Bavarian State Research Center
for Agriculture

Exploiting properties of H-Matrix as an aid to identify metafounders

Anglhuber C.[1,2], Edel C.[1], Pimentel ECG.[1], Emmerling R.[1], Götz K.-U.[1], Thaller G.[2]
[1]Institute for Animal Breeding, LfL Bayern
[2]Institute for Animal Breeding and Husbandry, CAU Kiel

Pedigree-Founder

- Assumption made by the IBD approach:
  - **Pedigree founders (PF) are unrelated**
- Reasonable assumption:
  - **There is a relationship between PF** (e.g. families, strata).
- If PF were genotyped
  - $\mathbf{G}_{\mathrm{PF,PF}}$ would be directly accessible.

Metafounder

- **Approximation of the true relationship between PF**
  - $\mathbf{A}^{\mathbf{\Gamma}}_{\mathrm{PF,PF}}$ *is an estimate of true* $\mathbf{G}_{\mathrm{PF,PF}}$.
  - $\mathbf{A}^{\mathbf{\Gamma}}_{\mathrm{PF,PF}}$ is *reduced* in dimensionality/complexity.
- In most real-world situations:
  - The information to establish $\mathbf{A}^{\mathbf{\Gamma}}_{\mathrm{PF,PF}}$ is derived from available GT of non-founders.

H-Matrix

- **Approximation of the true relationship between PF**
  - $\mathbf{H}_{\mathrm{PF,PF}}$ *is an estimate of true* $\mathbf{G}_{\mathrm{PF,PF}}$.
  - $\mathbf{H}_{\mathrm{PF,PF}}$ has *full* dimensionality/complexity.
  - $\mathbf{H}_{\mathrm{PF,PF}}$ is *shrinked* towards a prior.
- In most real-world situations:
  - The information to establish $\mathbf{H}_{\mathrm{PF,PF}}$ is derived from available GT of non-founders.

Objective of Investigation

- Use $\mathbf{H}_{\mathrm{PF,PF}}$ (or functions hereof) in cluster-analyses
  - to define MF and
  - assign PF to MF.
- Clarify:
  - Since $\mathbf{H}_{\mathrm{PF,PF}}$ already is an estimate of $\mathbf{G}_{\mathrm{PF,PF}}$:
    - ✓ Why and when is it advantageous to use MF in Single-Step?

Simulation

- Semi-stochastic simulation: sample of available GT from Fleckvieh.
  - Pedigree of 27.806 animals with available GT
  - 1.505 PF (≙ max. MF)

Simulation

- Semi-stochastic simulation: sample of available GT from Fleckvieh.
  - Pedigree of 27.806 animals with available GT
  - 1.505 PF (≙ max. MF)
    - ✓ Predominantly vertical stratification (PF span 20 birth years).
    - ✓ Resembles common genetic groups problem

Simulation: TBV

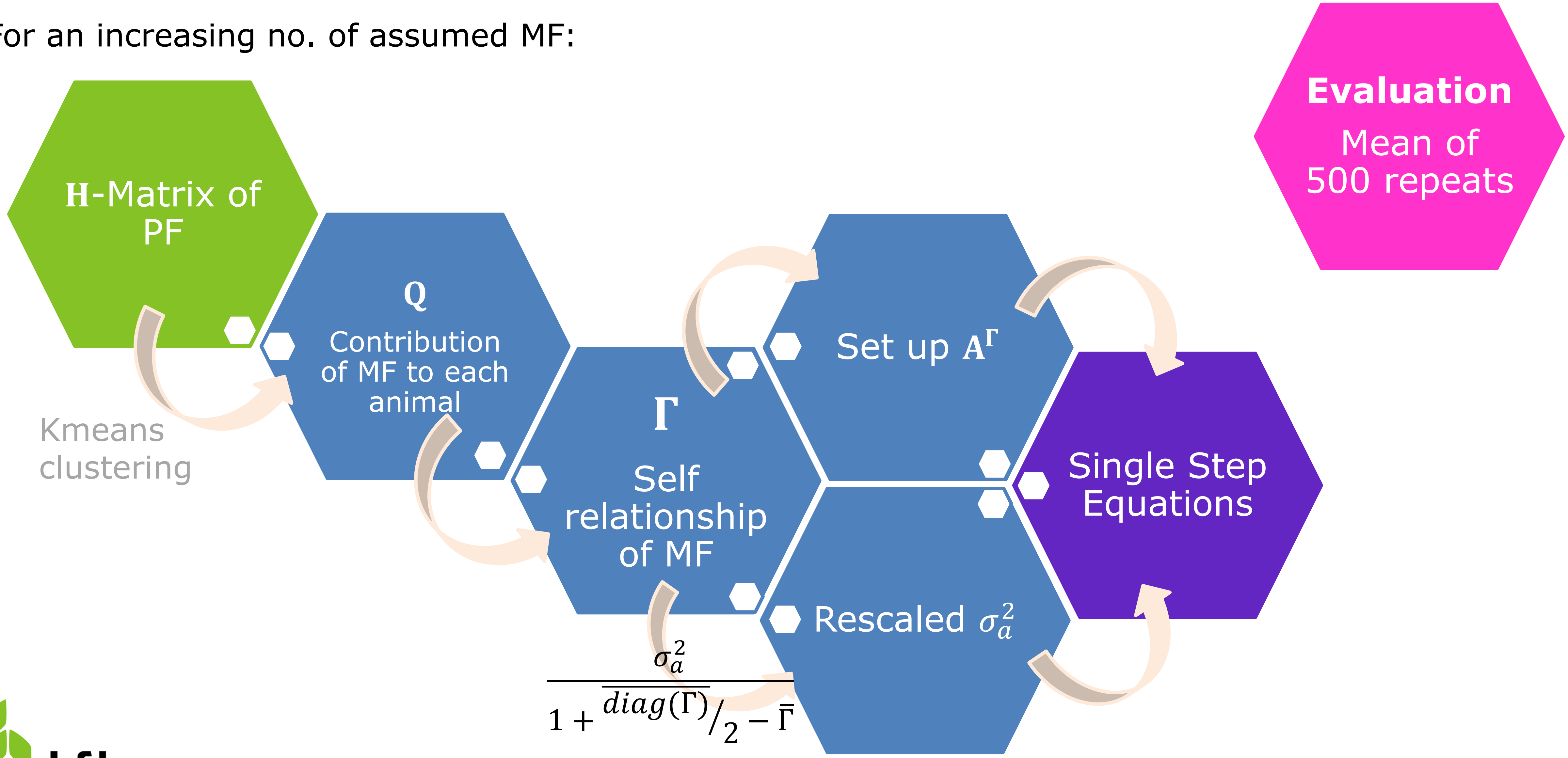
- Semi-stochastic simulation: sample of available GT from Fleckvieh.
  - Pedigree of 27.806 animals with available GT
  - 1.505 PF (≙ max. MF)
    - ✓ Predominantly vertical stratification (PF span 20 birth years).
    - ✓ Resembles common genetic groups problem
- Simulation of TBV of a trait with strong genetic trend ($h^2 = 0.25$).
  - Using true GT and simulated SNP-effects (for 1.000 of 40.859 SNPs).

Simulation: Phenotype

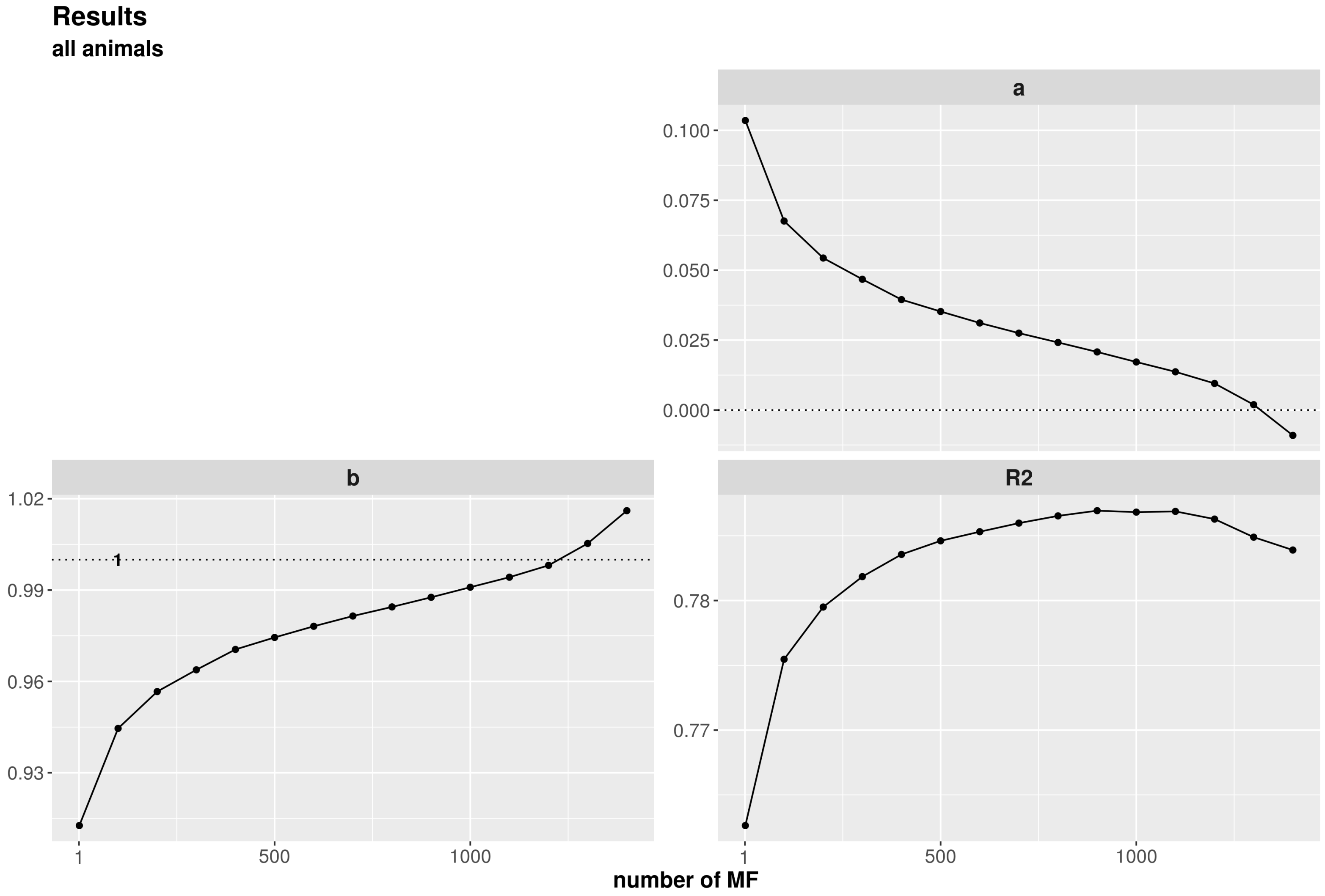
- Phenotypes generated based on TBV.
- Design assumptions for the estimation:
  - All GT used except PF and Gen1.
  - All PT used except animals without offspring (incl. PF).

Pipeline

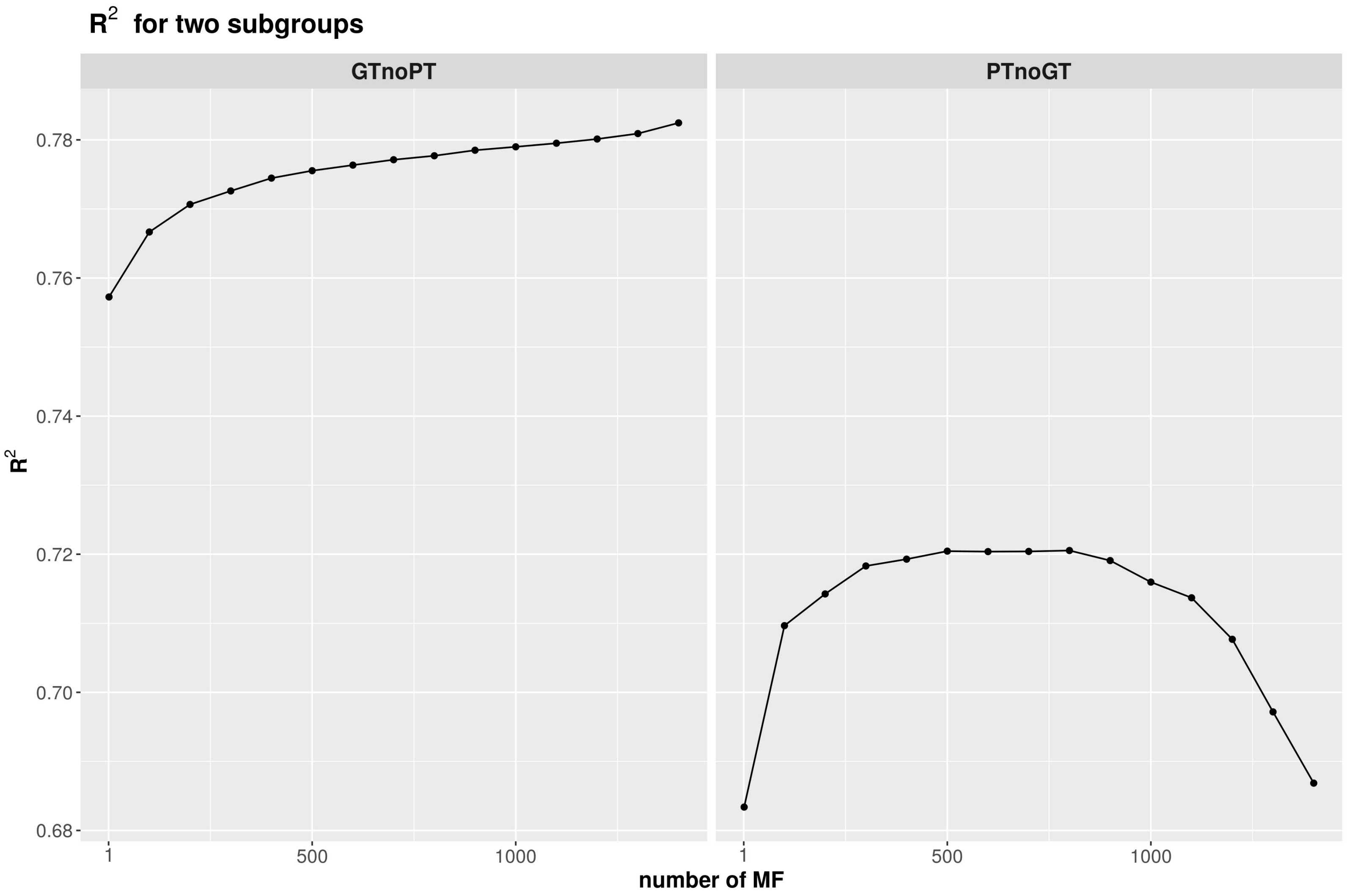
For an increasing no. of assumed MF:



Results


Results
all animals
a
0.100
0.075
0.050
0.025
0.000
b
1.02
0.99
0.96
0.93
1
R2
0.78
0.77
1
500
1000
1
500
1000
number of MF

Results

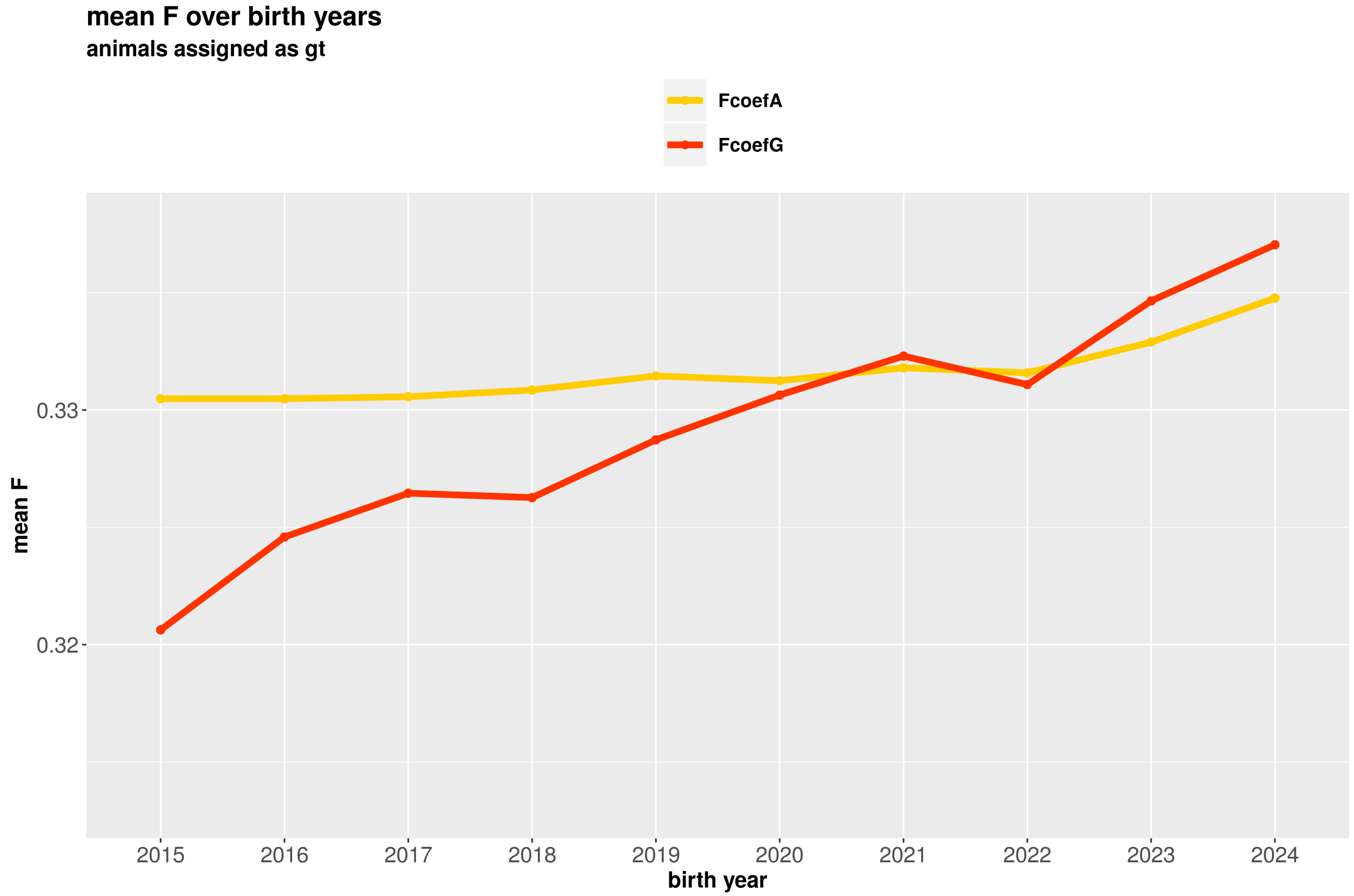


R^2 for two subgroups
GTnoPT
PTnoGT
R^2
0.78
0.76
0.74
0.72
0.70
0.68
1
500
1000
number of MF

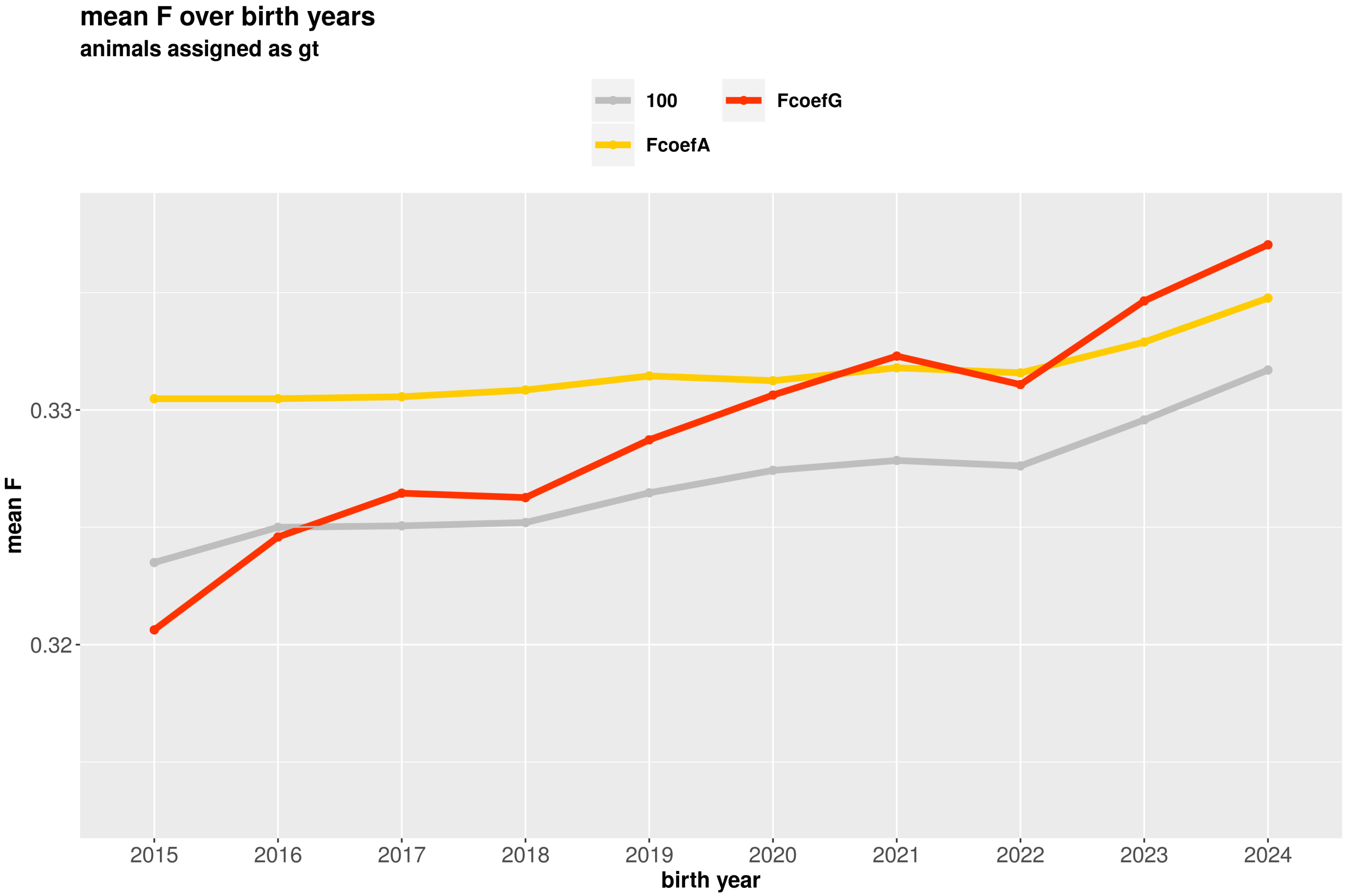
Results

Γ-Diagonal			
No. MF	Min	Mean	Max
1	0.650	0.650	0.650
100	0.643	0.959	1.463
200	0.643	1.106	1.627
300	0.646	1.201	1.646
400	0.649	1.295	1.923
500	0.650	1.352	1.922
600	0.650	1.405	1.980
700	0.652	1.458	2.079
800	0.652	1.508	2.335
900	0.655	1.564	2.396
1000	0.658	1.625	2.872
1100	0.661	1.694	2.873
1200	0.666	1.781	3.625
1300	0.689	1.942	6.565
1400	0.844	2.221	9.933

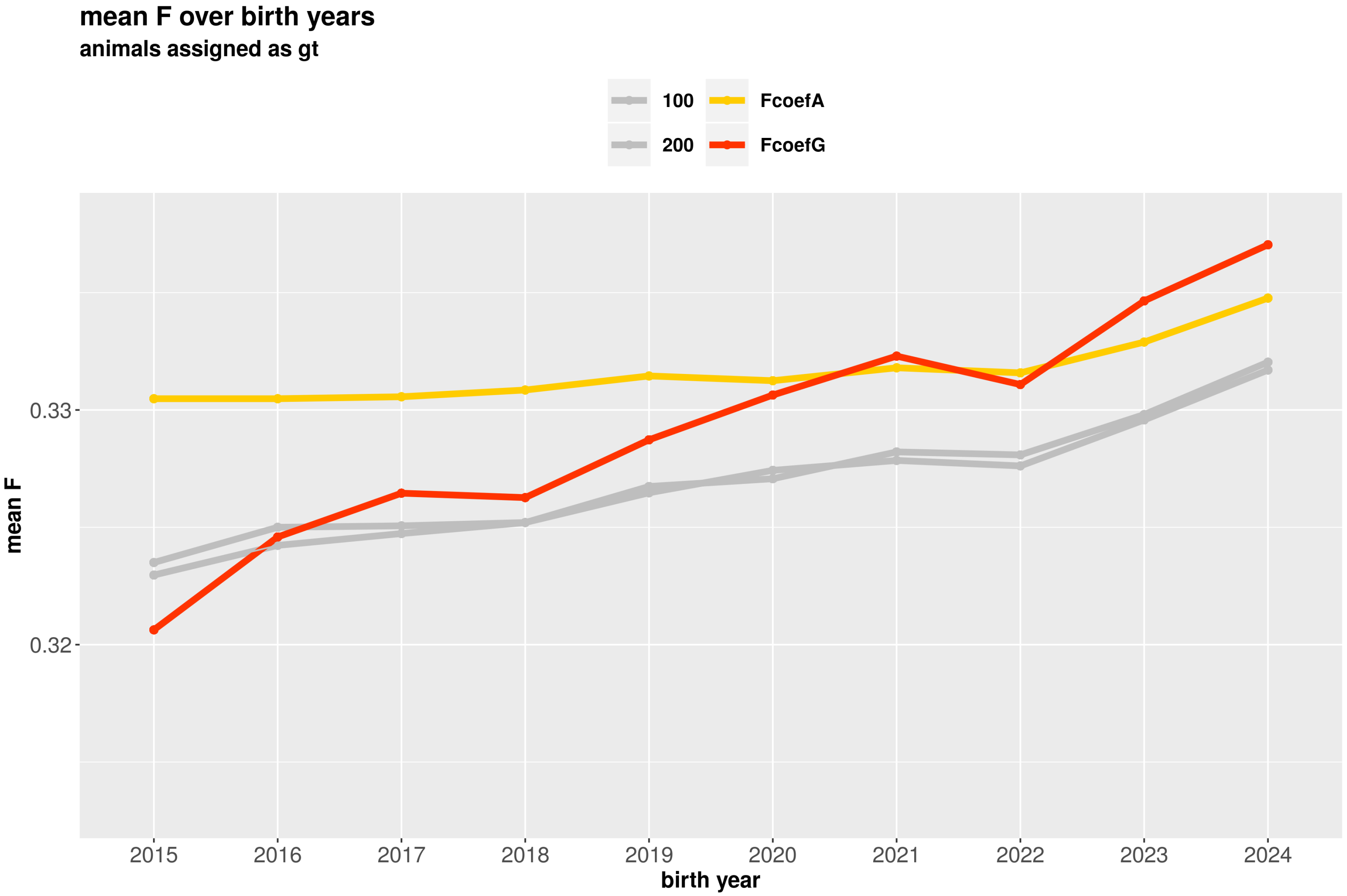
Results


mean F over birth years
animals assigned as gt
FcoefA
FcoefG
mean F
0.33
0.32
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results

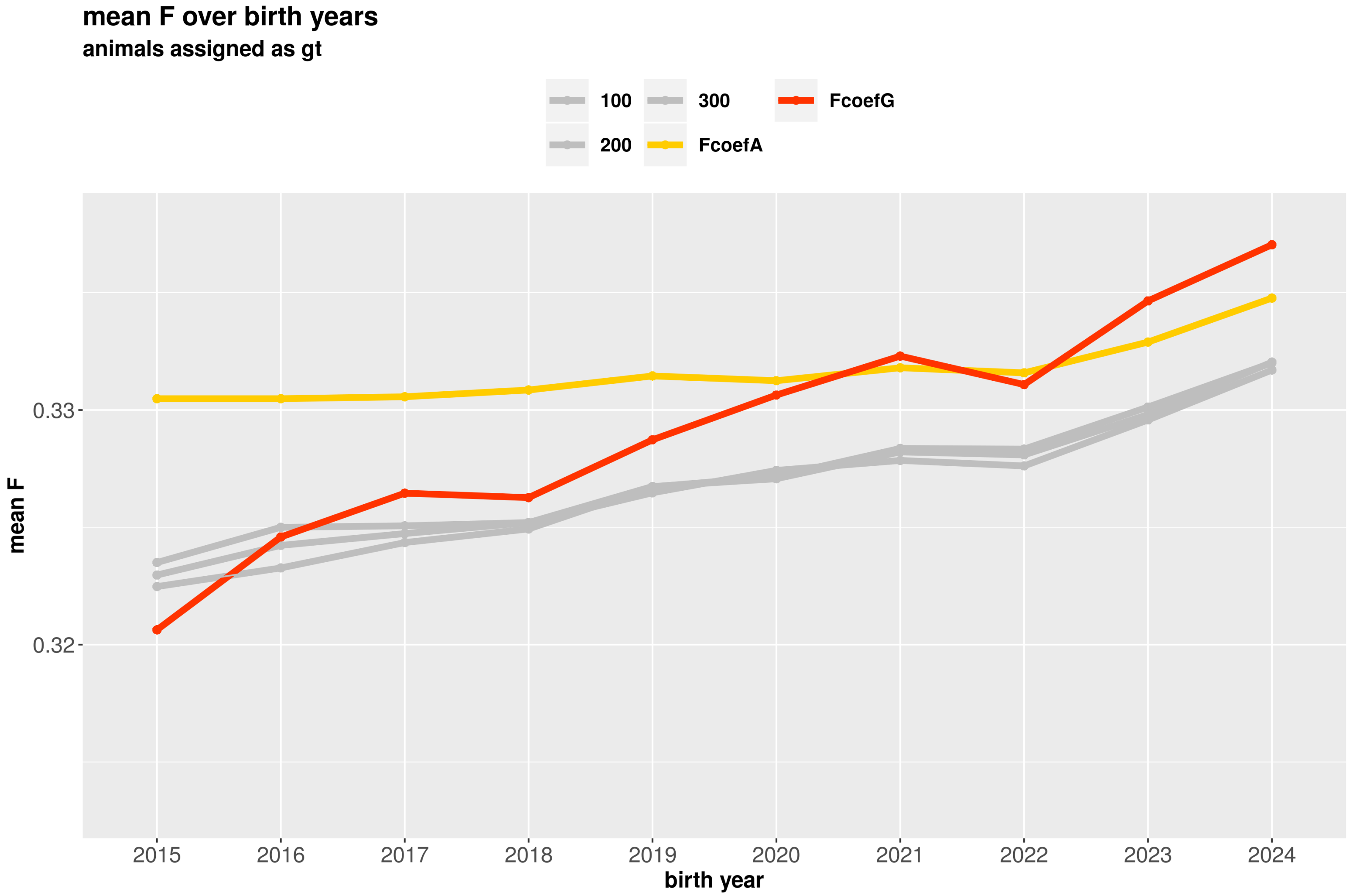

mean F over birth years
animals assigned as gt
100
FcoefG
FcoefA
0.33
0.32
mean F
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results



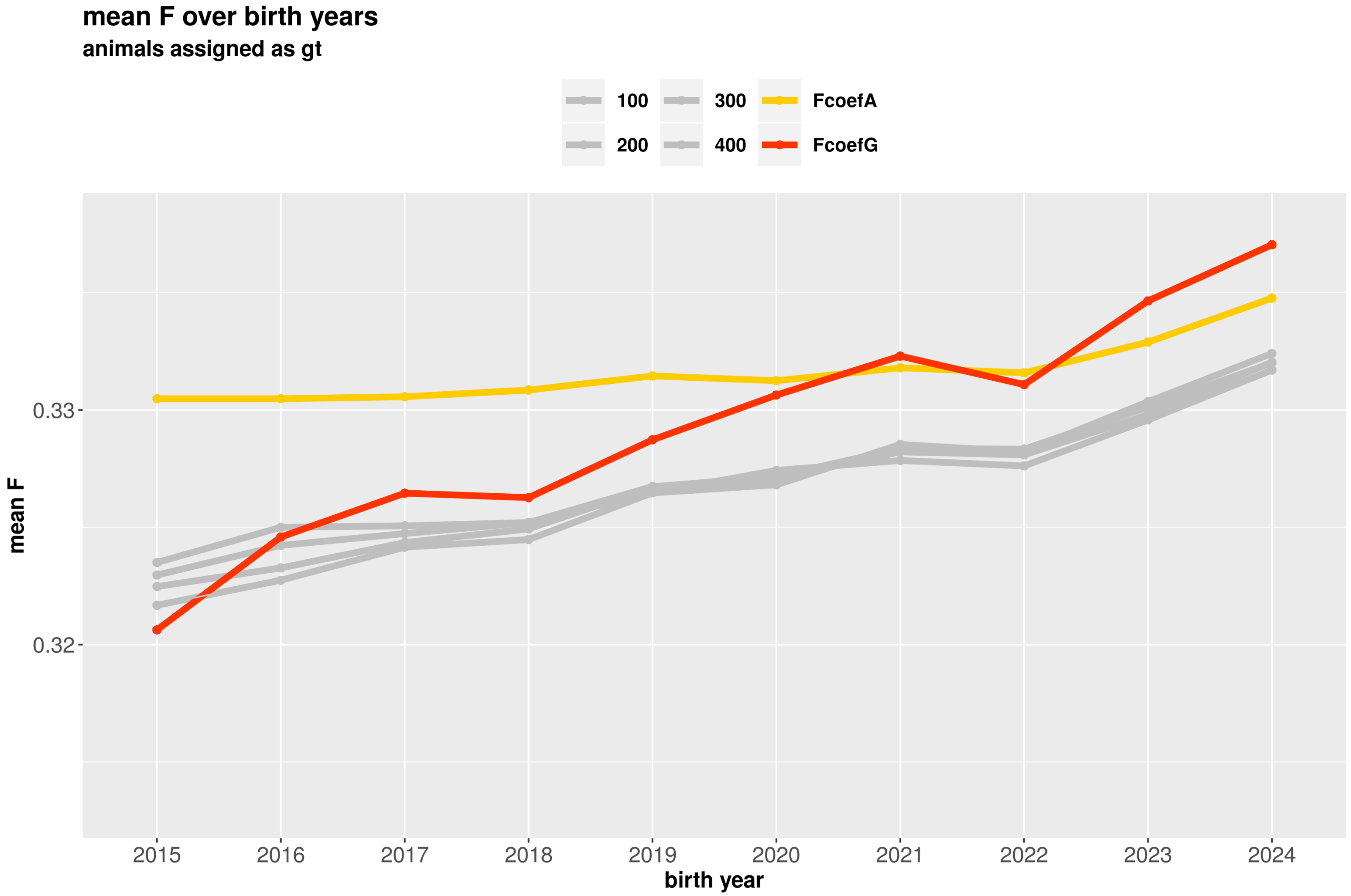
mean F over birth years
animals assigned as gt
100
200
FcoefA
FcoefG
0.33
0.32
mean F
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results

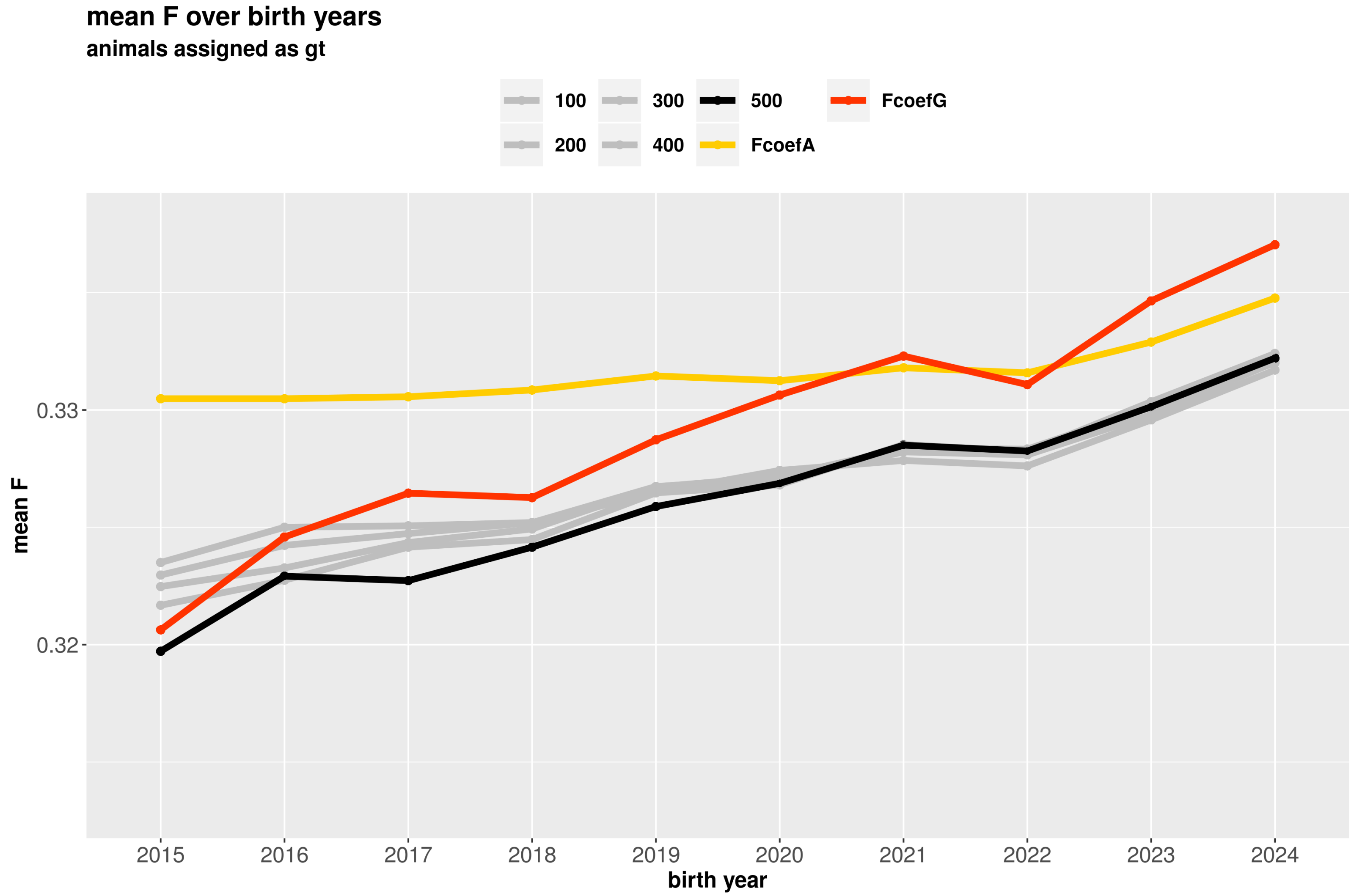


mean F over birth years
animals assigned as gt
100
300
FcoefG
200
FcoefA
0.33
0.32
mean F
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results

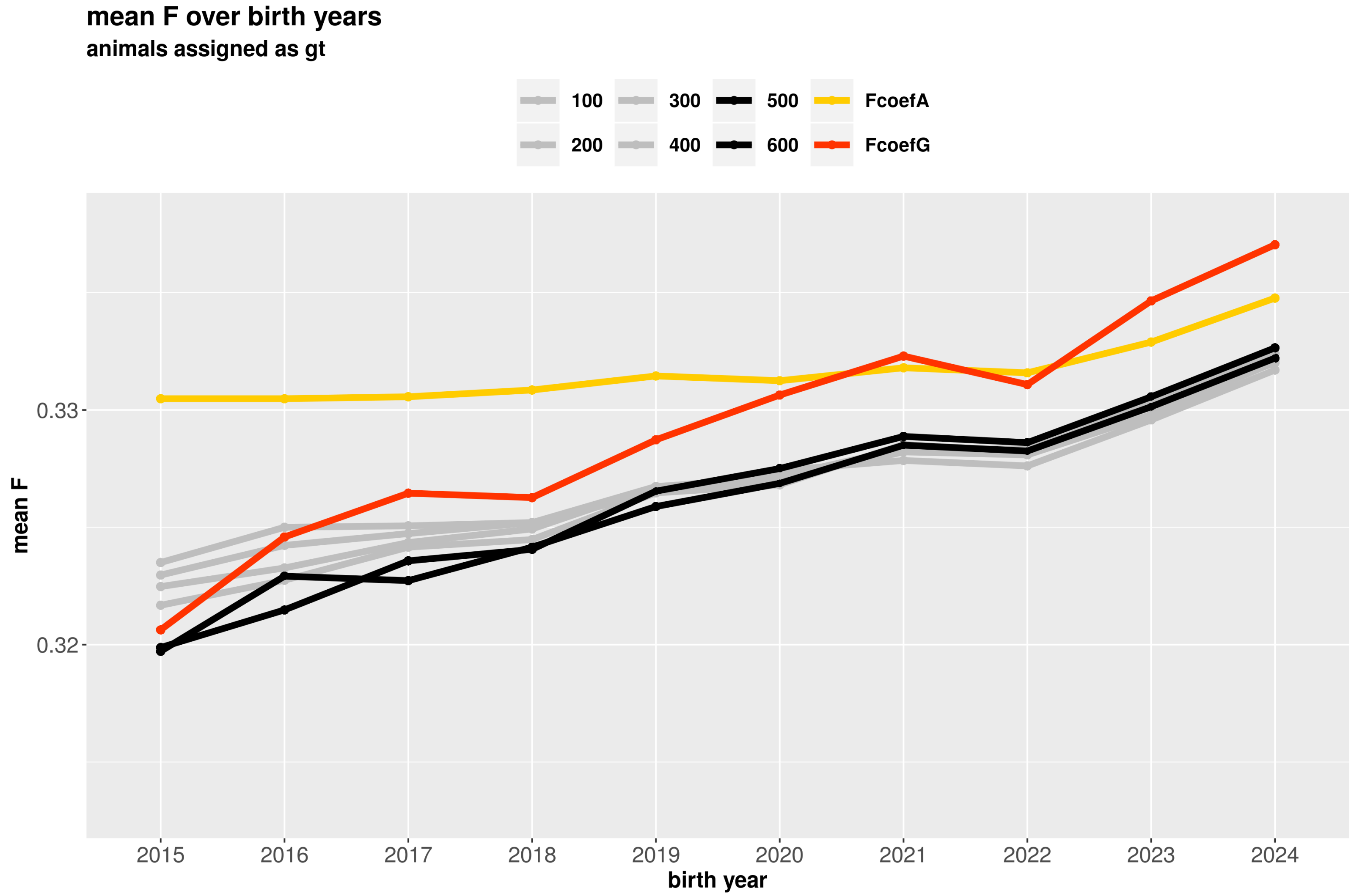

mean F over birth years
animals assigned as gt
100
200
300
400
FcoefA
FcoefG
mean F
0.33
0.32
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results

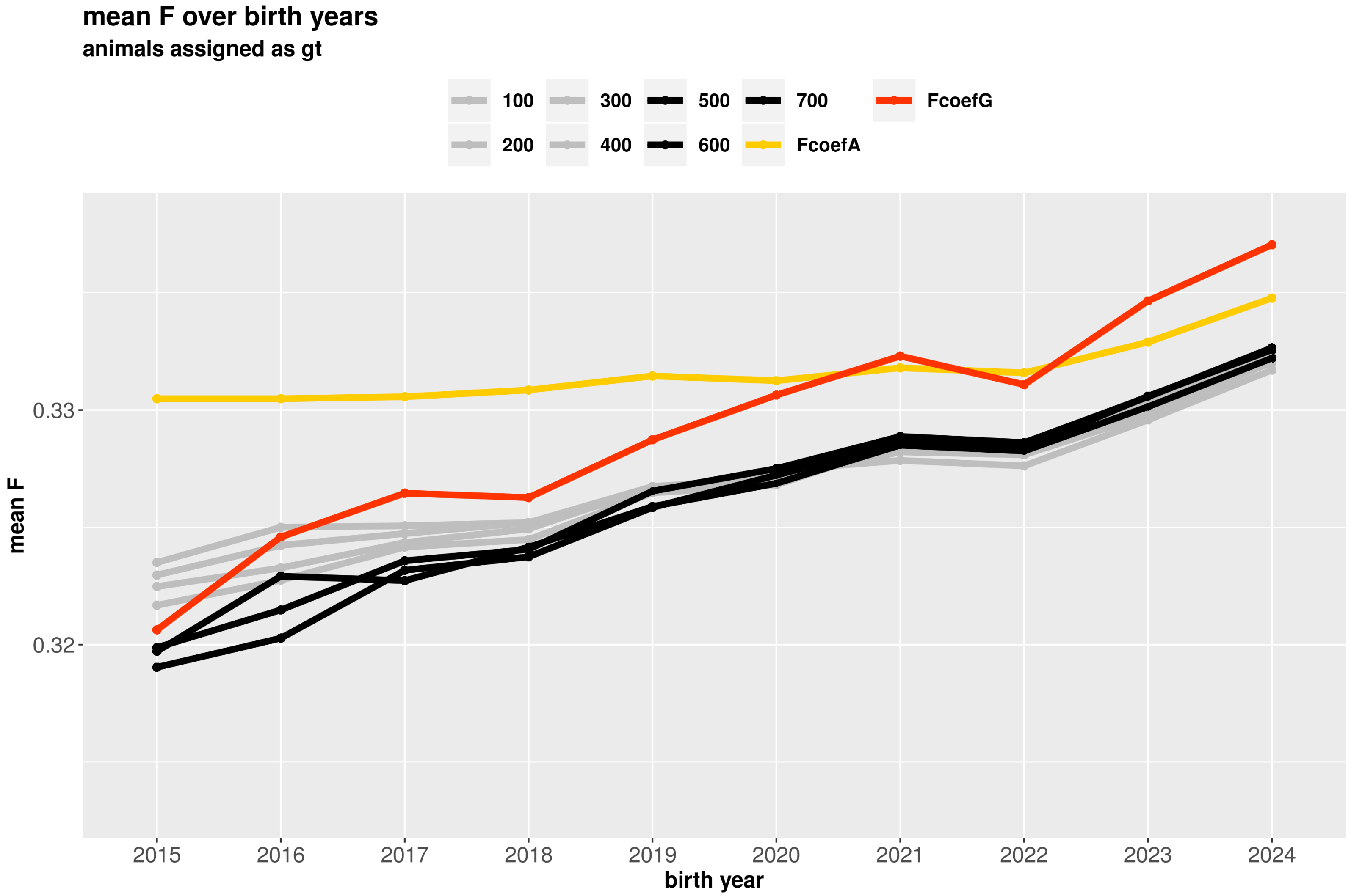


mean F over birth years
animals assigned as gt
100
300
500
FcoefG
200
400
FcoefA
0.33
0.32
mean F
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results


mean F over birth years
animals assigned as gt
100
200
300
400
500
600
FcoefA
FcoefG
mean F
0.33
0.32
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Results



mean F over birth years
animals assigned as gt
100
200
300
400
500
600
700
FcoefA
FcoefG
0.33
0.32
mean F
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
birth year

Discussion

- It is possible to define MF based on cluster-analysis of $\mathbf{H}_{\mathrm{PF,PF}}$.
- Such a clustering provides meaningful results.

Discussion

- Animals without GT
  - Clear optimum range (no. MF 500 to 800).
- Animals with GT
  - Model as many MF as possible (no. MF → no. PF).
- Even beyond the optimum number of MF derived:
  - Results (no. MF > optimum) > Results (no. MF = 1)

Discussion

- $\mathbf{H}^{\Gamma}_{\mathrm{PF,PF}}$ for no. MF = 1 → no. PF

No.MF 1 → PF

$$\mathbf{H}^{\Gamma}_{\mathrm{PF,PF}} = \underbrace{\mathbf{A}^{\Gamma}_{\mathrm{PF,PF}}}_{\text{Prior}} + \underbrace{\mathbf{A}^{\Gamma}_{\mathrm{PF,GT}}\left(\mathbf{A}^{\Gamma}_{\mathrm{GT,GT}}\right)^{-1}\left(\mathbf{G} - \mathbf{A}^{\Gamma}_{\mathrm{GT,GT}}\right)\left(\mathbf{A}^{\Gamma}_{\mathrm{GT,GT}}\right)^{-1}\mathbf{A}^{\Gamma}_{\mathrm{GT,GT}}}_{\text{Estimation}}$$

$\mathbf{G}$

Discussion

- $\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}$ for no. MF = 1 → no. PF

$$1 \xrightarrow{\text{No.MF}} \text{PF}$$

$$\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}} = \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}}_{\text{Prior}} + \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,GT}}\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\left(\mathbf{G} - \mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}}_{\text{Estimation}}$$

$\mathbf{G}$

Discussion

- $\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}$ for no. MF = 1 → no. PF

$$\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}} = \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}}_{\text{Prior}} + \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,GT}}\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\left(\mathbf{G} - \mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}}_{\text{Estimation}}$$

$\mathbf{G}$

Discussion

- $\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}$ for no. MF = 1 → no. PF

$$\mathbf{H}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}} = \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}}}_{\text{Prior}} + \underbrace{\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,GT}}\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\left(\mathbf{G} - \mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)\left(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}\right)^{-1}\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}}_{\text{Estimation}}$$

Conclusion

- It is advantageous to use MF in Single-Step:
- Why:
  - MF improve trend estimate (reduced trend bias).
  - MF improve $\Delta F_{\mathbf{A}^{\boldsymbol{\Gamma}}}$.
  - MF improve the prior part of $\mathbf{H}_{\mathrm{PF,PF}}$ $(\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{PF,PF}})$.
  - MF improve compatibility of $\mathbf{A}^{\boldsymbol{\Gamma}}_{\mathrm{GT,GT}}$ and $\mathbf{G}$.

Outlook

- Finding the optimum number of MF:
  - $\Delta F_{\mathbf{A}^{\boldsymbol{\Gamma}}} \approx \Delta F_{\mathbf{G}}$
  - Analyses on $\boldsymbol{\Gamma}$
  - ….


Thanks for
your attention!