for heritability 0.1 (black), 0.5 (grey) and 0.8 (white). S_i refers to sire of family i , O_i refer to offspring of family i .

REFERENCES

- Kinghorn B.P. (2008) "Application of Evolutionary Algorithms to solve complex problems in Quantitative Genetics and Bioinformatics" ed. Centre for Genetic Improvement of Livestock University of Guelph.
- Meyer, K. (1989) *Livestock Animal Production*. **21**:87
- Sonesson, A.K. and Meuwissen, T.H.E. (2000) *Genetics Selection Evolution* **32**: 231
- Storn, R. and Price, K. (1997) *Journal of global optimization*. **11**:341