MTT

Single step genomic evaluations for the Nordic Red Dairy cattle test day data

Minna Koivula, Ismo Strandén,
Jukka Pösö, Gert Pedersen Aamand,
Esa A. Mäntysaari*

Nordic Cattle Genetic Evaluation

faba

Introduction

- Most genomic evaluations are currently based on multi step -approach:
    1) traditional evaluation with an animal model
    2) extraction of pseudo-observations
    3) genomic model is used to predict direct genomic values (DGV) of candidate animals without own records

- In the single step analysis the phenotypic records are combined directly with genomic information, and the resulting genomic enhanced breeding value (GEBV) already combine both sources of information optimally

Objectives

- Random regression test-day (TD) model is currently used for the official genetic evaluation in Nordic Red Dairy Cattle (RDC)

- As more selection decisions are being made utilizing genomic information, it is becoming essential that all genomic information is included in national evaluations

- Objectives of this study:
    1) **evaluate feasibility of the TD single step model using phenotypic records of Nordic RDC cows**
    2) **calculate validation reliabilities for single step TD model**

Data sets and evaluation models:

- Nordic production test-day data from March 2012
    - 3.5 million cows with records
    - 4.8 million animals in the Nordic RDC pedigree
    - 95.6 million records
    - 184 million equations
- Nordic udder health trait data
    - 4.4 million cows with records
    - 5.4 million animals in the Nordic RDC pedigree
    - 77.3 million records
    - 146 million equations

- Multiple trait multi-lactation models:
    - Production evaluation: 27 biological TD traits/lactation functions
    - Udder health evaluation: 9 biological TD traits

Single step model:

- Pedigree extracted for 5,729 animals with genotypes
- $\mathbf{H}^{22} = [\mathbf{G}^*]^{-1} - \mathbf{A}_{22}^{-1}$

1) $\mathbf{A}^{-1}$constructed using full pedigree file with all animals
2) **G** -matrix scaled with $\Sigma 2pq$ and $\Sigma G_{ii} / \Sigma A_{ii}$
3) 0.20 weight for polygenic (A_{22}) in **G*** (Christensen and Lund, 2010)

- **Implementation**
- PCG iteration on data using full pedigree
- Additional covariance structure ($\mathbf{H}^{22}$) read in each iteration

- Procedure implemented in MiX99 software package

GEBV validation set up

Full run included all observations

1) Full data --> **EBV$_F$** and GEBV$_F$

Reduced run – data until Feb 2008 (4 years of observations removed)

2) Reduced data --> EBV$_R$ and **GEBV$_R$**

- EBVs and GEBVs for all animals
  - combined (G)EBVs for Milk, Protein, Fat, SCC and CM

- For validation purposes:
  1. EDCs calculated with ApaX in MiX99 package
  2. **Deregressed bull EBVs** from the EBV_F for
     Milk, Protein, Fat, SCC and CM

  3. Genotyped bulls divided into reference and candidate bulls
     - Candidate bulls
       - had no daughters with observations in reduced data
       - had EDC>20 in the full data
         (bulls were born between 2003-2007)

- The production evaluation had 748 candidate bulls
- The udder health evaluation had 737 candidate bull

Model solving

- Routine TD evaluation models without heterogeneous variance correction

- Production traits:
  - runs took ~12 h
  - Fixed number of iteration rounds 1500
    no time and convergence difference in TD and in single-step TD

- Udder health traits:
  - Runs took 17 – 31 h;
    single step models needed
    ~14 h more for convergence and
    ~1600 iterations more than models without genomic data

Correlations among GEBVs and EBVs in candidate bulls

– above diagonal for milk and below diagonal CM

	EBV_R(PA)	EBV_F	$GEBV_R$	$GEBV_F$
EBV_R(PA)	1	0.51	0.80	0.51
EBV_F	0.40	1	0.63	0.99
$GEBV_R$	0.70	0.51	1	0.67
$GEBV_F$	0.40	0.96	0.63	1

Validation results

Regression of DRP to GEBV_R or EBV_R(PA)

$$R^2_{\text{validation}} = R^2_{\text{model}} / r^2_{\text{DRP}}$$

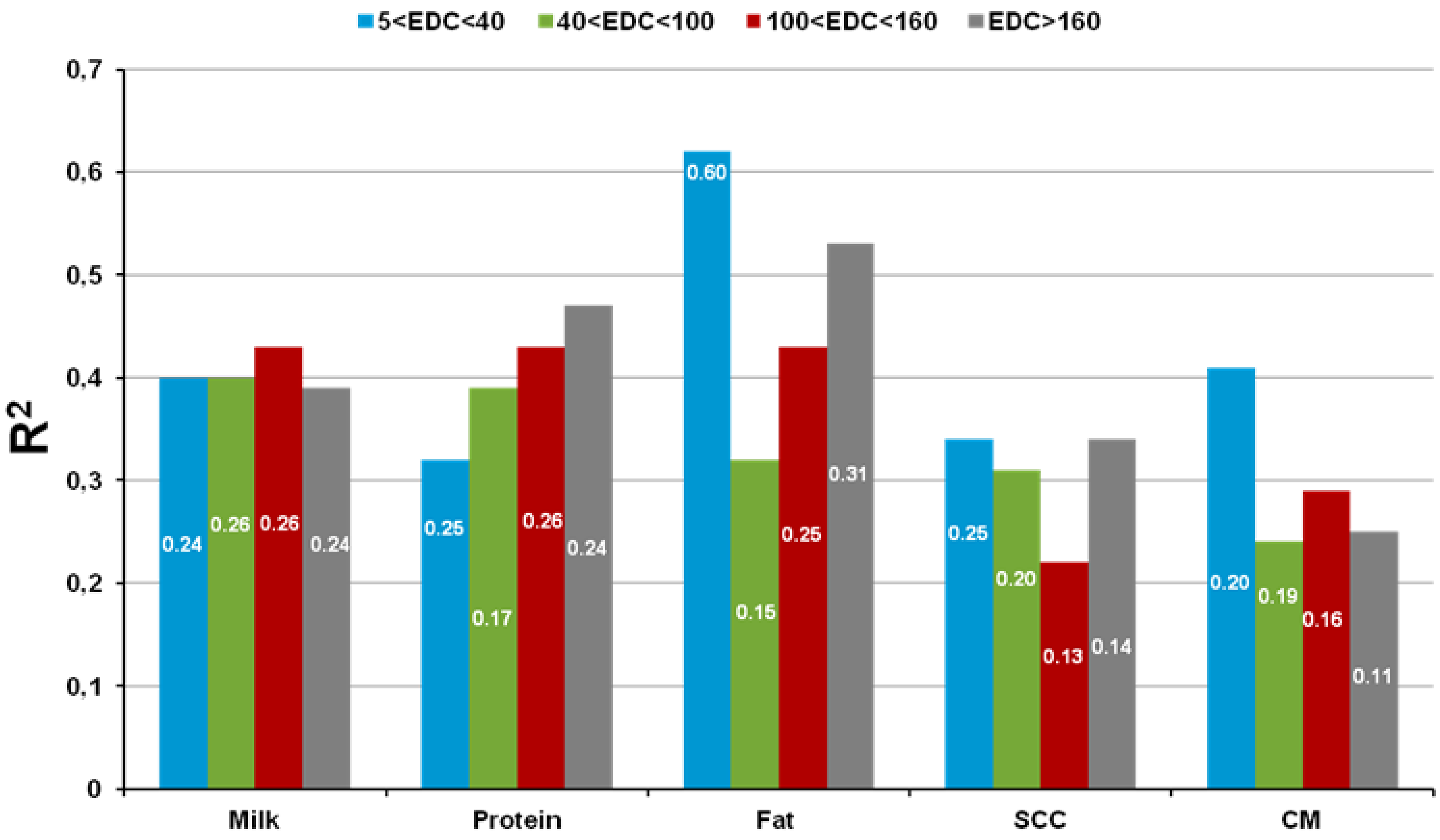
	PA		GEBV_R		r^2_{DRP} Candidate bulls
	b_1	R^2	b_1	R^2	
Milk	0.82	0.25	0.88	0.40	0.93
Protein	0.81	0.23	0.90	0.40	0.91
Fat	0.78	0.29	0.85	0.50	0.91
SCC	0.86	0.15	0.87	0.31	0.87
CM	0.77	0.13	0.76	0.27	0.80

Validation results from TD model single step, sire model and animal model deregressed proofs

	Milk		Protein		Fat	
	b_1	R^2	b_1	R^2	b_1	R^2
PA_{AM}	0.70	0.22	0.89	0.25	0.80	0.28
DGV	0.76	0.30	0.77	0.31	0.85	0.40
$GEBV_{SM}$	0.69	0.32	0.74	0.35	0.80	0.44
$GEBV_{AM}$	0.72	0.35	0.81	0.38	0.79	0.45
$GEBV_R$	0.88	0.40	0.90	0.40	0.85	0.50

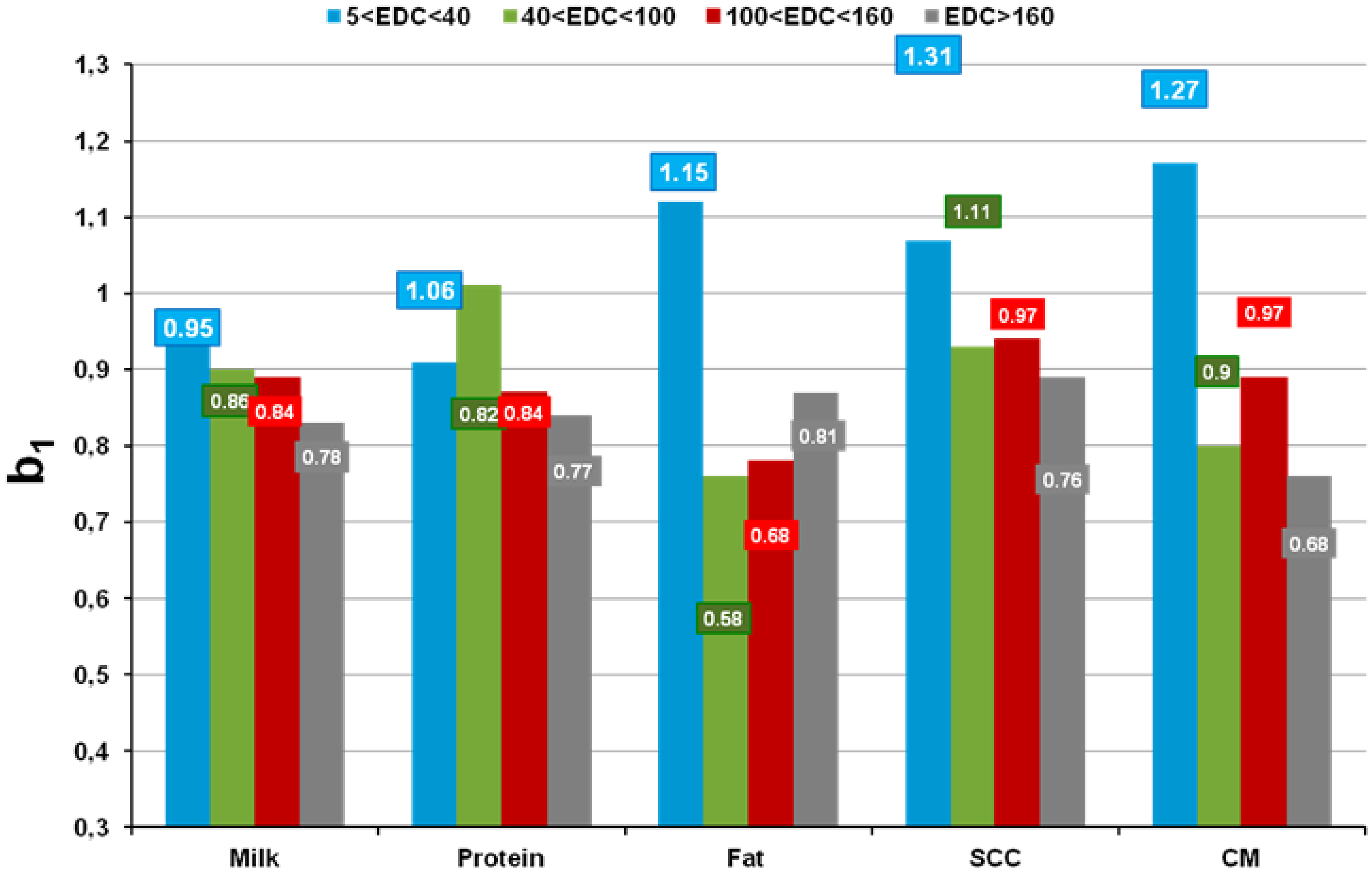
PA_{AM}	= parent average - animal model,
DGV	= direct genomic values - 2-step fit,
$GEBV_{SM}$	= GEBVs using single step genomic model with sire deregressed proofs (Koivula et al. JDS 2012).
$GEBV_{AM}$	= GEBVs using the parameters from animal model deregression (Mäntysaari et al. 2012, Interbull Bulletin 44)
$GEBV_R$	= TD model GEBV

Validation R^2 by EDC group for candidate bulls



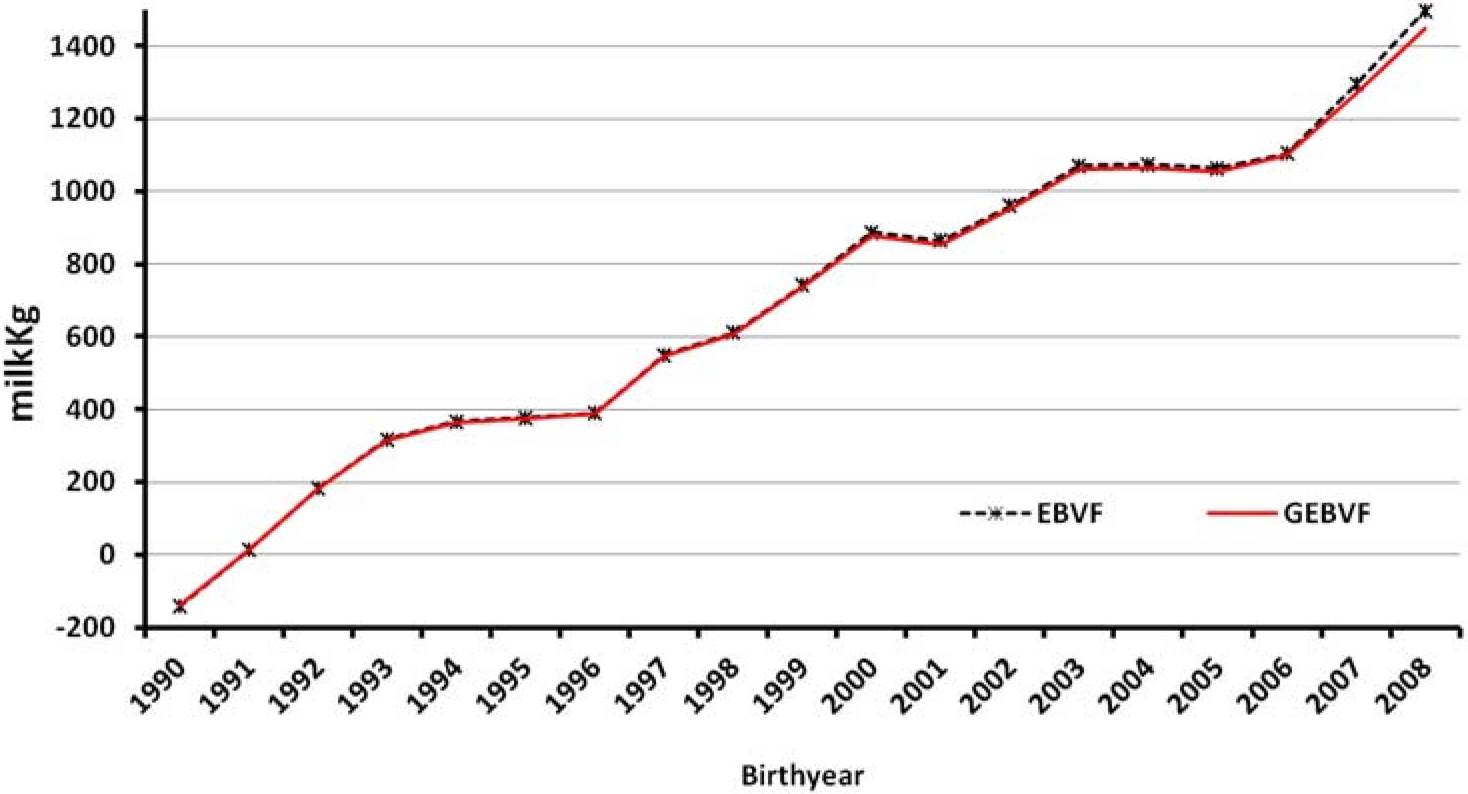
Values in bars show R^2 for PA

Validation b_1 by EDC group for candidate bulls

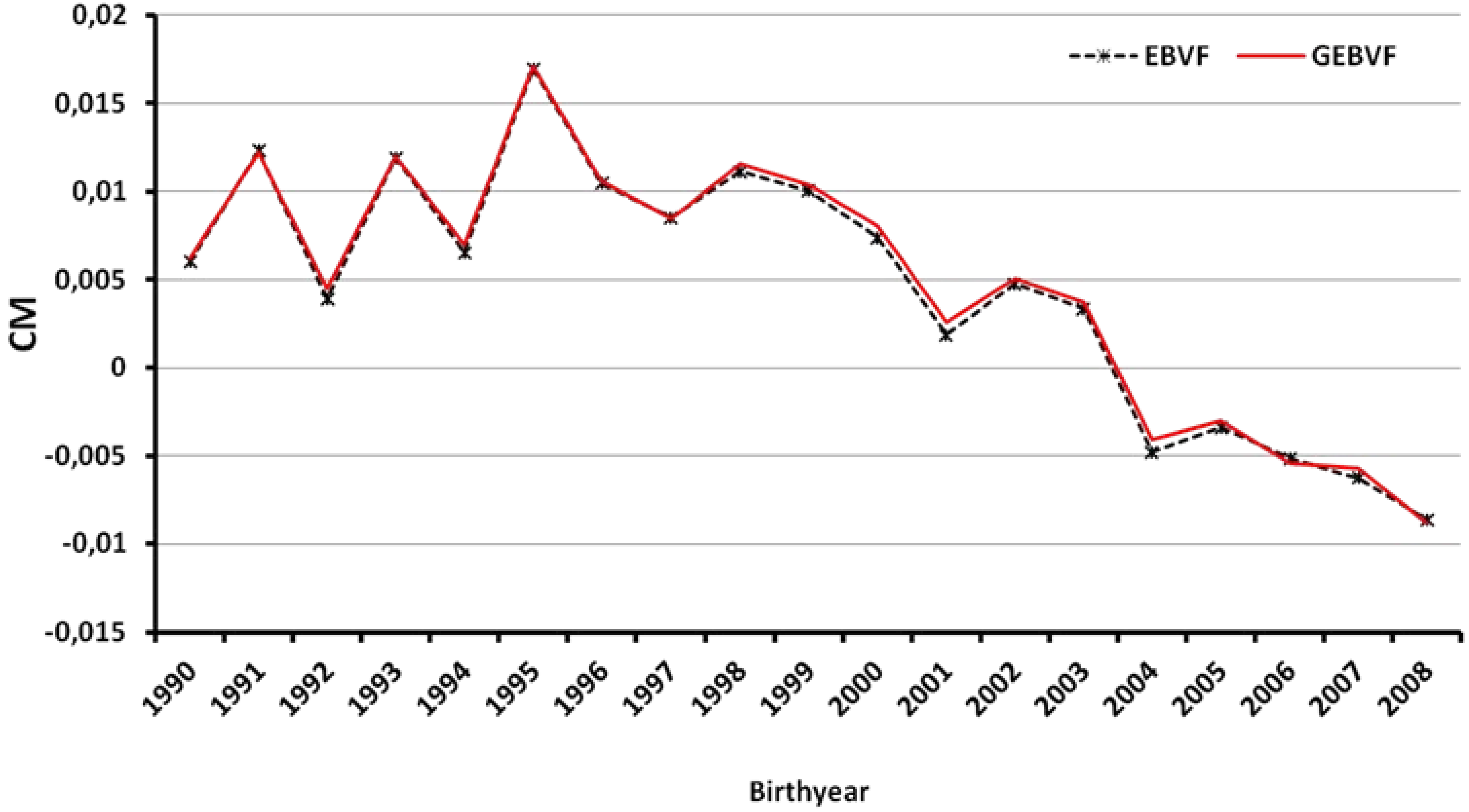


Values in bars show b_1 for PA

Genetic trends for milk EBV_F and $GEBV_F$



Genetic trend for CM EBV_F and $GEBV_F$



”Top 100” of candidate bulls

- 39/100 same with EBV_R (PA) and EBV_F
- 49/100 same with GEBV_R and EBV_F
- 93/100 same with GEBV_F and EBV_F

  - Genomic information cause some re-ranking of bulls with daughters
  - GEBV_R recognizes more top 100 bulls than PA

Conclusions

- Use of phenotypic test-day records in single-step analysis is feasible
  - The model is easy to implement
    - Build $\mathbf{A}_{22}$
    - Build **G***
    - Combine them into $\mathbf{H}^{22}$

  - GEBVs and EBVs are consistent for bulls with daughters

  - GEBV validation reliabilities are **higher** compared to earlier sire model validations
  - GEBV inflation **less** than with DGVs but still exists