



# Unknown parent groups and metafounders in genomic evaluation of Norwegian Red cattle

Tesfaye K. Belay<sup>1</sup>, A.Gjuvsland<sup>2</sup>, J.Jenko<sup>2</sup>, S.Eikje<sup>2</sup>, and T.Meuwissen<sup>1</sup>

<sup>1</sup>Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Norway

<sup>2</sup>Geno SA, Storhamargata 44, 2317 Hamar, Norway

The 74<sup>th</sup> EAAP Annual Meeting, 31<sup>st</sup> August 2023, Lyon, France



# Introduction

- Single-step genomic BLUP (ssGBLUP):
  - combines all data from genotyped and ungenotyped animals.
  - requires compatibility between **A** and **G** matrices for unbiased predictions.
    - ✓ i.e.,  $\mathbf{H}^{-1}$  used in ssGBLUP was developed assuming **a complete pedigree**.

(Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010)
- However, pedigree information in dairy cattle is often **incomplete**.
- The missing pedigree causes:
  - biased prediction due to selection,
  - missing inbreeding in **A** matrix, and
  - incompatibility between **G** and **A** matrices (Masuda et al., 2022).



## Introduction ...

- These missing pedigree and compatibility issues in ssGBLUP can be solved by:
  - “**J factor**” and other scaling parameters (Fernando et al., 2014; Vitezica et al., 2011)
    - J-factor estimates base-population differences between **A** and **G** matrices
  - Unknow parent groups (UPG) or genetic groups
  - Combination of UPG with **J factor**
  - Metafounders (MF)
- Belay et al.(2022) tested alternative ways of combining UPG with the **J factor**
  - Single **J factor** was extended to multiple per group **J factors** (“**Q<sup>+</sup>**»)
    - **Q** = genetic group contributions
  - A model fitting the difference between original (**Q**) and modified (**Q<sup>+</sup>**) group contributions (denoted as “**Q-Q<sup>+</sup>**”) showed the least bias and the highest stable genomic prediction.



## Introduction ...

- The  $Q^+$  matrix:

- original group contribution matrix of genotyped ids (2)

- predicted group contributions for ungenotyped (1)

- Computed in the same manner as  $J$  factor:  $J = \begin{bmatrix} J_1 \\ J_2 \end{bmatrix} = \begin{bmatrix} A_{12} A_{22}^{-1} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$

$$-Q^+ = \begin{bmatrix} Q_1^+ \\ Q_2^+ \end{bmatrix} = \begin{bmatrix} A_{12} A_{22}^{-1} Q_2 \\ Q_2 \end{bmatrix}$$

*Journal of Animal Science*, 2022, **100**, 1–13

<https://doi.org/10.1093/jas/skac227>

Advance access publication 25 June 2022

**Animal Genetics and Genomics**



## Correcting for base-population differences and unknown parent groups in single-step genomic predictions of Norwegian Red cattle

Tesfaye K. Belay,<sup>†,1</sup> Leiv S. Eikje,<sup>\*</sup> Arne B. Gjuvsland,<sup>\*</sup> Øyvind Nordbø,<sup>‡</sup> Thierry Tribout,<sup>||</sup> and Theo Meuwissen<sup>†</sup>



## Introduction ...

- The Q-Q+ model's performance on **national datasets is unknown** and
- Its performance in **comparison to existing models** for genetic evaluation in NRF cattle is yet to be determined.
  
- The MF approach (Legarra et al., 2015) is considered a comprehensive solution to missing pedigree issues (Masuda et al., 2022)
  - requires relationships between MF's
    - estimated from genotype data
  - However, MF's performance in NRF cattle is unknown.**
  
- Alternative: fit **UPG with relationships**
  - estimate relationships from pedigree (VanRaden, 1992, Aguilar and Misztal. 2008)



## Aims of this study

- to evaluate the performances of the Q-Q+, MF, and UPG (with relationships) models when applied to national datasets routinely used for the genomic evaluation of NRF cattle.
  
- to compare these models with each other and with a “routine model” currently used in the genomic evaluation of NRF cattle



## MODELS:

### 1) Routine (rtn) model

–52 UPG included in the pedigree

–Fitted as fixed effects

–**A** weighted into **G** with 10% weight (  $w=0.1$  )

–Allele frequency (AF) of 0.5 used in **G** construction

$$\bullet \mathbf{H}^{-1} = \mathbf{A}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{0.5}^{-1} - \mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

• The  $\mathbf{A}^*$  is  $\mathbf{A}^{-1}$  matrix including random UPG effects.



## 2) MF model

- NRF70, NRF80, NRF90, NRF00, NRF10, NRF20 (6 MF)
  - Norwegian Red animals born 1970 or earlier, 1971-1980, 1981-1990, 1991-2000, 2001-2010 and 2011-2020
- SRB, FAY, RDM, SLB, HOL, BSW, SIM, JER (8 MF)
  - Metafounders for other breeds
- Totally, 14 MF fitted, and **gamma** ( $\Gamma$ ) computed using Bpop (Stranden, 2020)

$$\Gamma = 8 * \text{cov}(\text{base AF}) \text{ (Christensen et al. (2015))}$$

$$\bullet \mathbf{H}_{\Gamma}^{-1} = \mathbf{A}_{\Gamma}^* + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{0.5}^{-1} - \mathbf{A}_{\Gamma 22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$



### 3) Q-Q+ models

- Q-Q+ matrix is fitted together with

- $\mathbf{G}_{\text{cal}}$ , which is **not weighted by  $\mathbf{A}$** , and **calculated AF** used

- Q-Q+

- $\mathbf{G}_{0.5}$ , which is **not weighted by  $\mathbf{A}$**  and used **AF = 0.5**

- Q-Q+ $_{\mathbf{G}_{0.5}}$

- $\mathbf{G}_{0.5w}$ , both **weighted by 10%  $\mathbf{A}$**  and used **AF = 0.5** (as in 'routine')

- Q-Q+ $_{\mathbf{G}_{\text{rtn}}}$

- $\mathbf{H}_Q^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_*^{-1} - \mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$ , where  $\mathbf{G}_*^{-1}$  is either  $\text{inv}(\mathbf{G}_{\text{cal}}, \mathbf{G}_{0.5}, \text{ or } \mathbf{G}_{0.5w})$



## 4) UPG models with relationships among groups

- Pedigree relationships among the 52 groups ( $\Sigma$ ):
  - computed using Aguilar and Misztal's (2008) algorithm.
- Inbreeding(F) with  $\Sigma$  computed using Meuwissen & Luo's (1992) algorithm
- $\mathbf{A}_{22}$  with  $\Sigma$  ( $\mathbf{A}_{\Sigma 22}$ ) was computed using a Julia program
- Different  $\mathbf{G}$  was used:
  - ✓ Weighted by 10%  $\mathbf{A}_{22}$  or not
  - ✓ 'aligned' to  $\mathbf{A}_{\Sigma 22}$  with  $\alpha$  (Vitezica et al., 2011) or not.

$$\bullet \mathbf{H}_{\Sigma}^{-1} = \mathbf{A}_{\Sigma}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_*^{-1} - \mathbf{A}_{\Sigma 22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \mathbf{A}_{\Sigma}^{-1} = \mathbf{A}^{-1} \text{ with group relationships } (\Sigma)$$



## Model acronyms

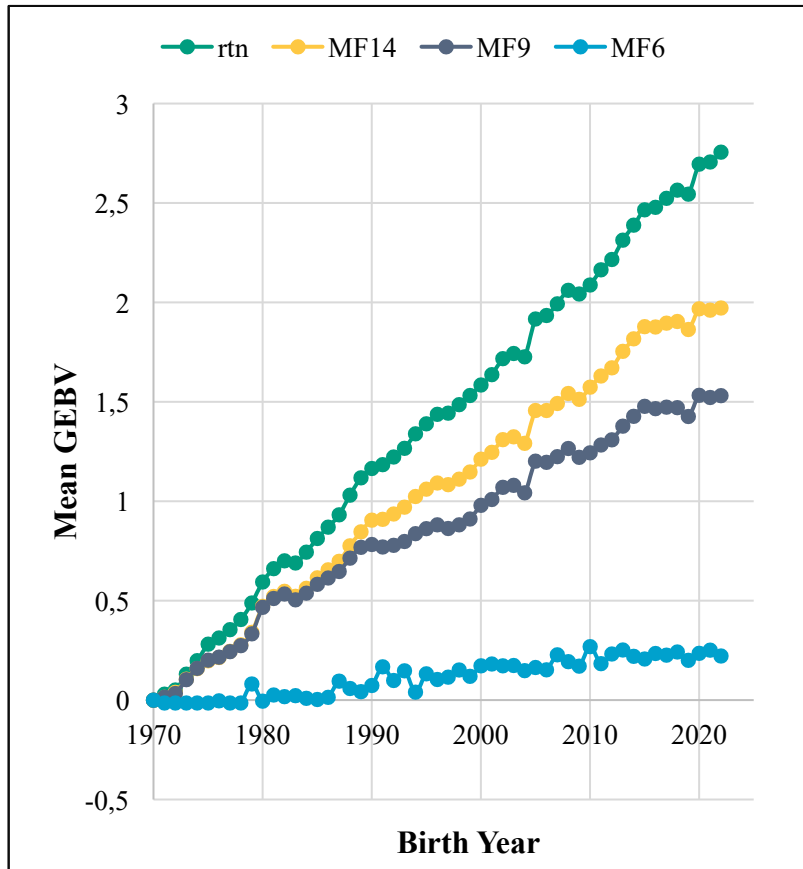
- $y = Xb + Wh + Za + ZQg + e$

Model type	Acronym	G matrix
Routine	rtn	AF=0.5 & w=10% of $A_{22}$
Metafounders	MF14	AF=0.5 & w=10% of $A_{22}$
<b>Combined UPG &amp; J</b>		
	Q-Q+	Calculated AF and w=0
	Q-Q <sup>+</sup> _G <sub>0.5</sub>	AF=0.5 and w=0
	Q-Q <sup>+</sup> _G <sub>rtn</sub>	AF=0.5 & w=10% of $A_{22}$
<b>UPG with <math>\Sigma</math></b>		
	UPG_G <sub>anw</sub>	Aligned but not weighted; AF=cal
	UPG_G <sub>aw</sub>	Both aligned and weighted; AF=cal
	UPG_G <sub>nanw</sub>	Neither aligned nor weighted; AF=cal
	UPG_G <sub>naw</sub>	Not aligned but weighted; AF=cal
	UPG_G <sub>rtn</sub>	AF=0.5 & w=10% of $A_{22}$

## Data

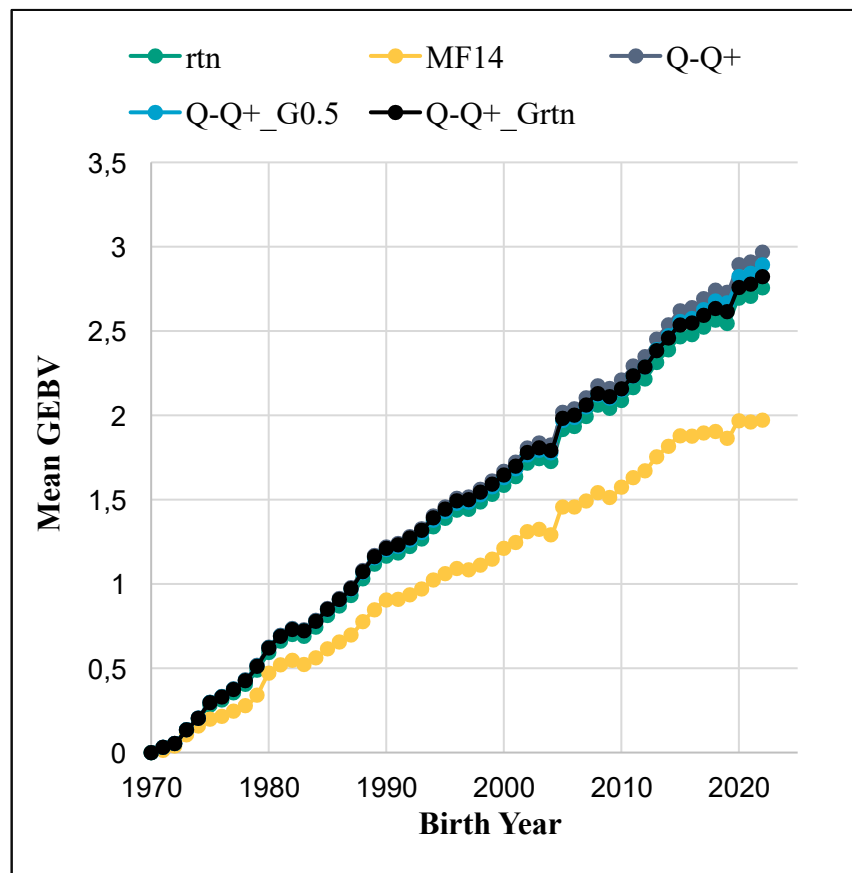
- Milk model from routine evaluations
  - 305d, single trait-repeatability-animal model
  - 8.4 million records
  - 4.9 million animals in the pedigree
  - 170,293 genotyped animals (121,740 SNPS)
- Validation
  - Masked phenotype of 5000 youngest genotyped cows
    - Create a reduced dataset
  - GEBV from the reduced dataset dataset
  - GEBV from the full dataset (Legarra & Reverter method)
- MiX99 (Lidauer et al., 2019) was used for prediction
- In all models,  $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$  computed using HGinv (Stranden and Mantysaari, 2018)

# Genetic trends – routine vs MF models



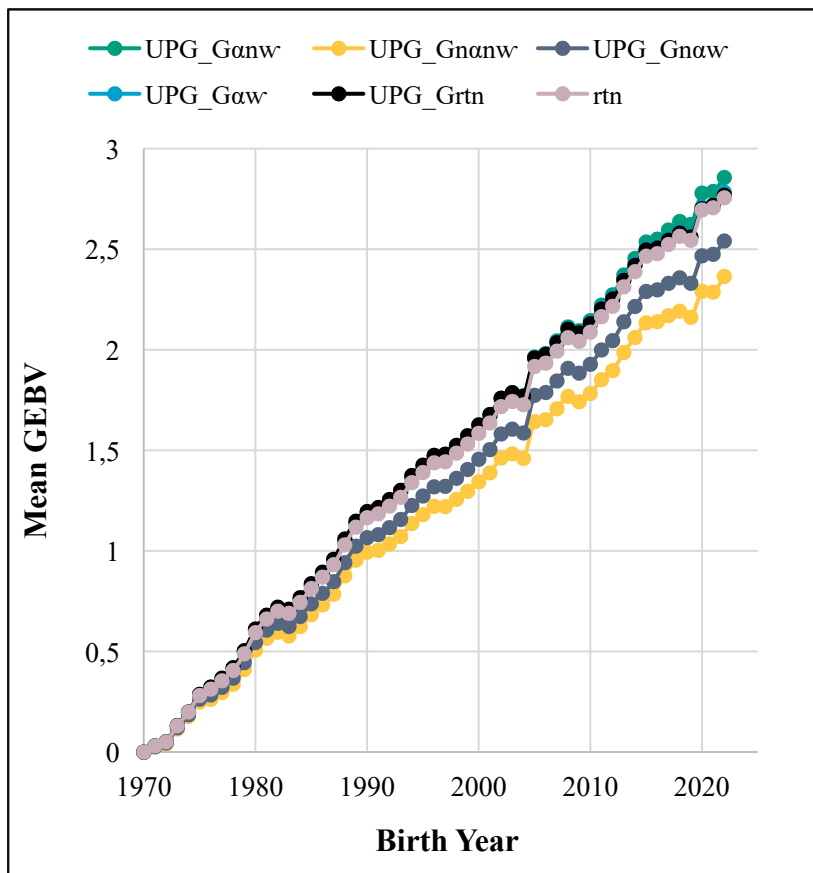
- In MF9:
  - Smaller breeds combined
  - Some time periods merged for NRF
- In MF6:
  - Grouping based only on breed
  - Smaller breeds combined
- MF models:
  - Lowest genetic trend
  - Flat in recent years
- Genetic trends reduced as MF definition got cruder.

## Genetic trends – Q-Q+ models



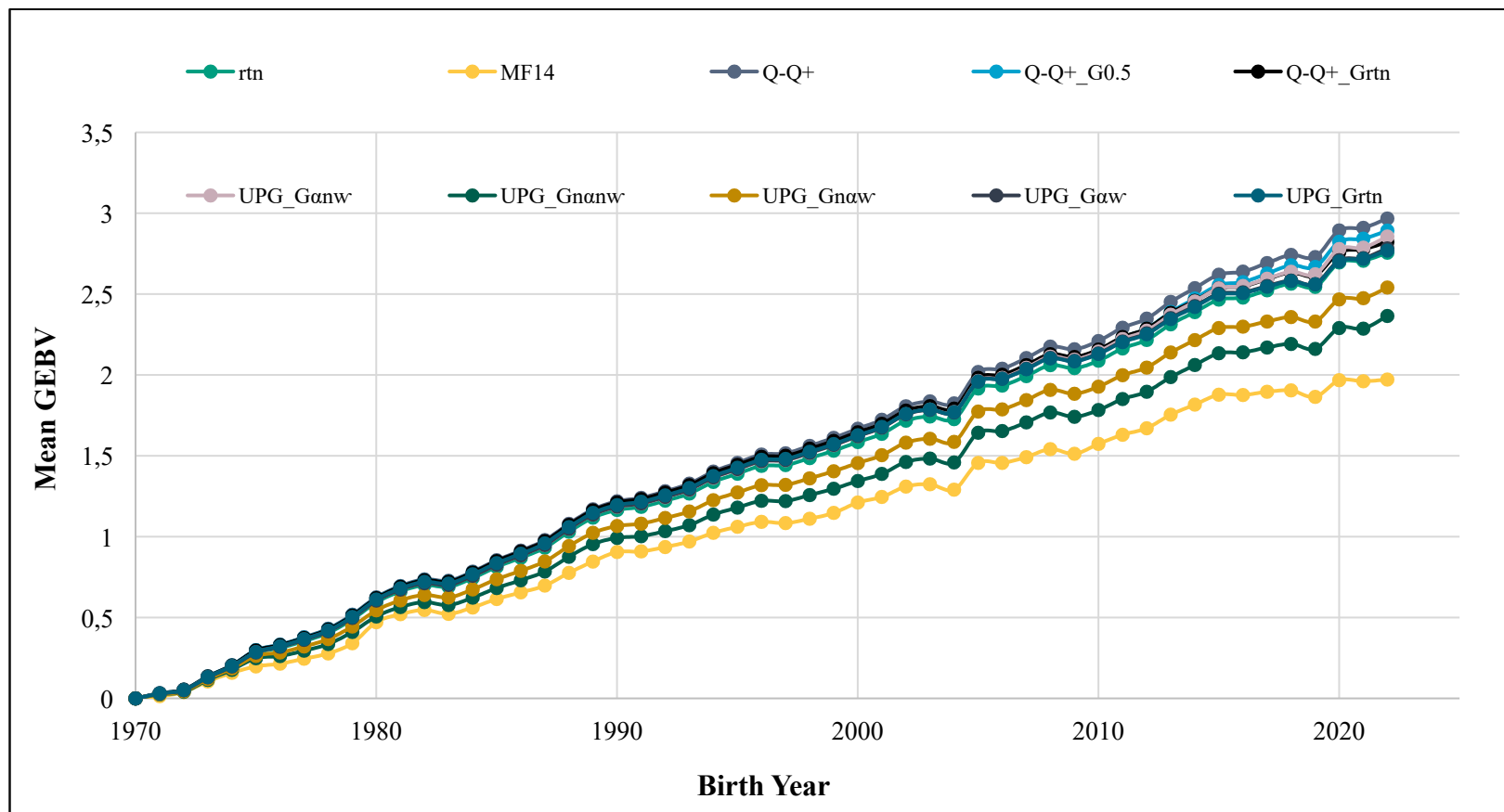
- Q-Q+ model resulted in a **higher trend**.
- Gap between 'routine' and Q-Q+ increased especially from 2016.
- Fitting combined effects of UPG and **J** with  $G_{0.5}$  or  $G_{rtn}$  reduced the gap.

## Genetic trends – UPG models



- Scaling  $\mathbf{G}$  by  $\alpha$  resulted in a higher trend (e.g., UPG\_ $G_{\alpha nwr}$  & UPG\_ $G_{\alpha w}$ ).
- Weighting  $\mathbf{G}$  by 10% A22 reduced genetic trends.
- Not scaling and weighting  $\mathbf{G}$  resulted in the lowest trend (UPG\_ $G_{n\alpha nwr}$ ).
- Trends for UPG\_ $G_{rtn}$  and routine (rtn) mostly overlap.

# Genetic trends – all models



## Model evaluation – bias, inflation, and stability

Model	Level-bias	Inflation	Correlation
Routine(rtn)	-0.0260	1.0055	0.9799
MF14	-0.0237	1.0282	0.9764
Q-Q+	<b>-0.0129</b>	0.9931	0.9825
Q-Q <sup>+</sup> _G <sub>0.5</sub>	<b>-0.0127</b>	<b>0.9995</b>	<b>0.9893</b>
Q-Q <sup>+</sup> _G <sub>rtn</sub>	-0.0244	1.0065	0.9798
UPG_G <sub>anw</sub>	<b>-0.0174</b>	0.9937	<b>0.9821</b>
UPG_G <sub>aw</sub>	-0.0269	0.9982	0.9724
UPG_G <sub>nanw</sub>	-0.0212	0.9946	0.9813
UPG_G <sub>naw</sub>	-0.0235	0.9944	0.9714
UPG_G <sub>rtn</sub>	-0.0263	1.0067	0.9796



## Summary

- Type of **allele frequency** and **weight** used in **G** affected the performance of the Q-Q<sup>+</sup> models:
  - When **G** was not weighted and AF=0.5 or calculated, Q-Q<sup>+</sup> models resulted in the
    - highest genetic trend (e.g., Q-Q<sup>+</sup>)
    - least bias and inflation (Q-Q<sup>+</sup>\_G<sub>0.5</sub>)
    - most stable genomic predictions (Q-Q<sup>+</sup>\_G<sub>0.5</sub>)
  - weighting **A** into **G** resulted in more biased genomic predictions.
  - Therefore, the use of weighted **G** in Q-Q<sup>+</sup> models is not recommended.



## Summary ...

- MF yielded biased prediction and unrealistic genetic trend
  - MF grouping (14 groups) is probably too crude
- Performances of UPG models with  $\Sigma$  were affected by  $\alpha$ , AF, and  $w$ .
  - Scaling is important but not weighting
- Generally, most models performed similarly to the routine method.
  - Except Q-Q<sup>+</sup>, Q-Q<sup>+</sup>\_G<sub>0.5</sub>, and UPG\_G<sub>anw</sub> that performed best.
- The Q-Q<sup>+</sup>\_G<sub>0.5</sub> model is recommended as it performed better across all parameters evaluated.



## Acknowledgments



Norges miljø- og  
biovitenskapelige  
universitet



The Research Council  
of Norway

For financial support through project # 309611

Thank you for your attention!!