

A SNP MACE model for international genomic evaluation:

- technical challenges and possible solutions

Zengting Liu and Mike E. Goddard

vit, IT solutions for animal production, Germany

University of Melbourne, Melbourne 3083, VIC, Australia

Overview

Current status of (inter)national conventional and genomic evaluations

- A SNP MACE model for international evaluation
- Solving algorithms for the SNP MACE model
- Approximation of prediction error (co)variances
- Further development and extension
 - Different SNP sets across countries

Countries use MACE info as phenotypes



International evaluation based on SNP effects

- Concept of a SNP based international evaluation (Goddard, 2011)
- A SNP-Focus Model replacing the Animal Model (Schaeffer, 2014)
- Interbull estimation of SNP effects (Goddard, 2016)
- A SNP MACE model proposed (Goddard, 2017)
 - Interbull Technical Workshop, Slovenia
- Interbull project on the SNP MACE model



A SNP MACE model

- A SNP genomic model for Multiple Across Country Evaluation

$$\mathbf{g}_i^N \implies \mathbf{g}_i \quad [1]$$

for country i ($i = 1, \dots, c$)

- A SNP BLUP model for national genomic evaluation

$$\mathbf{y}_i = \mu_i \mathbf{1} + \mathbf{Z}_i \mathbf{g}_i^N + \mathbf{e}_i \quad [2]$$

where \mathbf{y}_i is phenotype after absorbing all other effects (including the residual polygenic effect)

$$\text{var}(\mathbf{e}_i) = \mathbf{R}_i^{-1} = \text{diag}\{n_{ik} \sigma_{e_i}^{-2}\}$$

$$\text{var}(\mathbf{g}_i) = \mathbf{B}_i \sigma_i^2 \quad \text{with DGV variance } \sigma_i^2$$

$$\mathbf{B}_i = \frac{1}{\sum_j 2p_{ij}(1-p_{ij})} \mathbf{I} = \theta_i \mathbf{I}$$



SNP genetic (co)variances between countries (I)

- A (co)variance matrix for countries ($i = 1, \dots, c$) **for a single SNP marker**

$$\mathbf{G}_{cou} = \begin{bmatrix} \sigma_1^2 \theta_1 & r_{12} \sigma_1 \sigma_2 \sqrt{\theta_1 \theta_2} & \cdots & r_{1c} \sigma_1 \sigma_c \sqrt{\theta_1 \theta_c} \\ & \sigma_2^2 \theta_2 & \cdots & r_{2c} \sigma_2 \sigma_c \sqrt{\theta_2 \theta_c} \\ & & \ddots & \vdots \\ & & & \sigma_c^2 \theta_c \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1c} \\ & g_{22} & \cdots & g_{2c} \\ & & \ddots & \vdots \\ \text{symm.} & & & g_{cc} \end{bmatrix}$$

- Its inverse $\mathbf{G}_{cou}^{-1} = \begin{bmatrix} g^{11} & g^{12} & \cdots & g^{1c} \\ & g^{22} & \cdots & g^{2c} \\ & & \ddots & \vdots \\ \text{symm.} & & & g^{cc} \end{bmatrix}$



SNP genetic (co)variances between countries (II)

- Genetic (co)variance matrix for **ALL** SNP effects (**ordered by countries**)

$$\text{var} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_c \end{bmatrix} = \mathbf{G} = \mathbf{G}_{\text{cou}} \otimes \mathbf{I} = \begin{bmatrix} g_{11}\mathbf{I} & g_{12}\mathbf{I} & \cdots & g_{1c}\mathbf{I} \\ & g_{22}\mathbf{I} & \cdots & g_{2c}\mathbf{I} \\ & & \ddots & \vdots \\ \text{symm.} & & & g_{cc}\mathbf{I} \end{bmatrix}$$

- Its inverse $\mathbf{G}^{-1} = \mathbf{G}_{\text{cou}}^{-1} \otimes \mathbf{I} = \begin{bmatrix} g^{11}\mathbf{I} & g^{12}\mathbf{I} & \cdots & g^{1c}\mathbf{I} \\ & g^{22}\mathbf{I} & \cdots & g^{2c}\mathbf{I} \\ & & \ddots & \vdots \\ \text{symm.} & & & g^{cc}\mathbf{I} \end{bmatrix}$

- Inter-SNP genetic correlations: within or between countries are all 0
- Intra-SNP genetic correlations between countries to be estimated
 - Set to country correlations as in current MACE



Mixed model equations of the SNP MACE model

$$\begin{bmatrix} \ddots & & & & & & & & \\ & \dots & & & & & & & \\ & & \dots & & & & & & \\ & & & \dots & & & & & \\ & & & & \dots & & & & \\ & & & & & \dots & & & \\ & & & & & & \dots & & \\ & & & & & & & \dots & \\ & & & & & & & & \ddots \end{bmatrix}
 \begin{bmatrix} \left[\begin{array}{cc} \mathbf{1}'\mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{Z}_i\mathbf{1} & \mathbf{1}'\mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{Z}_i \\ \mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{Z}_i\mathbf{1} & \mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{Z}_i \end{array} \right] + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g^{ii}\mathbf{I} \end{bmatrix} \\ \vdots \\ \left[\begin{array}{cc} \mathbf{1}'\mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+}\mathbf{1} & \mathbf{1}'\mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} \\ \mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+}\mathbf{1} & \mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} \end{array} \right] + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g^{i^+i^+}\mathbf{I} \end{bmatrix} \\ \vdots \end{bmatrix}
 \begin{bmatrix} \left[\Psi_{ii^+} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g^{ii^+}\mathbf{I} \end{bmatrix} \right] \\ \vdots \\ \left[\begin{array}{cc} \mathbf{1}'\mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+}\mathbf{1} & \mathbf{1}'\mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} \\ \mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+}\mathbf{1} & \mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{Z}_{i^+} \end{array} \right] + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & g^{i^+i^+}\mathbf{I} \end{bmatrix} \\ \vdots \end{bmatrix}$$

$$\mathbf{X} \begin{bmatrix} \hat{\mu}_i \\ \hat{\mathbf{g}}_i \\ \vdots \\ \hat{\mu}_{i^+} \\ \hat{\mathbf{g}}_{i^+} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{1}'\mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{Z}'_i\mathbf{R}_i^{-1}\mathbf{y}_i \\ \vdots \\ \mathbf{1}'\mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{y}_{i^+} \\ \mathbf{Z}'_{i^+}\mathbf{R}_{i^+}^{-1}\mathbf{y}_{i^+} \\ \vdots \end{bmatrix}$$

Zero residual covariances between countries $\Psi_{ii^+} = \mathbf{0}$,
 if the countries do not use MACE EBV for national
 genomic evaluation



Solving the mixed model equations

- MME of the SNP MACE model have special structures:
 - Data contribution by country, zero residual covariances (off-diagonals)
 - SNP genetic contribution: only diagonal and sub-diagonals $\neq 0$
 - Block-diagonal matrix in the SNP-major order
- Identical processes for every country or every SNP \rightarrow parallel computing
- PCG algorithm using multiple cores

$$\mathbf{C} \mathbf{v} = \underbrace{\{\mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i\}}_{\text{for country } i} \mathbf{v} + \underbrace{\{\mathbf{G}_{cou}^{-1}\}}_{\text{for every SNP marker } j} \mathbf{v}$$

for country i

for every SNP marker j

parallelised by countries

parallelised by SNP markers

- Conditioner may be the inverted diagonal block for country i

$$\mathbf{M}_i = (\mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i + \mathbf{G}^{ii})^{-1}$$

the matrix \mathbf{M}_i is also used in reliability calculation.



Calculation of prediction error (co)variances

- Countries need to calculate reliabilities of DGV (sum of all SNP effects)
- Not only reliabilities of MACE SNP effect estimates but also (co)reliabilities between the SNP effect estimates
 - the whole PEC block of 50k x 50k

- Absorbing all the other countries into own SNP effects

$$\begin{aligned}
 \mathbf{C}_i &= (\mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i + \mathbf{G}^{ii}) - \sum_{j \neq i} \mathbf{G}^{ij} (\mathbf{Z}'_j \mathbf{R}_j^{-1} \mathbf{Z}_j + \mathbf{G}^{jj})^{-1} \mathbf{G}^{ji} \\
 &= (\mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i + g^{ii} \mathbf{I}) - \sum_{j \neq i} g^{ij} \mathbf{I} (\mathbf{Z}'_j \mathbf{R}_j^{-1} \mathbf{Z}_j + \mathbf{G}^{jj})^{-1} g^{ji} \mathbf{I} \\
 &= (\mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i + g^{ii} \mathbf{I}) - (g^{ij})^2 \sum_{j \neq i} \mathbf{M}_j
 \end{aligned}$$

- Invert the own block matrix \mathbf{C}_i^{-1}
- Provide the PEC matrix to countries \mathbf{C}_i^{-1}



Three methods for handling different sets of SNP markers

- Method 1: conversion of country SNP effects to a common set
- Method 2: conversion of SNP effects for GBLUP models
- Method 3: direct modelling heterogeneous sets of SNP markers



Method 1: Conversion of country SNP effects to a common set of SNP markers

- SNP effects of national set of SNP markers for i -th country: \mathbf{g}_i^N
- SNP effects of a common set of SNP markers: \mathbf{g}_c^N
- Define DGV of all reference animals with own set of SNP markers:

$$\mathbf{u}_i = \mathbf{Z}_i \mathbf{g}_i^N$$

- A SNP BLUP model is fitted to model the DGV of reference animals:

$$\mathbf{u}_i = \mathbf{Z}_i^c \mathbf{g}_c^N + \xi$$

$$(\mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} \mathbf{Z}_i^c + \sigma_i^{-2} \mathbf{B}_c^{-1}) \mathbf{g}_c^N = \mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} \mathbf{u}_i$$

$$\mathbf{g}_c^N = (\mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} \mathbf{Z}_i^c + \sigma_i^{-2} \mathbf{B}_c^{-1})^{-1} \mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} (\mathbf{Z}_i \mathbf{g}_i^N)$$

- Additional data needed for the conversion

$$\mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} \mathbf{Z}_i \quad \mathbf{Z}_i^{c'} \mathbf{R}_i^{-1} \mathbf{Z}_i^c$$

in addition to $\mathbf{Z}_i' \mathbf{R}_i^{-1} \mathbf{Z}_i$

- Back conversion of MACE SNP effect estimates to the own SNP set



Method 2: Conversion of country SNP effects for GBLUP models

- Country uses a GBLUP model with its own SNP set
- Assumption: equal GEBV for reference animals with both SNP sets
- For all reference animals: GEBV vector \mathbf{u}_i^*
- Genomic relationship matrix for all reference animals is **invertible**: \mathbf{G}_{rel}^{-1}
- Estimate SNP effects of the common set

$$\mathbf{g}_i^c = (1-k)\mathbf{B}_c \mathbf{Z}_i^c{}' \mathbf{G}_{rel}^{-1} \mathbf{u}_i^*$$

- Equal genomic relationship matrices

$$(1-k)\mathbf{Z}_i^c{}' \mathbf{B}_c \mathbf{Z}_i^c + k\mathbf{A}_i = (1-k)\mathbf{Z}_i{}' \mathbf{B}_i \mathbf{Z}_i + k\mathbf{A}_i$$

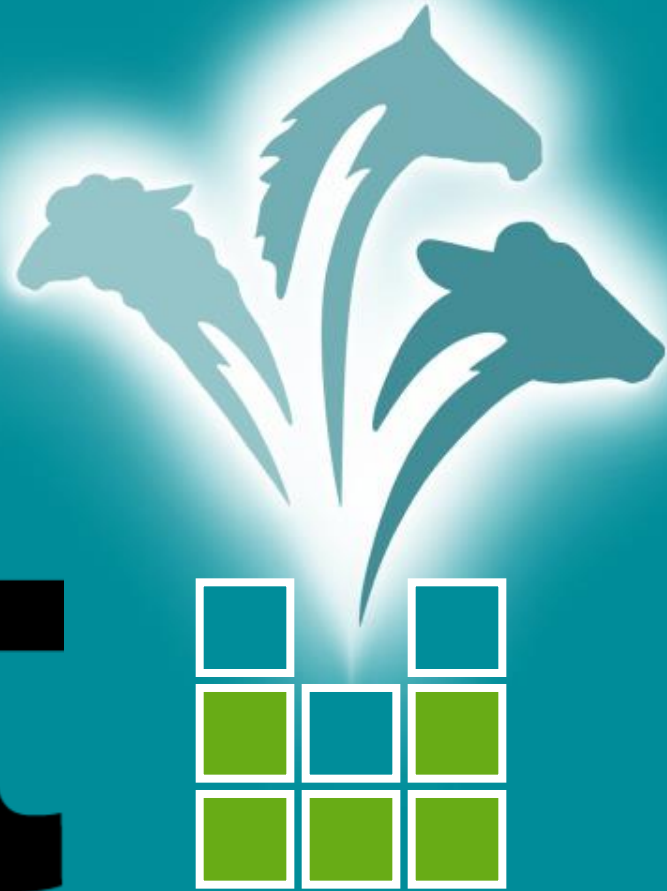


Summary

- The SNP MACE model is an efficient tool for utilizing phenotype info of foreign reference animals
 - Particularly useful for new traits with large-scale genotyped cows
- No requirement for direct access to original national genotype and phenotype data
 - Keep the current infra-structure of national evaluation systems
- Parallel computing for efficiently solving the SNP MACE equations
 - No more pedigree relationship matrix, difficult to be parallelized
- Direct modelling different sets of SNP reduces the need for conversion to a common set of SNP markers
- A gain in accuracy of prediction is expected, especially for novel traits



vit



IT-Solutions for Animal Production