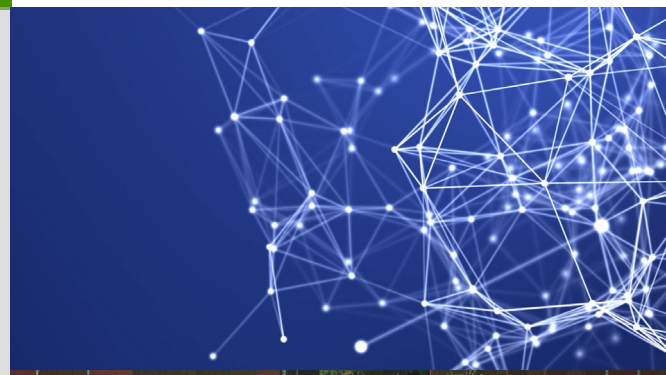
Faculty of Agricultural and Nutritional Science

Christian-Albrechts-University Kiel
Institute of Animal Breeding and Husbandry

Use of machine learning methods to predict feed efficiency traits in dairy cows using phenotypic and genomic information



A. Seidel[1], L. Stahmer[1] and G. Thaller[1]

[1] Institute of Animal Breeding and Husbandry, Kiel University

75th Annual EAAP Meeting Florence, Italy
September 1st – 5th, 2024
Session 43, Abstract number 2216204, aseidel@tierzucht.uni-kiel.de

Background

- By optimizing feed efficiency, farmers can achieve higher milk production with the same or even lower amount of feed

  → Plays an important role in the profitability and sustainability of dairy farms

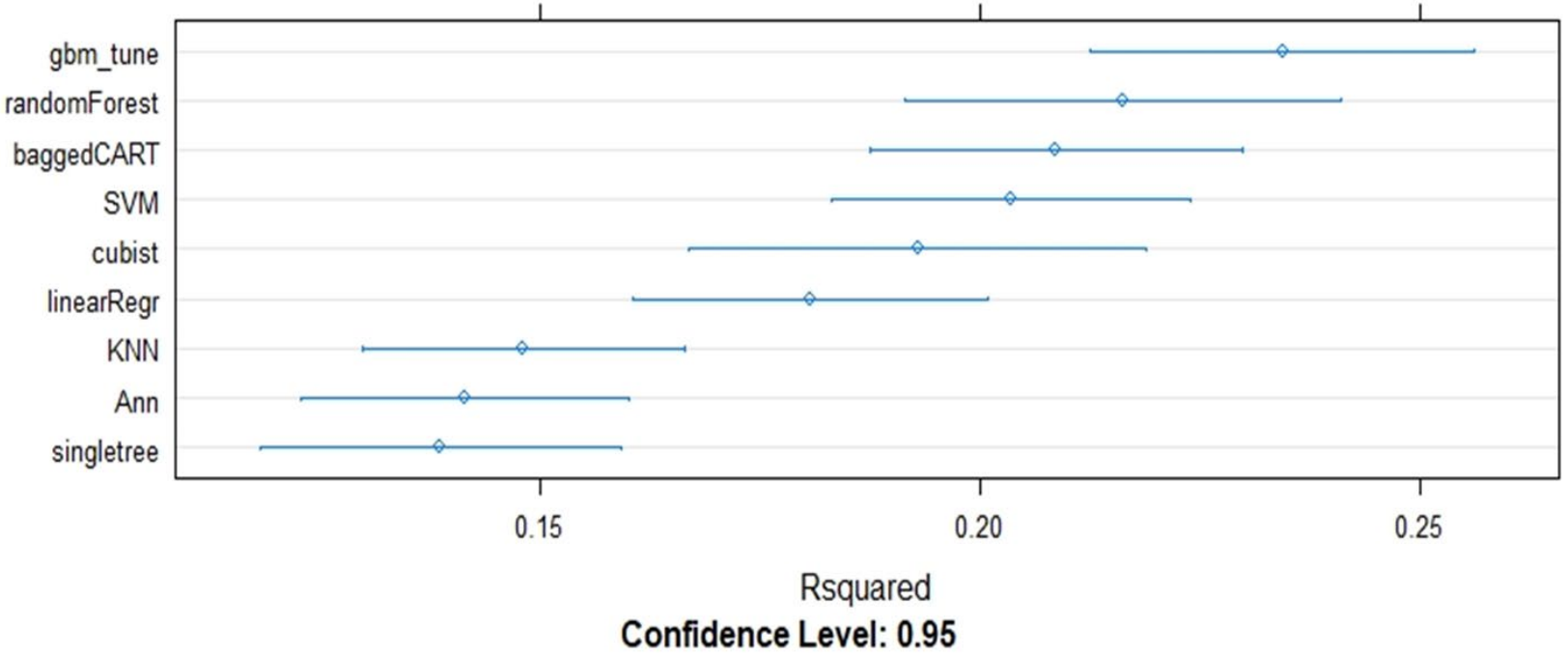
- Measuring feed efficiency traits on individual animal level is costly

  → Availability of on-farm data is limited

Is it possible to estimate direct production values of cows at farm level?

Background

- Feed efficiency traits: Feed intake, energy balance and feed efficiency
  - Preliminary study → Predicting feed intake based on confirmation and live weight data of cows



Objective of the study

Genotypic data (SNPs) + phenotypic data (production data)

↓

Supervised machine learning algorithms (Random Forest, Support Vector Machine, Extreme Gradient Boosting)

↓

Phenotypic performance (Feed intake, energy balance and feed efficiency)

Materials and Methods

Phenotypic data from 1684 dairy cows on 9 farms

(1-10 lactation, 100 - 150 day)

Feed efficiency traits	Mean	SD	Min	Max
Feed intake (kg DM / d)	23.17	3.41	9.76	34.02
Energy balance (MJ NEL / d)	8.06	19.53	-78.39	91.4
Feed efficiency (kg ECM / kg DM)	1.53	0.24	0.41	3.48

Production traits	Mean	SD	Min	Max
Weight (kg)	658.2	68.71	410.1	942.6
Milk yield (kg / d)	36.62	7.13	9.5	61.5
Protein content (%)	3.34	0.25	2.5	4.22
Fat content (%)	3.74	0.53	1.76	5.7

Materials and Methods

Phenotypic data from 1684 dairy cows on 9 farms

(1-10 lactation, 100 - 150 day)

Feed efficiency traits	Mean	SD	Min	Max
Feed intake (kg DM / d)	23.17	3.41	9.76	34.02
Energy balance (MJ NEL / d)	8.06	19.53	-78.39	91.4
Feed efficiency (kg ECM / kg DM)	1.53	0.24	0.41	3.48

Production traits	Mean	SD	Min	Max
Weight (kg)	658.2	68.71	410.1	942.6
Milk yield (kg / d)	36.62	7.13	9.5	61.5
Protein content (%)	3.34	0.25	2.5	4.22
Fat content (%)	3.74	0.53	1.76	5.7

eMissionCow

OptiKuh

Genotypic data

Imputed 50kSNP data

After QC→ 45,613 SNPs

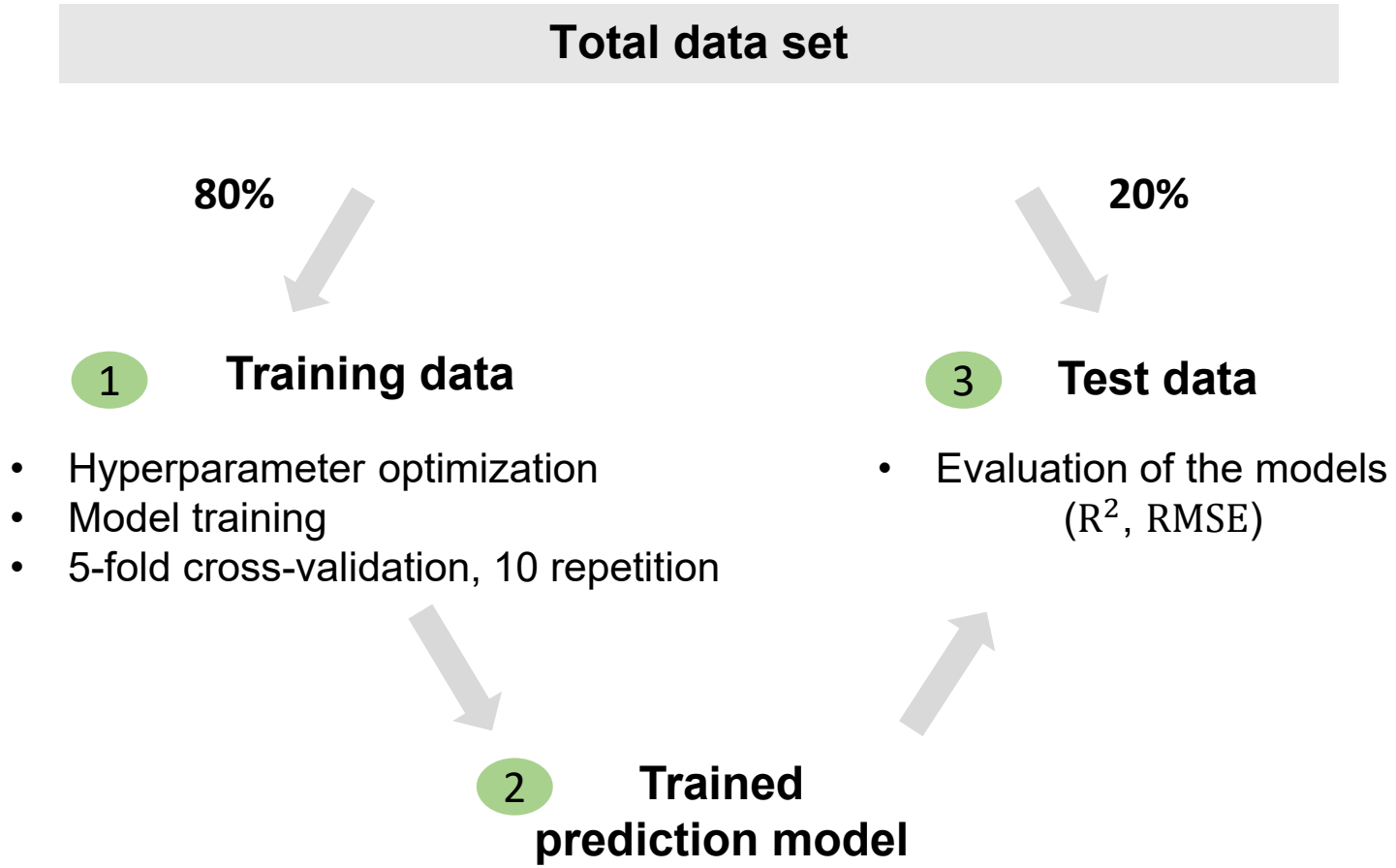
<u>Dimension reduction (nxn):</u>

G → Genomic relationship matrix (VanRaden, 2008)

UD → Principal component scores (Tussel et al. 2013)

Materials and Methods

Model training and testing



Materials and Methods

Feature set selection

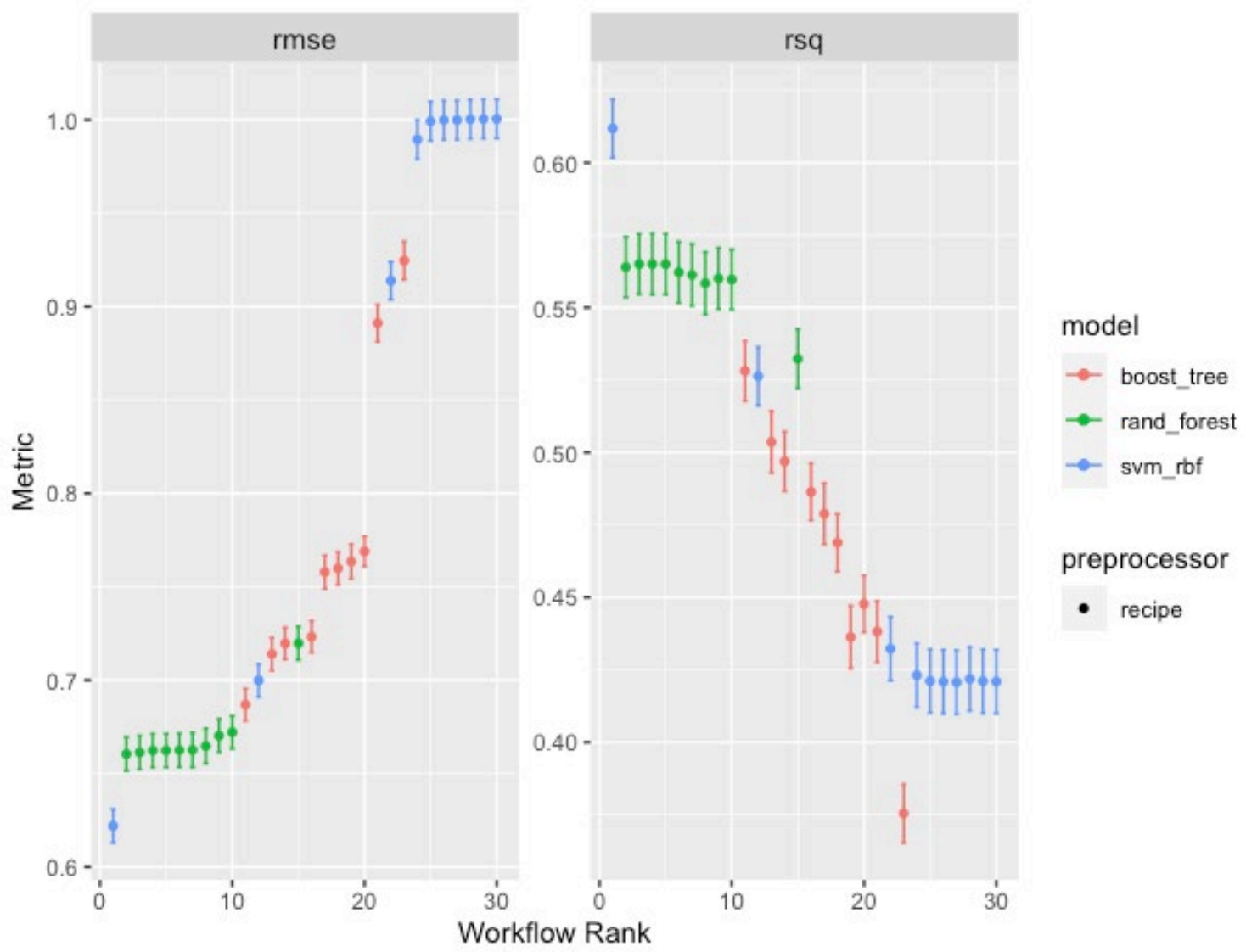
Feature-set	Weight	Milk yield	Protein content	Fat content	Lactation number	Herd-Year	G	UD
1 (P)	X	X	X	X	X	X		
2 (PG)	X	X	X	X	X	X	X	
3 (PUD)	X	X	X	X	X	X		X

G = genomic relationship matrix (dim: nxn); UD = Principal component scores of marker matrix (dim: nxn)

Materials and Methods

Hyperparameter opimization

Feature set 1 (P) : Feed intake



Results and Discussion

Prediction of feed intake (Test data):

	RF		SVM		XGBoost	
Feed intake	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Feature set 1 (P)	0.48	0.70	**0.60**	**0.62**	0.46	0.71
Feature set 2 (PG)	0.38	0.72	0.09	0.90	**0.43**	**0.69**
Feature set 3 (PUD)	0.35	0.74	0.36	0.91	**0.41**	**0.70**

RF = Random Forest; SVM = Support Vector Machine; XGBoost = Extreme Gradient Boosting

Results and Discussion

Prediction of energy balance (Test data):

	RF		SVM		XGBoost	
Energy balance	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Feature set 1 (P)	0.40	0.79	**0.45**	**0.75**	0.28	0.85
Feature set 2 (PG)	0.30	0.92	0.08	1.06	**0.33**	**0.89**
Feature set 3 (PUD)	**0.27**	**0.93**	0.17	1.08	0.25	0.94

RF = Random Forest; SVM = Support Vector Machine; XGBoost = Extreme Gradient Boosting

Results and Discussion

Prediction of feed efficiency (Test data):

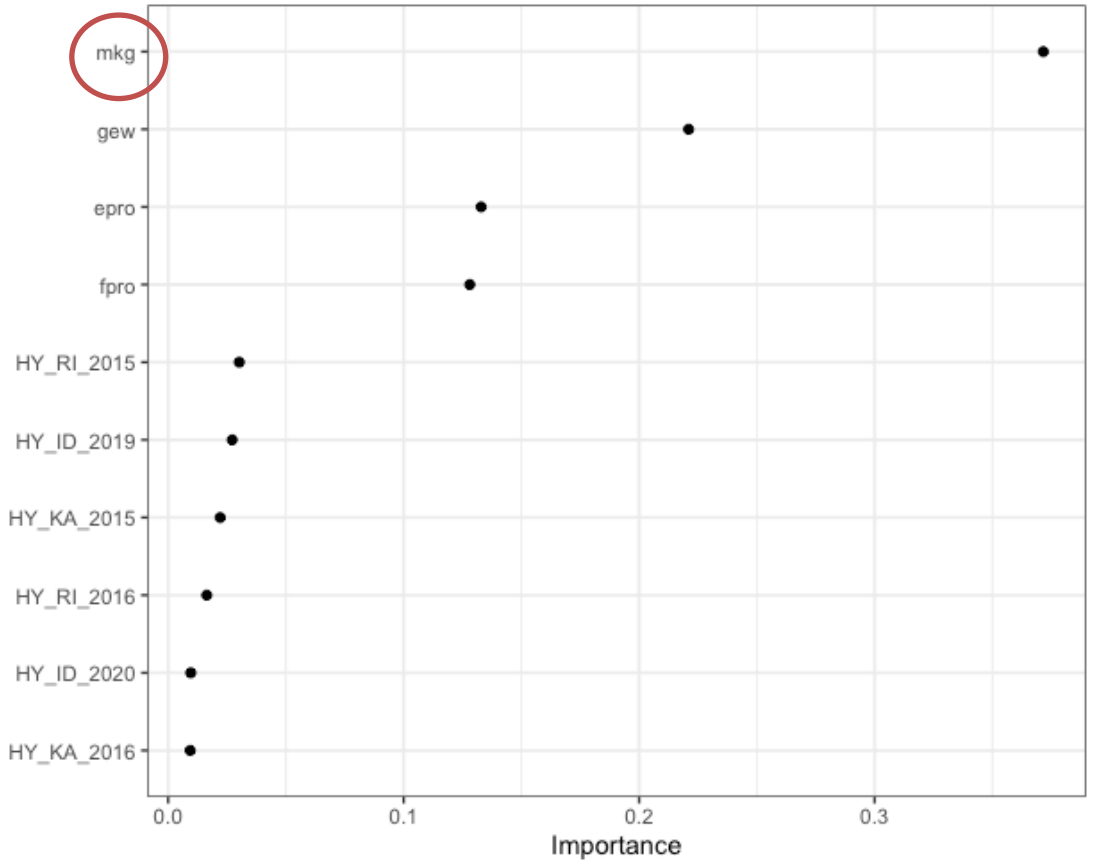
	RF		SVM		XGBoost	
Feed efficiency	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**	**RMSE**
Feature set 1 (P)	0.49	0.80	**0.55**	**0.74**	0.42	0.83
Feature set 2 (PG)	**0.45**	**0.87**	0.06	1.20	0.44	0.90
Feature set 3 (PUD)	**0.43**	**0.89**	0.3	1.13	**0.43**	**0.90**

RF = Random Forest; SVM = Support Vector Machine; XGBoost = Extreme Gradient Boosting

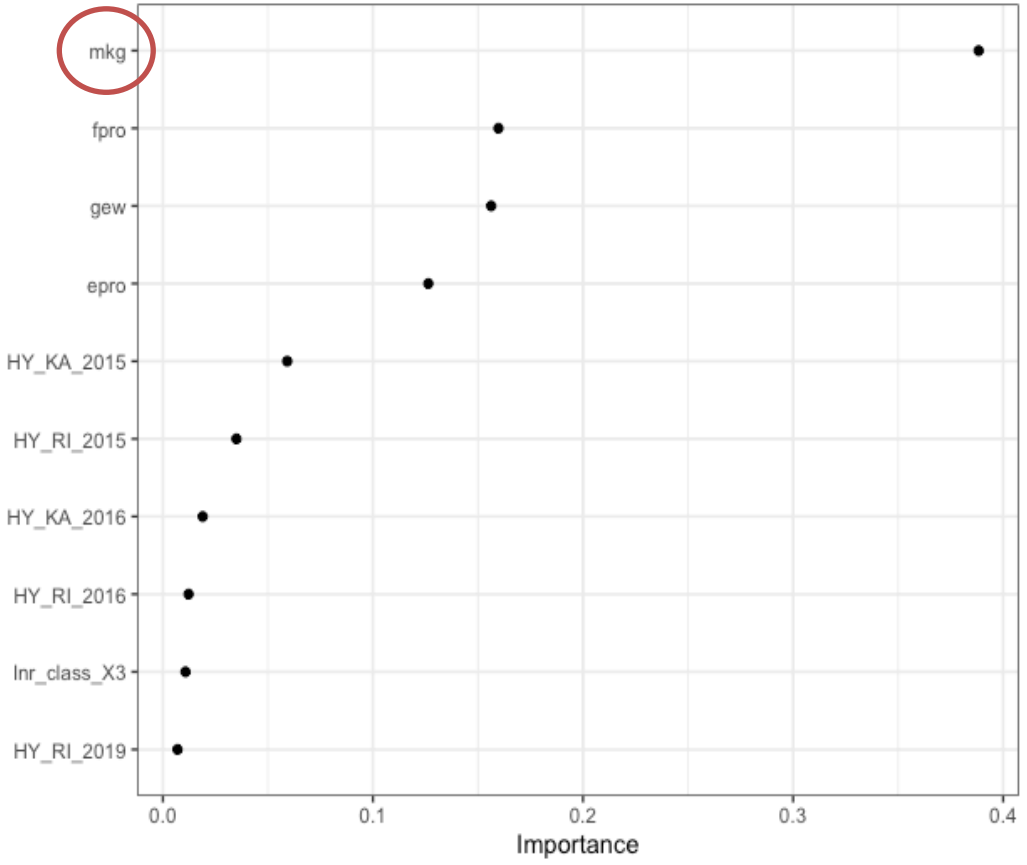
Results and Discussion

XGBoost Feature Importance

Feature set 1: Feed intake



Feature set 1: Feed efficiency



Conclusions

- General suitability of production data for prediction due to their availability on the farms (performance test + weight)
- Moderate to moderately high prediction accuracy of the models
- No improvement of predictions by using genotype information (G, UD)
- Varying variable import scores for the features in different models
- Particularly high information gain for the prediction of all three traits by milk yield
- Next steps:

    Models for further lactation stages or weeks

    Simple models with SNP markers (preselection?)

Thank you for your attention!