

Quantifying the effective contribution of phenotypic records to genetic evaluations: a case study on enteric methane emissions

J. Vandenplas, A.E. van Breukelen, B. Gredler-Grandl, R.G. Banks, R.F. Veerkamp
June 22, 2025



Context – creation of a database

What is the amount of information associated with a dataset relevant for a genetic evaluation?

- Depends on:
 - Number of records
 - Environmental factors (*aka* fixed effects)
 - Trait definition
 - Genetic parameters (e.g., heritability, repeatability)

Amount of information $\neq$ number of records

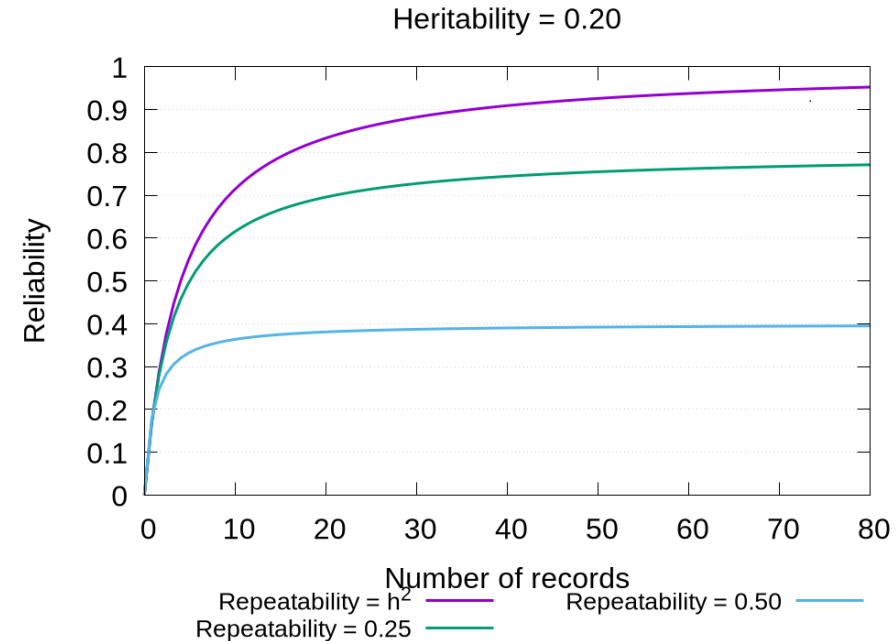
Based on Selection Index theory: $Reliability = \frac{nh^2}{1+(n-1)t}$

n : number of records, h^2 : heritability

t : repeatability

→ Number of records not appropriate!

→ Need of an estimate that quantifies the effective contribution of phenotypic data to genetic evaluations



Aim

To develop and validate a framework
that quantifies the amount of information
of a given dataset contributing to
the prediction of breeding values
for a trait of interest

Effective record contribution - concept

Total ERC for an individual:

$$ERC_{total} = \frac{1 - h^2}{h^2} \frac{REL}{1 - REL}$$

=

$ERC_{Relationships}$ (parents and progeny) + ERC_{own} (Own performances)

h^2 : heritability of the trait

REL : reliability of an estimated breeding value of an individual

Effective record contribution - concept

- ERC_{total} , $ERC_{Relationships}$

- Vary with addition of records through relationships

- ERC_{own}

- Independent of records of related individuals

→ ERC_{own} and $ERC_{Relationships}$ must be approximated

Quantification of the effective contribution of phenotypic data to genetic evaluations

- Based on the concept of effective record contributions
- Effective contribution of phenotypic data = sum of ERC_{own} associated with all phenotyped animals
- Development of a robust and flexible 3-step approach

3-step approach



1. Approximation of REL and ERC_{total} for the trait of interest using a pedigree-based evaluation and the original model
2. Estimation of ERC_{own} for all animals with records
 - Using a reverse reliability algorithm (Tier and Meyer, 2004)
 - Assuming a univariate model with only additive genetic and residual effects
3. Effective contribution of a dataset: $ERC_{data} = \sum ERC_{own}$

3-step approach - requirements

■ Data

- Phenotypic records
- Pedigree (e.g., for 5 generations)
- Original model
- Variances components
 - Observed trait and trait of interest

■ Software for approximating and reversing reliabilities

- e.g., MiXBLUP

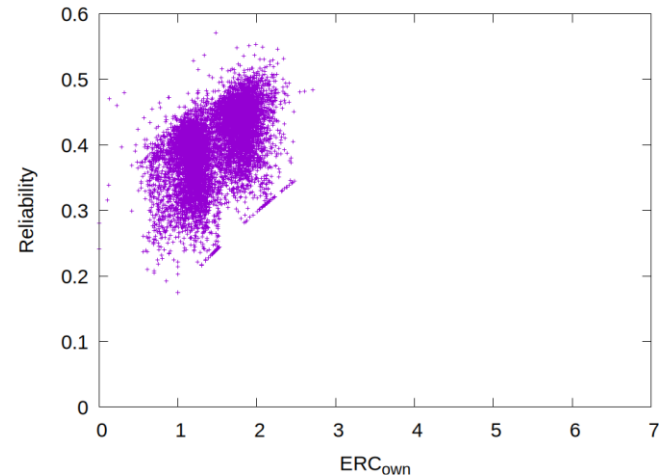
Example - data

- CH4 weekly emissions
 - 187,219 records (until June 2024)
- Individuals with records: 8668
- Pedigree (5 generations): 31,471
- Original model
 - herd + animal + within-lactation + across-lactation + residual
- h^2 : 0.18
- Repeatability: 0.47

Scenario 1 – whole dataset

	Minimum	Average	Maximum	Total
Number of records	1	21.6	80	187,219
ERC_{total}	1.0	3.2	6.3	25,724
ERC_{own}	0.0	1.5	2.7	12,671

$$ERC_{data} = \sum ERC_{own} = 12,671$$



Scenario 1 – whole dataset

- Trait of interest: CH4 weekly emissions

$$ERC_{data} = \sum ERC_{own} = 12,671$$

- Trait of interest: correlated trait

- h^2 : 0.20
- Genetic correlation: 0.80

$$ERC_{data} = \sum ERC_{own} = 3,287$$

Scenario 2 – 2 submissions

Submission 1

Up to Oct. 2023
8,034 individuals
125,169 records



$$ERC_{data-1} = 10,870$$

Submission 2

Oct. 2023-June 2024
634 new individuals
62,050 records



$$ERC_{data-2} = ERC_{data} - ERC_{data-1} = 1801$$


$$ERC_{data} = ERC_{data-1} + ERC_{data-2} = 12,671$$

Scenario 3 – 2 submissions

Submission 1

Up to June 2024
8,668 individuals
125,484 records



$$ERC_{data-1} = 12,137$$

Submission 2

No new individuals
61,735 records



$$ERC_{data-2} = ERC_{data} - ERC_{data-1} = 534$$

$$ERC_{data} = ERC_{data-1} + ERC_{data-2} = 12,671$$

Conclusions

- Robust and practical framework for quantifying the effective contribution of phenotypic data to genetic evaluations
- Three steps
 1. Approximation of reliabilities of a pedigree-based genetic evaluation using phenotypic information
 2. Approximation of ERCs due to own performances from reliabilities of phenotyped animals
 3. Calculation of the total effective contribution of phenotypic records as the sum of ERCs associated with all phenotyped animals

Thank you!

Questions?



References

- I. Misztal and G. R. Wiggans, 1988. Approximation of prediction error variance in large-scale animal models. J. Dairy Sci., 71, 27-32.
- B. Tier and K. Meyer, 2024. Approximative Schätzung der Standardfehler der Kovarianzen zwischen additiv genetischen Tiereffekten in Mehrmerkmals- und zufälligen Regressionsmodellen. J. Anim. Breed. Genet., 121, 77-89.
- P.M. VanRaden and G.R. Wiggans, 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci., 74, 2737-2746.
- H. Ben Zaabza, M. Taskinen, E.A. Mäntysaari, T. Pitkänen, G. Pedersen Aamand, and I. Strandén, 2022. Breeding value reliabilities for multiple-trait single-step genomic best linear unbiased predictor. J. Dairy Sci., 105, 6, 5221-5237.