

Single-step genomic BLUP in multiple breeding populations of Atlantic Salmon

Gebreyohans T. GebregiwerGIS, Solomon Antwi Boison, Tesfaye K. Belay, Binyam Sime Dagnachew and Theo Meuwissen



Forskningsrådet

MQWI



Nofima



**Norges miljø- og
biovitenskapelige
universitet**

Introduction

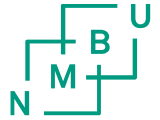
Single-step genomic BLUP (SSGBLUP):

- Combine all data from genotyped and ungenotyped animals
- Expected to yield unbiased GEBVs
- In practice, GEBVs from SSGBLUP are usually biased

(Legarra et al., 2009, Misztal et al., 2009)

Introduction...

- Possible causes of biases in GEBVs from SSGBLUP:
 - ✓ The incompatibility between $\mathbf{A}$ matrix and $\mathbf{G}$ matrix relationships
 - ✓ Disregarding or inappropriate fitting missing pedigrees in genetic evaluations



Introduction ...

- Missing pedigree are commonly modeled through:
 - ✓ Unknown parent group (UPG)
 - theory well established in PBLUP model (Quaas and Pollak, 1981)
 - Also expanded into SSGBLUP model (Miszta et al., 2013)
 - most suitable model for SSGBLUP remains unknown (Miszta et al., 2020; Masuda et al., 2022),
 - ✓ Metafounders, proxy groups representing animals in base populations (Legarra et al., 2015)
-

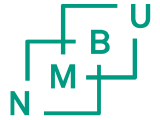
Introduction ...

➤ Among the solutions to the incompatibility problem of **A** and **G** relationships

✓ Fernando et al. (2014) proposed to fit a fixed covariate **J** with effect μ_g

$$\mathbf{J}_1 = -(\mathbf{A}^{11})^{-1} \mathbf{A}^{12} \mathbf{J}_2$$

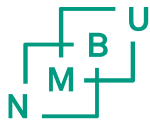
➤ Currently no or limited studies have explored the application of these solutions using aquaculture empirical data



Objectives of this study

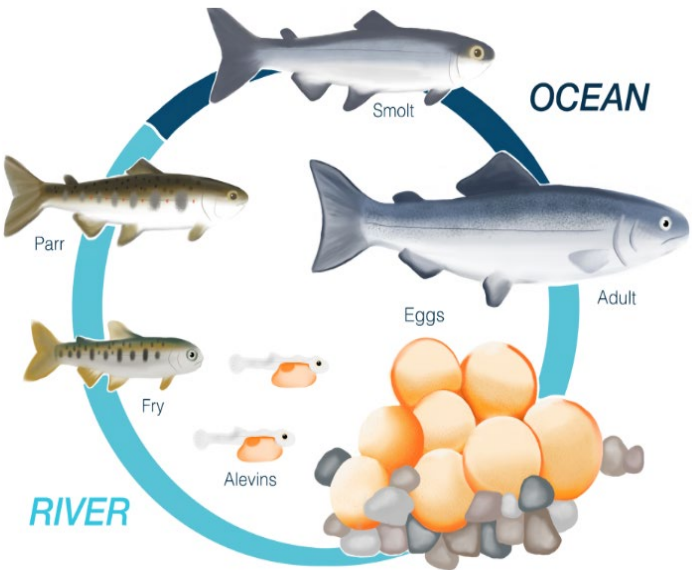
- To compare alternative methods for SSGBLUP in aquaculture breeding schemes for their genetic trend, biases and stability of the breeding values
 - To evaluate alternative approaches of fitting **J** factor with genetic groups on biases and stability of SSGBLUP evaluations in Atlantic salmon breeding with multiple population
-

Materials and Methods



Generations and population structure of Mowi broodstock

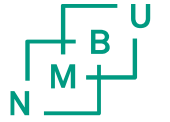
	P1	P2	P3	P4
G0	1997	1998	1999	2000
G1	2001	2002	2003	2004
G2	2005	2006	2007	2008
G3	2009	2010	2011	2012
G4	2013	2014	2015	2016
G5	2017	2018	2019	2020
G6	2021	2022	2023	2024



Life cycle of Atlantic Salmon

<https://nunatukavut.ca/kavisilik-conservation/>

MM:- Sires of smolt year 2014 (population 2)



2014 sires

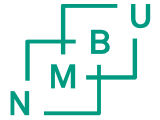
- 2008G= 6 sires, 107 individuals
- 2009G= 16 sires, 278 individuals
- 2010G= 178 sires, 3562 individuals

221 individuals with unknown sires

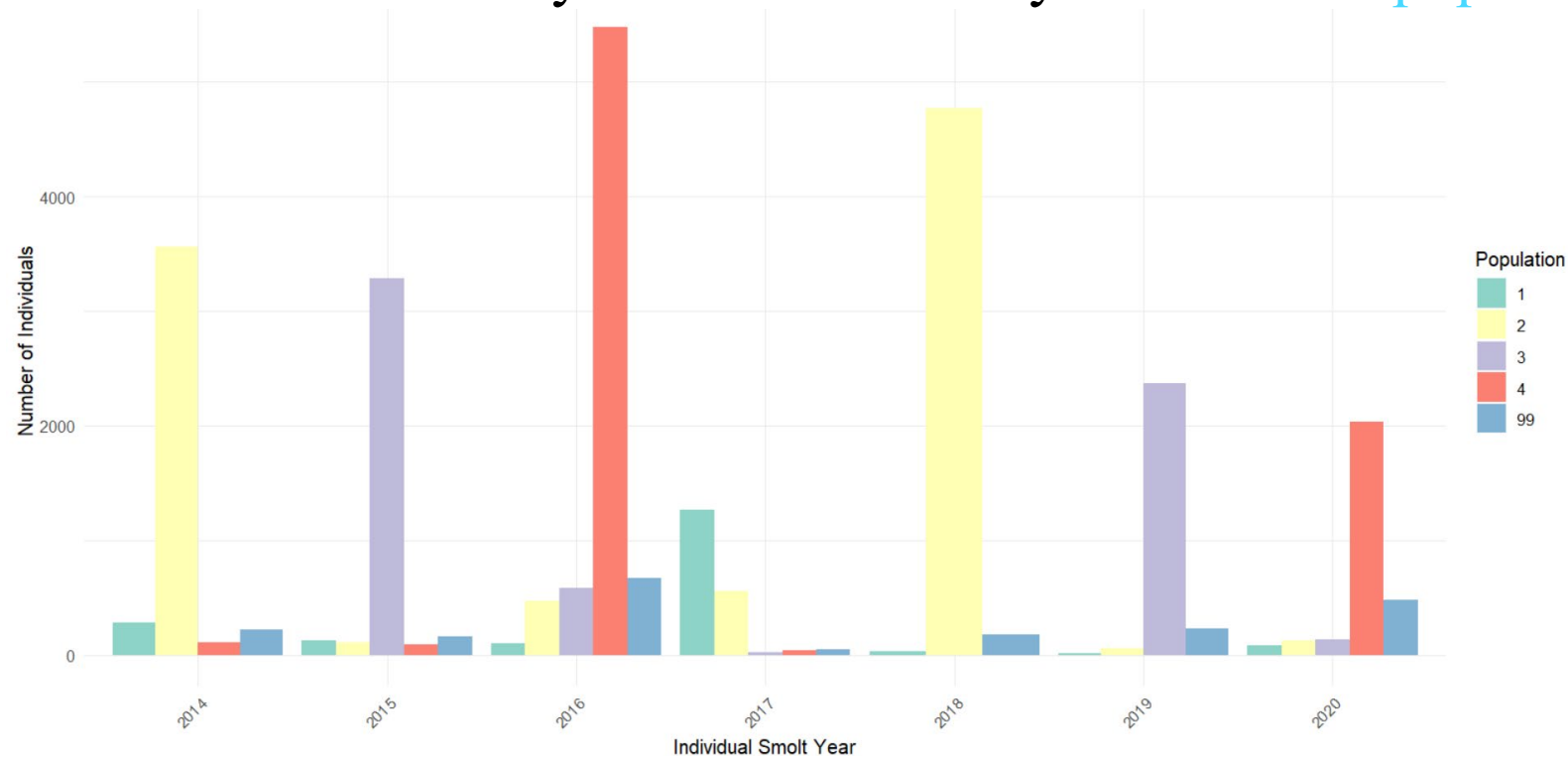
	P1	P2	P3	P4
G0	1997	1998	1999	2000
G1	2001	2002	2003	2004
G2	2005	2006	2007	2008
G3	2009	2010	2011	2012
G4	2013	2014	2015	2016
G5	2017	2018	2019	2020

There has been crossbreeding among the year classes originating from different population groups

MM:- Individuals smolt years vs their sires population



Number of individuals by individual smolt years and **sire population**

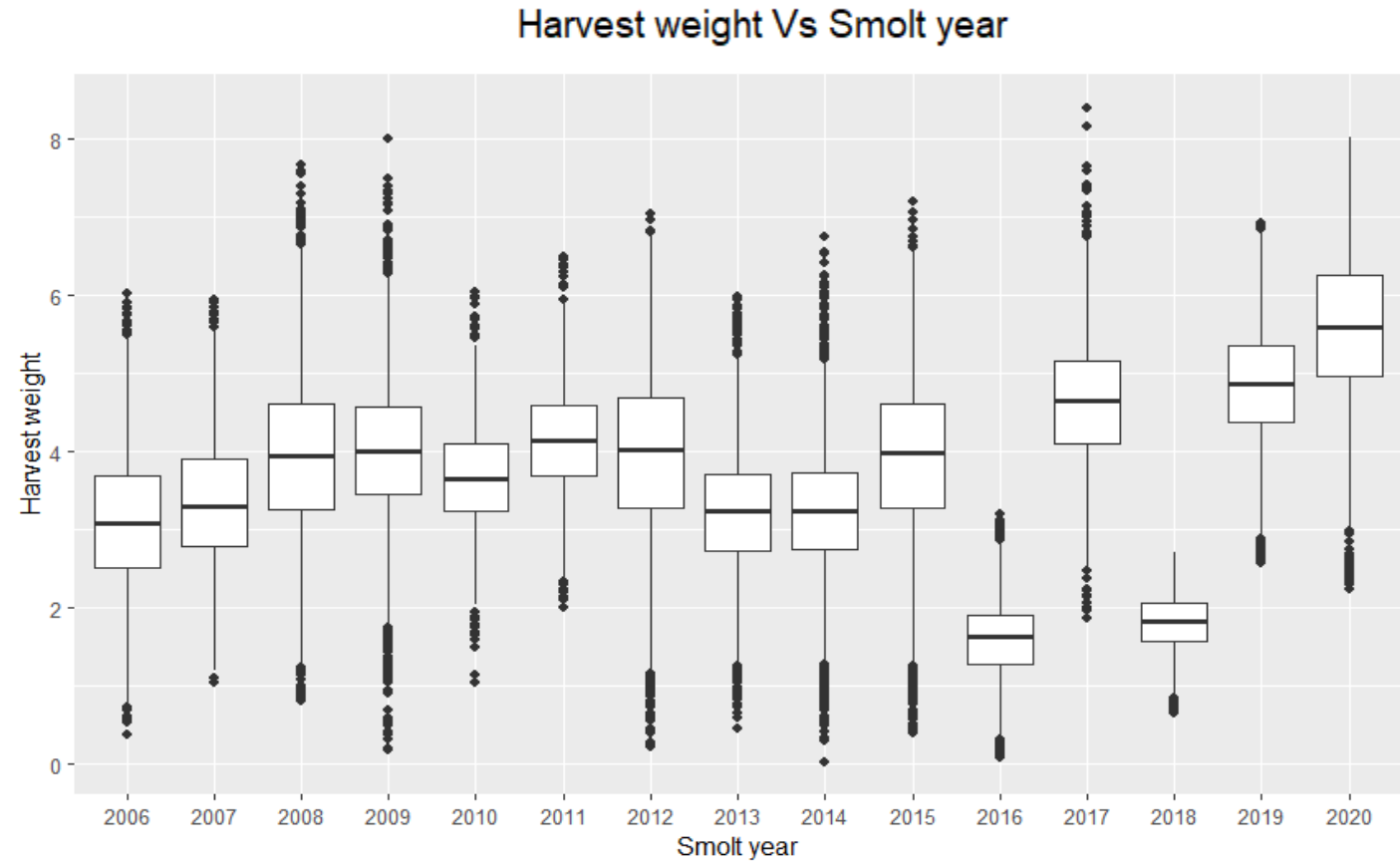


The populations are not completely genetically separated

MM:- Phenotype data

- Harvest weight > 2005
- 475,96 records

Descriptive statistics	Values (in kg)
Mean	3.365
SD	1.41
Min	0.030
Max	8.4



MM:- Pedigree and genotype data

➤ Pedigree

- ✓ 61927 (1996- 2020)

➤ Genomic data

- ✓ Two different SNP chips were used for genotyping
 - $\approx$ 55K SNPs genotyping array (NOFSAL03, Affymetrix axiom array)
 - $\approx$ 66K SNPs genotyping array (SALMOW1, Affymetrix axiom array)
- ✓ Imputed to the $\approx$ 70K SNPs chip using FImpute v2.2

(Sargolzaei, Chesnais, & Schenkel, 2014)

MM:- Phenotype and genotype data



Smolt years with genotypes and/or Phenotypes data

	P1	P2	P3	P4	
G0	1997	1998	1999	2000	→ No phenotype
G1	2001	2002	2003	2004	
G2	2005	2006	2007	2008	→ phenotype
G3	2009	2010	2011	2012	
G4	2013	2014	2015	2016	→ Genotype + phenotype
G5	2017	2018	2019	2020	

MM:- Creating groups

➤ Four groups

- ✓ 4 population

➤ Twenty groups

- ✓ Smolt year having 50 individuals or more with unknow parents

➤ Ten groups

- ✓ 4 population

- ✓ 2016 (4 groups using PCA)

- ✓ 2015 (2 groups using PCA)
-

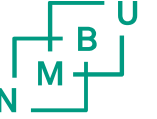
MM:- SSGBLUP Models

$Y = Xb + Zu + Z(m) + e$, where **m** is model options

#	Model name	Description
1	SSGBLUP	A model without genetic group and J
2	SSGBLUP G A*	J was fitted as random variable via G modification
3	SSGBLUP J	J was fitted as a fixed covariate
4	SSGBLUP Q	Q -group was fitted as a fixed covariate
5	SSGBLUP QJ	Q and J fitted in the model as fixed covariates
6	SSGBLUP QGA	Q was fitted as fixed covariates with random J

PBLUP/PBLUPQ

***G** modification= $\mathbf{G} + \mathbf{1}\mathbf{1}'\alpha$ (Vitezica et al., 2011)



MM:- Model performance evaluation

- 5 fold cross-validation used
 - ✓ In each cross-validation approximately 2400 phenotypes randomly masked
 - ✓ In total 11996 phenotypes masked
-

MM:- Validation parameter (Legarra and Reverter regression)

➤ Level-bias

- ✓ The difference in means between the $GEBV_{CV}$ and $GEBV_{all}$ scaled by the genetic standard deviation of the trait

➤ Inflation

- ✓ The regression of $GEBV_{all}$ on $GEBV_{CV}$

$$b_{cv} = \text{cov}(GEBV_{all}, GEBV_{CV}) / \text{var}(GEBV_{CV})$$

➤ Ratio of accuracies

- ✓ The correlation between $GEBV_{CV}$ and $GEBV_{all}$

(Legarra and Reverter, 2018)

Results :- Evaluation of the different prediction models without groups

Models	Inflation	Bias	Ratio of accuracies
SSGBLUP _b	1.006	0.001	0.960
SSGBLUP ^{GA}	1.008	0.008	0.959
SSGBLUP ^J	1.008	0.010	0.958
PBLUP	0.987	0.034	0.837

All models were without groups

Results - Level bias

Model	Number of UPG			
	0	4	10	20
SSGBLUP _b	0.001	0.002	0.007	0.006
SSGBLUPGA	0.008	0.012	0.017	0.023
SSGBLUPJ	0.010	0.015	0.020	0.039
PBLUP	0.034	0.046	0.031	0.008

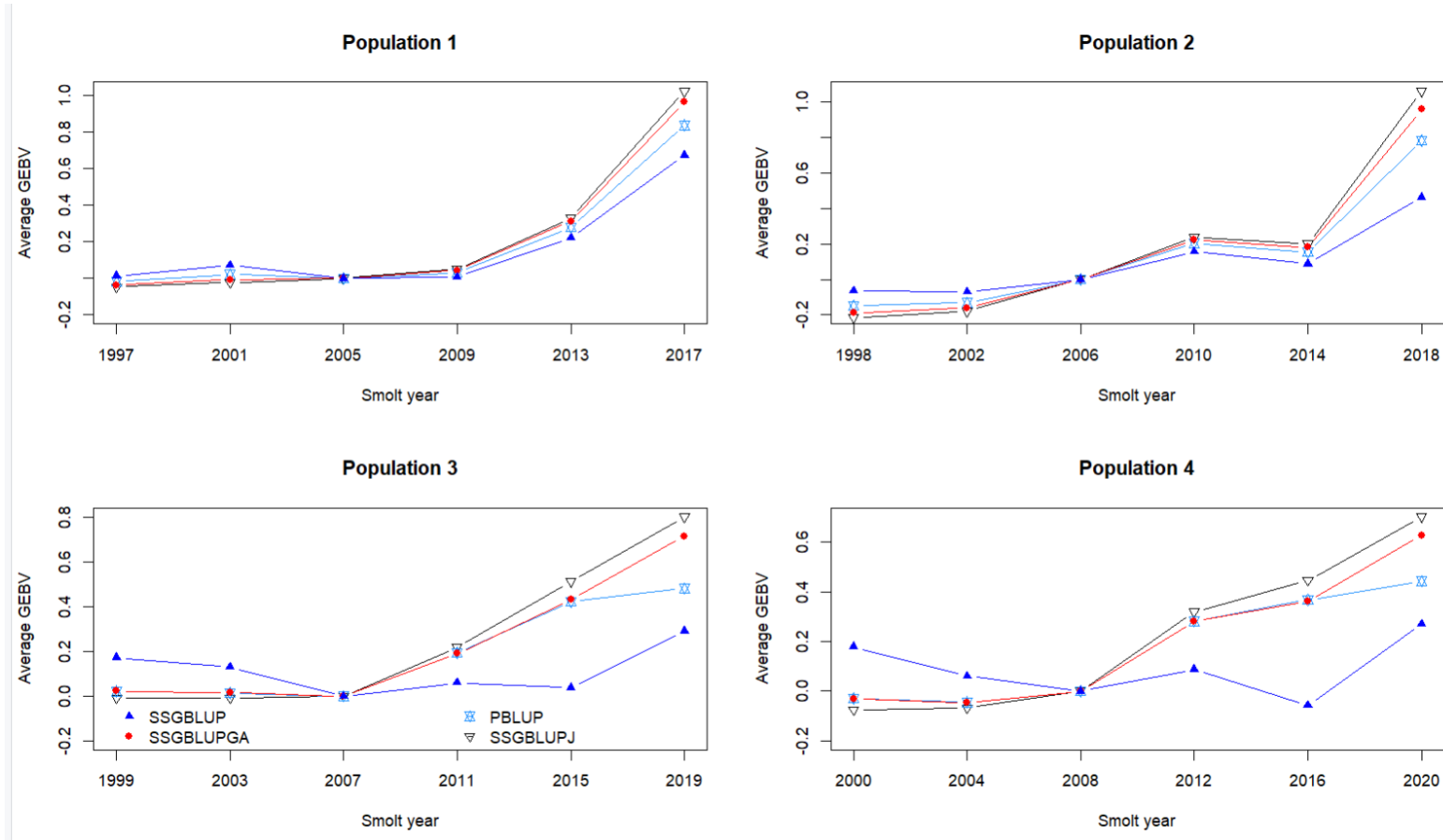
Results – Ratio of accuracies

Model	Number of UPG			
	0	4	10	20
SSGBLUP _b	0.960	0.958	0.957	0.955
SSGBLUPGA	0.959	0.960	0.959	0.958
SSGBLUPJ	0.958	0.961	0.959	0.960
PBLUP	0.837	0.843	0.840	0.840

Results – Inflation

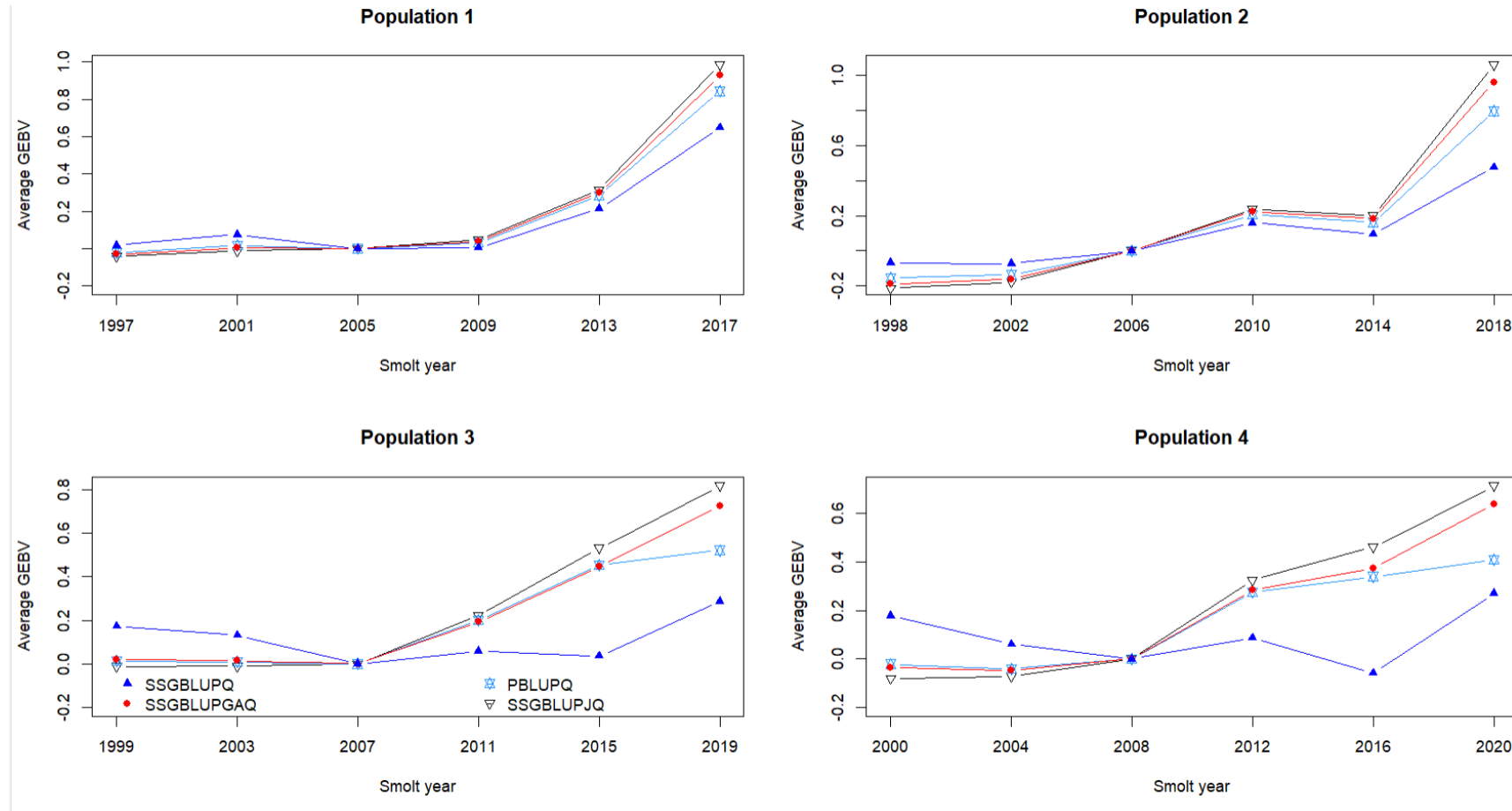
Model	Number of UPG			
	0	4	10	20
SSGBLUP _b	1.006	1.004	1.005	1.006
SSGBLUPGA	1.008	1.005	1.006	1.005
SSGBLUPJ	1.008	1.005	1.005	1.004
PBLUP	0.987	0.980	0.980	1.005

Results :- Genetic trend plot across smolt years



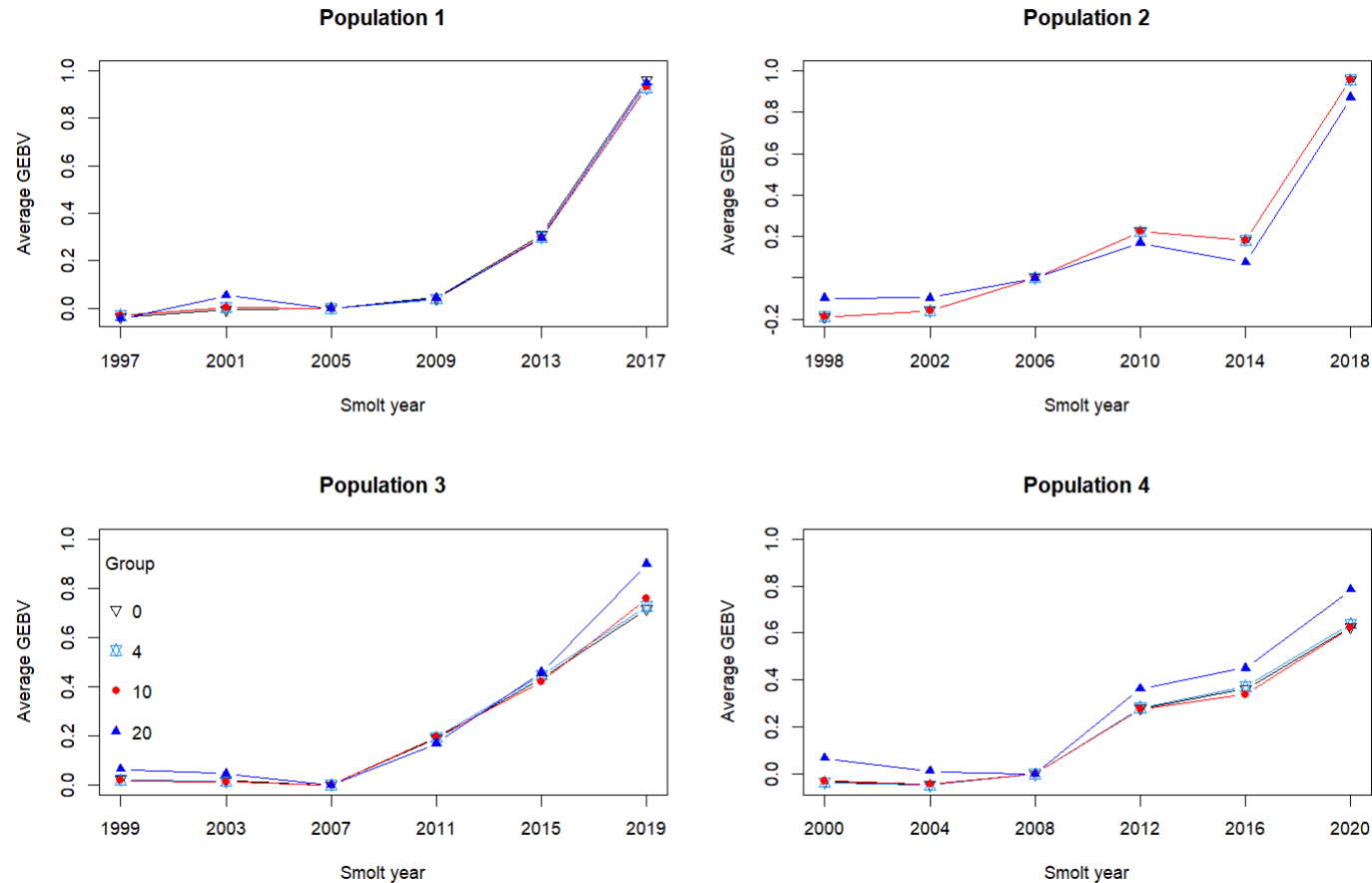
Estimate of genetic trends across smolt years for different SSGBLUP models **without UPG**

Results :- Genetic trend plot across smolt years



Estimate of genetic trends across smolt years for different SSGBLUP models with 4 UPG

Results :- Genetic trend plot across smolt years



SSGBLUPGA prediction method and with different groups

Take home messages

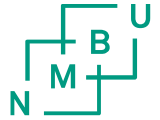


- Incorporating **J** factors into SSGBLUP models, whether as random or fixed covariates
 - ✓ Enables the estimation of genetic trends across generations
 - ✓ A marginal improvement in ratio of accuracies and maintains low inflation biases
 - ✓ Slightly increased the level-bias
- Implicit incorporation of the **J** factor as a random covariate (Vitezica et al., 2011) reduced level biases when compared to fitting **J** factor as fixed covariate effects

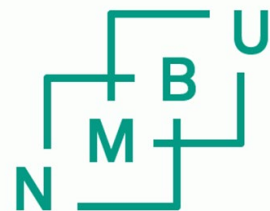
Take home messages

- Genetic groups
 - ✓ Inflation biases slightly lower with group than with out group
 - ✓ Highest ratio of accuracies was obtained with 4 genetic groups in models with **J** factor
 - ✓ level-biases tend to increase as the number of genetic groups increased

Acknowledgements

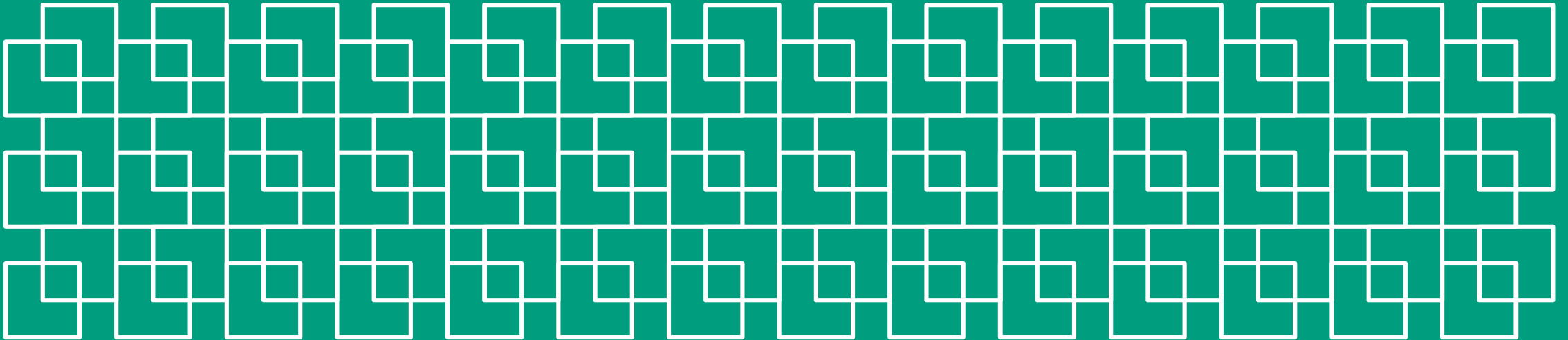


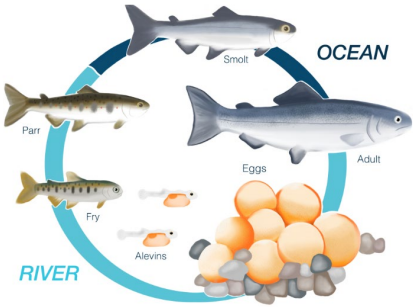
- ✓ All co-authors
- ✓ The Norwegian Research Council (project no 321539) for financial support



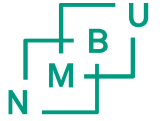
Norges miljø- og
biovitenskapelige
universitet

Thank you





Summary



Available data

- Harvest weight
- Multi-generation
- Multi-population
- Genotype (yes/no)
- Parents information (yes/no)

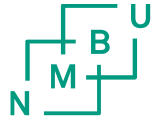
Prediction models

- SSGBLUP
- SSGBLUP + **J** factor
- SSGBLUP + groups
- SSGBLUP + **J** factor+groups

Evaluation criteria

- Inflations
- Biases
- Accuracies
- Genetic trends

Which prediction models perform better???



MM:- Phenotype and genotype data

Summary of the number of records with both phenotype and genotype

Population	Phenotype	Genotype (smolt year)	Both
1	8770	1866 (2017)	1866
2	12549	4845 (2018)	4845
3	11568	6317 (2015+2019)	6079
4	14709	9871 (2016+2020)	9543
Total	47596	22899	22333