
Sequential Evaluation of Longitudinal Conformation Data in Dairy Cows

(or How to Deal with Massive Multi-Trait Longitudinal Data)

N. Gengler^{1,2}, S. Vanderick¹, and C. Bastin¹

¹ University of Liège, Gembloux Agro-Bio Tech (GxABT), Gembloux, Belgium

² National Fund for Scientific Research (FNRS), Brussels, Belgium

Acknowledgements

- Walloon Breeding Association (AWE)
- National Fund for Scientific Research (FNRS):
2.4.623.08.F
- Ministry of Agriculture of the Walloon Region of Belgium (different projects)
- European Commission, Directorate-General for Agriculture and Rural Development, under Grant Agreement 211708 (project Robustmilk).

This study has been carried out with financial support from the Commission of the European Communities, FP7, KBBE-2007-1. It does not necessarily reflect its view and in no way anticipates the Commission's future policy in this area.

www.robustmilk.eu



- **Why still trying to improve “traditional genetic evaluations”?**
 - We have genomic evaluation !!!
- **However genomic evaluations also need accurate models to describe phenotypic records!**
 - Two step approach:
 - Genetic evaluations \Rightarrow first step \Rightarrow prediction equations
 - One step approach as basic model
 - Genetic evaluations \Rightarrow genomic evaluations (GBLUP)

- **Which issues still need work in genetic evaluation?**
 - First ↗ complexity of models
 - But also ↗ computing resources
- **However still potentially some bottlenecks**
 - In this study massive multiple-trait (MT) random regression models (RRM) for longitudinal traits:
 - Type data (as announced)
 - Extended to milk composition data (7FP RobustMilk)

- **Complex models**
 - Modified to simpler “equivalent” ones
- **A type of generic longitudinal model**

$$\mathbf{y} = \mathbf{H}\mathbf{h} + f[\Phi, \mathbf{t}] + \mathbf{e}$$

– where:

- \mathbf{h} = time-independent effects
- \mathbf{H} = incidence matrix of \mathbf{h}
- Φ = “time-dependent” effects
- \mathbf{t} = time
- $f[\Phi, \mathbf{t}]$ = function linking \mathbf{y} and Φ depending on \mathbf{t}

Some Theory

- **Please note:**
 - If $f[\Phi, t]$ time-dependent, Φ is not !
- **However Φ depend on other effects:**

$$\Phi = Xb + Zu + \varepsilon$$

- **Where:**
 - b fixed effects, u random effects, ε residual effects
 - X and Z being incidence matrices
- **Final model needs to be rewritten as:**

$$y = Hh + f[Xb + Zu + \varepsilon, t] + e$$

Some Theory

- **Models for longitudinal data \Rightarrow complex**
 - Time-dependent covariance structures
 - Often multi-trait (MT) models
- **However very useful**
 - Many traits, highly correlated, some missing
 - Two examples: type traits, milk composition data
- **Idea: rewriting model in two stages**
 - First stage estimating Φ from y
 - Then modeling Φ which has become **time-independent**

Some Theory

- **First stage:** $y_{ij} = H_{ij}h_{ij} + f[\Phi_{ij}, t_{ij}] + e_{ij}$
 - For every animal i having records along a given longitudinal time gradient j , this allows the estimation of a specific Φ_{ij}
 - Φ often called meta-data (meta-traits)
 - These models could remain single-trait (ST)
 - E.g., could be typically any type of regression coefficients per cow x lactation

Some Theory

- **Second stage :** $\Phi = Xb + Zu + \varepsilon$
 - Where estimates from first stage Φ and/or a function of these are modeled
 - These models typically multi-trait (MT) using the meta-data (meta-traits) as input
 - MT necessary to recover links across meta-traits

Example

-
- **Sounds exotic ?**



Example

- **Sounds exotic ?**
- **However very simple US example**
 - Stage 1: Best Prediction (BP)
 - Stage 2: Current USDA Animal Model (AM)
- **Interesting example because shows different hidden issues**
 - Will be used to give additional theoretical background

- **Complete equivalence**
 - Need complete BLUP and BLUE properties
 - Similarly to
Modified Contemporary Comparison \Leftrightarrow Mixed Models
 \Rightarrow Iterative solving required
 - Updating estimation of meta-traits in Stage 1 using results from Stage 2
 - For a two step RRM shown by Gengler et al. (2000*)
 - Can be considered as difference between BP + AM and full test-day model
 - even if persistency or lactation differences included in BP

* Gengler N., Tijani A., and G R Wiggans. 2000. Use of sequential estimation of regressions and effects on regressions to solve large multitrait test-day models. J Dairy Sci 83: 369

- **Distribution of meta-traits**
 - **Meta-traits: estimates**
 - **Two consequences**
 - Loss of variance as meta-trait being an estimate
 - Uneven weights as differences in information used to estimate (reliability of estimates different)
 - ⇒ **Expansion of meta-traits required to recover variance**
 - E.g., expansion of BP ⇒ AM
 - ⇒ **Weighting of meta-traits required to adjust for uneven weights**
 - E.g., lactation weights ⇒ AM

Example: Type

- **Type data from the routine performance recording in Walloon part of Belgium (01/2010)**
 - 102,875 records from first parity
 - 30,378 records from second or later parities
 - 117,013 classified Holstein cows
 - Repeated records 16,240
 - With repetitions within and across lactations
- **Request from the field**
 - Better use of available longitudinal data along age at classification
- **Use of this strategy to do (co)variance estimation**

- **Modification of current model**
 - Introduction of additional maturity effect
 - If lact = 1 \Rightarrow regression variable = 0 (LACT1)
 - If lact > 1 \Rightarrow regression variable = 1 (LACT2+)
- **Random regression model**
 - 33 traits \Rightarrow 66 random regressions (RR)
 - 2145 parameters per (co)variance matrix
 - Genetic and Non-genetic (stage 2)
 - Residual (stage 1) \Rightarrow simplified to single traits (ST)
- **Current results**
 - Without expansion and weightings
 - Using Multiple Diagonalization (CT) EM-REML

Type Results

- **Residual variances \Rightarrow close to old estimates**
- **h^2 dropped**
 - On average 0.02, largest drop 0.06
 - h^2 drop showed lost of overall variance
 \Rightarrow need expansion
- **Relative differences in phenotypic (genetic) correlations for LACT1 and LACT2+**
 - Based on Frobenius Norm ratios:
 - LACT1: 0.34 (0.42), LACT2+: 0.33 (0.42)
- **Very high genetic correlations LACT1, LACT2+**
 - Always > 0.99

- **Better example**
- **Large number of traits**
 - Potentially > 30
- **Highly correlated**
- **Only recent data**
 - Interest to include Indicator traits (e.g., fat, protein), as recorded since +30 year
- **Here results from an ongoing feasibility study by Catherine Bastin**

- **Data**

- 162,021 test-day records
- 44,885 cows
- 1029 herds
- Traits (all known to reduce need to weight):
 - Milk, fat and protein yields, saturated (SAT) and monounsaturated (MONO) fatty acids content in milk

- **Model**

- Same basic model as presented by Soyeurt et al. (2010)
- Stage 1
 - Meta-trait was defined as phenotypic animal effects
 - Regression coefficients expanded by dividing them by REL
- Stage 2
 - CT-EM-REML, no weighting of meta-traits

305 d Results

(average h^2 on diagonal, genetic correlations above, phenotypic below)

	Milk (kg)	Fat (kg)	Protein (kg)	SAT (%)	MONO (%)
MT-RRM estimates					
Milk (kg)	0.24	0.57	0.83	-0.42	-0.41
Fat (kg)	0.56	0.22	0.70	0.50	0.38
Protein (kg)	0.69	0.59	0.18	-0.11	-0.11
SAT (%)	-0.24	0.24	-0.08	0.44	0.80
MONO (%)	-0.19	0.13	-0.09	0.42	0.23
Two stage method estimates					
Milk (kg)	0.19	0.60	0.86	-0.51	-0.47
Fat (kg)	0.59	0.14	0.74	0.36	0.24
Protein (kg)	0.72	0.61	0.13	-0.22	-0.16
SAT (%)	-0.23	0.21	-0.10	0.33	0.71
MONO (%)	-0.47	0.09	-0.11	0.32	0.15

- **Type traits**
 - Based on the current results, limited interest
- **Milk composition traits**
 - Close estimates for correlations across traits
 - Still a certain lost of relative genetic variability
 - For the given situation
 - Hugh number of traits, MT
 - Random regression models RRM
 - **Best solution**
- **Some methodological improvement under development**

- **Proposed method showed large potential**
 - In these studies for VC estimation
 - Also very interesting for improved solving of mixed model equations (better convergence)
- **Easy to go further**
 - More traits
 - Rank reduction
 - Sequential and iterative solving (updating), could be asynchronous (not same moment)
 - Herd-level for Stage 1
 - Population level for Stage 2

Questions ?

