Modifying MACE for genomic pre-selection

Pete Sullivan  (Lactanet, Canada)
Esa Mäntysaari  (Luke, Finland)
Gerben deJong  (CRV, Netherlands)
Haifa Benhajali  (Interbull, Sweden)
• **Countries without genomics** use MACE for **selection**

• Countries with genomics use MACE as **input for SNP**
  - 2-step national GEBV systems
  - ssEBV systems with MACE integration
  - Intergenomics Brown Swiss and IG-HOL (2-step)

• Country **correlations from MACE** used in:
  - GMACE, SNP-MACE (Interbull and Eurogenomics)
Why change MACE?

- Genomic pre-selection effects

• Mendelian Sampling of AI sires no longer follows the centralized normal distribution assumed in MACE

\[ MS \sim N(0, \frac{1}{2}\sigma_g^2) \]

• MACE should accommodate effects of genomic pre-selection (GPS) on MS distributions for recent AI bulls:

\[ E(\text{MS}) > 0 \]

\[ V(\text{MS}) < \frac{1}{2}\sigma_g^2 \]

• National systems can also accommodate GPS effects to minimize GPS bias in the national EBV input to MACE
Graphical view of GPS bias

Distributions of Mend. Sampling

Protein
(Year of Birth = 2011)

Candidates (GPA)

AI-selected (GPA)

Progeny-tested (EBV)

Interbull meeting: Jan 2017
Graphical view of GPS bias

**BLUE should be** expanded from **RED**
but **BLUE is** expanded from **RED+GREEN**

Negative tail too heavy **and** wrongly centered on 0

**Interbull meeting: Jan 2017**
• **Data over-rides** bad model assumptions on MS
  • Biases *larger for culled* than for selected bulls
  • Biases *diminish with* more *daughters*
• Biases similar within a selected group
  • *relatively small* comparing bulls born in *same year* with similar *AI usage* (number daughters)
  • *larger between* birth years
• shows up as *under-estimated genetic trend*
• *Impact* of bias on MACE is *small?* ...*increasing*
GPS effects should be included as part of predicted genetic merits of genomically pre-selected bulls.

- GPS effects are included in GEBV but not in EBV.
- Missing GPS effects are referred to as GPS bias.

**GPS bias = EBV - GEBV**
Detecting GPS bias in EBV

- oGEBVr = *genomic* selection criterion, *before AI*
- GEBV = *genomic* estimate *after AI*
- EBV = *conventional* estimate *after AI*

Since 2010

Intensity of GPS increased dramatically, for MS>0
Expect obvious increasing trends in %MS>0
Compare trends in %MS>0 for GEBV and EBV
%Positive MS: **Conformation**

bull+sire+dam proven in **Canada**

![Graph showing the positive MS for conformation in bulls, sires, and dams proven in Canada over the years 2000 to 2013. The graph includes lines for oGEBVr, Expected, GEBV, and EBV. There is a highlighted area in 2011 indicating GPS bias.](image-url)
%Positive MS: **Protein**
bull proven in **DFS**

RDC bull mendelian percentage
genotyped bulls with $r^2 > 0.79$

![Graph showing positive protein distribution over years with a note on GPS bias and data from different sources.](image-url)
• Strong trends in %MS>0 from GPS effects are included in national GEBV
• The Trends in %MS>0 are reduced or eliminated in national EBV, because of GPS bias
• Interbull cannot compute MS as \((AN - \text{Sire}/2 - \text{Dam}/2)\)
  ➢ Countries do not submit EBV of cows (for Dam/2)
• Interbull can only look at mPI as \((AN - \text{Sire}/2 - \text{MGS}/4)\)
%Positive mPl: **Conformation**

bull+sire+dam proven in **Canada**

![Graph showing the percentage of positive mPl for conformation bull+sire+dam proven in Canada from 2000 to 2013. The graph includes lines for oGEBVr, Expected, GEBV, and EBV. The line for Expected is constant at 40%. The line for oGEBVr shows a consistent increase over the years. The line for GEBV shows a significant increase from 2007 onwards. The line for EBV shows a steady increase from 2007.](http://example.com/graph.png)
Average & %Positive mPI: **Milk**

bull+dam tested in **Netherlands**

**MS_{si} milk**
Detecting GPS from mPI

- Trends in %mPI more difficult to work with than %MS
- mPI includes genetic trends of females (MGD)
- Current MACE model works more with mPI distributions than MS distributions
  - MACE is Animal Model, but without using cow EBVs
- Can Interbull account for GPS effects in MACE with the data currently provided (only sire EBVs) ???
Interbull working groups

- GPS and Future MACE:
  Pete Sullivan (Canada), Esa Mäntysaari (Finland), Gerben deJong (Netherlands) Haifa Benhajali (Sweden)

  Develop and test international methods for genomics era

- Trend Validation Tests
  Paul VanRaden (USA), Pete Sullivan (Canada), Raphael Mrode (UK), Zengting Liu (Germany), Esa Mäntysaari (Finland), Valentina Palucci (Sweden)

  Update current tests and expand beyond trend validation

- National EBV and MACE (ssEBV input to MACE?)
- National GEBV (2-step) and ssEBV (MACE as input?)
- International GMACE and SNP-MACE
**Options for future MACE**

- Using national data provided currently

**Estimating GPS effects** without genotypes

1. Modified relationship matrix (3rd parent)
2. **Hyper-parameter** for GPS effects on MS averages
3. **Genetic groups** known animals (+ phantom parents)
4. Data augmentation with **pseudo records** for culls
   - Fill missing gaps to make MS/mPI ~ Normal ???

**Better underlying assumptions**

\[ V(\text{Animal} – \text{PA} – \text{GPS effect}) < V(\text{Animal} – \text{PA}) \]

GPS effects can be *in national EBV and MACE* models
Genetic groups (Qg) added to models in the 1970’s e.g. AI stud or regions (as genetic pre-selection) effects.

Early-1980’s, Kennedy and others showed: Groups improve accuracy only if group effects are large enough. If all ancestors known (A complete) genetic groups are not needed.

Late-1980’s, Westell, Quaas simplified the application: Replace $A^{-1}$ with $W$ to add unknown phantom parents groups. Became standard approach for genetic groups in dairy evaluations. But using $W$ restricts grouping to unknown, missing pedigree.

GPS groups are known animals with known parents!!
Methods are well-known (since 1970’s), but not standard option in EBV software, except using $W$.

**New programming** required to fit genetic groups for both **Unknown Parents AND** additionally for GPS effects on **known animals** with known parents.

- Unknown parent groups ($W$) account for **PA-selection**
- Known-animal GPS groups account for **MS-selection**
- Feasible to update national and MACE software
- Will require **custom programming**, especially for low variance, skewness, or multi-modal ~ MS / mPI
Holstein Conformation (Canada)
( Genomic PreSelection Group Effects )

December 2018 Evaluation

EBV\textsubscript{2018} – GPA\textsubscript{2014}

Val'n Bias

GPSgrp

Bull’s Year of Birth

2007 2008 2009 2010 2011 2012 2013 2014
Holstein Conformation (Canada)
( Genomic PreSelection Group Effects )

December 2018 Evaluation

- Bull’s Year of Birth
- EBV_{2018} – GPA_{2014}
- GPSgrp
- New Bias
- Val'n Bias
After adding GPS group effects to the model:

- **Bias** in genetic trend **partially removed** (~40-50% smaller)
- **Expecting further improvement with reduced variance** for selected MS (next step in this ongoing research)

Consistent with simulation results (Fikse, 2014)

GPS groups recommended by Patry and Ducrocq

Obvious implementation, however, might give only partial benefits
Success will be in the details … some ideas to be tested

Same Unknown Parent Groups in Deregression and MACE
- Across-country groupings
- Country of origin, Birth year, Selection path

GPS groups only in the MACE model, not in Deregression
- Within-country groupings
- Country where used in AI, Birth year
- Need to consider single-country versus multi-country bulls
If modeling GPS effects is not a good enough solution, another option is to use pre-corrected EBV input data

\[ \text{EBV}^* = \text{EBV} + f(\text{GEBV}-\text{EBV}) \]

\[ \text{MACE}^* \text{ uses } \text{EBV}^* \]

\[ \text{MACE} = f^{-1}(\text{MACE}^*) \]

Transformation function \( f() \) could be simple or complex

A simple \( f() \) example: add yearly averages of GEBV-EBV, correcting only for the GPS-biased genetic trend in EBV

Options for future MACE
- With new data: Countries provide EBV + GEBV
The idea with transformation: use EBV* in MACE that are significantly "less biased" by GPS, while still excluding individual genotype contributions, to continue MACE in SNP

- i.e. Not double-counting genotype effects in national GEBV

- Variances and correlations could be estimated from the "less biased" EBV, with modifications to allow truncated MS data

- Accommodate variation between countries, in methods for GPS effects, because transformations are relatively smaller (appropriately) for countries with more advanced methods
• Not yet clear if **EBV of females** will be helpful or if **EBV+GEBV** will be required for future MACE

• Picking the best approaches objectively will require good **validation testing** and **GPS simulation**