Changes in GEBV in ssGBLUP with inversion by the APY algorithm using different core animals

Ignacy Misztal, Shogo Tsuruta, Yutaka Masuda, Ivan Pocrnic, Andres Legarra, and Daniela Lourenco

University of Georgia

APY algorithm and different sets of core animals

- ssGBLUP used routinely in chicken, pigs and beef
- Inverse of G by APY to reduce costs
  - Up to 2.3 million genotyped animals
- Reports of GEBV changes with different core animals
- Why and how much?

APY algorithm

SNP matrix or GRM have limited dimensionality , < 20k in Holsteins

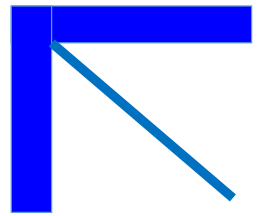
Decompose genotyped animals into N “core” animals $\mathbf{u}_c$ and noncore $\mathbf{u}_n$

Recursion of noncore on core animals

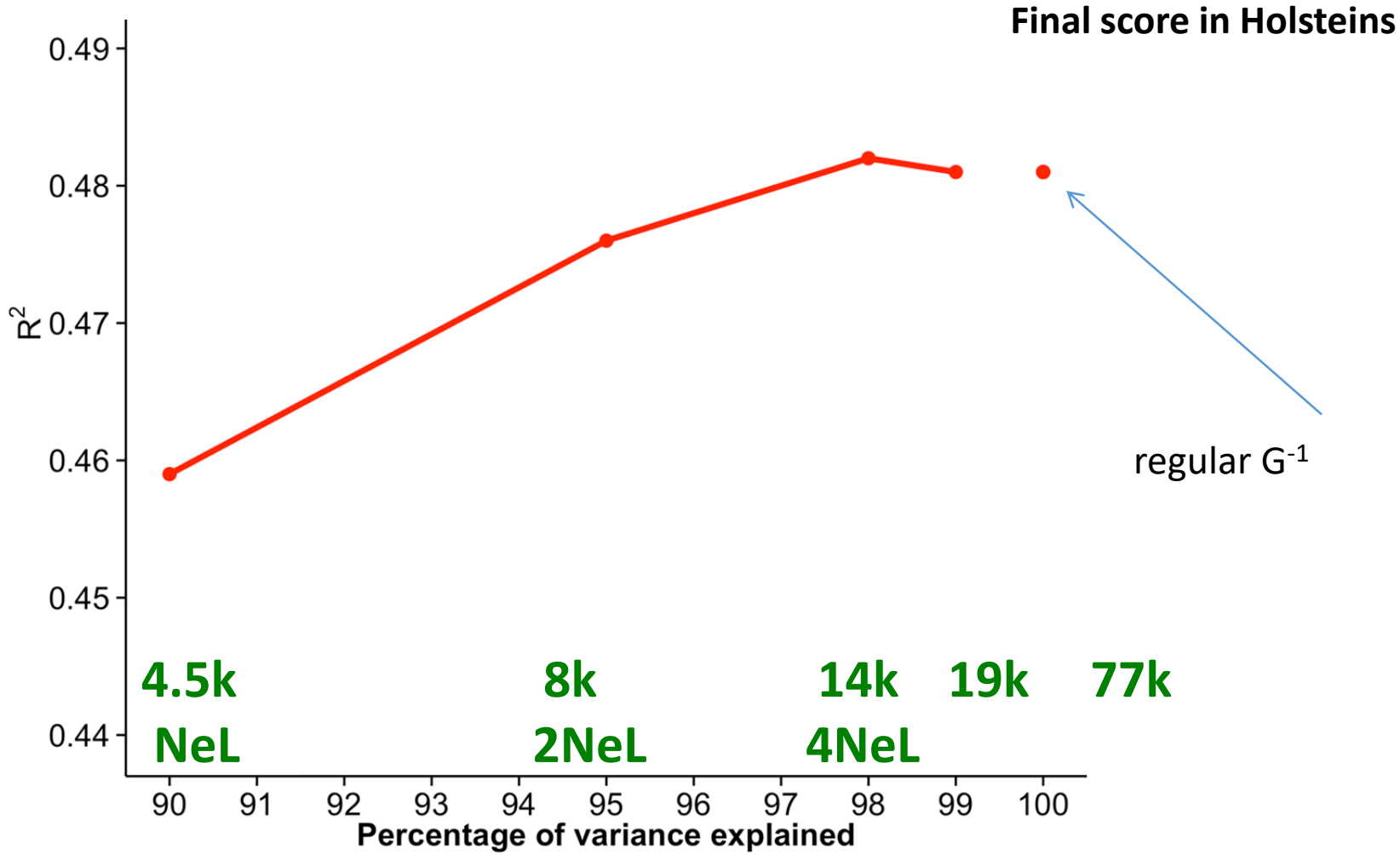
$$\mathbf{u}_n = \mathbf{P}_{nc}\mathbf{u}_c + \varepsilon_n$$

APY inverse

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} \mathbf{G}_{nc}'\mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$



Maximum reliability with the number of core animals at 98% of variance explained



Pocrnic et al., 2016b

Origin of changes

Genomic relationship matrix – information + noise

$$\mathbf{u_n} = \mathbf{P u_c} + \boldsymbol{\varepsilon}$$

$$\sigma_a^2 \qquad \eta\sigma_a^2 \qquad (1-\eta)\sigma_a^2$$

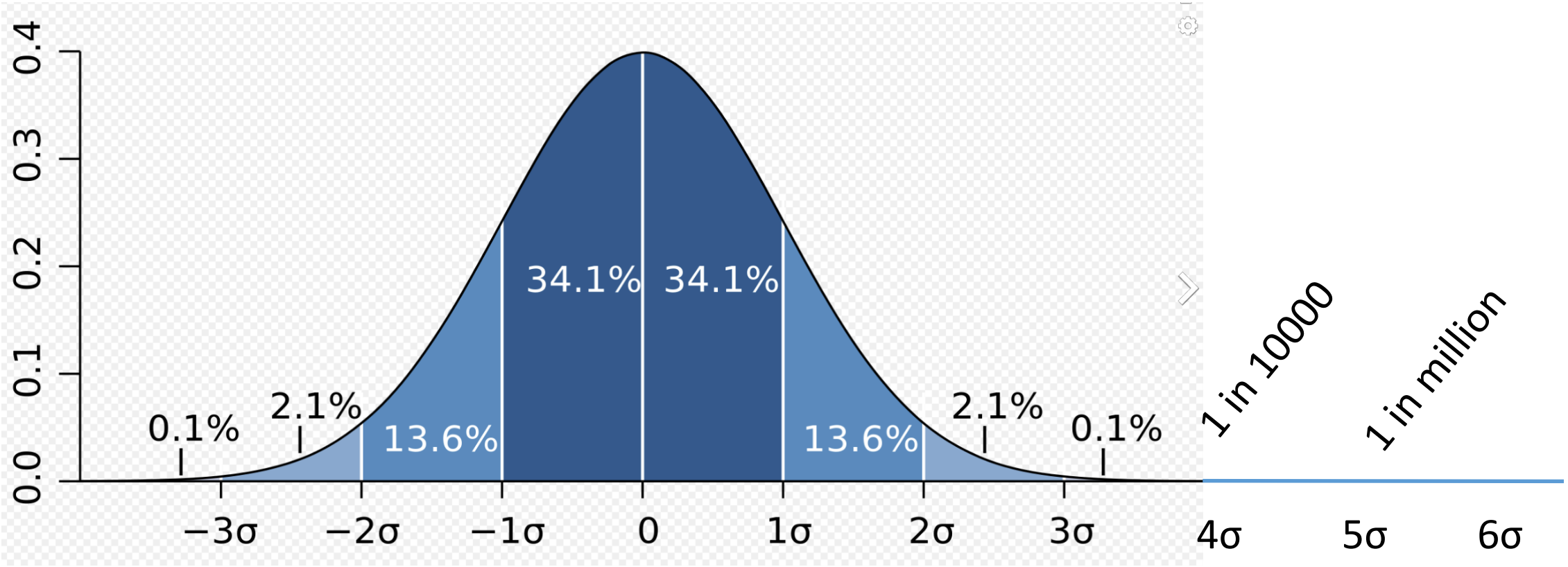
0.98 0.02

$$\mathbf{sd}(\boldsymbol{\varepsilon}) = \sigma_a \sqrt{(1-\eta)}$$ Main source of noise

Approx. difference between GEBV with 2 random cores:

$$\mathbf{sd}(\boldsymbol{\varepsilon_1} - \boldsymbol{\varepsilon_2}) \approx 1.4\, \sigma_a \sqrt{(1-\eta)}$$

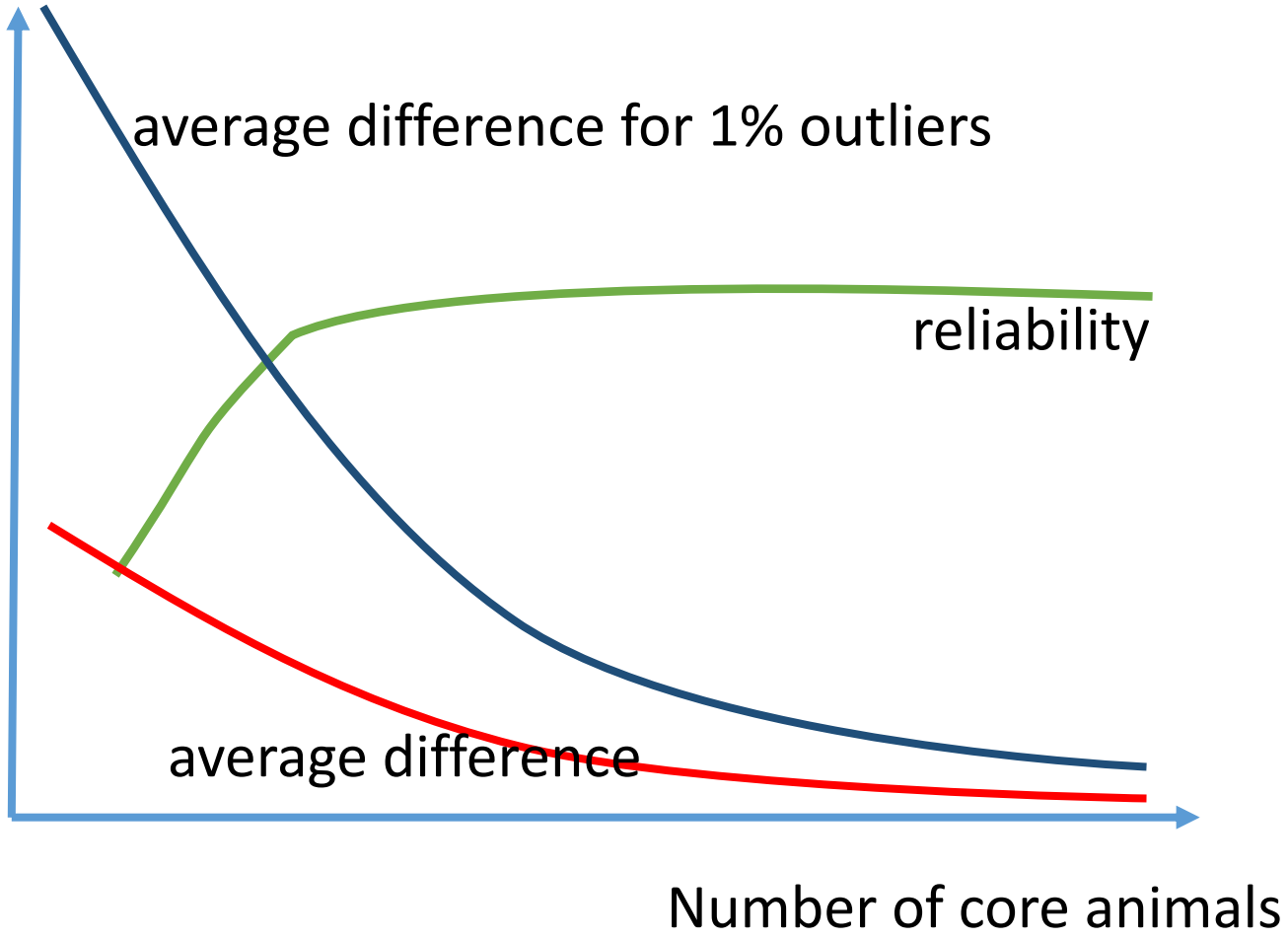
Normal distribution and outliers

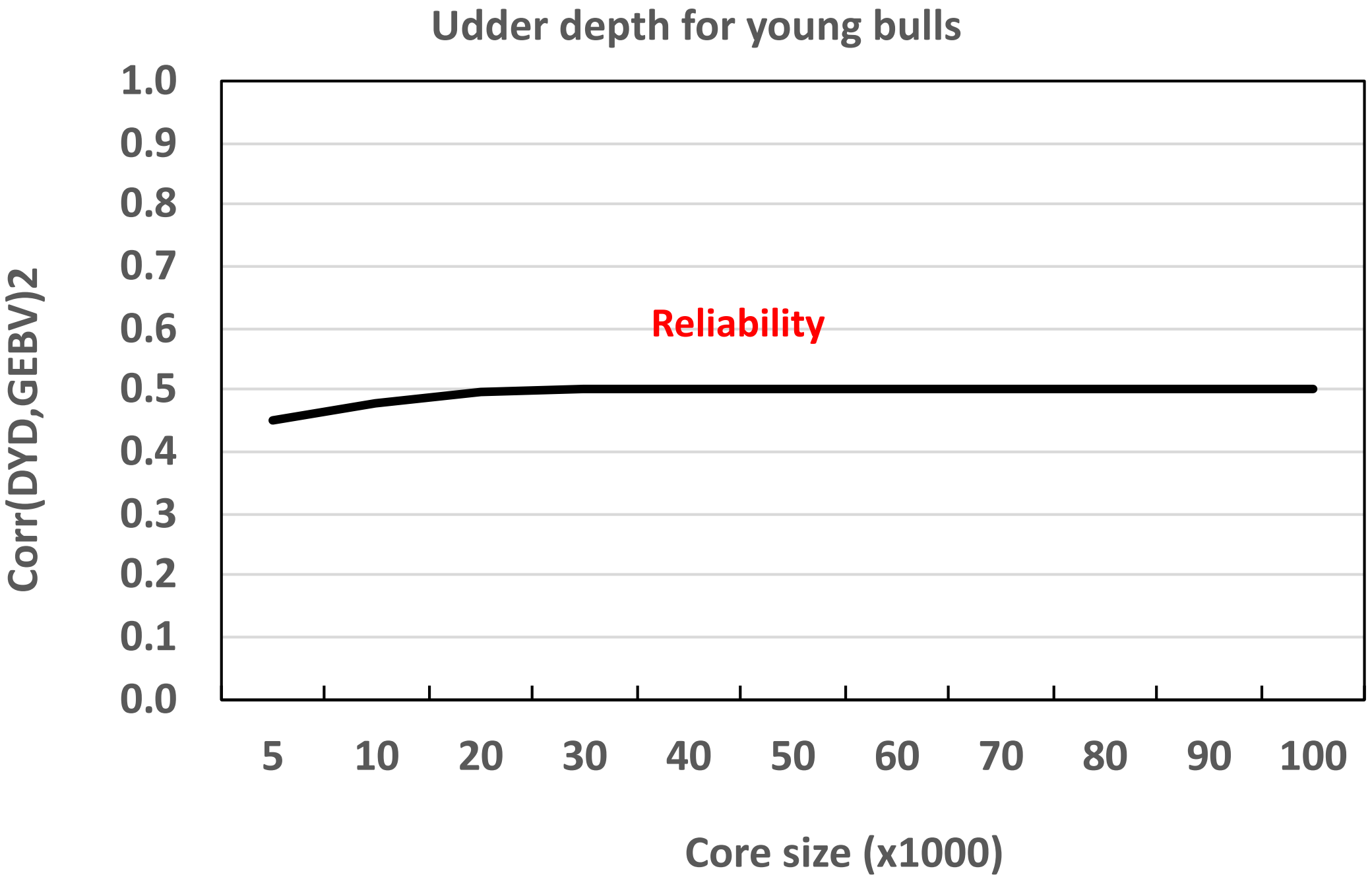


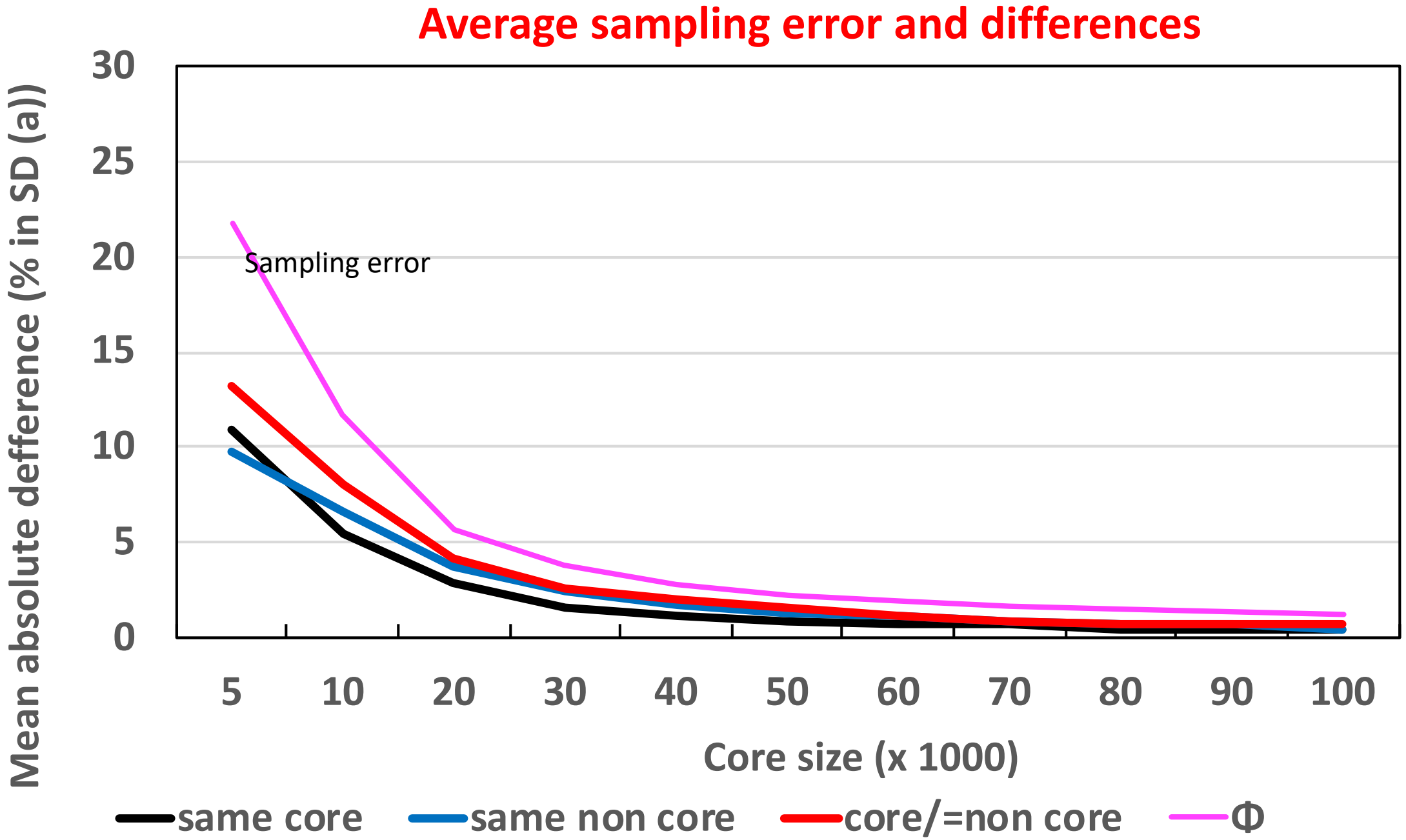
Position of outliers

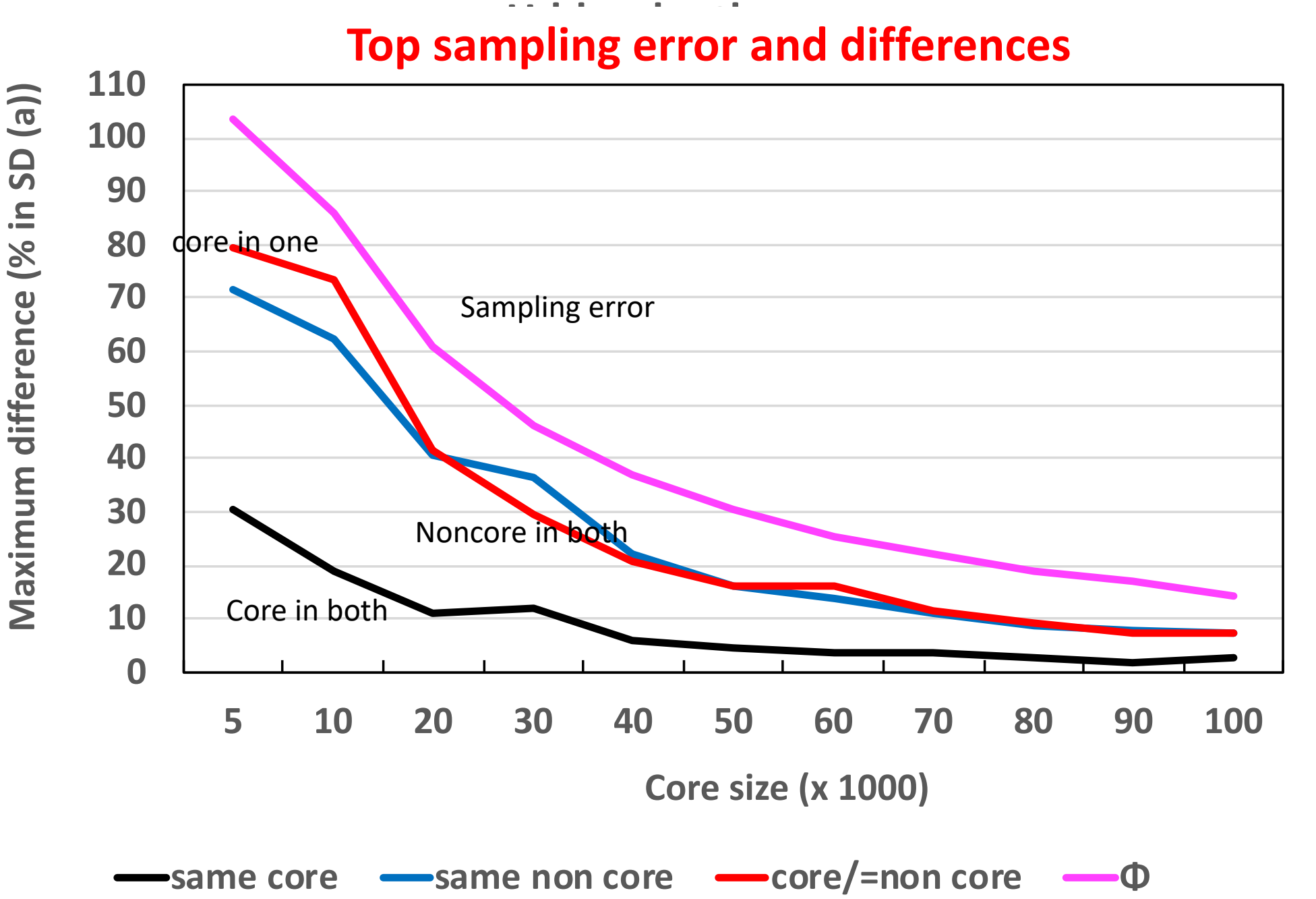
1 in 100: >2.6σ 1 in 1000” >3.3σ 1 in 10,000: > 3.9σ 1 in million: ≈ 4.9σ

Theoretical reliability and average differences




Udder depth for young bulls
Corr(DYD,GEBV)2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Reliability
5
10
20
30
40
50
60
70
80
90
100
Core size (x1000)


Average sampling error and differences
Mean absolute defference (% in SD (a))
30
25
20
15
10
5
0
Sampling error
5
10
20
30
40
50
60
70
80
90
100
Core size (x 1000)
same core
same non core
core/=non core
Φ


Top sampling error and differences
Maximum difference (% in SD (a))
110
100
90
80
70
60
50
40
30
20
10
0
core in one
Sampling error
Noncore in both
Core in both
5
10
20
30
40
50
60
70
80
90
100
Core size (x 1000)
same core
same non core
core/=non core
Φ

Average change for outliers

$$GEBV = rel * BV + rel(1 - rel)N(0, \sigma_a^2)$$

Reliability	Average deviation in additive SD			
	All	1 in 100	1 in 10,000	1 in million
0.70	0.45	1.17	1.76	2.20
0.80	0.40	1.04	1.56	1.96
0.90	0.3	0.78	1.17	1.47
0.99	0.1	0.26	0.39	0.49

How to minimize changes due to APY

- Increase number of core animals
- Treat important animals as core
- Keep core animals same for some period (e.g., 1 year)
- Use indirect prediction
- Use groups of bulls

Conclusions

- Fluctuations of GEBV with APY due to choice of core animals
- Little impact on accuracy/reliability with sufficient number of core animals (EIG98 to EIG99)
- Fluctuations in line with reliabilities and normal distribution

Acknowledgements

Tom Lawlor
Paul VanRaden

Smithfield
PIC
zoetis
CDCB
COUNCIL ON DAIRY CATTLE BREEDING
Cobb
ABCZ
ANGUS
THE BUSINESS BREED
THE MASCHHOFFS
Progressive Farming. Family Style.
DNA
GENETICS
Greater gains. Improved performance.
U.S. Registered Holsteins
FOR MAXIMUM PROFIT
HOLSTEIN ASSOCIATION USA, INC.
PLEMDAT
PFHBiPM
methodgenetics
USDA
United States Department of Agriculture
National Institute of Food and Agriculture