FACULTY OF VETERINARY MEDICINE
accredited by EAEVE

DEPARTMENT OF REPRODUCTION, OBSTETRICS AND HERD HEALTH
KUL - DEPARTMENT OF BIOSYSTEMS, BIOSYSTEMS TECHNOLOGY CLUSTER, LIVESTOCK TECHNOLOGY

MACHINE-LEARNING BASED PREDICTION OF TEST DAY MILK YIELD USING HISTORICAL DATA OF THE PREVIOUS LACTATION.

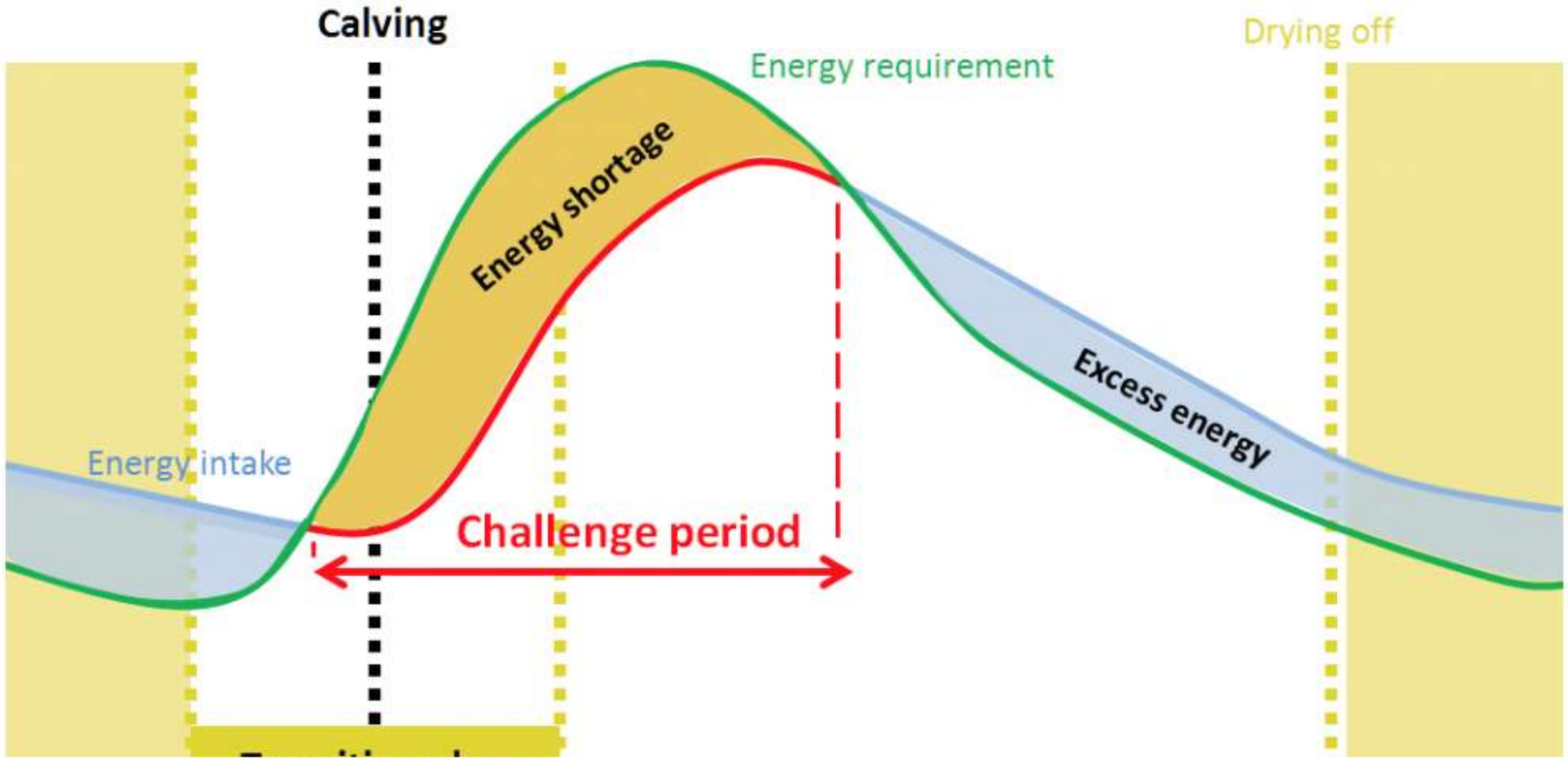
PhD Researcher: Matthieu Salamone

GHENT UNIVERSITY

KU LEUVEN

INTRODUCTION

THE INITIATION OF THE LACTATION



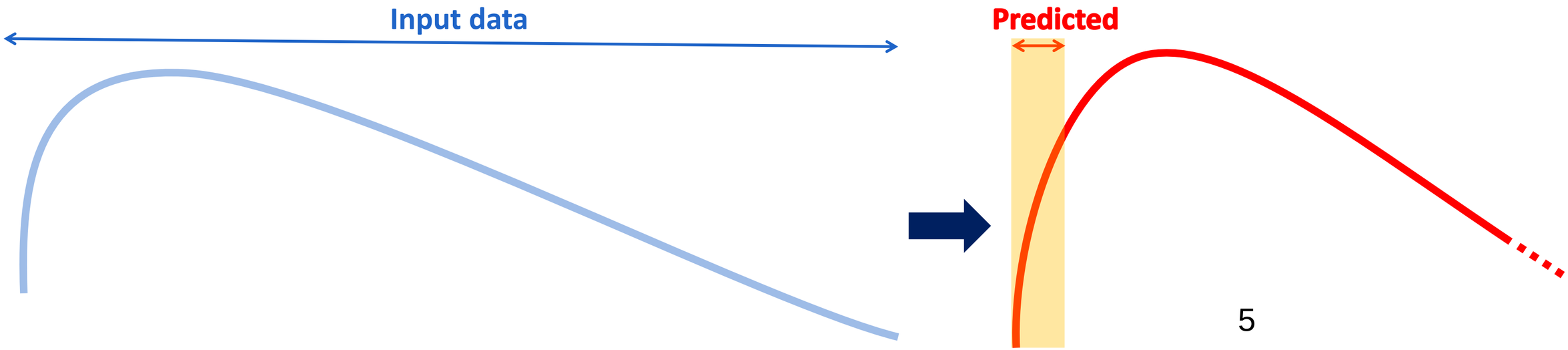
THE TRANSITION PERIOD

- Animal variation to response during challenge period
- Identification of at-risk animals: crucial
  - Animal welfare
  - Economic
  - Efficiency ~ ecological footprint

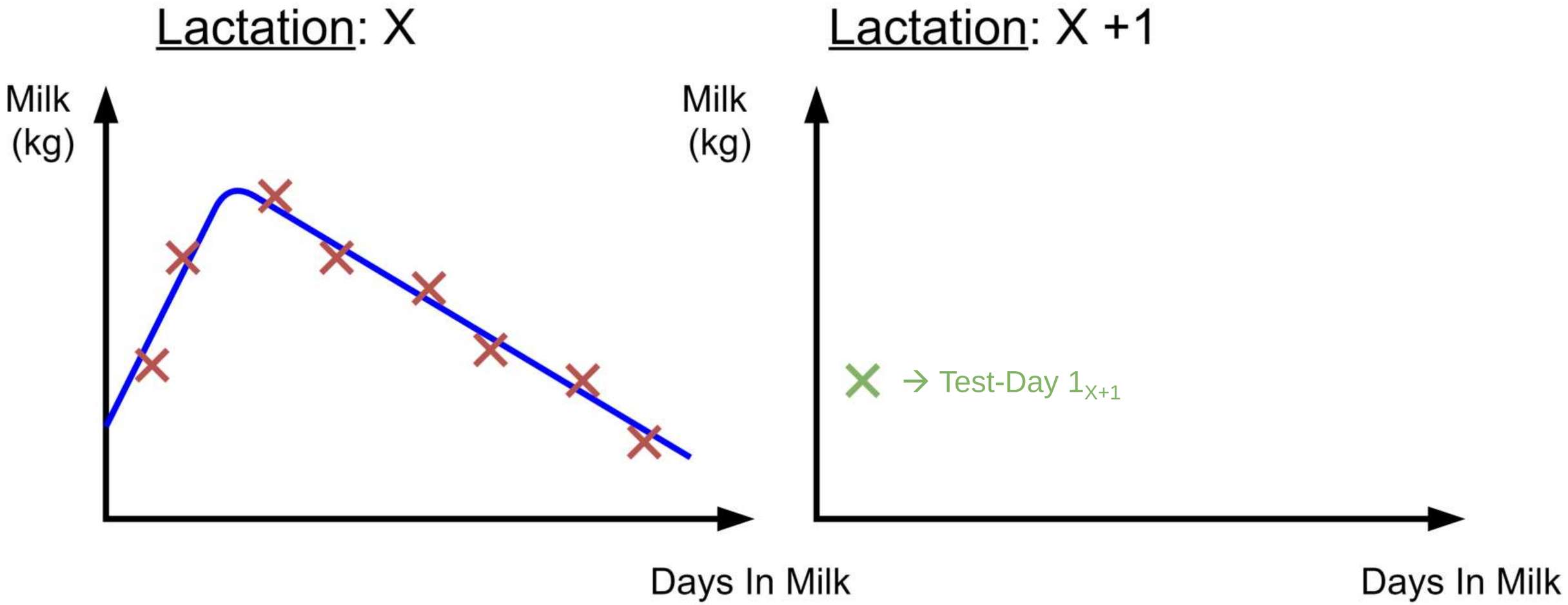
GHENT UNIVERSITY

RESEARCH QUESTION

- Data driven identification → Prediction of unperturbed production↔ Realised production
- Capacity to predict production in the transition period
  - Current models (Wood, Wilmink,…) → Possibly influenced by transition diseases
  - Usage of historic test day data → develop & validate a new predictive model



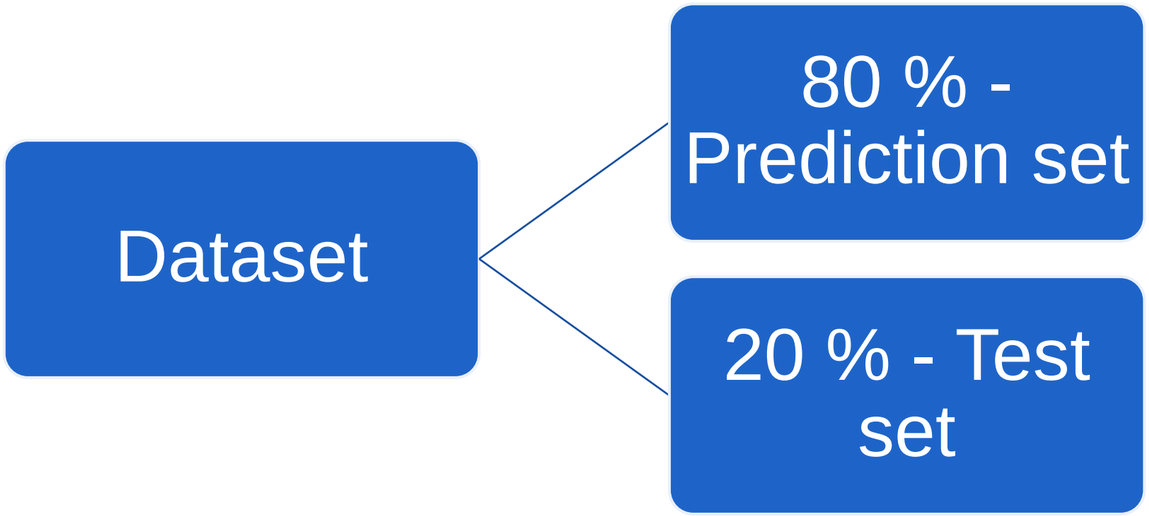
MATERIAL & METHODS


Lactation: X
Milk
(kg)
Days In Milk
Lactation: X +1
Milk
(kg)
→ Test-Day 1_{X+1}
Days In Milk

THE INPUT DATA

- Data Source →
  - Test-Day (TD) records, calving dates, breeding dates and dry-off dates, Lactation Curve Parameters (MilkBot):
    - 102 herds
    - 32 530 Animals (Multiparous)
    - 54 082 Lactation pairs

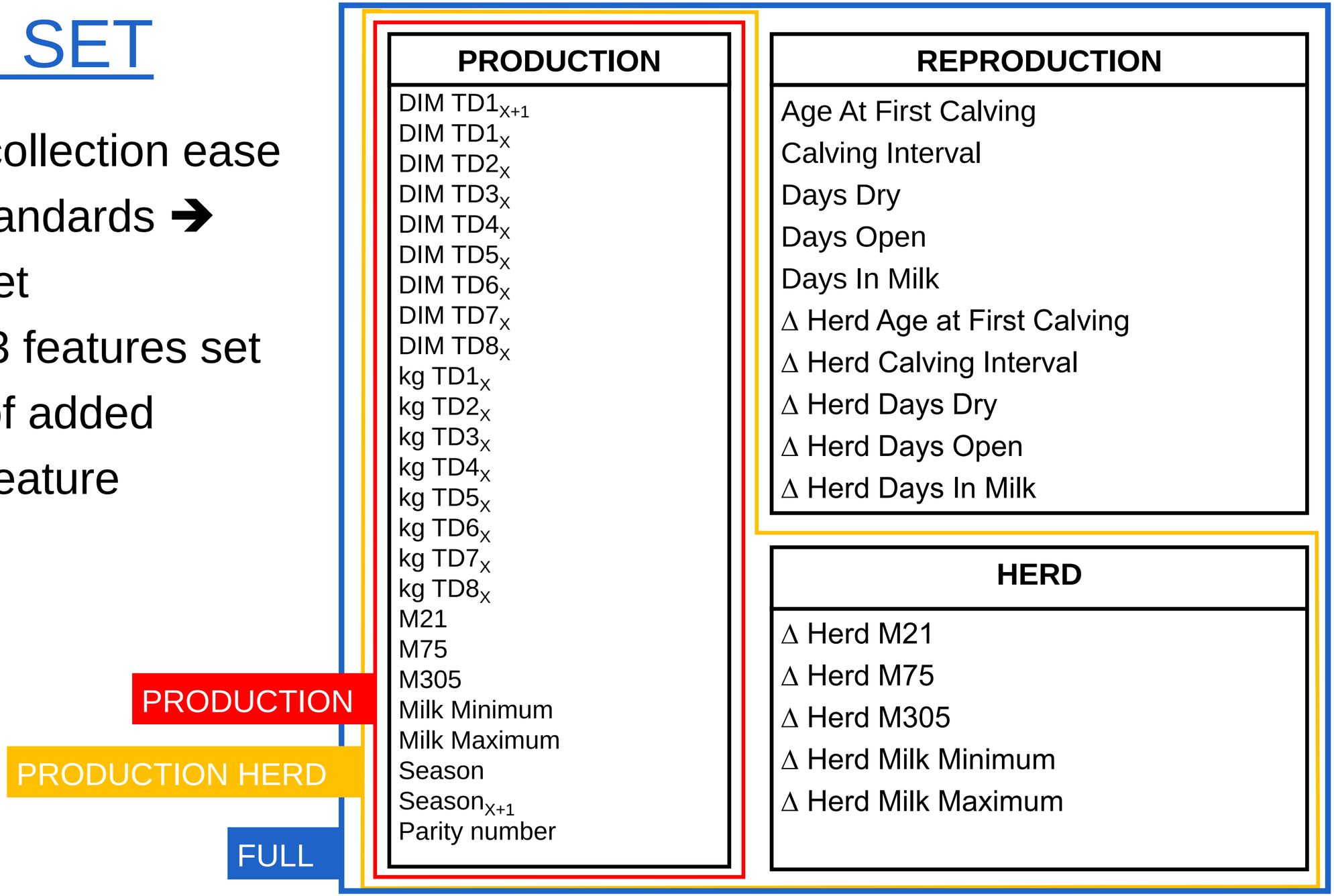
DATA SPLICING



- Splicing at Animal Level
  - Animal data either in Prediction or in Test Set
  - Reduces Possible bias in Performance evaluations

FEATURE SET

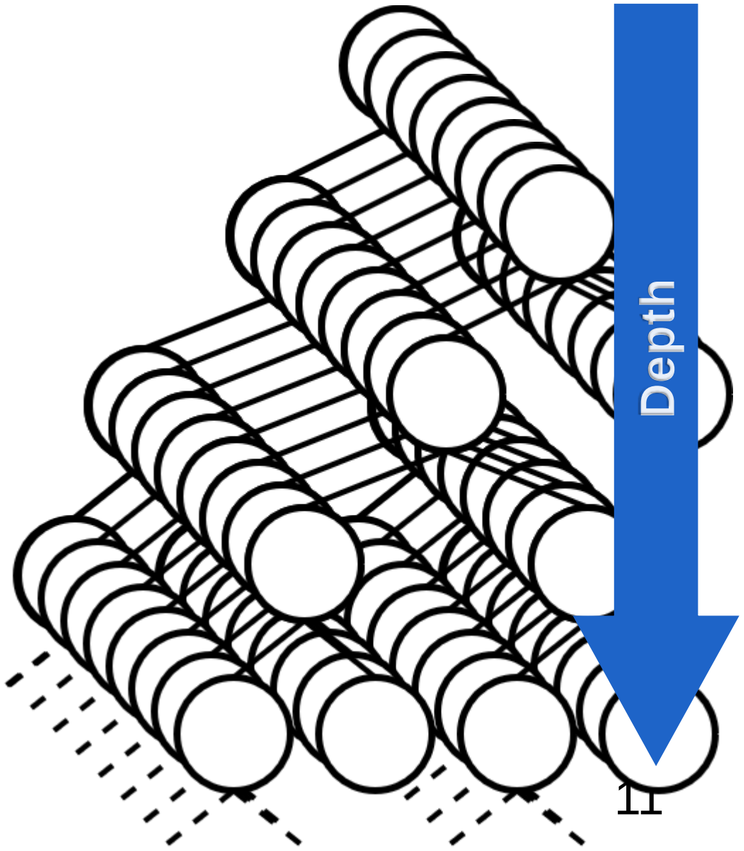
- Difference in collection ease and Quality standards ➔ Split feature set
- Working with 3 features set = Evaluation of added benefit more feature



GHENT UNIVERSITY

THE MACHINE LEARNING ALGORITHM

- Random Forest Regression (RF)
  - Creates set of decision trees
  - Parameters of the model:
    - Depth
    - # of trees



GHENT UNIVERSITY

KU LEUVEN

HYPERPARAMETER TUNING

- Grid search Hyperparameters=
  - Depth = 5, 10, 15, 20, 25
  - # Trees = 5, 25, 125, 250, 500
- 5 fold cross validation ➔ Optimal set = set with no significant difference with best performing set

FINAL MODELS NEXTMILK

- Train final set of models → **nextMILK** models :
  - The full training data set
  - Optimal set of hyperparameters
  - For each Feature Set

GHENT UNIVERSITY

KU LEUVEN

BENCHMARK MODELS

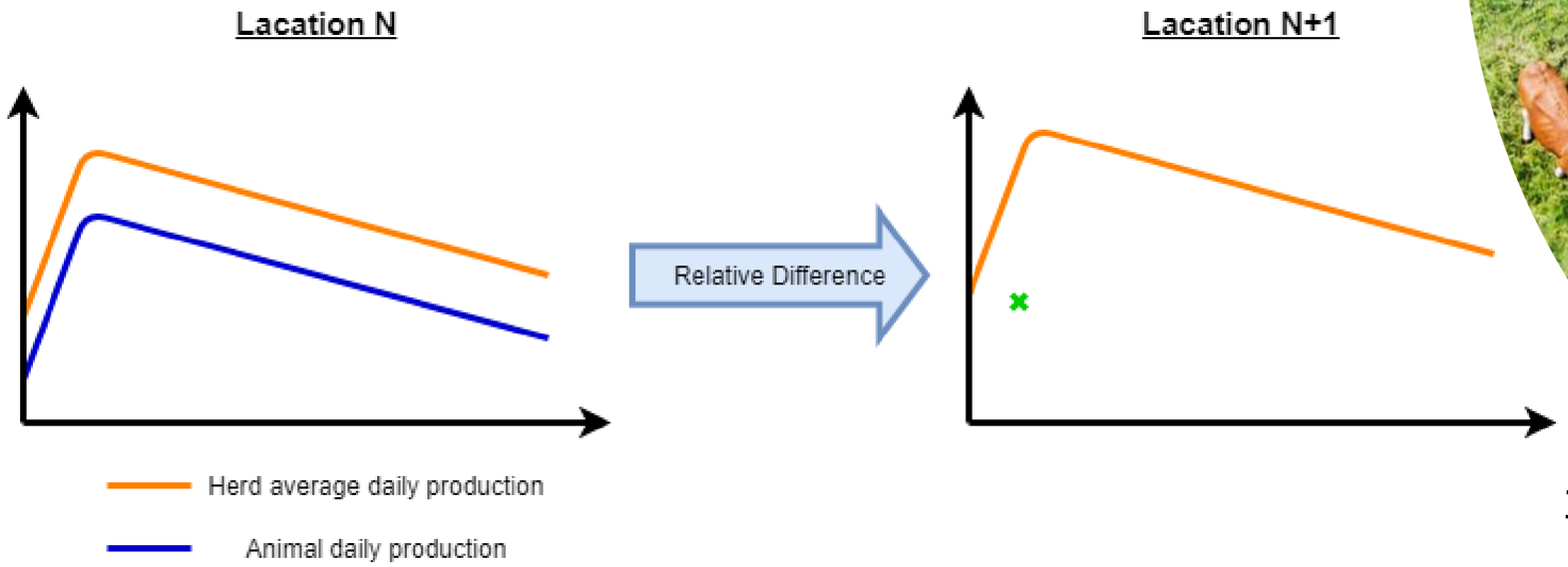
- Novel methodology
  - Evaluate our models
  - Performance of the models ↔ Complexity
  - Investigate complexity
    - Simpler models ↔ Performance
    - Significant difference?

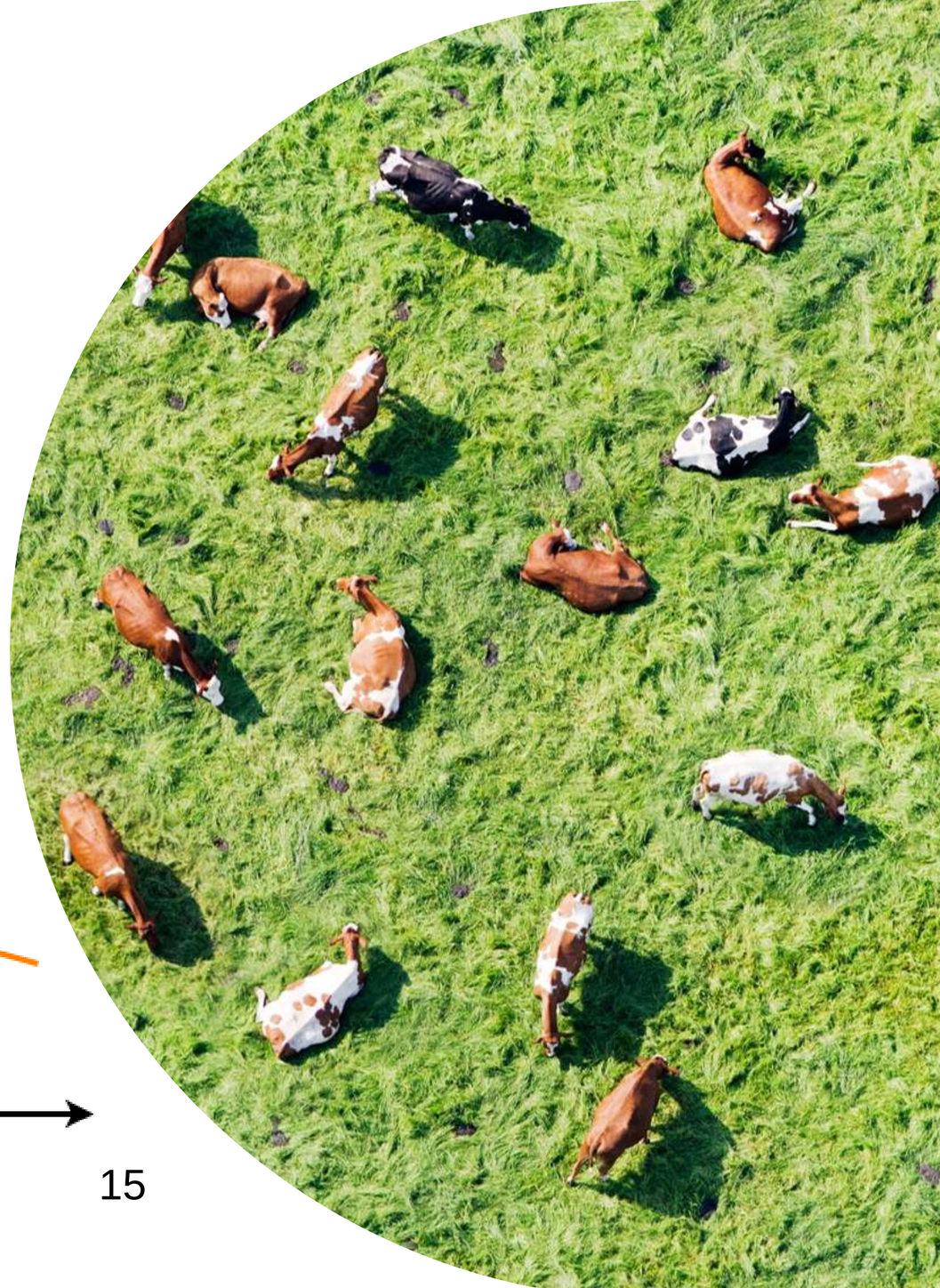
GHENT UNIVERSITY

BENCHMARK I - HERD

$$MILKP_{N+1} = MILKP_N * (1 + \%difference_{Herd(N \rightarrow N+1)})$$

- $\%\, difference_{Herd}$ calculated over timeframes 0d - 75d and 0d - 305d
- Corrected for DIM

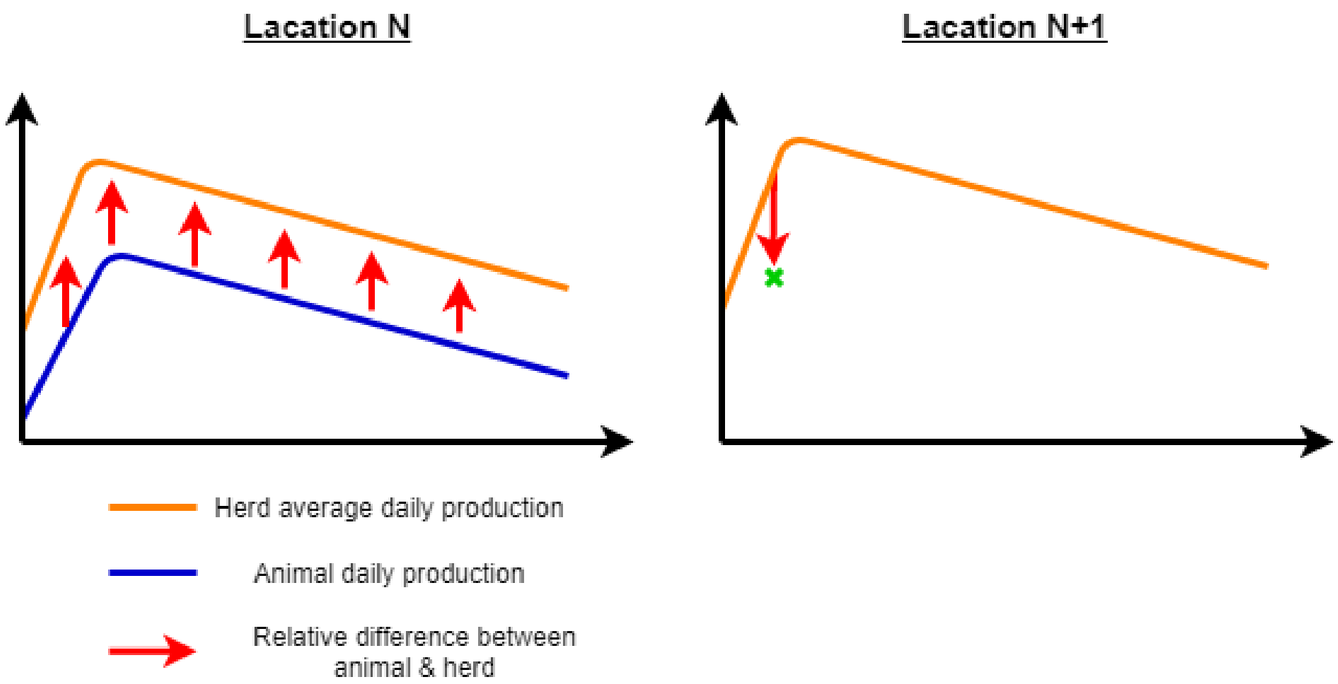




BENCHMARK II - ANIMAL

$$MILKP_{N+1} = \widehat{MILKP}_{N+1} * (1 + \%difference_{Animal \rightarrow Herd(N)})$$

- $\%difference_{Animal \rightarrow Herd(N)}$ calculated over timeframes 0d - 75d and 0d - 305d
- Corrected for DIM



ANALYSIS OF BENCHMARK

- Predictions on test set of **nextMILK**
  - |Residuals| → ANOVA → post-hoc, one-sided paired t-test
- Performance metrics – RMSE, R^2, MAE

GHENT UNIVERSITY

KU LEUVEN

RESULTS & DISCUSSION

HYPERPARAMETER TUNING

Optimal Set:

- **FULL:** 125 trees, 25 deep
- **PRODUCITON HERD:**
  250 trees, 25 deep
- **PRODUCTION:**
  250 trees, 25 deep

GHENT UNIVERSITY

FINAL MODELS: NEXTMILK

- Calculated on Test Set : 9 903 lacations

Model performance indicator:

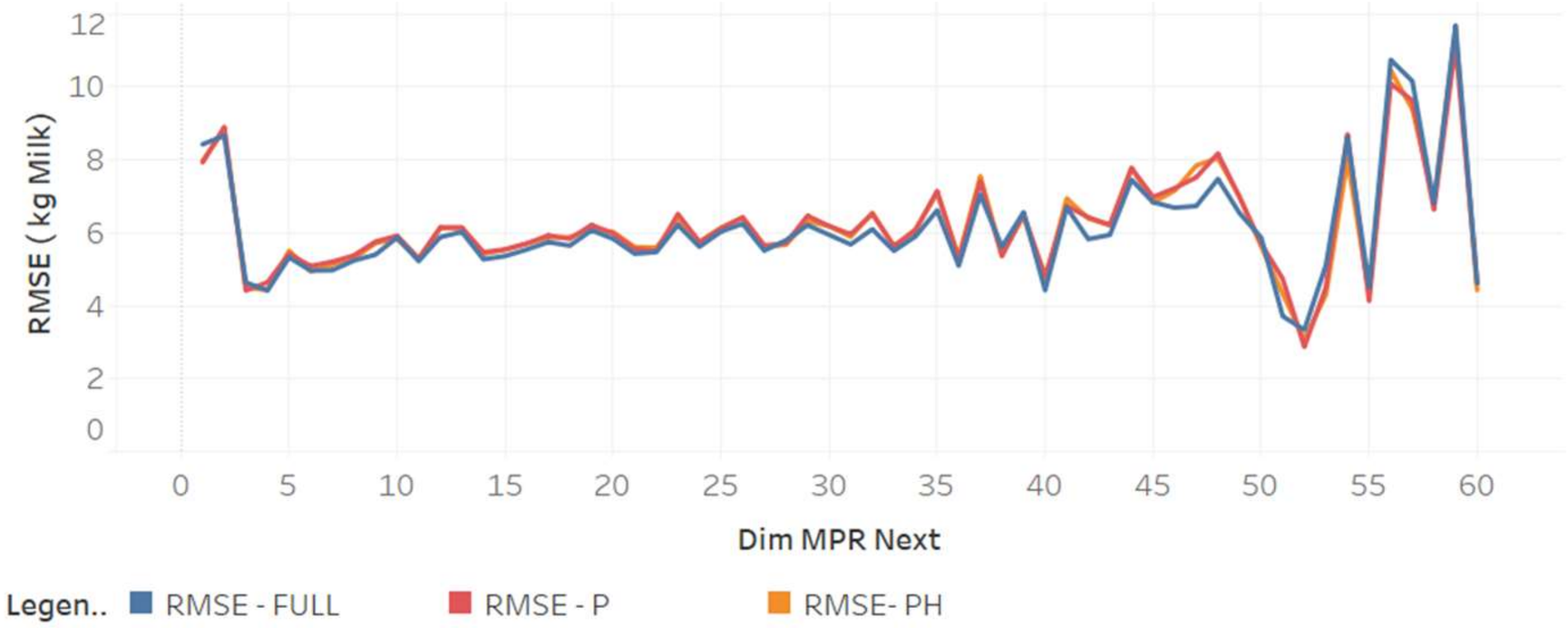
- **Root Mean Squared Error**
- **Mean Absolute Error**
- **R^2**

Model	RMSE	MAE	R^2
nextMILK - FULL	5.79	4.36	0.56
nextMILK – PRODUCTION HERD	6.01	4.53	0.52
nextMILK - PRODUCTION	6.03	4.52	0.52

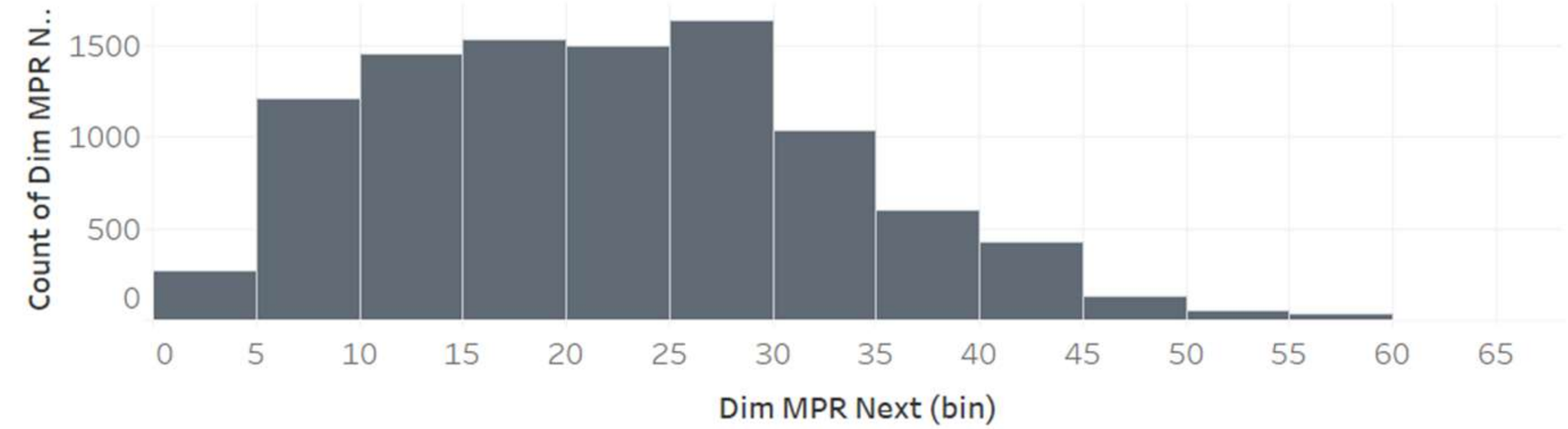
GHENT UNIVERSITY

KU LEUVEN

RMSE



DIM TD1 next

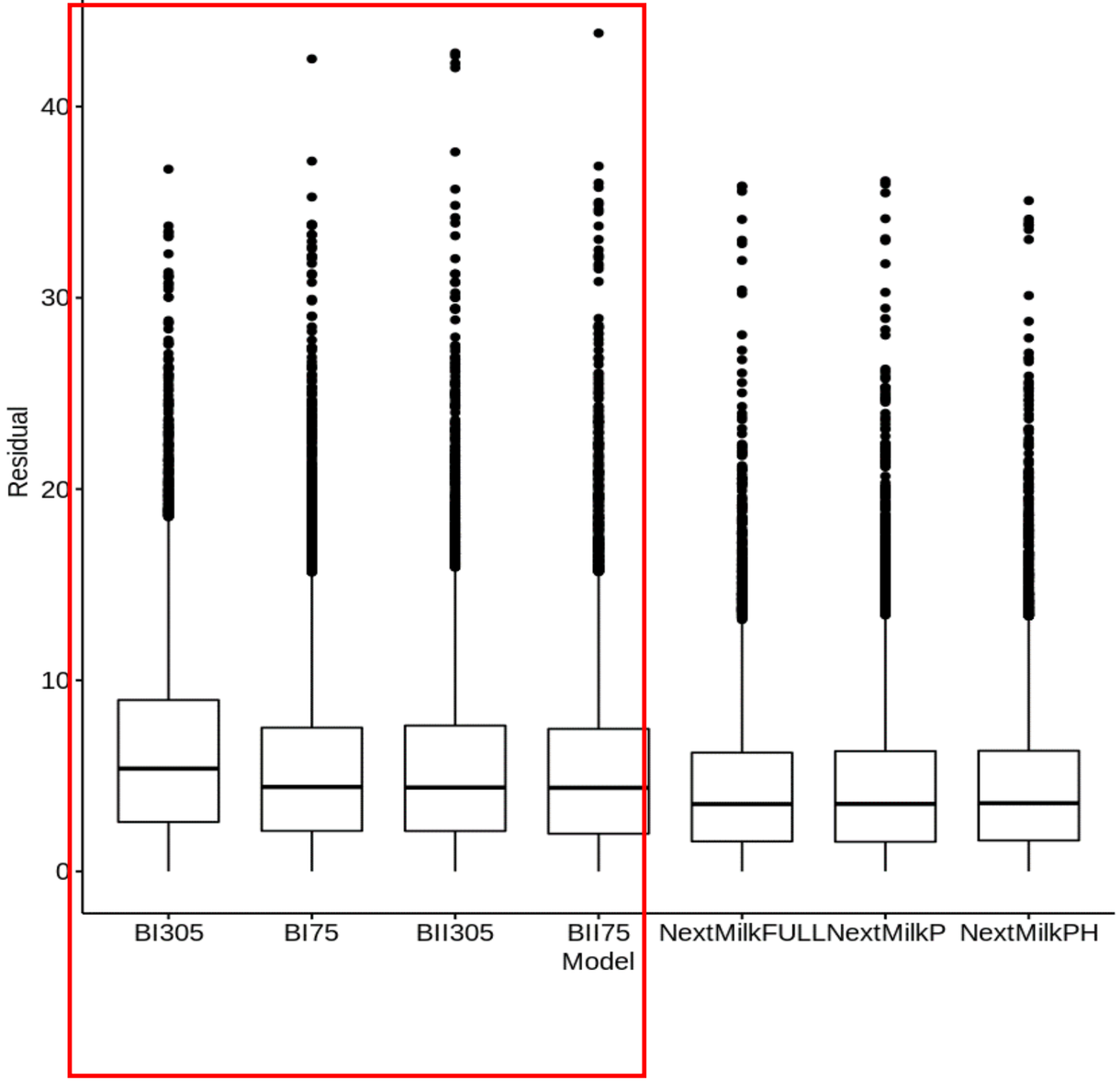


BENCHMARK ANALYSIS

Statistical analysis :

- Significant difference between models

  ANOVA: P-value < 0.001
- All benchmark models have significantly higher residuals
- No Significant difference between nextMILK models.



GHENT UNIVERSITY

KU LEUVEN

PERFORMANCE METRICS

Model	RMSE	MAE	R^2
Benchmark I – HERD - 75	7.09	4.55	0.33
Benchmark I – HERD - 305	7.51	4.84	0.16
Benchmark II – ANIMAL -75	7.00	4.53	0.35
Benchmark II – ANIMAL -305	7.09	4.54	0.34
nextMILK - FULL	5.79	4.36	0.56
nextMILK – PRODUCTION HERD	6.01	4.53	0.52
nextMILK - PRODUCTION	6.03	4.52	0.52

CONCLUSION

- nextMILK models → show potential in predicting production in next lactation
- Further validation necessary → Data driven identification tool

GHENT UNIVERSITY

FACULTY OF VETERINARY MEDICINE
accredited by EAEVE

Matthieu Salamone
PhD Student

DEPARTMENT OF REPRODUCTION,
OBSTETRICS AND HERD HEALTH

E matthieu.salamone@ugent.be
T +32 9 264 75 62

Universiteit Gent
Ghent University

www.ugent.be

GHENT UNIVERSITY

KU LEUVEN