Investigations on the metafounder concept in ssGBLUP based on a simulated cattle population

Judith HIMMELBAUER, Hermann SCHWARZENBACHER, Christian FUERST, Birgit FUERST-WALTL
Interbull meeting, Lyon, August 26th 2023

What is the aim of the study?

ssGBLUP for German-Austrian-Czech Fleckvieh population since April 2021

15 UPG and scaling of G to match A for fitness traits

one of the next steps in the national evaluation system: **metafounder**

Aim:

- test different methods for gamma estimation

- compare the difference between genetic evaluations with and without MF

for a very simple population structure with two base populations and without any unknown pedigrees

How to simulate two MF?

- common founder population: 2 500 generations evolution

- two traits (trait 1 and trait 2)

- population is split in to two subpopulations (A, B)

- 15 generations of positive/negative selection based on TBV of **trait 1**

- subpopulations are again merged

- 30 years of selection by PBLUP and 8 years of selection by ssGBLUP based on **trait 2**

- controlled mating of subpopulations and animals are selected separately by subpopulation
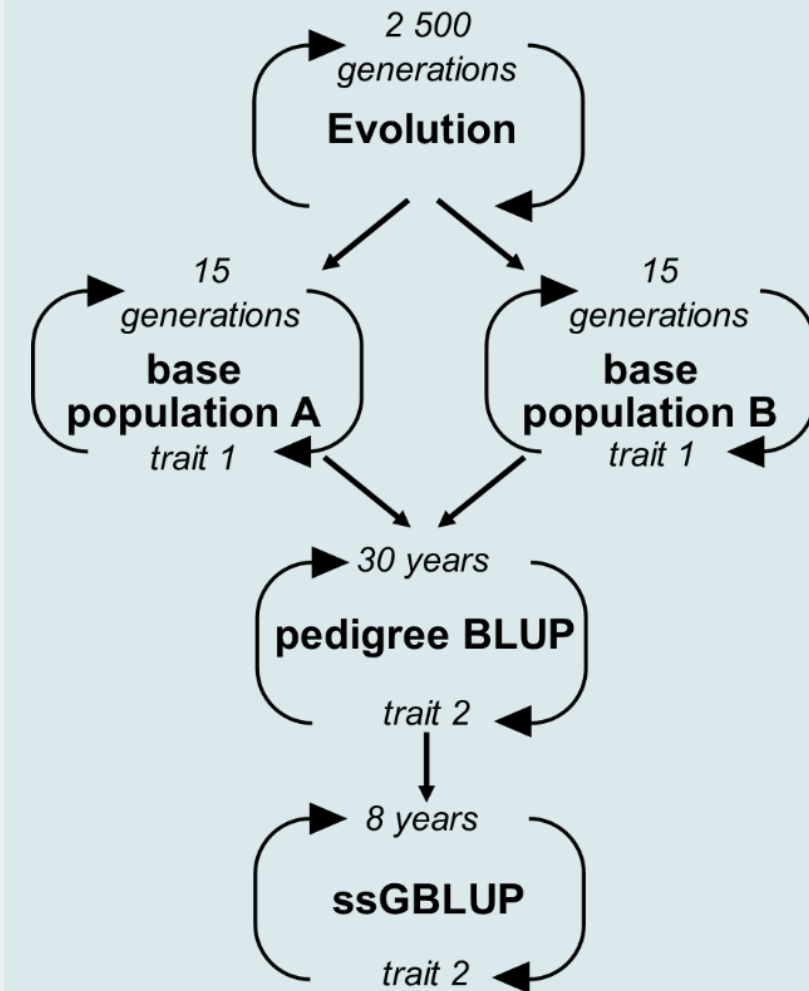
Dataset for analysis

dataset from last year of simulation

all females with offspring have phenotypes

90% of phenotypes of old animals (first 15 years) randomly deleted

final dataset:

- 1 105 500 animals

- 154 500 phenotypes

- 204 900 genotypes

results are based on 10 repetitions

Estimation of Gamma matrix

- **true:** $\Gamma = 8 * cov(p_A, p_B)$ with true base allele frequencies

- $\Gamma = 8 * cov(p_A, p_B)$ with estimated base allele frequencies (Garcia-Baccino et al., 2017)
  - **BFQ_pure:** base allele frequencies from purebred animals
  - **BFQ_all:** base allele frequencies from purebred and crossbred animals

- Method of moments based on summary statistics… (Legarra et al., 2015)
  - **MM_pure:** …for multiple pure populations
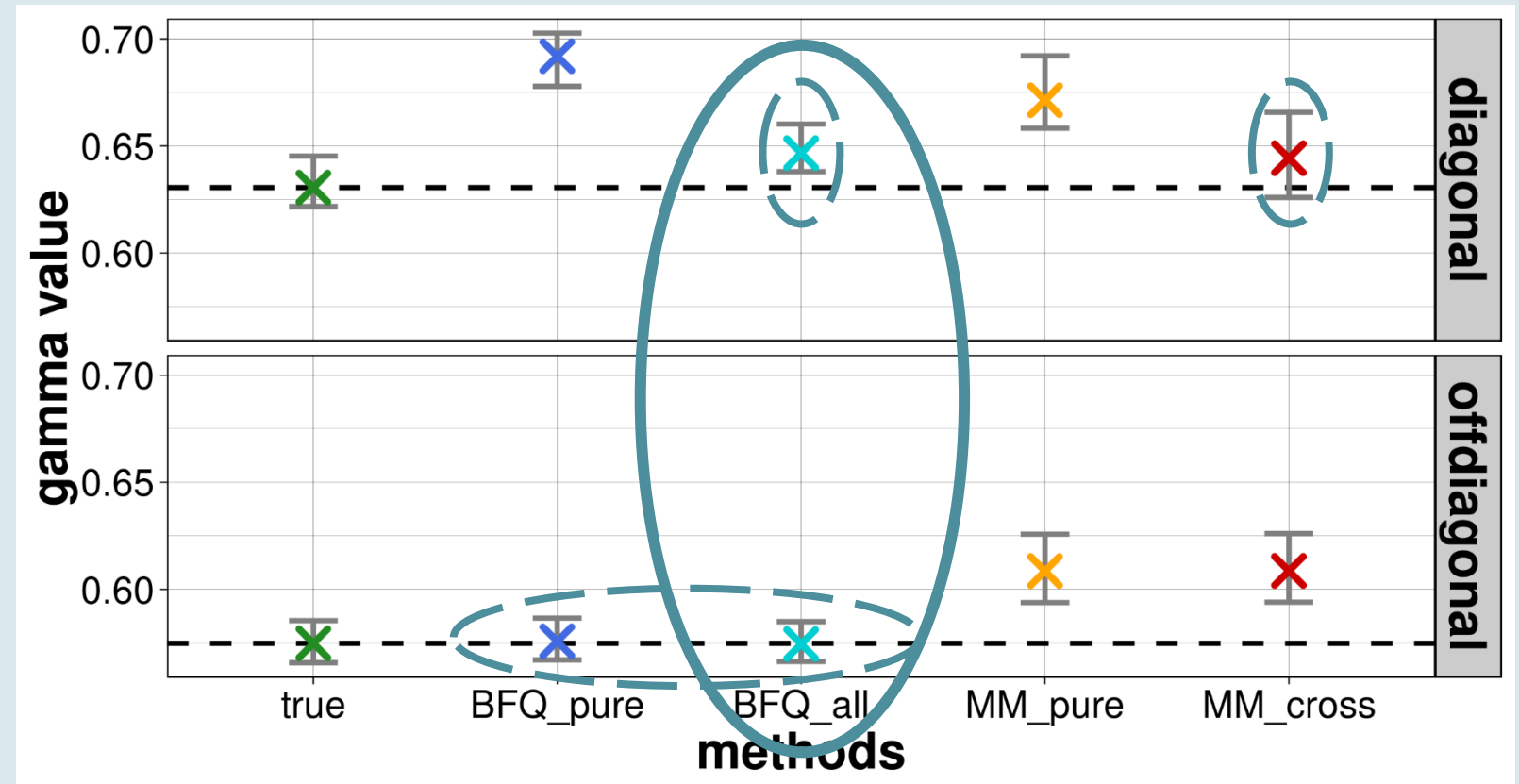 - **MM_cross:** …for populations with crosses

Estimation of Gamma matrix

true average Gamma:

$$\Gamma = \begin{pmatrix} 0.631 & 0.575 \\ 0.575 & 0.632 \end{pmatrix}$$

→ BFQ_all estimated true Γ most accurately in this situation

→ genotypes of crossbreeds are important

Genetic evaluations

	PED	no_UPG	UPG_qp	MF_true	MF_est	MF_sc
SNP information	✘	✔ (*)	✔ (*)	✔ (*)	✔ (*)	✔ (*)
accounting for 2 base populations	✔ (UPG)	✘	✔ (UPG)	✔ (MF)	✔ (MF)	✔ (MF)
relationships between and within base populations	✘	✘	✘	✔ (true)	✔ (estim.)	✔ (true)
scaled variance components	-	-	-	✘	✘	✔ (**)

(∗) ssGBLUP; G computed with APY

(**) $\sigma^2_{related} = \frac{\sigma^2_{unrelated}}{1 + \frac{\overline{diag(\Gamma)}}{2} - \overline{\Gamma}} = \frac{0.3}{0.713} = 0.421$ (Legarra et al., 2015)
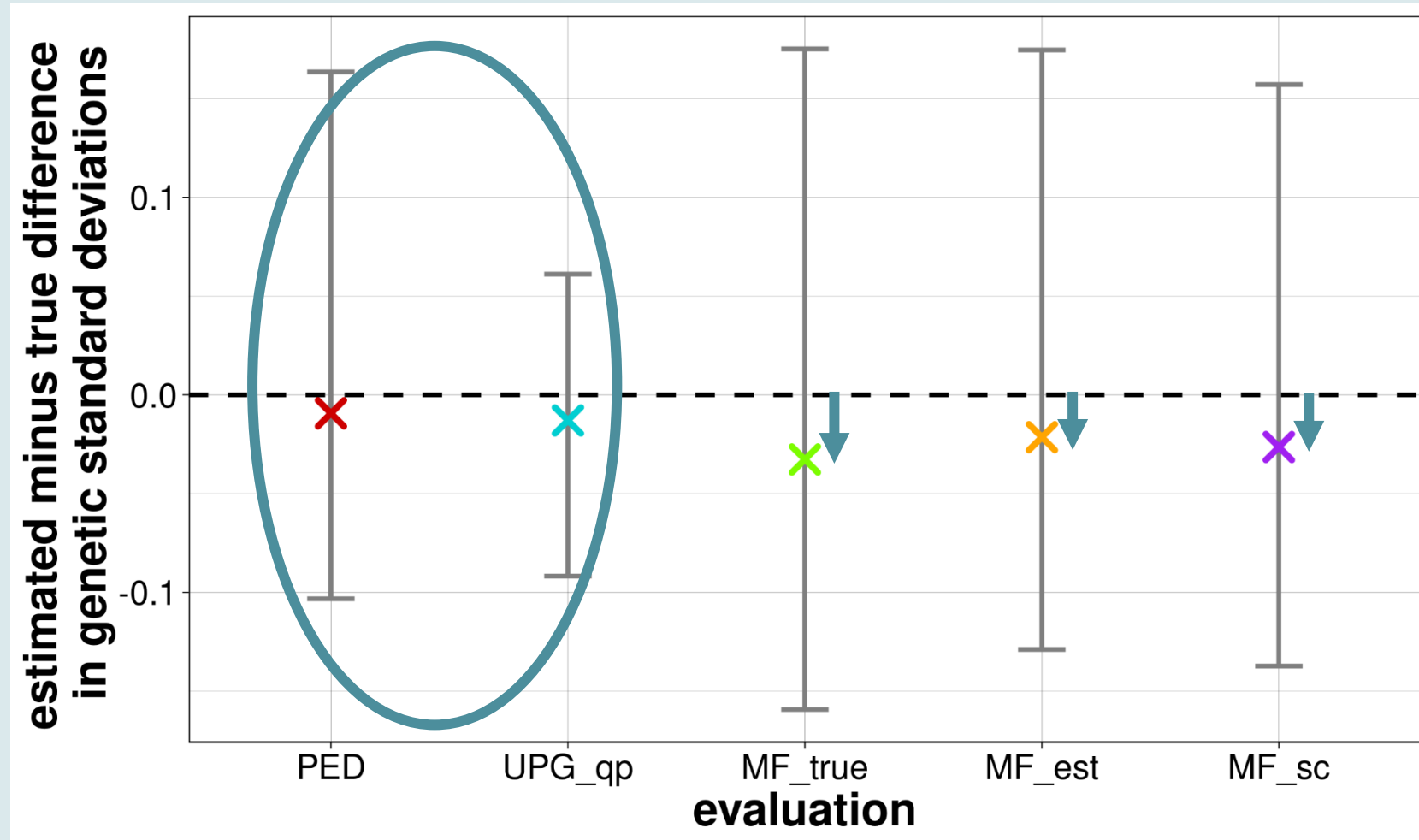
Results for UPG/MF estimates

true average difference: 0.834 genetic standard deviations

→ MF underestimate difference

→smallest bias: PED and UPG_qp

→smallest error variance: UPG_qp

Validation statistics

for animals of the last generation:

Correlation (c):
$$c = cor(TBV, EBV)$$

Bias (b):
$$b = \overline{EBV} - \overline{TBV}$$

Dispersion (b_1):
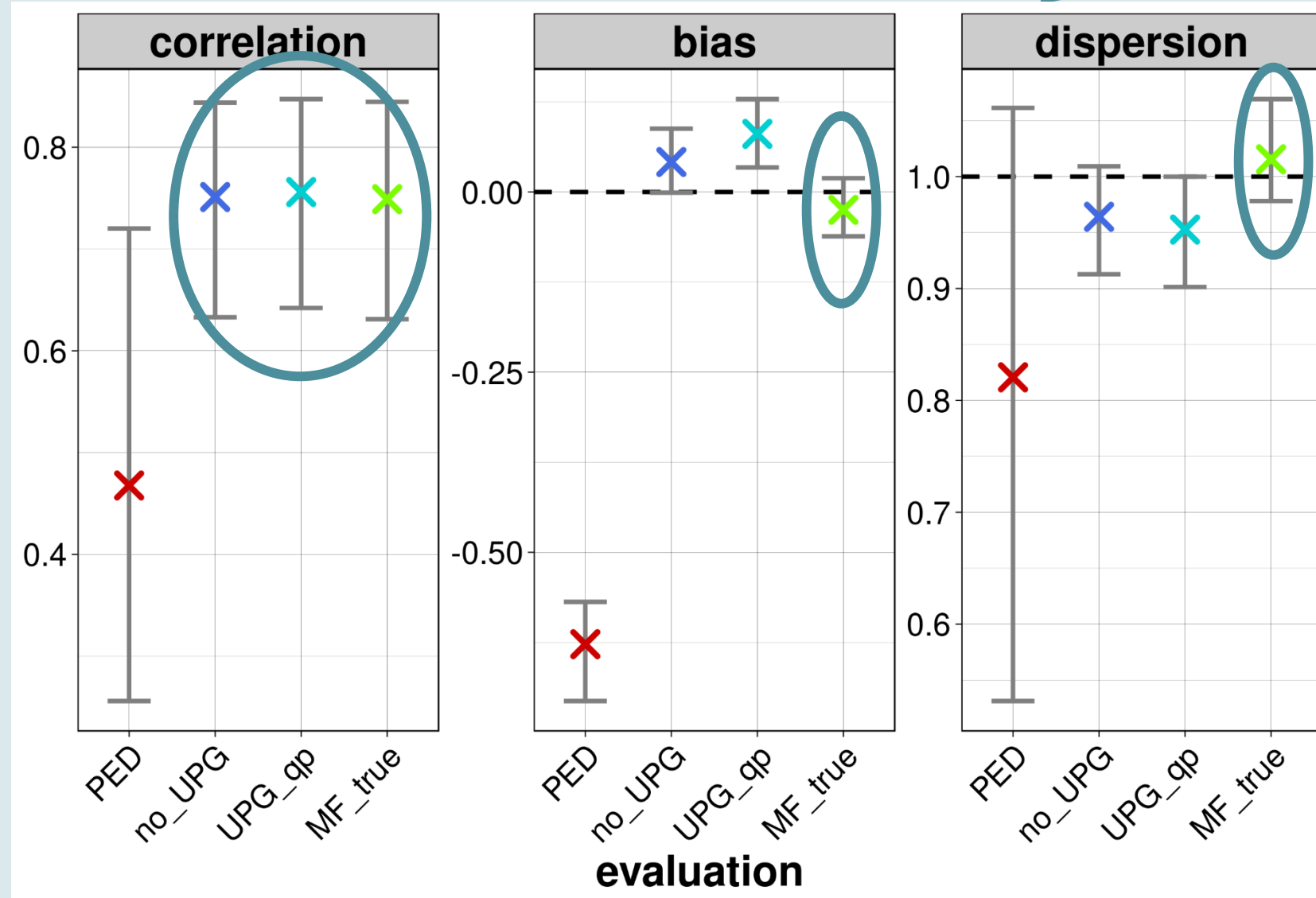$$TBV = b_0 + b_1 \cdot EBV + e$$

Validation statistics

Correlation:

→ no significant differences with/without MF

Bias:

→ less bias with MF

Dispersion:

→ no over-/under dispersion with MF
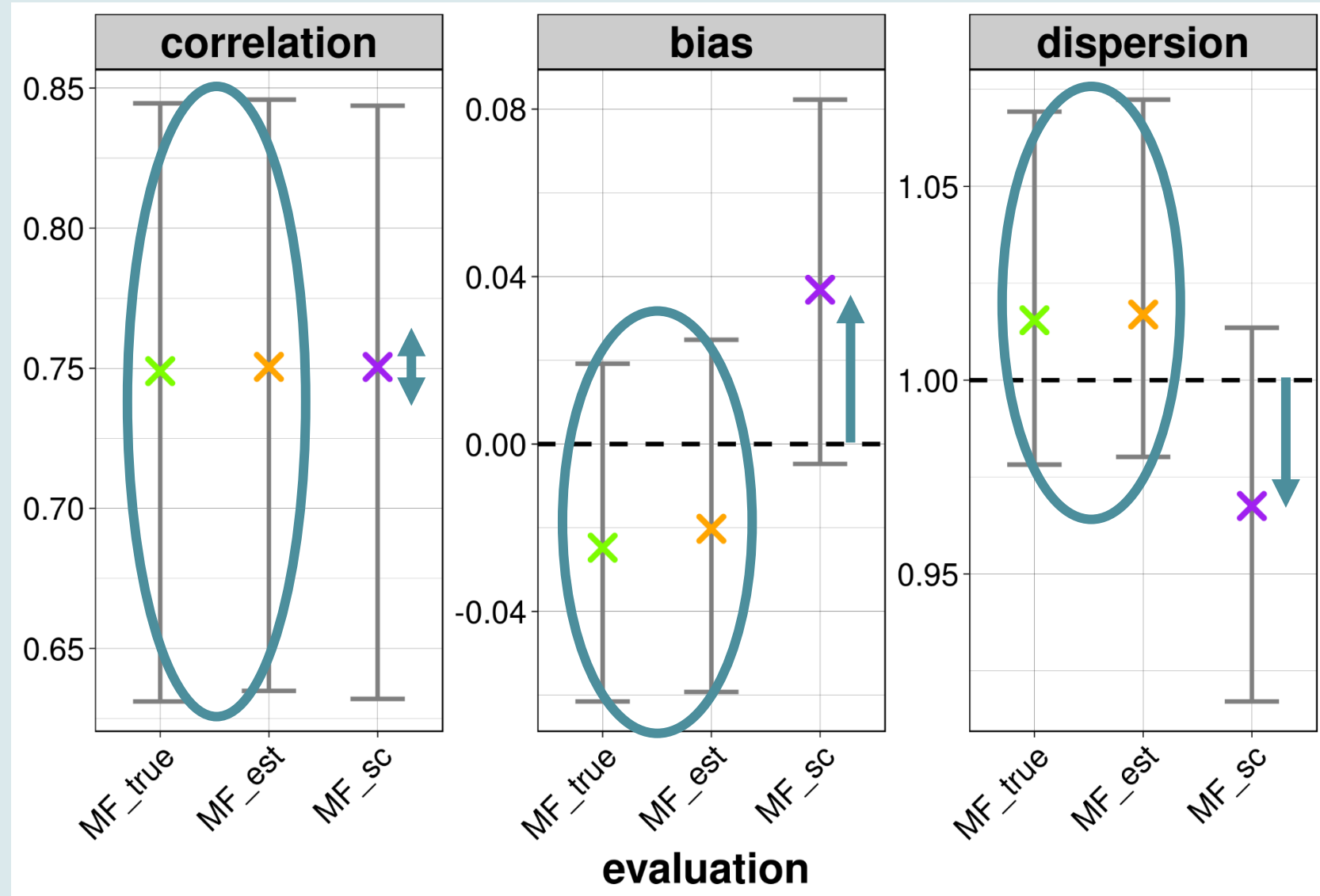
Validation statistics

Estimated Γ :

→ no significant
   difference

**Scaled variance
components:**

→ no difference in
   correlation

→ upward bias

→ overdispersion

Conclusion

- estimation of Gamma matrix based on estimated base allele frequencies works very good, if genotypes from crossbreed animals are used to estimate base allele frequencies

- metafounder have a positive effect on bias and dispersion in this simple situation

- scaling of variance components lead to worse validation statistics in the simulated scenario

→ investigations on more complex situations with missing pedigrees and more MF are necessary

Thank you for your attention!