

Note – Genomic evaluation of Metabolic efficiency

By Rasmus B. Stephansen and Ulrik S. Nielsen, SEGES

Introduction

The motivation of this note is documentation of data, editing procedures and genetic evaluation of metabolic efficiency (**RFI**), which is a part of the Saved feed index. The phenotypes that is required to estimate breeding values for RFI is feed intake measured individually and the most common energy sink traits for a modern dairy cow (typically production, body weight & change in energy status).

Data description

There are 3 data sources of individually measured feed intake data in this project, (1) the Efficient Dairy Genome Project (**EDGP**) on Holstein cows, (2) Luke on Red cows and (3) the Cattle Feed Intake (**CFIT**) by VikingGenetics with data from Holstein, Red cows and Jersey cows.

EDGP data

The EDGP dataset is a joint consortium of research farm data, where the consortium provide access to each other's feed intake data with the purpose of promoting a genetic evaluation for feed efficiency.

In this project there is used data from following research farms:

- Elora – Canada
- DRTC – Canada
- Foulum – Denmark
- Beltsville – USA
- Ellinbank – Australia

Data is not used from Strickhof and Posieux in Switzerland due to few cows and the data from Langhill in United Kingdom due to very heterogenous data.

Luke data

The dataset from Luke contains feed intake on daily level for 1st parity cows at the research farms Rehtijarvi, Minkio, Viikki and Maaninka in Finland.

CFIT data

CFIT data is provided by VikingGenetics on all three breeds, using 3D technology to identify the individual cows' daily feed intake and body weight. CFIT data is merged with the test day data from the Danish cattle database.

Data editing

All data sources must undergo data editing to ensure that outliers were removed. Only data from 1st to 6th parity and the first 330 days in milk is utilized in the genetic evaluation of RFI. Data delivered from Finland was edited according to research criteria and therefore Luke data was not edited.

Data was deleted in EDGP and CFIT data if

- Calving age at 1st calving in the range 20-34 month for HOL and 18-35 for JER
- Cows is moved to another herd with lactation
- Less than two test days and six feed intake days - only CFIT data

Daily observations were set missing for the specific phenotypes if:

- Daily dry matter intake not in the range 2-50 kg dry matter – only EDGP
- Daily feed intake not in the range 10-100 kg – only CFIT
- Body weight (**BW**) not in the range 300-1,100 kg for HOL+RDC and 100-900 kg for JER
- Daily milk production not in the range 5-100 kg
- Daily fat production not in the range 0.05-5.0 kg – only EDGP
- Fat content not in the range 2.5-10.0 for JER and 2.5-8.0 for HOL+RDC – only CFIT
- Daily protein production not in the range 0.05-4.0 kg - only EDGP
- Protein content not in the range 2.0-8.0 for JER and 2.0-6.0 for HOL+RDC – only CFIT
- Daily Δ BW not in the interval -15 to 15 for EDGP data and -7.5 to 7.5 kg for CFIT data

Afterwards outliers were defined as greater than the range $\text{mean} \pm 4 \times \text{SD}$ for the single phenotypes. Outliers was set missing.

Production, feed intake and BW data are interpolated by animal and parity to maximize the number of records. This is especially important for production data in CFIT data, since production data here only is measured once a month and feed intake is measured on weekly basis.

Genetic evaluation of Metabolic efficiency

The use of RFI as selections criteria for feed efficiency was firstly described by Koch et al. (1963), using the phenotypic covariance matrix. Kennedy et al. (1993) developed the theory of using genetic RFI, using the genetic covariance matrix. However, limiting literature describes how to use genetic RFI – therefor calculation of RFI in this report is performed using phenotypic RFI.

Li et al. (2017) demonstrated that the assumptions on having 1 regression for ECM and Δ BW might be wrong, because they fluctuate during lactation in her results. This assumption will be adapted in the model construction of RFI in this project.

Model construction for RFI in two step approach

The first step to calculate the phenotype for RFI. The model used to calculate the phenotype is for all breeds:

$$DMI_{lpP} = CA_p + CA_p^2 + LACP_P + HYS + YS_{LACP} + ECM_{lP} + MBW_P + \Delta BW_{lP} + e_{iw}$$

where

DMI is the average daily dry matter intake,

CA_p is the linear regression for age at calving nested within p parity ($p=1, 2, \dots, 6$),

CA_p^2 is the quadratic terms for age at calving nested within p parity ($p=1, 2, \dots, 6$),

$LACP$ is the fixed effect lactation period nested within P (1st or later lactations)

HYS is the fixed effect of Herd x Year x Season (quarters for date of observation)

YS_{LACP} is the fixed effect of Year x Season x Lactation period

ECM is the regression on energy corrected milk nested within l lactation period

($l=2,3,\dots,11$) and P (1st or later lactations)

MBW is the regression of metabolic body weight nested within P (1st or later lactations)

ΔBW is the regression of change in body weight nested within l lactation period

($l=2,3,\dots,11$) and P (1st or later lactations)

e is the residual of the model that is used for genetic evaluation.

This model creates a residual/phenotype for each test day across lactation and parities. Normally such a trait would be treated as a trait by parity; however, the data size did not allow that, since there are very few records. Therefore it is crucial to correct for heterogeneous variance to avoid that the effect of lactation stage and parity goes into the breeding values.

Heterogeneous variance correction

For each breed there is calculated an average standard deviation for RFI across parities, lactation stages and year x season. The next step is to standardize all residuals to the average variance by:

$$\hat{y} = \mu_{PIY} + ((\mu_{PIY} - y_i) \times (\hat{\sigma} / \sigma_{PIY}))$$

Where

$\hat{y}$ is the standardized phenotype,

μ_{PIY} is the mean RFI level in P (1st or later lactations), l lactation period and Y year x season,

y_i is the residual for the i observation of a cow

$\hat{\sigma}$ is the standard deviation that is standardized for,

σ_{PIY} is the standard deviation level in P (1st or later lactations), l lactation period and Y year x season.

After the standardization the fixed effects of HYS and YS_{LACP} from the RFI model, is added to the heterogeneous corrected phenotype. The reason for doing that is to avoid that additive variance is lost in herd and seasonal effects. The effect of HYS and YS_{LACP} will be included in the genetic model.

Validation of genomic breeding values for RFI

The validation test of GEBVs for RFI could not be performed traditionally as described by Mäntysaari et al. (2010), since we have too few data. However, it was decided to use yield traits as an indicator for RFI, to evaluate the expected genomic reliability. The procedure was to compare to deregressed proofs from the operation run with genomic breeding values from the EDGP data set. This was performed for:

- Bulls born from 2011 to 2015
- Effective daughter contribution is GT 5 in the operation
- Bulls from the operation had no offspring in EDGP data

For all traits a Proc Reg procedure was used as described by Mäntysaari et al. (2010) and the reliability results are presented in the table below. The results shows for all traits a gain in reliability at 2 to 8% compared to the parent average (calculated from EDGP data as $0.5 \times \text{EBV}_{\text{sire}} + 0.25 \times \text{EBV}_{\text{grandsire}}$). It should be noticed that all production traits used in these calculations has a higher heritability than RFI. Therefore the expected genomic reliability is for RFI is 3% based on this validation study.

	Milk yield		Fat		Protein	
	PA	GEBV	PA	GEBV	PA	GEBV
HOL (1,303 bulls)	0.05	0.13	0.05	0.07	0.02	0.05

Genomic evaluation of residual feed intake

It has been a challenge that genomic prediction with Single step has required a lot of memory in the inversion of the G-matrix. Therefore, for computational reasons it was decided to use ssGTaBLUP in MiX99 (MiX99 Development Team, 2019) with 10% weight on the residual polygenetic effect. This method uses a T-matrix which is like the G-matrix, however this does not require inversion, that is memory demanding. For detailed explanation about Single Step methods, find more explanation in Mäntysaari et al. (2020)

All animals with data was traced for 3 generations (5 generations for JER). Afterwards all genotyped animals (candidates) were included in the pedigree and genotypes was imputed based on the animals in the pedigree. In the pedigree genetic groups are included as 5-birth year classes, country and breed.

The repeatability model used for the genetic evaluation for all breeds are as follows:

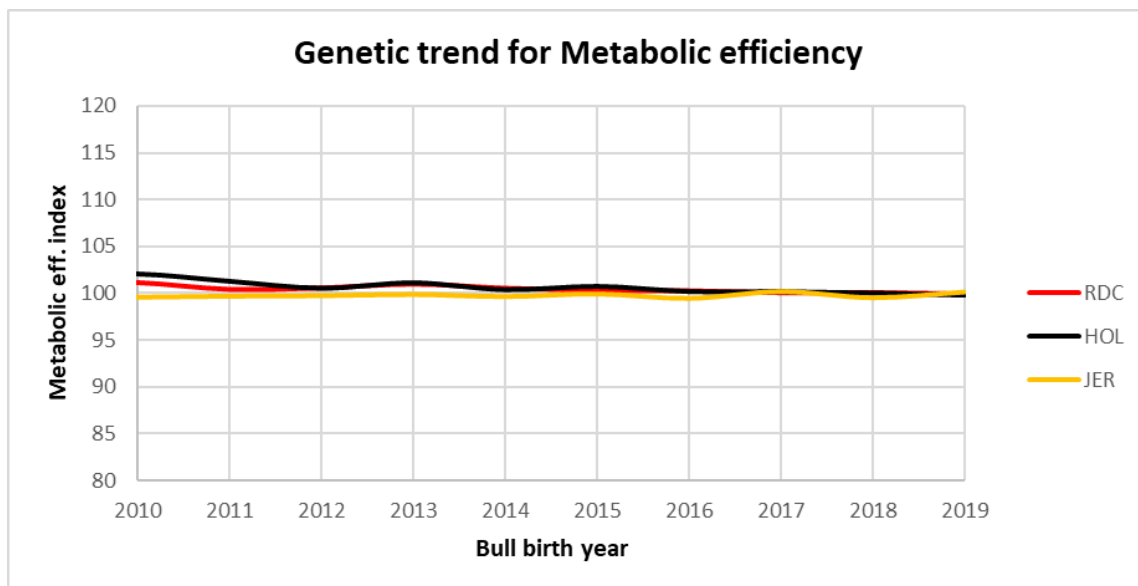
$$RFI = HYS + YS_{LACP} + a + e$$

The model was performed with following variance component estimates:

	HOL	RDC	JER
Phenotypic SD ¹	1.65	1.84	3.0
Heritability, %	15%	15%	15%
Genetic SD for RFI	0.64	0.71	1.165

Genetic trend and effect breeding values

The genetic trend for bulls GEBVs is plotted in the Figure below. It clearly appears that there is no trend for this trait for all breeds, as expected for a non-selected trait.



The results after the standardization is presented below and shows that cows with data and bulls with offspring gets the highest SD on RFI indices in all breeds. Candidates born from 2017-2019 has a much lower SD on RFI indices as expected, since the genomic reliability for this trait is very low at 3% compared to for example yield at 35% (pure genomic reliability).

Breed	Animal group	N animals	Mean (SD) index
RDC	Phenotyped cows	725	102.0 (4.0)
RDC	Candidate heifers	59,728	100.2 (1.8)
RDC	Candidate bulls	7,648	100.0 (1.8)
RDC	Bulls with offspring	272	101.5 (3.5)
HOL	Phenotyped cows	799	101.2 (8.1)
HOL	Candidate heifers	117,134	100.2 (2.0)
HOL	Candidate bulls	9,757	100.0 (1.9)
HOL	Bulls with offspring	141	103.4 (5.5)
JER	Phenotyped cows	505	99.9 (3.1)
JER	Candidate heifers	32,170	99.8 (1.8)
JER	Candidate bulls	1,454	100.0 (1.8)
JER	Bulls with offspring	115	99.9 (2.3)

The expected effect of 1 index unit RFI are equal to the effect for maintenance at:

- RDC = 9.8 kg DMI per annual cow
- HOL = 8.2 kg DMI per annual cow
- JER = 6.7 kg DMI per annual cow

Meaning the offspring from at bull with a RFI index at 120 and a cow with a RFI index at 100 will eat 67-98 kg less dry matter in an average lactation.

References

Kennedy, B., J. Van der Werf, and T. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. *Journal of Animal Science* 71(12):3239-3250.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of Feed Use in Beef Cattle 1. *Journal of Animal Science* 22(2):486-494.

Li, B., B. Berglund, W. Fikse, J. Lassen, M. H. Lidauer, P. Mäntysaari, and P. Løvendahl. 2017. Neglect of lactation stage leads to naive assessment of residual feed intake in dairy cattle. *Journal of Dairy Science* 100(11):9076-9084.

MiX99 Development Team. 2019. MiX99: A software package for solving large mixed model equations. Release XI/2019., Natural Resources Institute Finland (Luke). Jokioinen, Finland.

Mäntysaari, E., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. *Journal of Dairy Science*.

Mäntysaari, E., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull bulletin* (41):17-17.