

**Genetic heterogeneity:
Is large scale breeding value
estimation possible?**

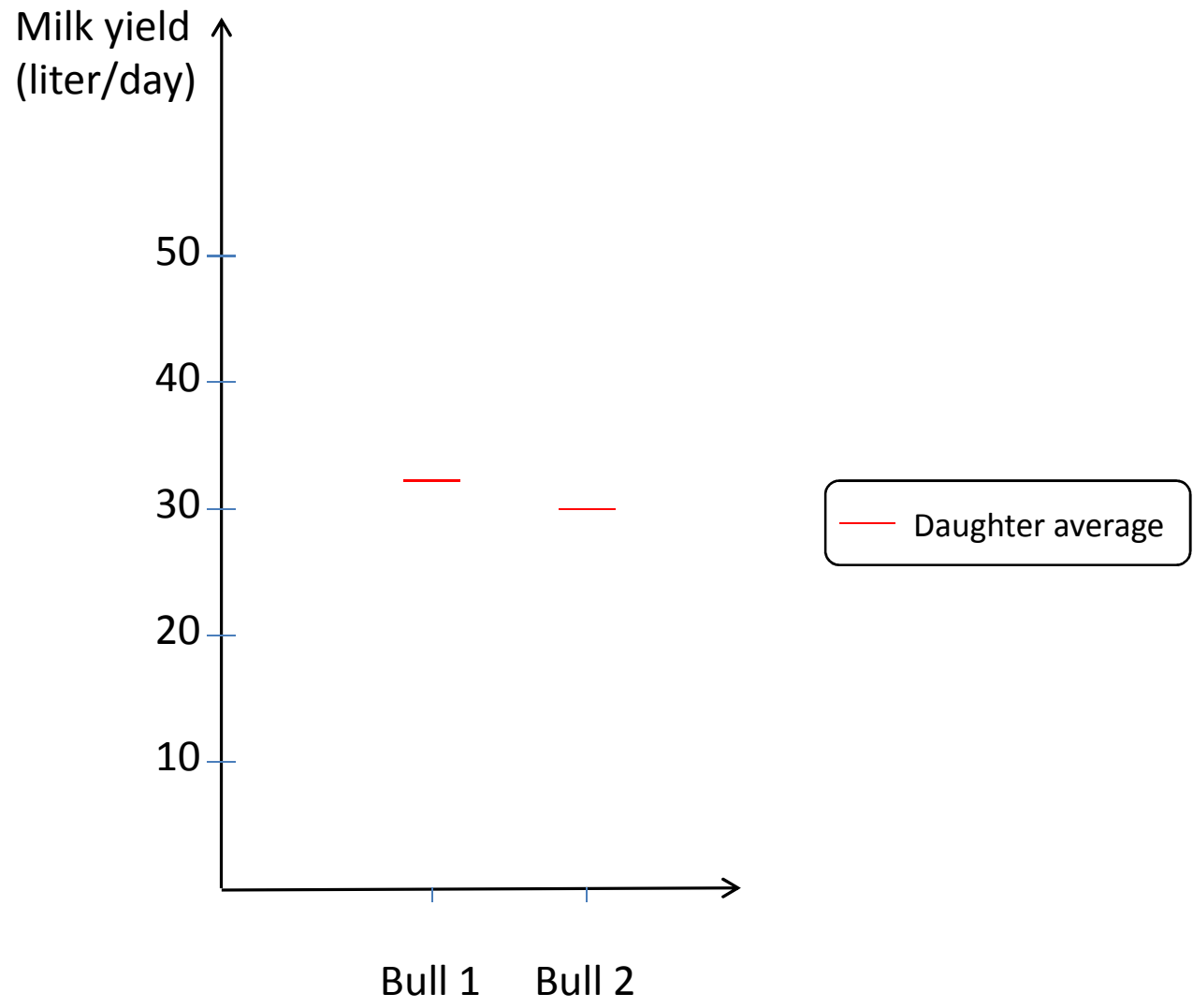
Lars Rönnegård

SLU, Uppsala

Collaborators

- Majbritt Felleki,
- Freddy Fikse,
- Han Mulder,
- Erling Strandberg

Financed by the EU project RobustMILK



Milk yield
(liter/day)

50

40

30

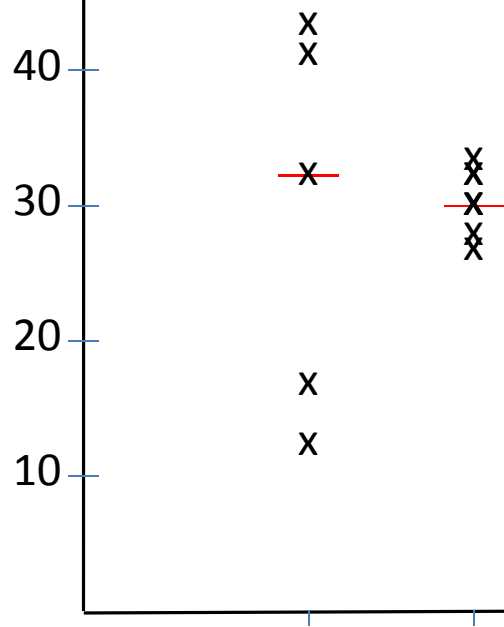
20

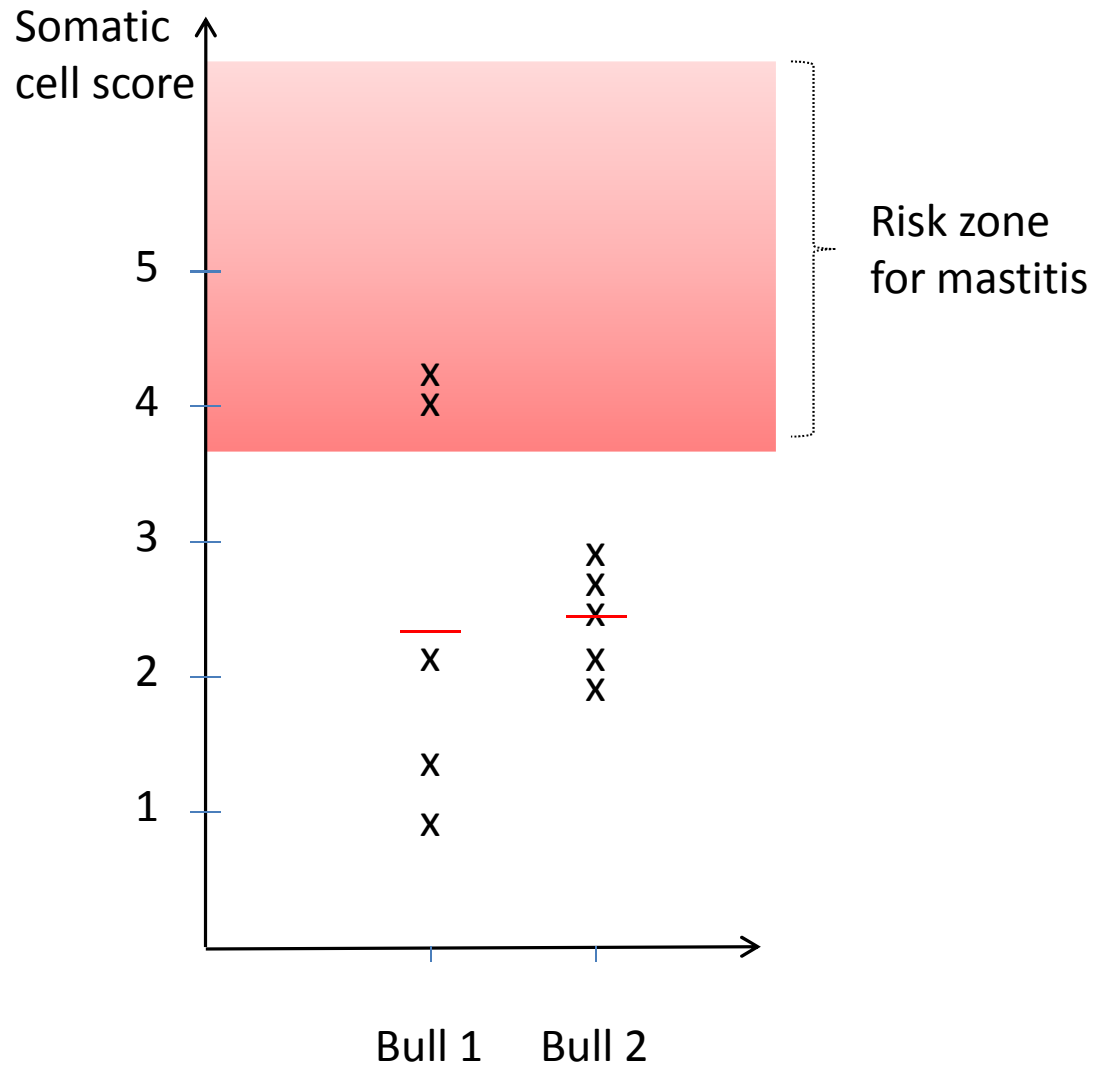
10

Bull 1

Bull 2

— Daughter average





Aim

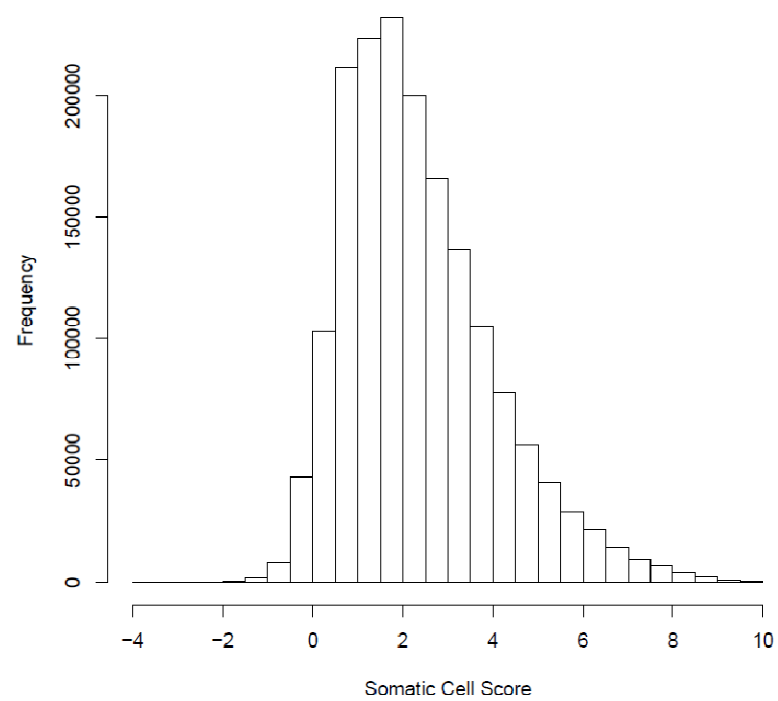
To estimate variance components, on a large dairy cattle data set, using a model having breeding values explaining part of the residual variance.

Data

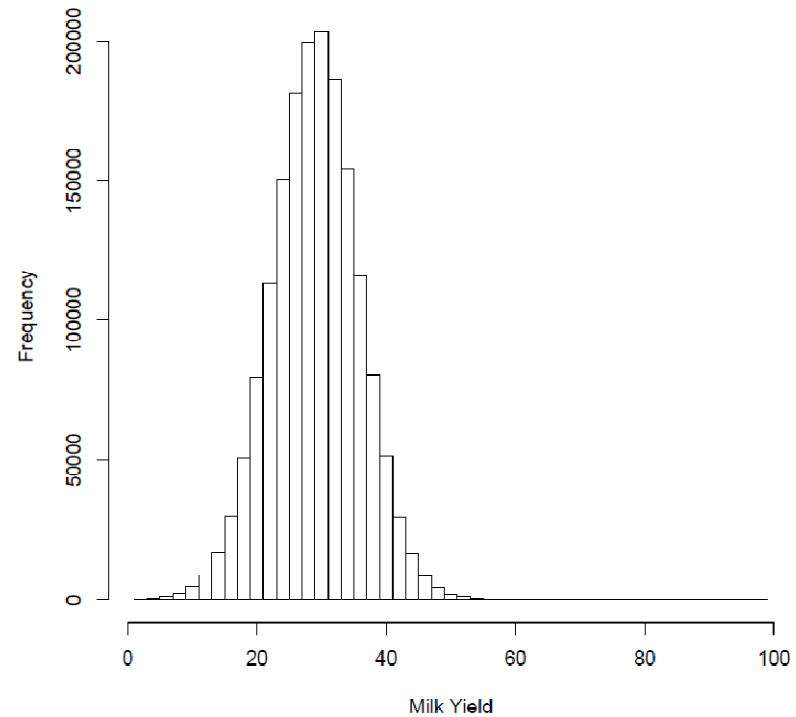
- Swedish Holstein
- Records: 1,693,154
- Animals: 177,411
- Years: 2002-2009

Histograms of studied traits

Histogram of Somatic Cell Score



Histogram of Milk Yield



Statistical Model

$$y = \mathbf{X}\beta + \mathbf{Z}a + \mathbf{W}u + e$$

$$a \sim N(0, \mathbf{A}\sigma_a^2) \quad \text{Breeding values}$$

$$u \sim N(0, \mathbf{I}\sigma_u^2) \quad \text{Permanent environmental effect}$$

Statistical Model

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}a + \mathbf{W}u + e$$

$$a \sim N(0, \mathbf{A}\sigma_a^2) \quad \text{Breeding values}$$

$$u \sim N(0, \mathbf{I}\sigma_u^2) \quad \text{Permanent environmental effect}$$

$$V(e_i) = \exp(\mathbf{X}_d\boldsymbol{\beta}_d + \mathbf{Z}a_d + \mathbf{W}u_d)$$

$$a_d \sim N(0, \mathbf{A}\sigma_{a_d}^2) \quad \text{Breeding values controlling the residual variance}$$

$$u_d \sim N(0, \mathbf{I}\sigma_{u_d}^2)$$

Statistical Model

$$y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}a + \mathbf{W}u + e$$

$$a \sim N(0, \mathbf{A}\sigma_a^2) \quad \text{Breeding values}$$

$$u \sim N(0, \mathbf{I}\sigma_u^2) \quad \text{Permanent environmental effect}$$

$$V(e_i) = \exp(\mathbf{X}_d\boldsymbol{\beta}_d + \mathbf{Z}a_d + \mathbf{W}u_d)$$

$$a_d \sim N(0, \mathbf{A}\sigma_{a_d}^2) \quad \text{Breeding values controlling the residual variance}$$

$$u_d \sim N(0, \mathbf{I}\sigma_{u_d}^2)$$

Fixed effects include year-season, age at calving, days in milk

Estimation Method

Double hierarchical generalized linear model
(DHGLM)

Implemented using ASReml

(Rönnegård et al. 2010 GSE 42:8)

Estimation Method

Double hierarchical generalized linear model
(DHGLM)

Implemented using ASReml

(Rönnegård et al. 2010 GSE 42:8)

Basically an “ordinary heterogeneity correction”
with a genetic effect added to commonly used
fixed effects such as herd, year and season.

Results

- Computation time
 - Iterated between 20 ASReml runs.
 - 10 days in total on a Linux server
- Somatic Cell Score: $\sigma_{a_d}^2 = 0.20$
- Milk Yield: $\sigma_{a_d}^2 = 0.27$

Conclusion

Possible to fit a model with genetic heterogeneity on a large dairy cattle data set

Discussion

- Possible to programme directly in ASReml from version 3.1 => speed up
- Estimating a correlation between breeding values in the mean and variance, ie $cor(a, a_d)$