Sequential Evaluation of Longitudinal Conformation Data in Dairy Cows
(or How to Deal with Massive Multi-Trait Longitudinal Data)

N. Gengler\textsuperscript{1,2}, S. Vanderick\textsuperscript{1}, and C. Bastin\textsuperscript{1}

\textsuperscript{1} University of Liège, Gembloux Agro-Bio Tech (GxABT), Gembloux, Belgium
\textsuperscript{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
Introduction

• 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 ⇒ first step ⇒ prediction equations
  – One step approach as basic model
    • Genetic evaluations ⇒ genomic evaluations (GBLUP)
Introduction

• 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)
Some Theory

• Complex models
  – Modified to simpler “equivalent” ones
• A type of generic longitudinal model

\[ y = Hh + f[\Phi, t] + e \]

– where:
  • \( h \) = time-independent effects
  • \( H \) = incidence matrix of \( h \)
  • \( \Phi \) = “time-dependent” effects
  • \( t \) = time
  • \( f[\Phi, t] \) = function linking \( y \) and \( \Phi \) depending on \( t \)
Some Theory

• Please note:
  – If $f[\Phi,t]$ time-dependent, $\Phi$ is not!

• However $\Phi$ depend on other effects:

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

  – Where:
    • $b$ fixed effects, $u$ random effects, $\varepsilon$ 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 ⇒ 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 $\Phi$ from $y$
  - Then modeling $\Phi$ 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 \( \Phi_{ij} \)
- \( \Phi \) 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: $\Phi = Xb + Zu + \epsilon$
  
  – Where estimates from first stage $\Phi$ 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
Hidden Issues

- **Complete equivalence**
  - Need complete BLUP and BLUE properties
  - Similarly to Modified Contemporary Comparison ⇔ Mixed Models

⇒ 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

Hidden Issues

- **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
Type Model

- **Modification of current model**
  - Introduction of additional maturity effect
    - If lact = 1 ⇒ regression variable = 0 (LACT1)
    - If lact > 1 ⇒ regression variable = 1 (LACT2+)

- **Random regression model**
  - 33 traits ⇒ 66 random regressions (RR)
  - 2145 parameters per (co)variance matrix
    - Genetic and Non-genetic (stage 2)
    - Residual (stage 1) ⇒ 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
Milk Composition Data

- 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
Example: Milk Composition

• **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)

<table>
<thead>
<tr>
<th></th>
<th>MT-RRM estimates</th>
<th>Two stage method estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk (kg)</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>-0.42</td>
<td>-0.51</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>-0.41</td>
<td>-0.47</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>0.56</td>
<td>0.59</td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>0.22</td>
<td>0.14</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>0.50</td>
<td>0.36</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>0.38</td>
<td>0.24</td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>-0.11</td>
<td>-0.22</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>-0.11</td>
<td>-0.16</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>-0.24</td>
<td>-0.23</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>-0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>0.44</td>
<td>0.33</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>0.38</td>
<td>0.71</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>-0.11</td>
<td>0.32</td>
</tr>
<tr>
<td>SAT (%)</td>
<td>0.80</td>
<td>0.71</td>
</tr>
<tr>
<td>MONO (%)</td>
<td>-0.11</td>
<td>0.15</td>
</tr>
</tbody>
</table>

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

• **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**
Conclusions

- 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 ?