Hybrid one-step genomic evaluation system in Italian Simmental

L. Degano[1], G. Jansen[2], D. Vicario[1], R. Finocchiaro[3], A. Rossoni[4]

*[1]*ANAPRI: Via Ippolito Nievo 19, 33100 Udine (UD)-Italy; tel. +39 043222411; email: vicario@anapri.it*
[2] Dekoppel Consulting, Casale Rovera 10, Chiaverano(TO), Italy*

*[3]*ANAFI: Via Bergamo 292, 26100 Cremona (CR)-Italy*

*[4]*ANARB: 204, Localita' Ferlina, 37012 Bussolengo (VR)-Italy*

Interbull Meeting 2016, Puerto Varas-Chile, 26th October

INTRODUCTION

Introduction

One-step project: is a collaborative project of ANAFI, ANAPRI and ANARB (2014)

Aims of Project: develop a joint genetic-genomic evaluations tool for bulls, cows and young animals, for all traits.

Issue: different kind of traits, statistical models, dimension of population...

Consistency of the data-sets used for genetic evaluation of the production traits			
	Italian Simmetal	**Italian Brown**	**Italian Holstein**
N° Test day	5,300,000	22,000,000	75,000,000
N° Pedigrees	600,000	3,800,000	17,400,000
N° Genotype	4,200	21,000	175,000
TD Model	Repeat. Tdm	Repeat. Tdm	R.R. Tdm
Genomic evaluation before project	PC-reduced SNP Blup	Intergenomics	GS3-SNPBLUP with poligenic effect

Home Pagina personale Progetti Aiuto

One-step

Panoramica   Attività   Segnalazioni   Nuova segnalazione   **Wiki**   Repository

Onestep Project Wiki

✎ Modifica ⭐ Osserva ↱ Rinomina 🗑 Elimina «

The Onestep project is a collaborative project of Anarb, Anafi and Anapri whose objective is to develop a system of joint genetic-genomic evaluations for bulls, co young animals, for all traits, with the eventual aim of replacing all current conventional and genomic evaluations in publications, databases and selection tools.

Meetings: minutes and supporting docs for meetings

SoftwareBlupf90 | SoftwareSnpblup and DesignNotes: system design and software usage notes

FileFormats: file formats for onestep inputs/outputs
InputFileNotes: extra notes on preparation of input data files
DataChecks: initial data checks performed by OSM software
Clones | ParentageErrors: handling of clones and parentage errors

Deregression: documentation for deregression method

InternalValidation: notes on internal validation
InterbullValidation: Interbull validation with GEBVtest software

TotalMeritIndices: proposal to compute approximate TMIs (see also AnimalStatus)

BaselineTests: initial testing with first datasets

LitReview: literature review about genomic evaluation and especially the so-called single-step or one-step approach
EvaluationSystems: links to country docs about national evaluation systems

Forum: interesting posts from original forum

Nuovo file

INTERACTIVE PROJECT DEVELOPING ACROSS R&D STAFF OF EACH BREED ASSOCIATION THROUGH A WEB-BASED SOFTWARE COLLABORATION PLATFORM (http://redmine.anafi.it/projects/one-step)

The pipeline
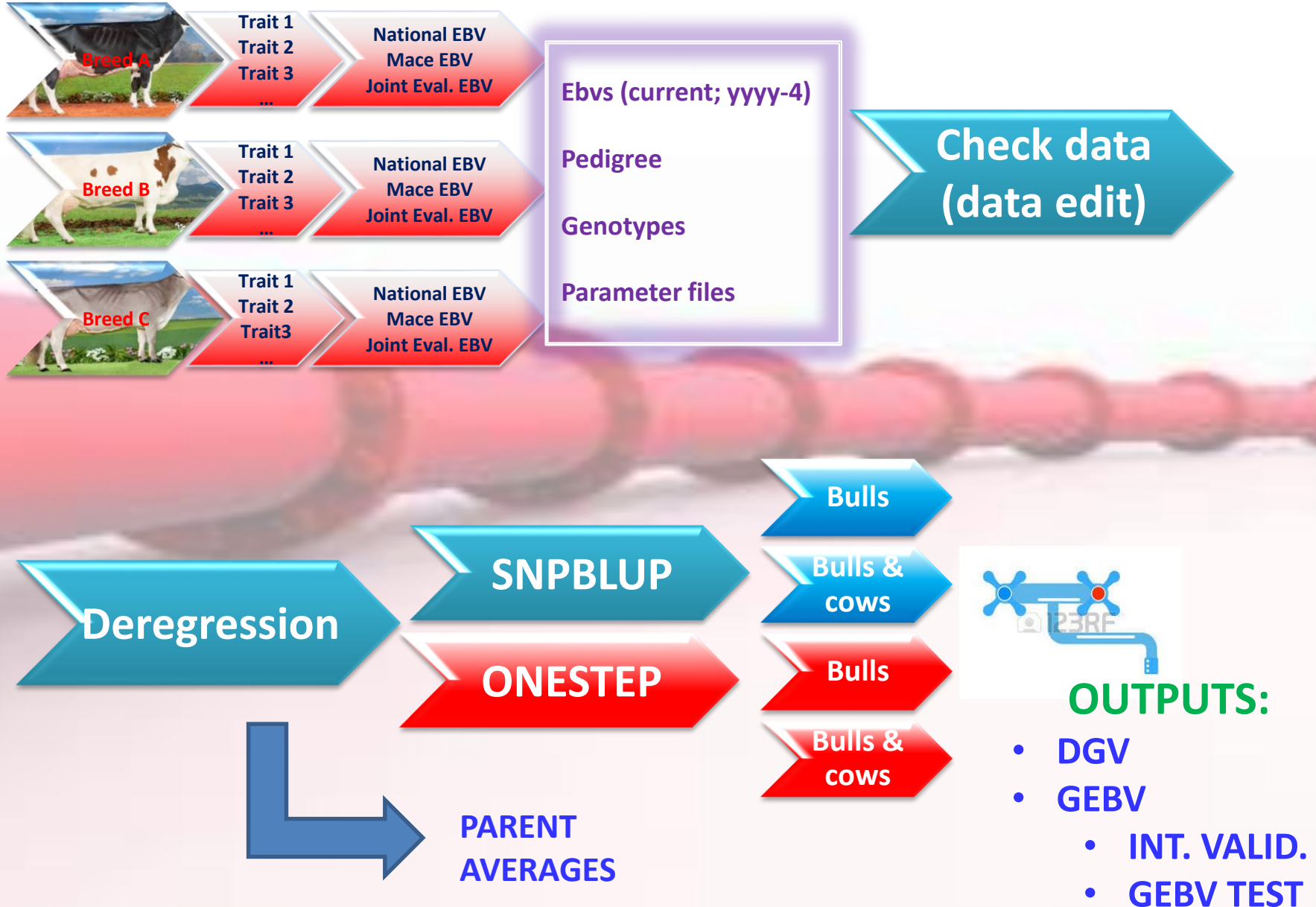
Data

Data:
- ✓ **286,311 EBVs for productive traits** (national EBVs for cows and NI bulls, national and MACE EBVs for AI bulls)
- ✓ **212,223 EBVs for SCS** (100±12; cows and NI bulls national EBVs, AI bulls national and MACE EBVs)
- ✓ **143,652 EBVs for muscling and udder** (100±12; EBVs from the international join evaluation that involve Simmental populations of DE, AT, CR, IT)

Pedigree:
581,521 Animals

Genotypes:
4,226 genotypes (3,083 M; 1,143 F); mostly 54k, others ranging from GGP_LD to IlluminaHD . Genotypes from all chips have been imputed to a standard set of 40,200 selected SNPs (Pedimpute developed by E.Nicolazzi & G.Jansen, 2013).

Source	Males	Females
International exchanges	1079	
Research projects	899	963
Routine breeding program	1105	180

Methodology

- EVBs DEREGRESSION
- DGV / GEBV ESTIMATION
- VALIDATION
- RELIABILITY COMPUTATION

Deregression

Deregression to achieve EDPs has been carried out according the underlying mixed model equations can be represented as:

$$(Z'DZ \ + \ A^{-1} k) \ \{EBVi\} \ = \ 2 \ Z'Dy$$

Where:
   D is a diagonal matrix of unknown EDC (Effective Daughter Contribution)
   A is the traditional relationship matrix
   k = (4-h2)/h2 (variance ratio)
   y is a vector of unkown EQUIVALENT DAUGHTER PERFORMANCE (EDP) values
   Z is an incidence matrix relating animals to EDP.

Deregression (3 steps)

PRUNING on EBVS & PEDIGREES (over 50% reduction) to reduce computing resourses

$$(Z'DZ + A^{-1}k)\{EBVi\} = 2\, Z'Dy$$

1. only domestic cow and bull EBV and REL were considered

2. Including MACE EBV and REL (by replacing domestic bulls EDVs and any other foreign bull in pedigree)

3. Cow EDP adjustment for MS variance

Re-computing EBVs and REL for all animals (BLUP AM solver) (Parent Averages as by-product from this kind of MACE EBVs blending)

Libro Genealogico
Pezzata Rossa Italiana
A.N.A.P.R.I. - Udine

DGV /GEBV estimation

ONE-STEP genomic evaluation

Single-step GBLUP evaluations were computed using simple animal model with mean as only fixed effect, with matrix A^{-1} replaced by matrix H^{-1} as in (Legarra et al., 2014) with a polygenic component incorporated through

$$G_w = 0.9\ G + 0.1\ A_{22}$$

The BLUPF90 suite of software of I. Misztal and collaborators was used to form the genomic relationship matrix with default scaling parameters and to solve the MME (http://nce.ads.uga.edu/wiki)

SNPBLUP genomic evaluation

For the purposes of comparison, DGV were also estimated with a SNPBLUP model followed by simple blending of the PA from the final deregression step, according to relative EDC, to obtain GEBV

$$DGV = \mu + \sum z_i a_j$$

Own developed BLUP solver has been used for computing SNP solutions

Validation

Validation

Performed by following the guidelines of the Interbull GEBV test.

Reducing dataset of phenotipic records 4 years back:

- **PRODUCTION /SCS -> domestic EBVs from reduced dataset were re-computed removing last 4 years of TD records.**
- **CONFORMATION -> the last four birth years of current cow and bull EBVs were removed and for all traits the last four birth years of current MACE EBV were removed.**

RE-RUN OF DEREGRESSION ON REDUCED DATASET

Approximated reliability of GEBV (Van Raden et al. 2009)

1. $$EDP_{full} = b * GEBV_{rd} \left(R^2\ gebv\right)$$

2. $$EDP_{full} = b * PA_{rd} \left(R^2\ PA\right)$$

$$\left(R^2\ gebv\right) - \left(R^2\ PA\right) = GEBV\ reliability\ gain\ due\ to\ DGV$$

…. <u>UNDERESTIMATED</u> …. because of loss of information by reducing dataset for validation, especially in small reference populations

RELIABILITY computing

Approximation of one-step GEBV reliabilities

A) Validation study

1. observed REL_GEBV: R^2 (EDP_f / GEBV_r) / REL_EDP_f

2. projection from reduced data to full data, based on increase in size of calibration (sum(REL_EDP))

3. add difference in published and observed REL_PA

4. yields average REL_GEBV for young bulls and extra EDC over PA

B) Published individual GEBV_REL

➤ EDC from conventional REL + extra EDC from genotype
➤ small increase for non-genotyped relatives too

Via Ippolito Nievo 19, 33100 Udine (UD) - Italy; tel. +39 0432 22411; email: anapri@anapri.it

COWS genotypes or NOT in Calibration

GEBVs estimation:

- ✓ GEBVs were obtained using the procedure developed by the consortium.
- ✓ Four different scenarios have been considered.

Scenario	Model	Reference population
Scenario 1	SNPBLUP	Only Bulls
Scenario 2	SNPBLUP	Bulls & Cows
Scenario 3	ONE-STEP	Only Bulls
Scenario 4	ONE-STEP	Bulls & Cows

RESULTS

Descriptive statistics of EBVs

	N° records	Mean±SD	Min	Max
Milk kg	286,311	-447.7±557.4	-2,170	2,248
Fat kg	286,311	-18.1±22.1	-70	69
Protein kg	286,311	-16.1±18.4	-69.3	66.6
SCS	212,223	95.8±10.0	50	148
Muscling	143,652	101.6±9.6	50	149
Udder	143,652	95.2±8.9	50	138

Internal Validation report

C:\Users\vicario\Downloads\validate_ONEMF.log - Notepad++

File   Modifica   Cerca   Visualizza   Formato   Linguaggio   Configurazione   Macro   Esegui   Plugin   Finestra   ?

validate_ONEMM.log | validate_SNPMM.log | validate_ONEMF.log | validate_SNPMF.log | correlazioni e medie.lst

```
 1  Internal validation for SIM 1611gj1   (run on 2016-Oct-11 11:25:12)
 2   - onestep blending (with preGSf90 and blupf90)
 3   - EDP/EDC for validation from full data using bulls-only deregression
 4
 5  Dereg options:  edc_delta=3.0 edp_delta=3.5  ped_upg=True  cow_rescale=2
 6  Solver options:  minmaf=0.01 tau=1.0 omega=1.0 alpha=0.9 beta=0.1
 7  Validation options:  weighted=True  rescale=False  minEDC=20.0  type2x=N  file300=Y
 8  Total genotypes: 4187
 9
10
11                    ___ N ___           _____ Rel _____                   ___ GEBV ___    R2     R2
12  Trait        Train Test Year    h2   EDP   PA GEBV GAIN   EDCG  EDCGd     a     b    GEBV    PA
13  mil latte    2179  119 2007 0.193  85.2 38.3 47.6  9.2   17.9    5.6 -176.68 0.592  0.136  0.082
14  fat kggras   2182  119 2007 0.129  83.3 38.0 49.4 11.4   29.3   10.9   -9.55 0.719  0.214  0.137
15  pro kgprot   2179  119 2007 0.170  84.0 37.9 50.9 12.9   23.3    9.6   -8.26 0.681  0.197  0.107
16  scs cell     2265  119 2007 0.120  79.9 36.5 48.5 12.0   30.4   11.8   -0.48 1.001  0.381  0.327
17  msp milgs    1169   77 2007 0.152  52.5 27.7 27.6 -0.2    9.6   -0.1   49.74 0.527  0.079  0.074
18  tag taglia   1934   95 2007 0.439  87.1 40.5 48.9  8.4    7.8    2.2   21.72 0.767  0.401  0.342
19  mus muscol   1926   95 2007 0.214  75.4 37.5 68.7 31.2   38.9   28.3    6.68 0.931  0.516  0.440
20  ofl arti     1907   95 2007 0.108  60.6 35.2 42.6  7.5   26.8    7.2  -30.84 1.314  0.255  0.311
21  ocs mammel   1929   95 2007 0.235  77.0 37.8 64.6 26.8   29.3   19.5  -19.72 1.203  0.494  0.510
22  dpd fpt       537  117 2007 0.049  26.5 14.9  1.0 -13.8    0.9  -13.2   94.93 0.048  0.000  0.072
23  tap tagpt     741  334 2014 0.250  23.6  6.9  9.0  2.1    1.5    0.4    0.13 0.988  0.012  0.012
24  mup muspt     751  359 2014 0.560  55.6  9.7  7.4 -2.3    0.5   -0.2   96.43 0.031  0.010  0.230
25  app artpt     760  359 2014 0.230  28.4  7.3 17.9 10.6    3.6    2.3  -13.13 1.124  0.045  0.047
26  img imgpt     745  342 2014 0.310  28.7  7.6 19.5 11.8    2.9    1.9  -47.11 1.462  0.050  0.042
27  ----------------------------------------------------------------------------------------------
28  All Average  1514  174 2009 0.000  60.6 26.9 36.0  9.1   15.9    6.2   -2.58 0.813  0.199  0.195
29  All Median   1916  119 2007 0.000  68.0 35.8 45.1  9.3   13.8    4.0   -4.37 0.849  0.167  0.122
30
31
32                    _____ Mean _____     _____ SD _____   r(PA,   ___ PA ___   R2    R2
33  Trait        EDP      PA    GEBV    EDP      PA    GEBV   GEBV)    a     b    PA PA+GEBV
34  mil latte   95.66  384.92  387.32  542.54  235.29  286.69  0.774 -171.35 0.550  0.082  0.136
```

Normal text file   length : 4174   lines : 51   Ln : 38   Col : 99   Sel : 0 | 0   UNIX   UTF-8 w/o BOM   INS

Pezzata Rossa Italiana
A.N.A.P.R.I. - Udine

Internal validation: SNPBLUP MODEL

Reference population: bulls; Model: SNPBLUP										
	N° Animals		% h²	Reliability %			EDP/GEBV		EDP/PA	
	Reference	Validation		PA	GEBV	Gain	b	% R²	b	% R²
Milk kg	1,644	121	19.3	38.3	46.7	**8.4**	0.67	9.1	0.48	6.0
Fat kg	1,646	121	12.9	38.0	49.3	**11.3**	0.83	16.8	0.64	10.9
Protein kg	1,645	121	17.0	37.9	51.9	**14.0**	0.86	14.8	0.57	8.0
SCS	1,700	120	12.0	36.4	39.2	**2.8**	1.10	30.0	1.07	29.5
Muscling	1,520	95	21.4	37.2	54.0	**16.8**	0.79	36.8	0.86	39.5
Udder	1,523	94	23.5	37.6	65.4	**27.8**	1.15	42.5	1.45	49.0

Reference population: bulls and cows; Model: SNPBLUP										
	N° Animals		% h²	Reliability %			EDP/GEBV		EDP/PA	
	Reference	Validation		PA	GEBV	Gain	b	%R²	b	%R²
Milk kg	2,170	121	19.3	38.3	48.8	**10.5**	0.64	8.7	0.48	6.0
Fat kg	2,172	121	12.9	38.0	50.9	**12.9**	0.82	16.4	0.64	10.9
Protein kg	2,171	121	17.0	37.9	52.2	**14.3**	0.81	13.6	0.57	8.0
SCS	2,253	120	12.0	36.4	40.9	**4.5**	1.17	30.5	1.07	29.5
Muscling	1,929	95	21.4	37.2	54.4	**17.2**	0.80	36.7	0.86	39.5
Udder	1,932	94	23.5	37.6	65.5	**27.9**	1.19	42.7	1.45	49.0

Via Ippolito Nievo 19, 33100 Udine (UD) - Italy; tel. +39 0432 22411; email: anapri@anapri.it

Internal validation: ONESTEP MODEL

Reference population: bulls; Model: ONE-STEP										
	N° Animals		%h²	% Reliability			EDP/GEBV		EDP/PA	
	Reference	Validation		PA	GEBV	Gain	b	%R²	b	%R²
Milk kg	1,655	116	19.3	38.3	44.1	**5.8**	0.67	15.6	0.64	11.0
Fat kg	1,658	116	12.9	38.0	45.5	**7.5**	0.85	24.9	0.89	19.0
Protein kg	1,655	116	17.0	37.9	44.9	**7.0**	0.77	23.1	0.82	17.5
SCS	1,714	116	12.0	36.5	58.2	**21.7**	1.20	46.5	1.24	37.0
Muscling	1,518	92	21.4	37.5	73.0	**35.5**	1.11	55.0	1.03	43.8
Udder	1,521	92	23.5	37.8	60.0	**22.2**	1.14	46.0	1.37	52.0

Reference population: bulls and cows; Model: ONE-STEP										
	N° Animals		%h²	% Reliability			EDP/GEBV		EDP/PA	
	Reference	Validation		PA	GEBV	Gain	b	% R²	b	% R²
Milk kg	2,179	116	19.3	38.3	46.1	**7.8**	0.62	15.4	0.48	6.0
Fat kg	2,182	116	12.9	38.0	51.7	**13.7**	0.79	26.4	0.64	10.9
Protein kg	2,179	116	17.0	37.9	51.9	**14.0**	0.77	25.6	0.57	8.0
SCS	2,265	116	12.0	36.5	60.5	**24.0**	1.22	48.0	1.07	29.5
Muscling	1,926	92	21.4	37.5	74.5	**37.0**	1.12	56.1	0.86	39.5
Udder	1,929	92	23.5	37.8	70.9	**33.1**	1.26	54.4	1.45	49.0

Via Ippolito Nievo 19, 33100 Udine (UD) - Italy; tel. +39 0432 22411; email: anapri@anapri.it

Descriptive statistics of GEBV

(970 young genotyped animals without EDP)

Scenario	Model	Ref. pop.	Milk kg	Fat kg	Protein kg	SCS	Muscling	Udder
1	SNPBLUP	Bulls	340.1±**222.5**	12.0±**9.0**	11.7±**7.4**	102.5±**6.4**	98.3±**7.8**	107.7±**6.8**
2	SNPBLUP	BullS & C.	355.3±**224.6**	12.6±**8.8**	12.5±**7.5**	102.6±**6.1**	98.5±**7.6**	107.4±**6.4**
3	ONESTEP	Bulls	439.4±**332.4**	15.2±**12.8**	15.2±**10.5**	102.9±**8.2**	99.0±**8.1**	106.1±**7.2**
4	ONESTEP	Bulls & C.	430.1±**337.1**	15.0±**13.0**	14.8±**10.6**	102.8±**8.4**	99.0±**8.2**	106.0±**7.2**

Correlation GEBV: protein Kg, SCS

(970 young genotyped animals without EDP)

Protein kg	Model	Ref. popul.	Scen. 1	Scen. 2	Scen. 3	Scen. 4
Scenario 1	SNPBLUP	Bulls	1	0.99	0.83	0.81
Scenario 2	SNPBLUP	Bulls & C.		1	0.82	0.81
Scenario 3	ONESTEP	Bulls			1	0.96
Scenario 4	ONESTEP	Bulls & C.				1

SCS	Model	Ref. popul.	Scen. 1	Scen. 2	Scen. 3	Scen. 4
Scenario 1	SNPBLUP	Bulls	1	0.99	0.90	0.87
Scenario 2	SNPBLUP	Bulls & C.		1	0.89	0.88
Scenario 3	ONESTEP	Bulls			1	0.97
Scenario 4	ONESTEP	Bulls & C.				1

Correlation GEBV: muscling, udder

(970 young genotyped animals without EDP)

Muscling	Model	Ref. popul.	Scen. 1	Scen. 2	Scen. 3	Scen. 4
Scenario 1	SNPBLUP	Bulls	1	0.99	0.91	0.91
Scenario 2	SNPBLUP	Bulls & C.		1	0.90	0.89
Scenario 3	GBLUP	Bulls			1	0.96
Scenario 4	GBLUP	Bulls & C.				1

Udder	Model	Ref. popul.	Scen. 1	Scen. 2	Scen. 3	Scen. 4
Scenario 1	SNPBLUP	Bulls	1	0.99	0.89	0.88
Scenario 2	SNPBLUP	Bulls & C.		1	0.87	0.86
Scenario 3	GBLUP	Bulls			1	0.99
Scenario 4	GBLUP	Bulls & C.				1

CONCLUSIONS

Conclusions-1

✓ **An stream-lined, flexible and user-friendly pipeline for computing GEBVs for different breeds and different traits has been developed for the main dairy breeds in Italy. This was feasible only by using a hybrid one-step approach based on deregressed EBV instead of rough phenotypes.**

✓ **By doing that, blending MACE ebvs into domestic evaluation together with cows in an efficient way, allows inclusion in calibration of foreign bulls with no daughters in such population as well as cows.**

✓ **Considering that in the future the estimation of GEBVs will be more frequent (weekly ?), this easy friendly pipeline improves the routine of GE.**

✓ **Italian Simmental has implemented officially this procedure (SNPBLUP) for genomic evaluation recently validated by GEBV test at ITB.**

Conclusions-2

✓ **In medium size populations as the Italian Simmental, with a low n° of genotyped bulls and without a borderless GE, inclusion of genotyped cows in one-step GE is expected to give substantial advantage in terms of reliability gain as the n° of genotyped cows increases.**

✓ **GEBV ITB test validation has been also passed. Obtained results in terms of R^2 and b-coefficient are not unfortunely enough meaningful because small size of validation bulls group. Further investigations are needed to explain these values.**

✓ **Moving from a SNPBLUP model to a ONESTEP, makes also sense in this case, giving an advantage especially in terms of GEBVs variability, much more comparable to conventional EBVs . Shortly, it is going to switch to ONESTEP officially, already internal validated.**

✓ **Next advances of this tool are going to be extending application possibly to other italian breeds and to new traits i.e. functional traits, beef traits, animal behaviour, etc.**

Thank you for your attention

Acknowledgements

RELIABILITY computing

$$REL\_DGV = w *[T*h2 / (T*h2 + Me)]$$

(E.Mantysaari, 2014, Berlin ITB meeting)

where

REL_DGV = avg. squared accuracy of DGV in young animals without phenotypes

w = fraction of genetic variance that can be explained by the marker panel

T = number of animals in the training (or reference) set

h2 = squared accuracy of "phenotypes" on training animals

Me = effective number of chromosome segments segregating in the population (=6,000)

By applying this formula either in FULL and REDUCED datasets we extrapolated the validation reliability as a function of size of training population and average REL_EDP in the reduced validation data (R) and in the full data (F) used to calculate routine genomic evaluations.

In practice, $T*h^2$ is the sum of REL_EDP of training animals.

Let's say SumRel{R|F}

RELIABILITY computing

SumRel can be interpreted roughly as the effective size of the training set if all animals had EDP accuracy = 1,
For example, SumRel = 2000 could be obtained :

- 2500 bulls with REL_EDP=0.8 (>100 daus. , h^2 =0.25)
- 3000 bulls with REL_EDP=0.67
- 8000 cows with REL_EDP=0.25

Therefore, theoretical reliabilities of DGV can be calculated as follows and converted in EDCs:

$\color{red}{\textbf{reldDGV\_R = SumRelR / (SumRelR + Me)}} \rightarrow \color{red}{\textbf{edcdR = k * reldR / (reldR - 1)}}$

$\color{green}{\textbf{reldDGV\_F = SumRelF / (SumRelF + Me)}} \rightarrow \color{green}{\textbf{edcdF = k * reldF / (reldF - 1)}}$

[edcdF – edcdR]

can be added to EDC from the empirical REL_DGV in the internal validation in order to extrapolate the genomic EDCs and RELs to values expected from the full data.