SNPMace: A meta-analysis to estimate SNP effects across countries

Abdulqader Jighly, Haifa Benhajali,
Mike E. Goddard

The accuracy of GEBVs depends on the size of the training population

Small breeds and all breeds for some traits cannot construct a large enough training population

The solution is to combine across countries (and possible breeds)

But countries may not want to share raw data

Therefore, we propose to mimic an analysis of the combined data using summary statistics provided by each country

A meta-analysis

Combining the results from >1 analysis rather than combining the raw data

Why do we need it?

- To increase power
- To increase robustness
- Cant combine raw data

The SNPMace project

The aim of this project is to test the proposal for a meta-analysis that Interbull could run as a service

Using data from 6 countries because Interbull have the individual animal data in this case

Therefore we can test whether a meta-analysis gives the same EBVs as an analysis of the individual animal data

Design of the SNPMace project

Interbull calculates single country SNP effects from the individual animal data

And send SNP effects to Melbourne

Melbourne carries out meta-analysis to estimate multiple country SNP effects and sends these to Interbull

Interbull compares the Melbourne SNP effects with those from a multi-country analysis of the individual animal data

Both Melbourne and Interbull analysis is multi-trait assuming the Interbull genetic correlations between yield in different countries

Single country equations to estimate SNP effects (g)

$(Z_1'Z_1 + \lambda I)\, g_1 = Z_1'y_1$

Two countries

$(Z_1'Z_1 + Z_2'Z_2 + \lambda I)\, g = Z_1'y_1 + Z_2'\, y_2$

If individual countries provide g_i and $Z_i'Z_i$ we can construct the multi-country BLUP and solve for g

Extensions: include $r_g < 1$ between countries and weights for records

$$\begin{bmatrix} Z_1'R^{-1}\mathbf{Z}_1 + \mathbf{G}^1 & \mathbf{G}^{12} \\ \mathbf{G}^{12} & Z_2'R^{-1}\mathbf{Z}_2 + \mathbf{G}^2 \end{bmatrix} \times \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix} = \begin{bmatrix} Z_1'R^{-1}y_1 \\ Z_2'R^{-1}y_2 \end{bmatrix}$$

$$\mathbf{B} = \frac{1}{\sum_j 2p_i(1-p_i)}\mathbf{I}$$

(VanRaden 2008)

$$G = \mathrm{var}\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_1^2\mathbf{B}_1 & r_{12}\sqrt{\sigma_1^2\mathbf{B}_1 \times \sigma_2^2\mathbf{B}_2} \\ r_{12}\sqrt{\sigma_1^2\mathbf{B}_1 \times \sigma_2^2\mathbf{B}_2} & \sigma_2^2\mathbf{B}_2 \end{bmatrix}^{-1}$$

Liu Z, Goddard ME. 2018. A SNP MACE model for international genomic evaluation: technical challenges and possible solutions. Page 11.393 in Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zeland

$$\begin{bmatrix} \ddots & \cdots & \cdots & \cdots & \cdots \\ \vdots & [Z_i'R^{-1}\mathbf{Z}_i + \mathbf{G}^{ii}] & \cdots & [\mathbf{G}^{ii^+}] & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & [\mathbf{G}^{ii^+}] & \cdots & [\mathbf{Z}_{i^+}'R^{-1}\mathbf{Z}_{i^+} + \mathbf{G}^{i^+i^+}] & \vdots \\ \ddots & \cdots & \cdots & \cdots & \ddots \end{bmatrix} \times \begin{bmatrix} \cdots \\ [\hat{\mathbf{g}}_i] \\ \vdots \\ [\hat{\mathbf{g}}_{i^+}] \\ \cdots \end{bmatrix} = \begin{bmatrix} \cdots \\ [Z_i'R^{-1}\mathbf{y}_i] \\ \vdots \\ [\mathbf{Z}_{i^+}'R^{-1}\mathbf{y}_{i^+}] \\ \cdots \end{bmatrix}$$

Liu Z, Goddard ME. 2018. A SNP MACE model for international genomic evaluation: technical challenges and possible solutions. Page 11.393 in Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zeland

SNPMace project

May 2018 – November 2019

Preliminary report on project

We wrote efficient software to run the SNPMace model

We tested the software on Australian data

We are analyzing data provided by Interbull for 6 countries

Currently using 50k SNP data

We wrote software in C++ to carry out the meta-analysis

And tested it on Australian Jersey and Holstein data

      (2 breeds instead of 2 countries)

Australian Dataset

- Main data
  - 1071 Jersey born before 2010
  - 4105 Holstein born before 2010
- Validation
  - 107 Jersey born since 2010
  - 522 Holstein born since 2010
- Three traits
  - Milk yield
  - Milk fat
  - Milk protein
- MTG2 (Lee and van der Werf 2016) was used to run the ST and MT models

Trait	Correlation
Milk yield	0.54
Milk fat	0.36
Milk protein	0.33

Correlation of SNP effects

Milk yield	ST	MT	SNPMace
ST	(0.386 /0.198)	0.66	0.69
MT	0.96	(0.394 /0.284)	0.98
SNPMace	0.97	0.99	(0.399 /0.285)

Milk fat	ST	MT	SNPMace
ST	(0.012 /0.007)	0.79	0.77
MT	0.97	(0.012 /0.009)	0.98
SNPMace	0.98	0.99	(0.012 /0.009)

- Above diagonal: Jer
- Below diagonal: hol
- Diagonal: SD of SNP effects (hol/jer)

Correlation of DGVs

Milk yield	ST	MT	SNPMace
ST	(209.8 /194.0)	0.993	0.996
MT	0.999	(208.6 /195.0)	0.995
SNPMace	0.999	0.999	(209.8 /194.0)

Milk fat	ST	MT	SNPMace
ST	(6.05 /7.7)	0.997	0.999
MT	0.999	(5.98 /8.0)	0.998
SNPMace	0.999	0.999	(6.05 /7.8)

- Above diagonal: Jer
- Below diagonal: hol
- Diagonal: SD of DGVs (hol/jer)

Validation

		ST		MT		SNPMace	
		Jer	Hol	Jer	Hol	Jer	Hol
Milk	Jer	0.52	0.32	0.50	0.46	0.53	0.50
	Hol	0.05	0.51	0.49	0.52	0.46	0.52
Fat	Jer	0.34	0.18	0.37	0.36	0.37	0.36
	Hol	0.00	0.52	0.31	0.53	0.30	0.53
Protein	Jer	0.55	0.15	0.54	0.39	0.54	0.40
	Hol	0.08	0.48	0.39	0.53	0.40	0.53

Interbull SNPMace Project

We are analyzing data for six countries provided by Interbull for milk protein

Only reference bulls were included

No overlap between countries data

One phenotype for each bull

No foreign genotypes or phenotypes

No polygenic effect

Multi-trait model (i.e. genetic correlations between countries)

Different from the official Intergenomics evaluation

Interbull calculated the single-trait (ST) model for each country and sent $Z'R^{-1}Z$ matrices and g estimations

We used the SNPMace model to calculate SNP effects

Interbull calculated a multi-trait (MT) model using all countries which is equivalent to our SNPmace analysis

We sent the calculated SNP effects to Interbull so they can compare the SNP effect estimates and the GEBVs (using ST, MT and SNPMace)

Interbull Data

- $Z'R^{-1}Z$ matrices and correlations across countries were provided for six countries:

Country	No Animals	CHE	DEA	FRA	ITA	SVN	USA
CHE	1748	1	0.916	0.846	0.855	0.813	0.827
DEA	2490	0.916	1	0.813	0.891	0.81	0.82
FRA	167	0.846	0.813	1	0.818	0.814	0.863
ITA	1275	0.855	0.891	0.818	1	0.81	0.832
SVN	227	0.813	0.81	0.814	0.81	1	0.821
USA	482	0.827	0.82	0.863	0.832	0.821	1

Correlation of DGVs and their SD across models

Country	NoAnimals	Var	ST/MT	ST / SNPMace	MT / SNPMace	ST SD	MT SD	SNPMace SD
CHE	1748	0.397	0.959	0.981	0.984	0.630	0.629	0.668
DEA	2490	0.284	0.911	0.985	0.955	0.533	0.534	0.517
FRA	167	0.002	0.689	0.866	0.943	0.044	0.415	0.047
ITA	1275	0.346	0.962	0.977	0.987	0.589	0.590	0.619
SVN	227	0.016	0.791	0.908	0.960	0.127	0.421	0.140
USA	482	0.014	0.808	0.916	0.968	0.116	0.408	0.111

Country	NoAnimals	Var	ST/MT	ST / SNPMace	MT / SNPMace	ST SD	MT SD	SNPMace SD
CHE	1748	1	0.959	0.994	0.972	0.630	0.629	0.621
DEA	2490	1	0.911	0.995	0.942	0.533	0.534	0.501
FRA	167	1	0.689	0.743	0.959	0.044	0.415	0.186
ITA	1275	1	0.962	0.995	0.979	0.589	0.590	0.563
SVN	227	1	0.791	0.906	0.946	0.127	0.421	0.198
USA	482	1	0.808	0.912	0.938	0.116	0.408	0.173

*VAR is the genetic variance assumed in the SNPMace analysis

- MT Above diagonal and SNPMace below diagonal (assuming all Var =1)

	CHE	DEA	FRA	ITA	SVN	USA
CHE	.	0.989	0.997	0.985	0.994	0.995
DEA	0.916	.	0.993	0.990	0.994	0.994
FRA	0.846	0.813	.	0.994	0.999	0.999
ITA	0.855	0.891	0.818	.	0.994	0.995
SVN	0.813	0.810	0.814	0.810	.	0.999
USA	0.827	0.820	0.863	0.832	0.821	.

$$G = \mathrm{var}\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_1^2 \mathbf{B}_1 & r_{12}\sqrt{\sigma_1^2 \mathbf{B}_1 \times \sigma_2^2 \mathbf{B}_2} \\ r_{12}\sqrt{\sigma_1^2 \mathbf{B}_1 \times \sigma_2^2 \mathbf{B}_2} & \sigma_2^2 \mathbf{B}_2 \end{bmatrix}^{-1}$$

Considering all $r = 1$ did not improve the results

Correlation SNP effects across models

	ST / SNPMace	ST / MT	MT / SNPMace
CHE	0.81	0.83	0.86
DEA	0.84	0.83	0.85
FRA	0.24	0.37	0.76
ITA	0.85	0.86	0.88
SVN	0.36	0.46	0.73
USA	0.35	0.53	0.74

Meta-analysis software seems to work

But

We don't have perfect agreement between the meta-analysis and the multi-trait analysis on individual data

Before November

Fix the bug

Deliver software to Interbull so they can offer a service to members

Deliver final report

Thank you