

# Genomic prediction of feed intake using predictor traits

C. I. V. Manzanilla-Pech<sup>1,2</sup>, R. F. Veerkamp<sup>1</sup>, Y. de Haas<sup>1</sup>, M. P. L. Calus<sup>1</sup>, J. ten Napel<sup>1</sup>

<sup>1</sup>Animal Breeding and Genomics, Wageningen University and Research, PO Box 338, 6700 AH Wageningen, the Netherlands

<sup>2</sup>Mococha Research Station, National Institute of Forestry, Agriculture and Livestock Research, 97454 Mococha, Yucatan, Mexico

## Summary

A total of 77,640 weekly records on dry matter intake (DMI), 64,443 on fat and protein corrected milk (FPCM) and 73,415 on live weight (LW) were analysed from 3,188 Dutch dairy cows in 6,820 lactations (first to third lactation) from 1980 to 2015. The objective of this study was to compare the accuracies of the genomic estimated breeding values (GEBV) for DMI, with or without predictor traits included (FPCM and LW) with a single step method (SS-GBLUP). Accuracies of GEBV for DMI was 0.36 when FPCM and LW were included as reference traits, 0.37 when DMI was the only reference trait, and 0.38 when all 3 traits (DMI, FPCM and LW) were included as reference traits. When only using predictor traits in the reference population, the accuracies of estimated GEBV for DMI, were lower than in the scenarios using DMI as LW and FPCM can only explain 53% of the variation in DMI. Moreover, there was very little benefit of adding information on predictor traits to the reference population when DMI was already included on the same animals. However, in the absence of DMI records, having records on FPCM and LW from different lactations is a good way to obtain GEBV with a relatively good accuracy.

*Keywords: genomic prediction, dry matter intake, predictor traits*

## Introduction

Optimization of dairy cattle breeding goals for feed efficiency requires the availability of breeding values for dry matter intake (DMI), as this is an important component of feed efficiency. To estimate accurate DMI breeding values, a large number of records is required. However, DMI is labour-intensive and therefore expensive to measure. Also, as DMI is not a trait typically recorded in commercial herds, the amount of available data is limited. The difficulty in recording DMI has hampered direct selection for DMI previously, since insufficient records were available on daughters of progeny-tested bulls. This difficulty might be overcome by using both DMI and predictor traits (Veerkamp and Brotherstone, 1997, Berry and Crowley, 2013). Some traits have been suggested to predict DMI; e.g., production or traits reflecting maintenance costs. Readily available predictor traits that are easier and cheaper to measure compared to DMI are fat and protein corrected milk (FPCM) and live weight (LW). Both traits are known to be strongly correlated with DMI (Korver, 1988, Van Arendonk, 1991, Veerkamp, 1998). Moreover, with the use of genomic prediction using single nucleotide polymorphism (SNP) markers and DMI in a reference population, combining it has become easier to accurately predict (Meuwissen et al., 2001) for it might be able to predict GEBV for DMI. This genomic approach combined with predictor traits, could contribute to increase the accuracy of GEBV for DMI.

## Materials and methods

### Phenotypes and Genotypes

A unique large dataset was available with a total of 77,640 weekly records on DMI, 64,443 on fat and protein corrected milk (FPCM) and 73,415 on live weight (LW) from 3,188 Dutch dairy cows in 6,820 lactations (first to third lactation) recorded in the Netherlands between

1980 and 2015. Recording frequencies of DMI varied by experiment: it was recorded either one, two, three or five times per week. Daily DMI records were averaged to weekly records in order to homogenize the data across experiments. An overview of the experiments, treatments and diets is given in Manzanilla-Pech et al. (2014). The following formula was used to calculate FPCM (FAO, 2010):

$$\text{FPCM (kg)} = \text{raw milk (kg)} * (0.337 + 0.116 * \text{fat content (\%)} + 0.06 * \text{protein content (\%)})$$

A total of 1,496 cows were genotyped (1,421 of them with records on DMI), either with 50K Illumina, 80K Geneseek, or 15K Eurogenomics SNP chips. The information from these three chips (missing genotypes and animals with 15K and 50K) was imputed to a customized set of 76,439 SNP using Beagle 3 (Browning and Browning, 2009). After editing based on phenotype records of DMI, in terms of standard deviations ( $\pm 5$ ), and the number of cows per experimental treatment (5), 1,313 of 1,421 animals with genotypes were retained.

### Estimation of genetic parameters and GEBV

A trivariate analysis was performed to estimate the variance and covariance components between DMI, FPCM, and LW using ASReml4 (Gilmour et al., 2009). The model used to estimate the genetic parameters for DMI, FPCM and LW included as fixed effects: herd, year and month, fraction of Holstein (5/8 to 8/8) by parity, age of cow in months, days in milk; and as random effects: the additive genetic effect, permanent environmental effect and the residual.

The GEBV for DMI were estimated using the MiXBLUP software (ten Napel et al., 2016) with a SS-GBLUP model using the inverted relationship matrix  $H^{-1}$ . The model was as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{c} + \mathbf{e} \quad [1]$$

where  $\mathbf{y}$  is the vector of phenotypes;  $\mathbf{X}$  and  $\mathbf{Z}_1, \mathbf{Z}_2$  are incidence matrices relating observations with fixed, direct additive genetic and permanent environmental effects;  $\boldsymbol{\beta}$  is the vector of fixed effects;  $\mathbf{a}$  is the vector of direct additive genetic effects; and  $\mathbf{c}$  is the vector of permanent environmental effects and  $\mathbf{e}$  is the vector of residual effects. Distributions of the random effects are  $\text{var}(\mathbf{a}) = \mathbf{H}\sigma_a^2$  for the SS-GBLUP method,  $\text{var}(\mathbf{c}) = \mathbf{I}_c\sigma_c^2$  where  $\mathbf{I}_c$  is identity matrix of an order equal to the number of individuals with records, and  $\text{var}(\mathbf{e}) = \mathbf{I}_E\sigma_e^2$ , where  $\mathbf{I}_E$  is an identity matrix of an order equal to the number of observations.

### Cross-validation groups and calculation of accuracies

The population of genotyped individuals was divided into four validation groups. The assignment to the groups was made by sire, using stratified random sampling, the sires of genotyped animals were sorted from having the largest to the smallest number of daughters. For each of the validation groups, GEBV were predicted after excluding the respective phenotypes from the analysis, using phenotypes of the other three validation groups only.

The accuracies were calculated via cross-validation per scenario: 1) DMI; 2) FPCM and LW; and 3) DMI, FPCM, LW as the correlation between the adjusted phenotype for DMI and the GEBV for DMI. Adjusted phenotypes for DMI were calculated per animal in ASReml4 (Gilmour et al., 2009) as the solutions of a model using all fixed effects in the model described above, but excluding the genetic animal effect. The accuracies were calculated as the above correlation divided by the approximated accuracy of the adjusted phenotypes computed with the formula for repeated records of Falconer and Mackay (1996).

## Results and Discussion

### Genetic parameters and percentage of the variance of DMI explained by predictor traits

Estimated heritabilities were 0.17 for DMI, 0.24 for FPCM, and 0.43 for LW. The estimated heritability of DMI was within the range (0.11 to 0.35) of previous studies presenting heritabilities, reviewed by Berry and Crowley (2013). Estimated heritabilities of FPCM in the current study were slightly lower than previously reported heritabilities ranging between 0.27 and 0.47 (Van Arendonk et al., 1991, Pszczola et al., 2013). Estimated heritabilities for LW in the current study were in the lower range of estimates in the literature ranging from 0.43 to 0.65 (Veerkamp and Brotherstone, 1997, Koenen and Veerkamp, 1998, Berry et al., 2003).

Moderate to high (0.4 to 0.8) genetic correlations between DMI and predictor traits, i.e., FPCM and LW, have been estimated in several studies (Veerkamp and Brotherstone, 1997, Cooper et al., 2010, Vallimont et al., 2011, Berry and Crowley, 2013). In the current study, the genetic correlation of DMI and FPCM was 0.59 and between DMI and LW was 0.43. Given that a considerable part of feed intake is used for milk production, maintenance (Veerkamp, 1998). An important question is how much genetic variation in DMI can be explained by these predictor traits. Based on the genetic correlations obtained in this study (0.59 and 0.43), 35% ( $0.59^2$ ) of the genetic variation in DMI is attributable to FPCM, while only 18% ( $0.43^2$ ) is attributable to LW. The genetic correlation between FPCM and LW was close to zero indicating that and LW explained independent parts of the genetic variation in DMI. Thus, together they explained 53% (35+18), meaning that 53% of the genetic variance in DMI could be explained by FPCM and LW as predictor traits, which will the maximum accuracy of prediction for DMI in this case is 0.73 ( $\sqrt{0.53}$ ).

### Accuracies of GEBV for DMI

Accuracy of GEBV for DMI was the lowest (0.36) when FPCM and LW were included as reference traits, slightly higher (0.37) when only DMI was included as reference trait, and was the highest (0.38) when all 3 traits (DMI, FPCM and LW) were included as reference traits. This suggests that the value of adding additional information from predictor traits, when DMI is already recorded in the reference population on the same animals, is limited. However, in this study it was not information available for the validation animals. Pszczola et al., (2013) reported higher accuracies ( $>0.47$ ) when predictor traits were recorded in the validation population. When only using predictor traits in the reference population, the accuracies of estimated GEBV for DMI, were lower than in the scenarios using DMI. The difference was not large, indicating that, in absence of DMI records, FPCM and LW could still predict a large and useful part of DMI. However, as the genetic correlations suggested, only the genetic variation in DMI can be predicted from predictors. These results are in agreement with Pszczola et al. (2013), who observed no benefit when adding FPCM to DMI in the reference population, while only a slight increase in accuracy (0.02) was observed when adding LW to DMI and FPCM in the reference population.

In conclusion, there was very little benefit of adding information on predictor traits to the reference population when DMI was already included. However, in the absence of DMI records, having records on FPCM and LW from different lactations is a good way to obtain GEBV with a relatively good accuracy.

## List of References

- Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. *J. Dairy Sci.* 86(11):3704-3717.
- Berry, D. P. and J. J. Crowley. 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91(4):1594-1613.
- Browning, B. L. and S. R. Browning. 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am J Hum Genet* 84(2):210-223.
- Cooper, A. J., C. L. Ferrell, L. V. Cundiff, and L. D. Van Vleck. 2010. Prediction of genetic values for feed intake from individual body weight gain and total feed intake of the pen. *J. Anim. Sci.* 88(6):1967-1972.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Pearson, Prentice Hall, Upper Saddle River, NJ.
- FAO. 2010. Food and agriculture organization of the United Nations. Greenhouse gas emissions from the dairy sector.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. *ASREML User Guide Release 3.0* VSN International Ltd, Hemel Hempstead, UK.
- Koenen, E. P. C. and R. F. Veerkamp. 1998. Genetic covariance functions for live weight, condition score, and dry-matter intake measured at different lactation stages of Holstein Friesian heifers. *Livest. Prod. Sci.* 57(1):67-77.
- Korver, S. 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. *Livest. Prod. Sci.* 20(1):1-13.
- Manzanilla-Pech, C. I. V., R. F. Veerkamp, M. P. L. Calus, R. Zom, A. van Knegsel, J. E. Pryce, and Y. De Haas. 2014. Genetic parameters across lactation for feed intake, fat-and protein-corrected milk, and liveweight in first-parity Holstein cattle. *J. Dairy Sci.* 97(9):5851-5862.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819-1829.
- Pszczola, M., R. F. Veerkamp, Y. de Haas, E. Wall, T. Strabel, and M. P. L. Calus. 2013. Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal* 7(11):1759-1768.
- ten Napel, J., M.P.L. Calus, M. Lidauer, I. Strandén, E. Mäntysaari, H.A. Mulder, R.F. Veerkamp. 2016. *MiXBLUP, user-friendly software for large genetic evaluation systems – Manual V2.0-2016-01*. Wageningen, the Netherlands. Accessed via [www.mixblup.eu](http://www.mixblup.eu).
- Vallimont, J. E., C. D. Dechow, J. M. Daubert, M. W. Dekleva, J. W. Blum, C. M. Barlieb, W. Liu, G. A. Varga, A. J. Heinrichs, and C. R. Baumrucker. 2011. Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. *J Dairy Sci* 94(4):2108-2113.
- Van Arendonk, J. A. M., G. J. Nieuwhof, H. Vos, and S. Korver. 1991. Genetic aspects of feed intake and efficiency in lactating dairy heifers. *Livest. Prod. Sci.* 29(4):263-275.
- Veerkamp, R. F. 1998. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: A review. *J. Dairy Sci.* 81(4):1109-1119.
- Veerkamp, R. F. and S. Brotherstone. 1997. Genetic correlations between linear type traits, food intake, live weight and condition score in Holstein Friesian dairy cattle. *Anim Sci* 64:385-392.