

# Dairy x Dairy genomic prediction

Huiming Liu, Ulrik Sander Nielsen and  
Emre Karaman

2023-10-17

# Background

- Genetic evaluation for crossbred animals
  - Official genetic evaluation: **BOM** - SNP solutions of purebreds + breed origin of alleles of crossbreds (Eiriksson et al., 2021)
  - Method under testing: **BOA** - Using purebred and crossbreds to estimate SNP effects + breed origin of alleles (Karaman et al., 2021)
  - BOM is not possible for crosses including MON
  - BOA may be feasible for providing genetic merit prediction for MON crosses

# Comparison of BOM and BOA models

## BOM model

- Breed of origin at each locus
- Summary statistics from purebred animals
  - SNP solutions
- Breed difference – phenotypic averages of purebreds
- No crossbred animals' data

## BOA model

- Breed of origin at each locus
- Summary statistics from purebred animals
  - SNP solutions
  - PEV ...
- Breed difference – estimated from the model for crossbreds
- Crossbred animals' phenotypes

# Genotype and Phenotype Data

	# animals with pheno	# animals with geno	# animals with both
HOL	5 905 415	502 912	108 177
JER	1 088 720	136 544	54 781
RDC	865 475	249 223	22 514
XXX	748 132	13 705	8 481

\* The individuals with missing phenotypes were removed from the data

\* XXX animals have HOL, JER or RDC sire and maternal grandsire, and HOL, JER, RDC or XXX dam.

# Imputation

- Super map – 47 586 markers
- Imputation and phasing (FImpute3)
  - Purebred animals were imputed separately based on the super map
  - Crossbred animals were imputed using imputed purebred animals as reference
- Breed origin estimation (AllOr)

# Fixed effects for corrected phenotypes

Pre-correction of phenotypes (DMU)

- Effects including in the official genetic evaluation

Name in phenotype	Explanation
MAN_GR (management group)	Herd*time*lactation
KLV_A_M	Calving*year*month
K_ALDER	Calving*age
Tothet	Heterozygosity

# Models for crossbred animals

- Scenarios:
  - 1: SumStat HOL | JER | RDC (BOM)
  - 2: SumStat HOL | JER | RDC +XXX (BOA)
  - 3: SumStat HOL | RDC + XXX (BOA\_MON) [JER-> MON]
- BOA - Reference and validation populations
  - REF = animals with a calving year before 2021 Apr (6948 animals)
  - TEST = animals with a calving year after 2021 Apr (1533 animals)

# Breed proportions

	HOL	JER	RDC
ALL	0.52	0.15	0.33
Ref	0.51	0.15	0.34
test	0.57	0.14	0.29



# Milk – Genomic prediction - cor (corrected phenotype and GEBV)

Group	Number of animals	BOM	BOA	BOA_MON	
		sumStat HOL JER RDC	sumStat HOL JER RDC +XXX	sumStat HOL RDC +XXX	
All test animals	1533	0.60	0.60	0.60	
> 50% JER	252	0.60	0.60	0.54	- 10%
(JER x HOL)	(107)	0.59	<b>0.58</b>	<b>0.48</b>	- 17%
(JER x RDC)	(6)	0.88	<b>0.90</b>	<b>0.75</b>	- 17%
(JER x XXX)	(139)	0.51	<b>0.50</b>	<b>0.47</b>	- 6%
> 50% HOL	1128	0.55	0.54	0.54	
(HOL x RDC)	(364)	0.52	0.50	0.50	
> 50% RDC	605	0.51	0.52	0.52	

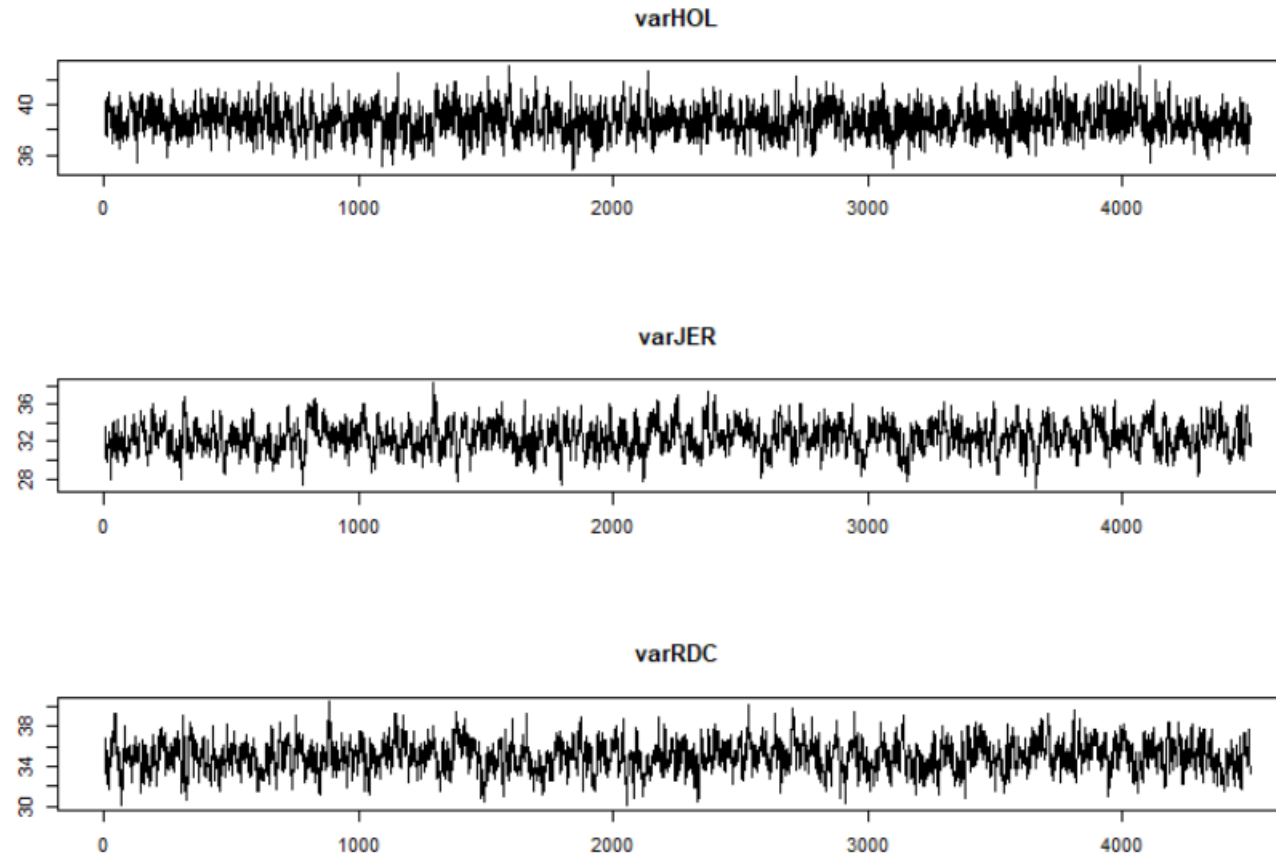
# Protein – Genomic prediction - cor (corrected phenotype and GEBV)

Group	Number of animals	BOM	BOA	BOA_MON	
		sumStat HOL JER RDC	sumStat HOL JER RDC +XXX	sumStat HOL RDC +XXX	
All test animals	1533	0.46	0.47	0.46	
> 50% JER	252	0.45	0.44	0.39	- 6%
(JER x HOL)	(107)	0.47	<b>0.46</b>	<b>0.36</b>	- 22%
(JER x RDC)	(6)	0.91	<b>0.86</b>	<b>0.78</b>	- 10%
(JER x XXX)	(139)	0.34	<b>0.31</b>	<b>0.30</b>	- 3%
> 50% HOL	1128	0.44	0.44	0.42	
(HOL x RDC)	(364)	0.41	0.41	0.41	
> 50% RDC	605	0.42	0.43	0.42	

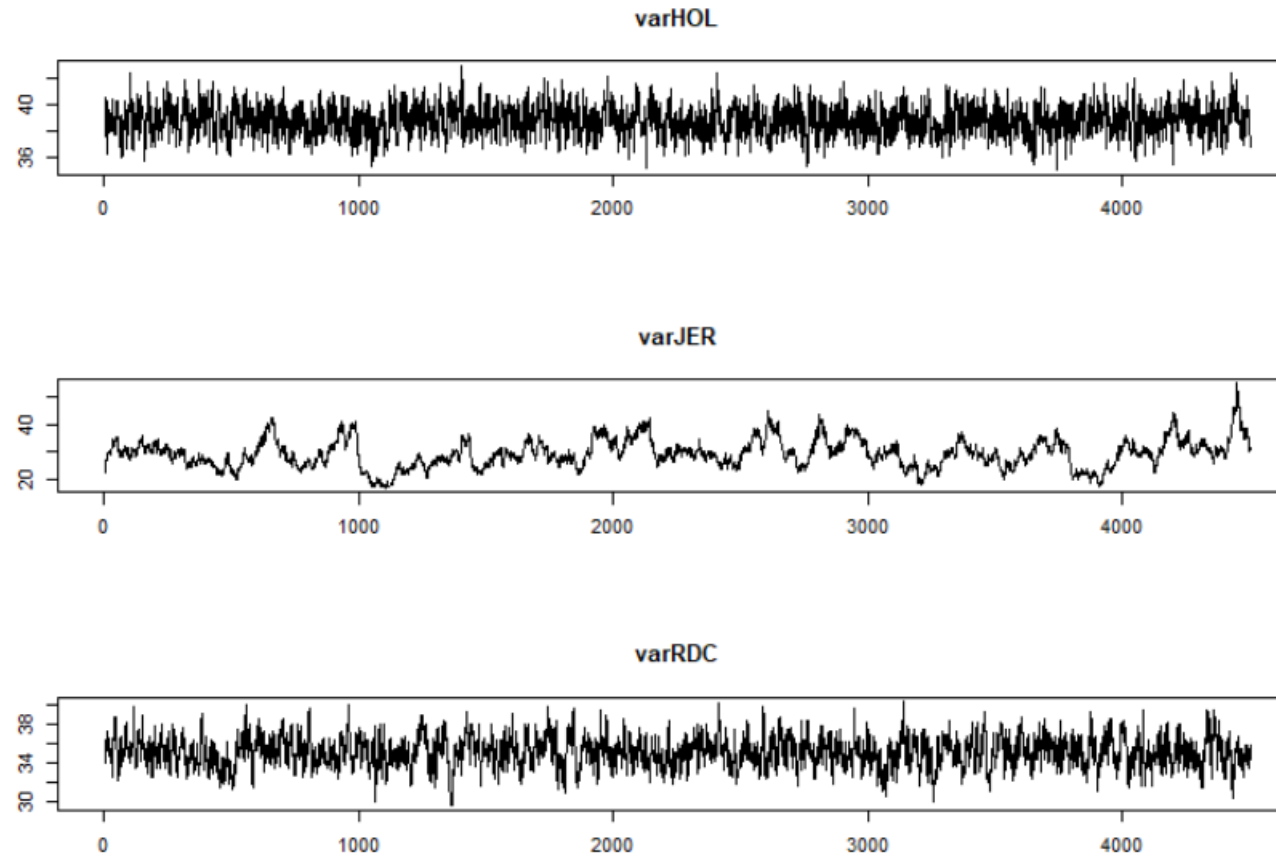
# Fat – Genomic prediction - cor (corrected phenotype and GEBV)

Group	Number of animals	BOM	BOA	BOA_MON	
		sumStat HOL JER RDC	sumStat HOL JER RDC +XXX	sumStat HOL RDC +XXX	
All test animals	1533	0.44	0.42	0.42	
> 50% JER	252	0.35	0.34	0.26	- 24%
(JER x HOL)	(107)	0.45	<b>0.46</b>	<b>0.39</b>	- 15%
(JER x RDC)	(6)	0.92	<b>0.83</b>	<b>0.57</b>	- 31%
(JER x XXX)	(139)	0.25	<b>0.24</b>	<b>0.20</b>	- 17%
> 50% HOL	1128	0.45	0.43	0.42	
(HOL x RDC)	(364)	0.40	0.37	0.37	
> 50% RDC	605	0.43	0.42	0.42	

# Trace plots SNP variances for milk - BOA



# Trace plots SNP variances for milk – BOA\_MON



# Conclusions

- BOA is like BOM when we have summary statistics from all purebreds and only limited number of crossbred animals
- BOM cannot be used to evaluate MON crosses, but BOA can be used

Gert suggested Huiming:

- To test the effect of using only 100 pure jersey genotypes in the BOA\_MON scenario.
- Correlate GEBVs from BOA and BOM models for candidates

## **We added a scenario with only 140 JER genotypes**

**140 JER genotypes were used in AllOr to trace the breed of origin, and the rest of the procedure is the same as BOA\_MON scenario. The scenario was BOA\_140JER.**



# Models for crossbred animals

- Scenarios:

- 1: SumStat HOL | JER | RDC (BOM)
- 2: SumStat HOL | JER | RDC +XXX (BOA)
- 3: SumStat HOL | RDC + XXX (BOA\_MON) [JER-> MON]
- 4: **only140 JERgeno** + SumStat HOL | RDC + XXX (BOA\_100JER) [JER-> MON]

- BOA - Reference and validation populations

- REF = animals with a calving year before 2021 Apr (6948 animals)
- TEST = animals with a calving year after 2021 Apr (1533 animals)

# The difference between official and 100JER

The difference between estimated breed proportions from using all Jersey genotypes and 140 Jersey genotypes

HOL		JER		RDC	
-0.07	1	-0.1	1	-0.05	1
-0.02	1	-0.08	2	-0.02	1
-0.01	4	-0.07	1	-0.01	5
0	13835	-0.06	2	0	13855
0.01	10	-0.05	4	0.01	6
0.02	6	-0.04	3	0.02	2
0.03	1	-0.03	2	0.05	1
0.04	6	-0.02	5		
0.05	4	-0.01	5		
0.07	1	0	13410		
0.08	2	0.01	381		
		0.02	45		
		0.03	5		
		0.04	2		
		0.06	1		
		0.07	2		

The difference between fixed effects from different scenarios

	(Intercept)	HOL	JER	RDC
BOA	1574.4012	987.3643	-1129.4949	145.4012
BOA_MON	1574.6019	983.5004	-1133.9906	147.6616
BOA_100JER	1574.4604	982.2833	-1138.8102	147.0297

# Milk – JER 100

Group	Number of animals	BOM	BOA	BOA_MON	BOA_140JER
		sumStat HOL JER RDC	sumStat HOL JER RDC +XXX	sumStat HOL RDC +XXX	sumStat HOL RDC Only 100 JER geno +XXX
All test animals	1533	0.60	0.60	<b>0.60</b>	<b>0.59</b>
> 50% JER	252	0.60	0.60	<b>0.54</b>	<b>0.53</b>
(JER x HOL)	(107)	0.59	0.58	<b>0.48</b>	<b>0.49</b>
(JER x RDC)	(6)	0.88	0.90	<b>0.75</b>	<b>0.75</b>
(JER x XXX)	(139)	0.51	0.50	<b>0.47</b>	<b>0.45</b>
> 50% HOL	1128	0.55	0.54	<b>0.54</b>	<b>0.54</b>
(HOL x RDC)	(364)	0.52	0.50	<b>0.50</b>	<b>0.50</b>
> 50% RDC	605	0.51	0.52	<b>0.52</b>	<b>0.52</b>

BOA\_140JER works as fine as BOA\_MON

# Correlate GEBVs from BOA and BOM models for candidates

All test animals

	BOA	BOA_MON	BOA_100JER
BOM	0.96	0.93	0.91
BOA		0.98	0.96
BOA_MON			0.99

# Correlate GEBVs from BOA and BOM models for candidates

>50% JER

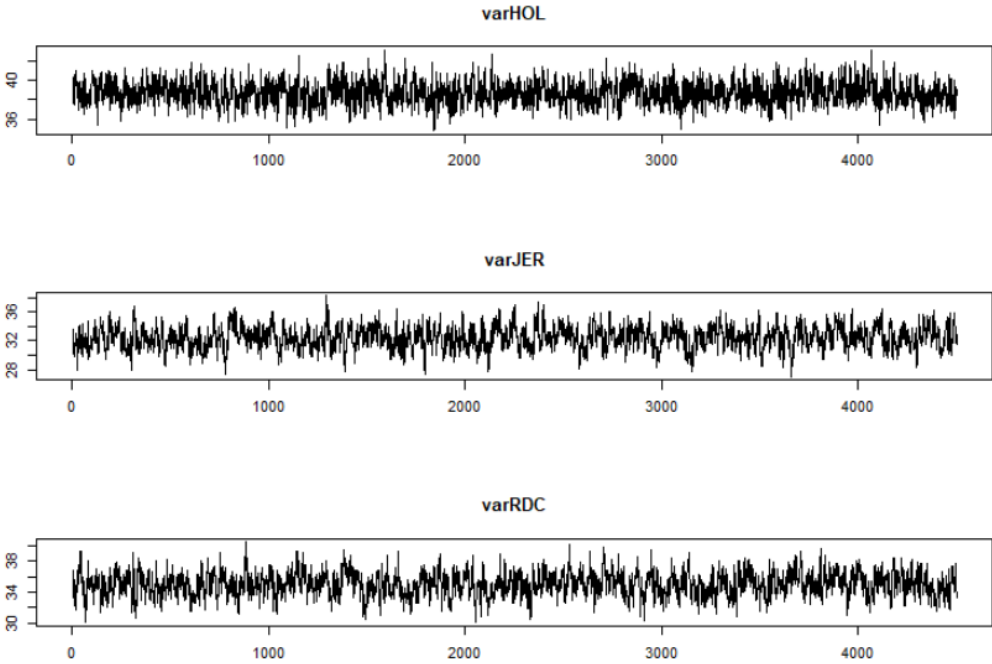
	BOA	BOA_MON	BOA_100JER
BOM	0.96	0.86	0.84
BOA		0.91	0.90
BOA_MON			0.99

# Conclusions

- BOA performs as well as BOM when we have summary statistics from all purebreds and only limited number of crossbred animals
- BOM cannot be used to evaluate cross when purebred info is missing, but BOA can be used
- In reality we only have 140 MON genotypes as purebred reference to trace the breed-of-origin. 140JER results show that BOA can definitely be used to evaluate MON crosses even though we don't have any MON phenotypes.

# back-up slide

BOA



BOA\_MON

