

# Selection for robustness: exploring the value of genomic prediction, reaction norm models and phenotyping strategies

M. Ghaderi-Zefreh<sup>1\*</sup>, A. Doeschl-Wilson<sup>1</sup>, V. Riggio<sup>1,2</sup>, O. Matika<sup>1,2</sup> and R. Pong-Wong<sup>1</sup>

<sup>1</sup>The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK; <sup>2</sup>Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK; \*[mghaderi@ed.ac.uk](mailto:mghaderi@ed.ac.uk)

## Abstract

Animals with high production potential and simultaneously high resilience to environmental challenges are vital for sustainable livestock production. A simulation study was carried out to (i) evaluate the influence of genomic data, statistical models, and phenotyping strategies on prediction accuracies for resilience and production potential, and (ii) assess the impact of different selection strategies on genetic improvement in these traits. Prediction accuracies were found to be compromised when families were clustered in similar environments or when the environmental challenge level was unknown. Selection on individuals' performance alone could simultaneously improve resilience if performance records of individuals were obtained across a wide range of environments, whereas a narrow range may decrease resilience. Simultaneous genetic improvement in resilience and production potential strongly benefits from the use of genomic evaluations, reaction norm models, and phenotyping in a wide range of environments.

## Introduction

Robust animals are highly resilient (i.e., remain productive in challenging conditions) and have a high production potential (Knap and Doeschl-Wilson, 2020). To genetically improve robustness in animals, accurate estimates of breeding values for production potential and resilience are needed. These can be estimated using reaction norm models based on performance records from different environmental conditions. For linear reaction-norms, the estimated breeding value (EBV) for the intercept represents the production potential, and the inverse of the slope represents resilience (Knap and Su, 2008). One of the impediments in improving robustness using such models is that the environmental challenge level is often unknown and that proxies need to be used instead (Knap and Doeschl-Wilson, 2020).

The objectives of this study were: firstly, to evaluate the influence of genomic data, statistical models, and phenotyping strategies on the accuracy of genetic prediction for resilience and production potential. Secondly, to assess the impact of different selection strategies on genetic improvement in these traits. The data are simulated assuming a sheep population, but the results can be generalised to other species.

## Materials & Methods

**Simulation of data.** The phenotypes were simulated assuming a linear reaction norm model where the performance is affected by two components: one related to the performance under ideal rearing condition (i.e., no challenge) and another influenced by the ability of the animal to withstand a challenge. Here, these components will be denoted as the production potential and resilience, and both are assumed to be under genetic and environmental control.

Hence, the performance ( $y$ ) of individual  $j$  in environment  $k$  with a challenge level  $X_k$  is:

$$y_{j,k} = (\mu_0 + a_{0j} + e_{0j,k}) + X_k (\mu_R + a_{Rj} + e_{Rj,k}) \quad (1)$$

where  $\mu_0$  is the mean performance of the population in the absence of any challenge and,  $\mu_R$  is the average reduction of performance due to the challenge level  $X_k$ ;  $a_{0j}$  and  $a_{Rj}$  are the breeding values of individual  $j$  for production potential and resilience;  $e_{0j,k}$  and  $e_{Rj,k}$  are the respective random environmental deviations for each trait. Both traits were assumed with a heritability of 0.1, phenotypic variance of 1.0, zero environmental correlation and an antagonistic genetic correlation of -0.5. Values for  $\mu_0$  and  $\mu_R$  were arbitrary chosen to be 10 and -3, respectively and  $X$  values ranged between 0 and 2 (i.e., average performance in best and worst environment differ by six phenotypic deviation).

The true breeding values were sampled by simulating the genome of each individual in the population (see info about population structure below). To assume a sheep species, the genome was distributed on 26 chromosomes, each with 1,500 segregating biallelic loci in linkage disequilibrium (500 QTL affecting both traits, and 1,000 in the SNP panel used to calculate the genomic relationship matrix). Linkage disequilibrium was simulated by sampling the base animals using a mutation-drift algorithm (Meuwissen et al., 2001). The offspring' genome was simulated by gene dropping haplotypes following Mendelian inheritance rules, given their parents' genotypes. The environmental deviations ( $e_0$  and  $e_R$ ) for each individual and environment  $k$  were sampled from a normal distribution.

The level of the environmental challenge that an individual was exposed to ( $X$ ) was simulated by assuming ten farms, each farm being a subset (10%) of the entire range of challenge levels. Once individuals were allocated to a farm, their specific challenge level were sampled within the range of that farm (i.e., all individuals in the same farm have different challenge level), and their phenotype were calculated from equation (1).

**Genetic evaluation.** Best linear unbiased predictor (BLUP) or genomic BLUP (GBLUP) (Strandén and Garrick, 2009) assuming the random regression model as in equation (1) was used to obtain genomic estimated breeding values (GEBV) for both production potential and resilience. If the level of the environmental challenge was assumed to be unknown (i.e., not measured), a reaction norm approach (Knap and Su, 2008) was implemented where the unknown challenge levels are first estimated using the average farm effects. The farm effect is estimated with the same data using a simpler model ignoring the resilience component. The farm effect estimates are rescaled, and all individuals with performance within a farm are assumed to have the same challenge level, equal to the scaled value of the effect of that farm.

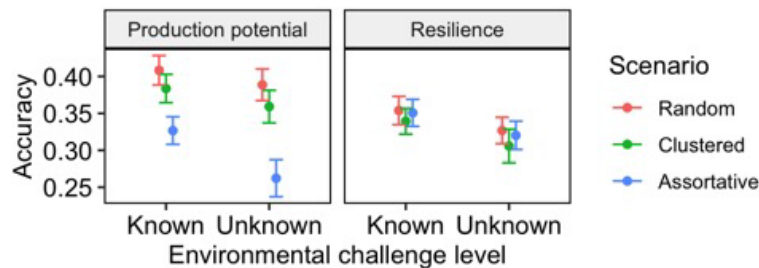
**Scenarios compared.** To assess the factors affecting the GEBV accuracy, a population of half sib families was assumed, where 100 grandsires were mated to 1,100 granddams to produce 100 sires and 1000 dams, which in turn, were mated to produce 1,000 offspring. Only the offspring generation were assumed to have performance records. We considered three different scenarios how families were distributed across environmental conditions: random, clustered, and assortative. In the random scenario, offspring of a sire are randomly allocated to different farms. In the clustered scenario, all offspring of a sire are allocated to the same farm. In the assortative scenario, a correlation between genetic merit and environment challenge was assumed, so that offspring of sires with the highest EBV were allocated to farms with the lowest challenge levels. The effect of knowing or not the challenge level was also assessed.

To assess the response to selection for resilience, we assumed a population of 8,100 individuals (2,700 males and 5,400 females) per generation, where the best 270 males and 2,700 females are selected to produce the next generation. Individuals were randomly distributed across the entire range of environments. The genetic evaluation was done via GBLUP implementing the reaction norm model (where the estimated farm effects were used

as challenge levels, see genetic evaluation section above). All candidates were genotyped and had one phenotypic record. Thereafter, a selection index was defined as  $I=(1-\alpha)*\hat{a}_0 + \alpha*\hat{a}_R$ , where  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is the weight given to resilience. The selection scheme was done over 10 discrete generations. We also considered scenarios where performance records were evaluated in a narrow range of environment (i.e.,  $X$  values between 0 and 1). The results were compared with the case where candidates were evaluated using a conventional genetic model that does not account for the term influenced by the level of environmental challenge.

## Results

***Impact of genomic information, unknown environmental challenge level and distribution of families across environments on prediction accuracies.*** As expected, genetic predictions for production potential and resilience were significantly more accurate with GBLUP than with BLUP (70-163% and 79-114%, respectively). Figure 1 shows the effect of unknown challenge level and the distribution of families on the prediction accuracies for production potential and resilience. Both traits have reduced accuracy when the environmental challenge level is unknown. In general, the accuracy of GEBVs for production potential are higher than that of GEBVs for resilience but also more sensitive to the distribution of individuals across environmental conditions. When the individuals were allocated at random across the whole range of available environmental conditions, the accuracy of GEBVs was highest (between 6-75% higher for production potential and 4-17% for resilience).

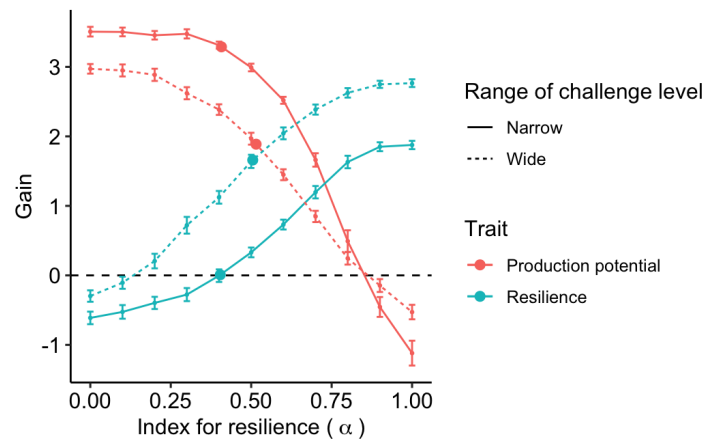


**Figure 1. Effect of unknown environmental challenge level and distribution of families across environments on the prediction accuracy of production potential and resilience.**

***Response to selection.*** Figure 2 shows the genetic gain for production potential and resilience after ten generations of selection using different strategies and when the data are obtained in a wide or a narrow range of environmental challenge levels. When the selection index has all the weight on one component only (i.e.,  $\alpha=0$  or  $\alpha=1$ ), a genetic loss for the other component was observed due to the assumed negative genetic correlation. When the information comes from a wide range of environmental conditions, selection based on the conventional model (which ignores the resilience component) yielded genetic gain for both traits, despite their negative genetic correlation (solid circles in Figure 2). However, when the range of environmental conditions were narrowed down, selection ignoring resilience yielded genetic gain for production potential, but resilience deteriorated. Using a selection index based on the GEBVs of the reaction norm analysis, both traits can be improved regardless of the range of environmental conditions where the animals were measured.

The trend of the selection response when varying  $\alpha$  depended on the range of environments where the animals were phenotyped. Phenotypes obtained from a narrower range of environmental conditions generally led to higher gain for production potential (except for large index weight on resilience,  $\alpha > 0.84$ ), but reduced genetic gain for resilience. Genetic

gain in resilience could only be obtained when the index weight  $\alpha$  was sufficiently large ( $\alpha > 0.12$  and  $\alpha > 0.42$  corresponding to the wide and narrow environmental range, respectively).



**Figure 2. Genetic gain for production potential and resilience after ten generations of selection with different index weight  $\alpha$  for resilience (error-bars indicate 95% C.I.). The results for the conventional model, which ignores the resilience component, are shown with solid circles.**

### Discussion and conclusion

The study showed that genomic prediction is beneficial to improve genetic prediction for production potential and resilience. The GEBV accuracies for both traits are strongly affected by the distribution of phenotyped individuals across environments and knowledge of the challenge level. Whilst the prediction accuracies for resilience was generally lower than for production potential, they were less sensitive to the distribution of phenotypes. Our results on the response to selection show the potential benefits for improving resilience using a random regression/reaction norm model in the evaluation. Conventional genetic evaluations ignoring resilience can improve both resilience and production potential if candidates are evaluated under a wide range of environmental conditions, even when their genetic correlation is negative. However, this is not the case when the performance data come from a narrow range of environmental conditions. This problem can be avoided with the reaction norm model. Hence, reaction norm model and selection index with appropriate weights have great potential for simultaneous improvement of both production potential and resilience.

### Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER) and from the Roslin Institute Strategic Programme grant (ISP3.1, BBS/E/D/30002275).

### References

- Knap P.W., and Doeschl-Wilson A. (2020) *Genet. Sel. Evol.* 52:60.  
<https://doi.org/10.1186/s12711-020-00580-4>
- Knap P.W., and Su G. (2008) *Animal* 2(12):1742-1747.  
<https://doi.org/10.1017/S1751731108003145>
- Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) *Genetics*. 157(4):1819-1829.  
<https://doi.org/10.1534/genetics.116.189795>
- Strandén I., and Garrick D.J. (2009) *J. Dairy Sci.* 92(6):2971-2975.  
<https://doi.org/10.3168/jds.2008-1929>