



# Prediction of blood parameters of buffaloes from milk mid-infrared spectra

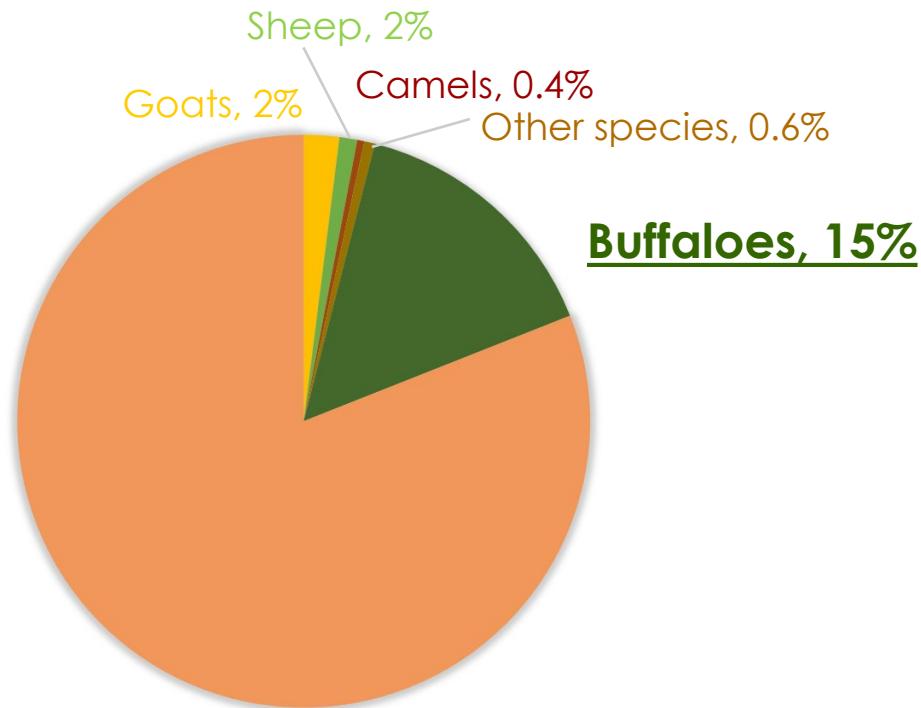
Silvia Magro<sup>1</sup>, Sergio Esposito<sup>2</sup>, Gianluca Neglia<sup>2</sup>, Giovanni Niero<sup>1</sup>, Massimo De Marchi<sup>1</sup>

<sup>1</sup> Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Legnaro, Italy

<sup>2</sup> Department of Veterinary Medicine and Animal Production, Federico II University, Napoli, Italy

# INTRODUCTION

World milk production:

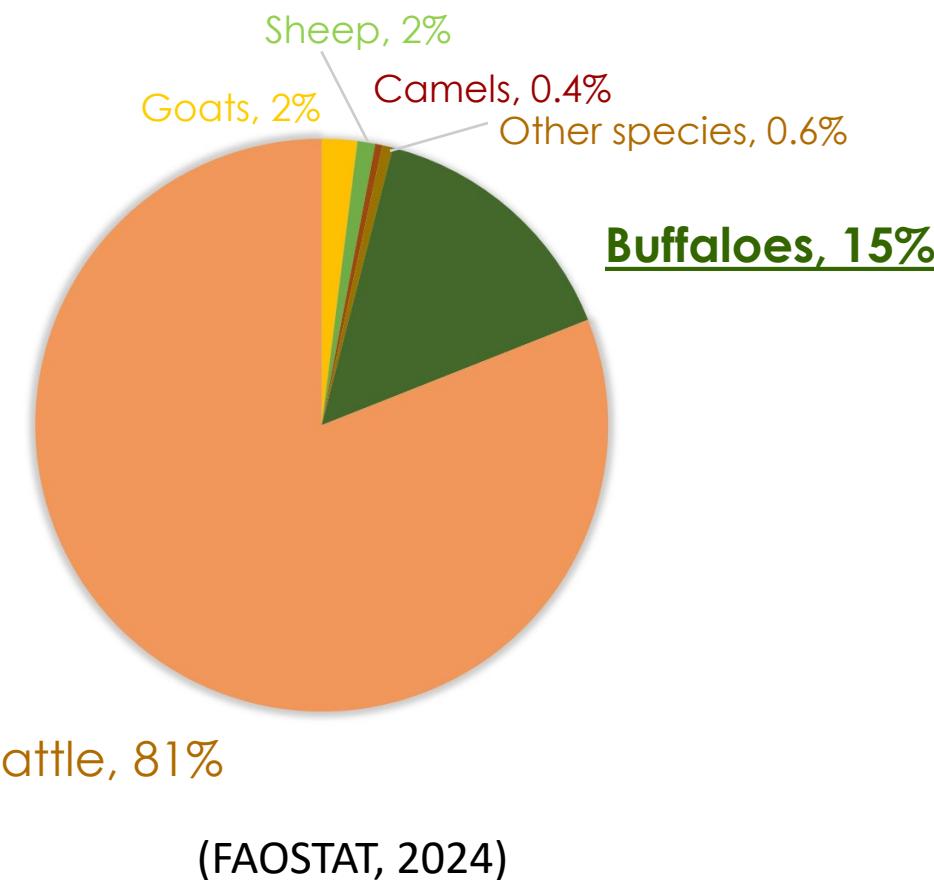


Cattle, 81%

(FAOSTAT, 2024)

# INTRODUCTION

World milk production:



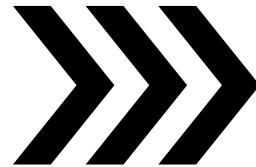
In Italy:

'Mozzarella di Bufala Campana' cheese

Protected Designation of Origin (PDO) by the European Community  
(Regulation 1107/96 of 12 June 1996)



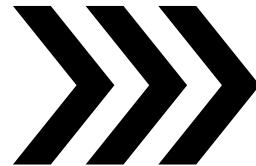
# INTRODUCTION



Progressive intensification of the system:

- an increase in the size of the herd
- the adoption of intensive/semi-intensive farming techniques (originally intended for cattle)

# INTRODUCTION



Progressive intensification of the system:

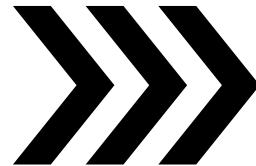
- an increase in the size of the herd
- the adoption of intensive/semi-intensive farming techniques (originally intended for cattle)



- ✓ maximize the efficiency of the dairy sector
- ✓ improve the yield and quality of milk and cheese



# INTRODUCTION



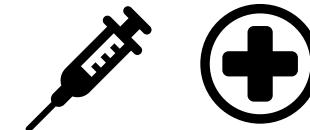
Progressive intensification of the system:

- an increase in the size of the herd
- the adoption of intensive/semi-intensive farming techniques (originally intended for cattle)

- ✓ maximize the efficiency of the dairy sector
- ✓ improve the yield and quality of milk and cheese



Management issues that also impact animal welfare and health



# INTRODUCTION

Blood metabolic profile testing allows for monitoring **metabolic health** and **nutritional status**.



- high costs
- labour
- animals' stress



conducting extensive blood analyses on a large scale

# INTRODUCTION

Blood metabolic profile testing allows for monitoring **metabolic health** and **nutritional status**.

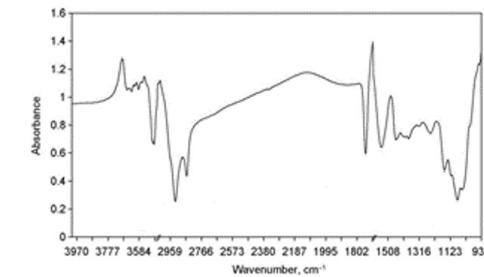
- high costs
- labour
- animals' stress



conducting extensive blood analyses on a large scale



Utilizing mid-infrared (MIR) spectra of milk to predict blood traits may present an effective opportunity.



# AIM

The present study aims to test the ability of MIR to predict blood traits using milk spectra collected from the same buffalo.



# MATERIAL & METHODS



310 buffaloes



9 farms

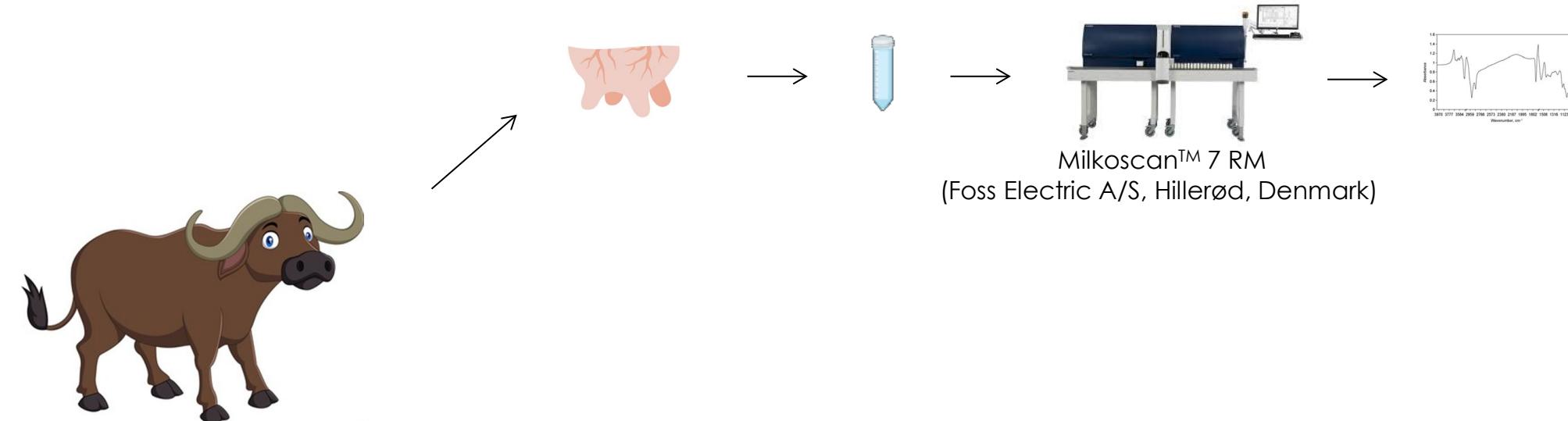


From February to September 2022

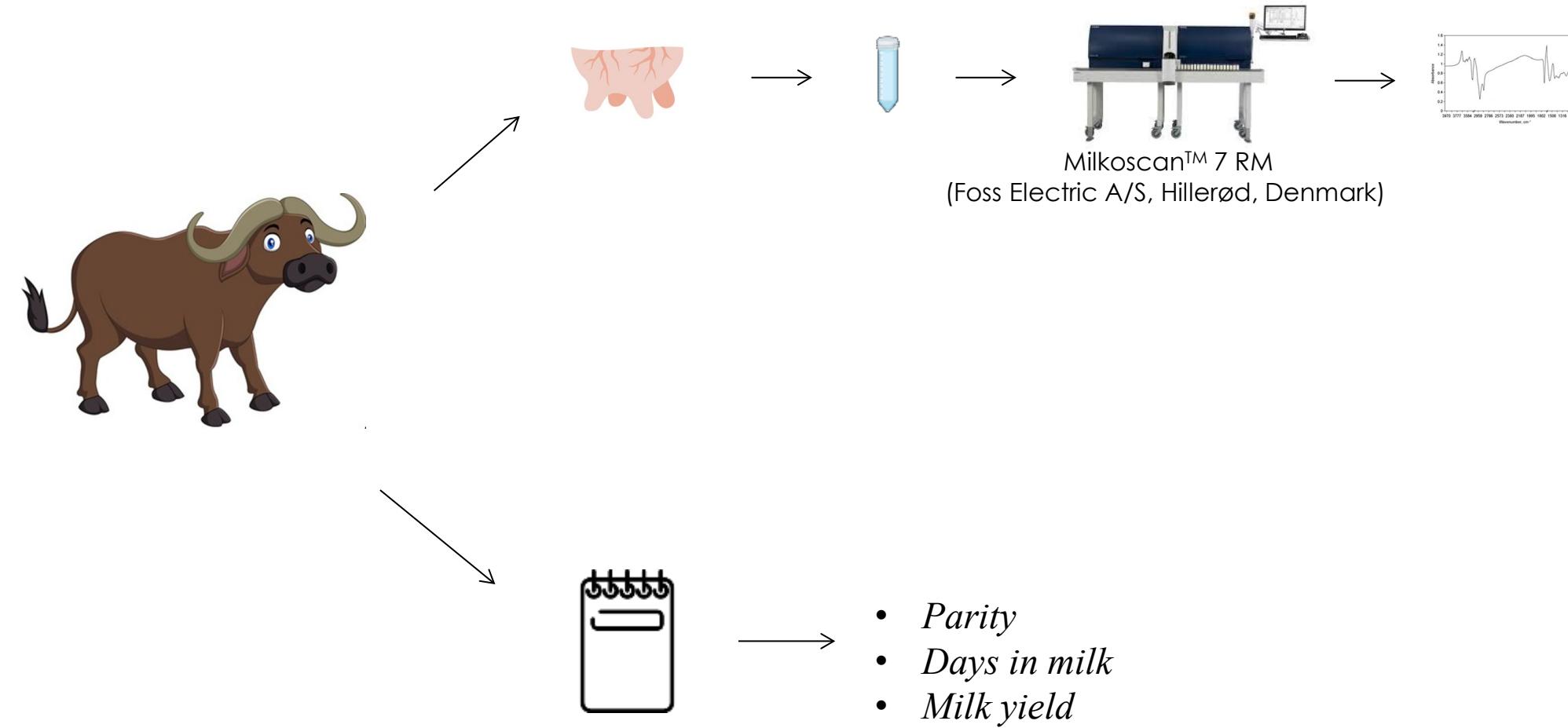
Individual blood and milk samples  
were collected during the morning milking.



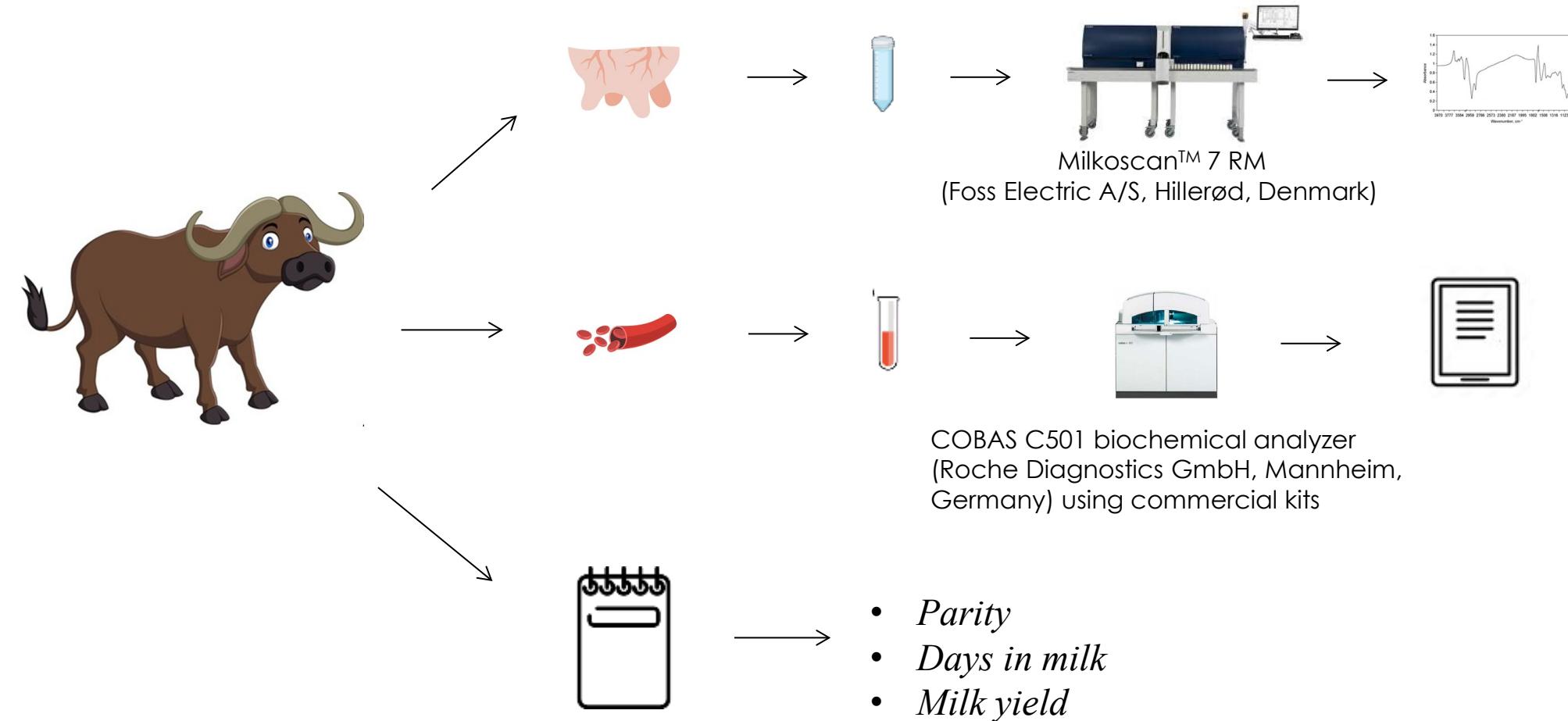
# MATERIAL & METHODS



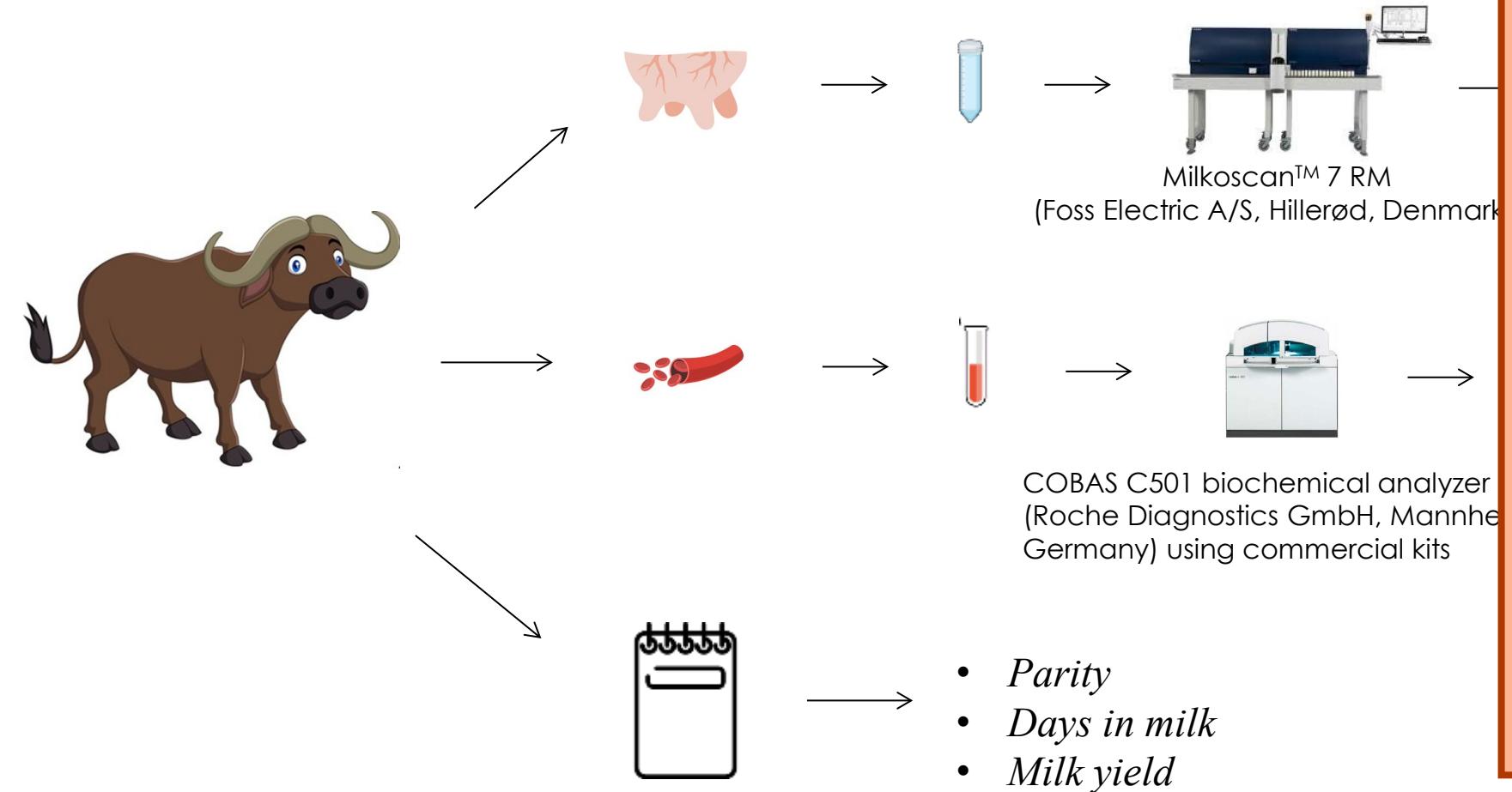
# MATERIAL & METHODS



# MATERIAL & METHODS



# MATERIAL & METHODS



## Enzyme activities:

- Alanine transaminase (ALT, U/L)
- Aspartate aminotransferase (AST, U/L)
- Gamma-glutamyl transferase (GGT, U/L)
- Alkaline phosphatase (ALP, U/L)
- Lactate dehydrogenase (LDH, U/L)
- Creatine kinase (CK, U/L)

## Blood traits:

- Urea, mg/dL
- Creatinine, mg/dL
- Glucose, mg/dL
- Non-esterified fatty acids (NEFA, mmol/L)
- Total bilirubin (TBIL, mg/dL)
- Cholesterol, mg/dL
- Triglyceride, mg/dL

## Protein profile

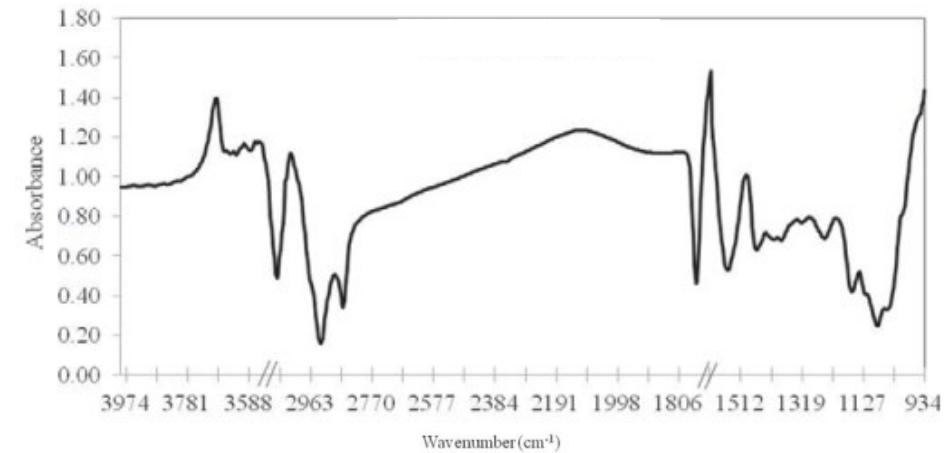
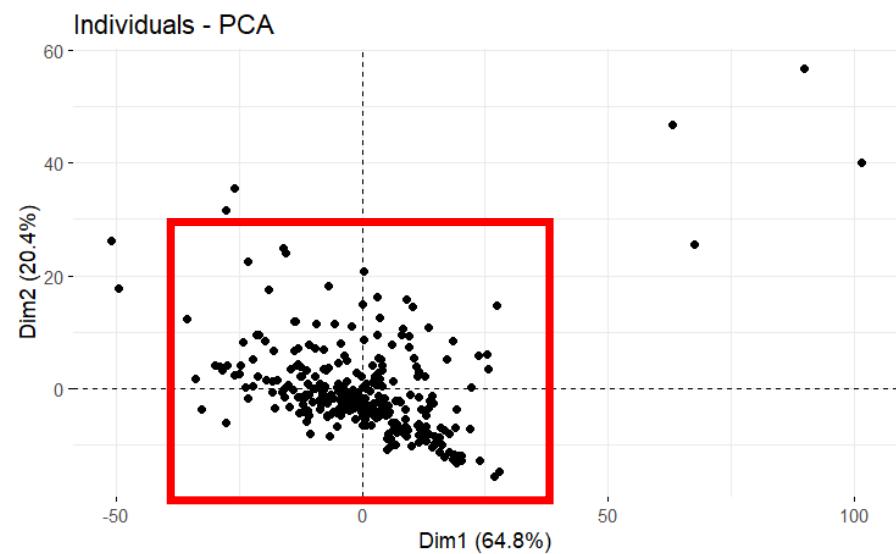
- Total protein, g/dL
- Albumin, g/dL
- Globulin, g/dL
- Albumin to globulin

# MATERIAL & METHODS

## Mid-infrared prediction models

Spectral regions known to be associated with noisy water absorption wavelengths were discarded

-> **leading to 338 wavelengths**



Preliminary analysis of the spectral data was conducted using principal component analysis (PCA) to identify anomalous samples in terms of the MIR spectrum -> **8 spectra were discarded**

For non-distributed blood traits, a logarithmic transformation was applied (CK, NEFA; TBIL).

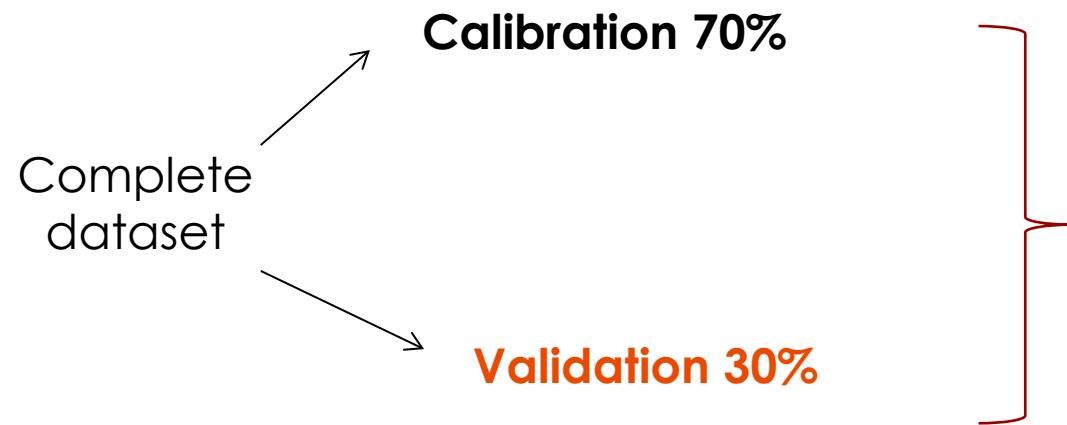


# MATERIAL & METHODS

## Mid-infrared prediction models

The **partial least squares (PLS)** analysis was carried out using the 'trainControl' function available in the R package 'caret' :

- ✓ The models were fine-tuned using leave-one-out cross-validation, and the number of latent variables was set automatically but capped at a maximum of 15 to avoid overfitting.
- ✓ Spectral data points were centered and scaled.



Fitting statistics of the PLS model:

- Coefficient of determination ( $R^2$ )
- Standard error

(Kuhn, 2008)



# RESULTS

## Descriptive statistics

Blood trait	Mean (SD)	Range	CV, %
<i>Enzyme activities</i>			
ALT, U/L	45.17 (10.97)	17.30 - 94.00	24.3
AST, U/L	130.50 (28.95)	63.60 - 327.00	22.2
ALP, U/L	287.94 (189.90)	19.74 - 1,528.00	66.0
GGT, U/L	30.52 (8.06)	13.00 - 66.02	26.4
LDH, U/L	1,256.22 (205.12)	652.00 - 1,845.00	16.3
CK, U/L	178.00 (76.65)	53.00 - 793.00	43.1
<i>Blood traits</i>			
Urea, mg/dL	60.01 (16.83)	21.00 - 129.90	28.0
Creatinine, mg/dL	1.62 (0.32)	0.90 - 2.70	19.8
Glucose, mg/dL	67.34 (11.92)	25.30 - 115.00	17.7
NEFA, mmol/L	0.95 (0.94)	0.07 - 3.16	98.9
TBIL, mg/dL	0.27 (0.30)	0.00 - 4.34	111.1
Cholesterol, mg/dL	122.22 (32.50)	38.90 - 283.00	26.6
Triglyceride, mg/dL	18.57 (12.45)	0.00 - 45.10	67.0
<i>Protein profile</i>			
Total protein, g/dL	6.93 (0.72)	4.73 - 8.73	10.4
Albumin, g/dL	3.40 (0.26)	2.55 - 4.07	7.6
Globulin, g/dL	3.53 (0.57)	2.18 - 5.05	16.1
Albumin-to-globulin	0.99 (0.15)	0.59 - 1.57	15.2

CV, coefficient of variation

# RESULTS

## Descriptive statistics

Blood trait	Mean (SD)	Range	CV, %
<i>Enzyme activities</i>			
ALT, U/L	45.17 (10.97)	17.30 - 94.00	24.3
AST, U/L	130.50 (28.95)	63.60 - 327.00	22.2
ALP, U/L	287.94 (189.90)	19.74 - 1,528.00	66.0
GGT, U/L	30.52 (8.06)	13.00 - 66.02	26.4
LDH, U/L	1,256.22 (205.12)	652.00 - 1,845.00	16.3
CK, U/L	178.00 (76.65)	53.00 - 793.00	43.1
<i>Blood traits</i>			
Urea, mg/dL	60.01 (16.83)	21.00 - 129.90	28.0
Creatinine, mg/dL	1.62 (0.32)	0.90 - 2.70	19.8
Glucose, mg/dL	67.34 (11.92)	25.30 - 115.00	17.7
NEFA, mmol/L	0.95 (0.94)	0.07 - 3.16	98.9 
TBIL, mg/dL	0.27 (0.30)	0.00 - 4.34	111.1 
Cholesterol, mg/dL	122.22 (32.50)	38.90 - 283.00	26.6
Triglyceride, mg/dL	18.57 (12.45)	0.00 - 45.10	67.0
<i>Protein profile</i>			
Total protein, g/dL	6.93 (0.72)	4.73 - 8.73	10.4
Albumin, g/dL	3.40 (0.26)	2.55 - 4.07	7.6 
Globulin, g/dL	3.53 (0.57)	2.18 - 5.05	16.1
Albumin-to-globulin	0.99 (0.15)	0.59 - 1.57	15.2

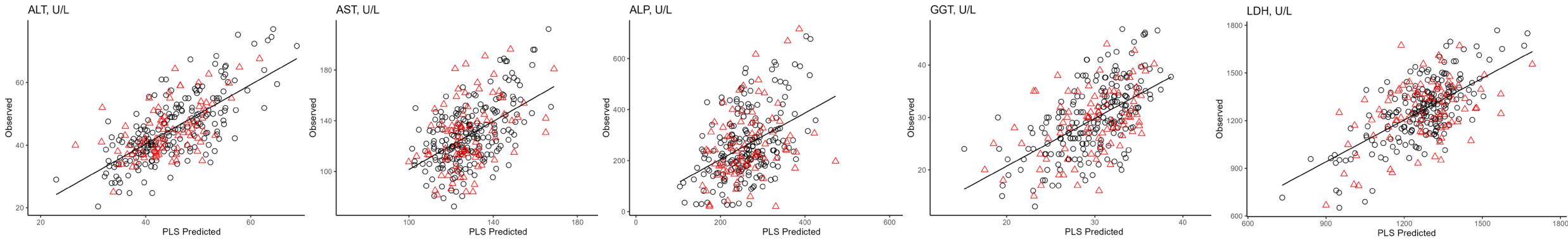
CV, coefficient of variation

# RESULTS

## Enzyme activities

Trait	LV	Calibration (70%)		Validation (30%)	
		$R^2_C$	$SE_C$	$R^2_V$	$SE_V$
ALT, U/L	10	0.51	6.87	0.35	7.90
AST, U/L	9	0.33	19.83	0.15	23.92
ALP, U/L	6	0.20	118.21	0.14	127.53
GGT, U/L	6	0.36	5.28	0.21	6.08
LDH, U/L	11	0.47	138.73	0.26	176.44
$\log_{10}$ CK	8	0.37	0.11	0.14	0.13

LV = Latent Variables



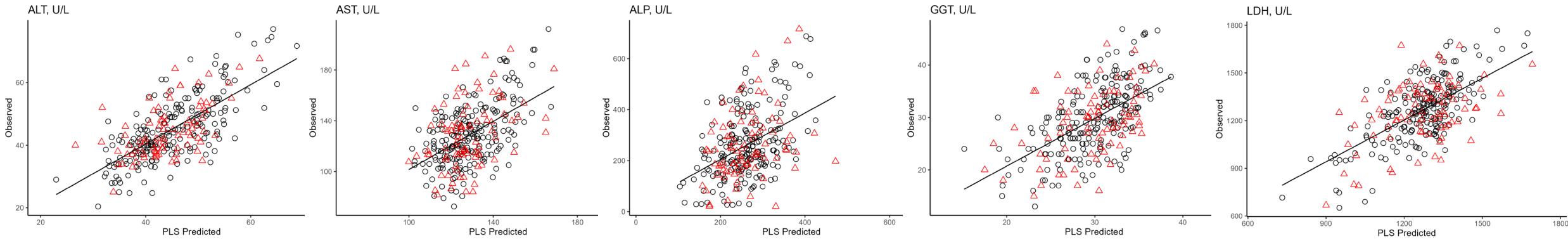
● calibration set, ▲ validation set

# RESULTS

## Enzyme activities

Trait	LV	Calibration (70%)		Validation (30%)	
		$R^2_C$	$SE_C$	$R^2_V$	$SE_V$
ALT, U/L	10	0.51	6.87	0.35	7.90
AST, U/L	9	0.33	19.83	0.15	23.92
ALP, U/L	6	0.20	118.21	0.14	127.53
GGT, U/L	6	0.36	5.28	0.21	6.08
LDH, U/L	11	0.47	138.73	0.26	176.44
$\log_{10}$ CK	8	0.37	0.11	0.14	0.13

LV = Latent Variables



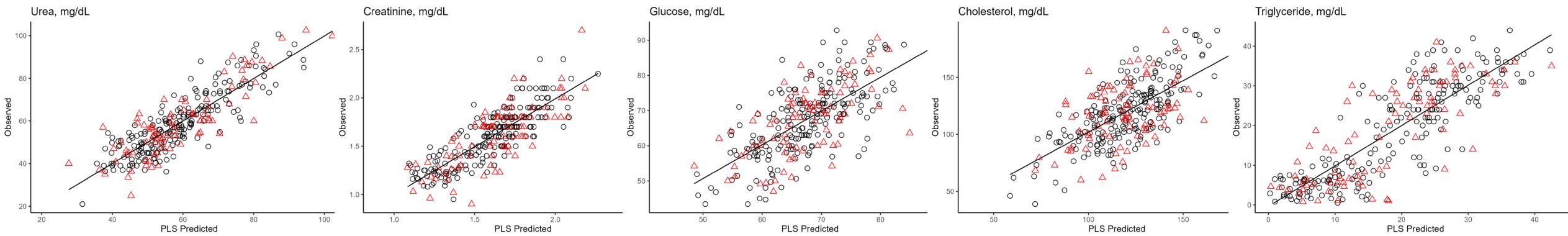
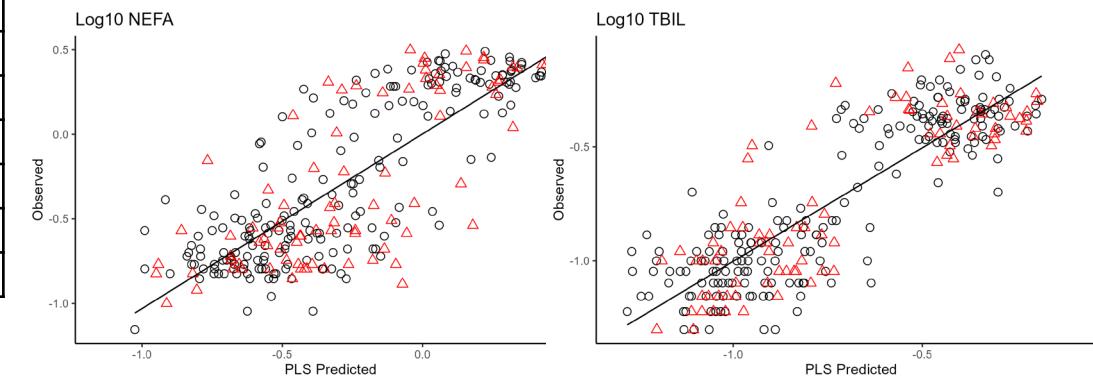
● calibration set, ▲ validation set

# RESULTS

## Blood trait

Trait	LV	Calibration (70%)		Validation (30%)	
		$R^2_C$	$SE_C$	$R^2_V$	$SE_V$
Urea, mg/dL	13	0.76	7.11	0.72	8.31
Creatinine, mg/dL	14	0.71	0.16	0.51	0.22
Log <sub>10</sub> NEFA	9	0.68	0.27	0.55	0.33
Log <sub>10</sub> TBIL	11	0.76	-0.27	0.69	0.20
Glucose, mg/dL	14	0.44	7.63	0.29	8.67
Cholesterol, mg/dL	15	0.58	18.08	0.24	25.06
Triglyceride, mg/dL	9	0.73	6.39	0.60	7.71

LV = Latent Variables



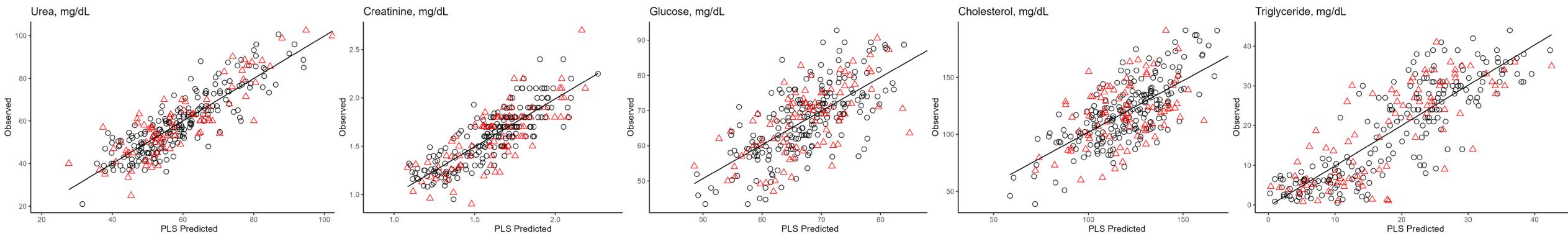
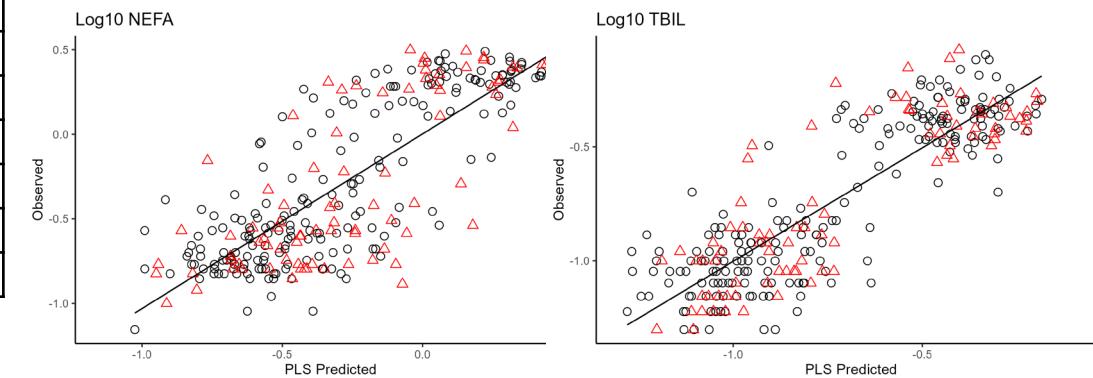
● calibration set, ▲ validation set

# RESULTS

## Blood trait

Trait	LV	Calibration (70%)		Validation (30%)	
		$R^2_C$	$SE_C$	$R^2_V$	$SE_V$
Urea, mg/dL	13	0.76	7.11	0.72	8.31
Creatinine, mg/dL	14	0.71	0.16	0.51	0.22
Log <sub>10</sub> NEFA	9	0.68	0.27	0.55	0.33
Log <sub>10</sub> TBIL	11	0.76	-0.27	0.69	0.20
Glucose, mg/dL	14	0.44	7.63	0.29	8.67
Cholesterol, mg/dL	15	0.58	18.08	0.24	25.06
Triglyceride, mg/dL	9	0.73	6.39	0.60	7.71

LV = Latent Variables



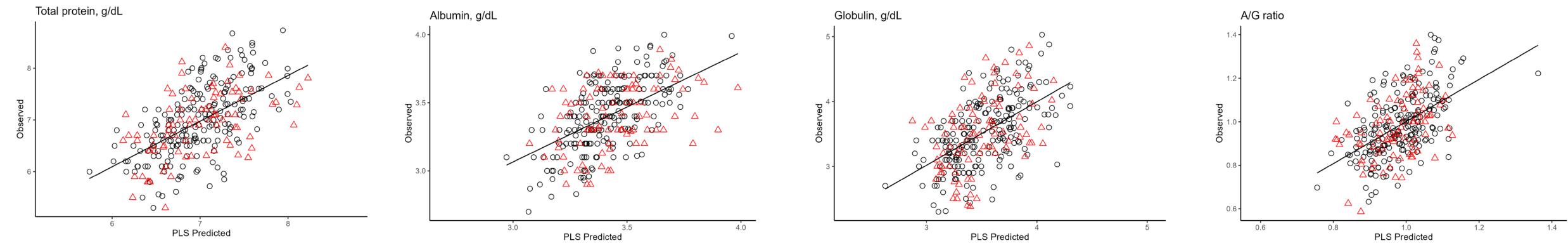
● calibration set, ▲ validation set

## Protein profile

# RESULTS

Trait	LV	Calibration (70%)		Validation (30%)	
		$R^2_C$	$SE_C$	$R^2_V$	$SE_V$
Total protein, g/dL	6	0.20	118.21	0.14	127.53
Albumin, g/dL	10	0.37	0.20	0.15	0.23
Globulin, g/dL	6	0.32	0.45	0.24	0.49
A/G ratio	7	0.32	0.45	0.24	0.49

LV = Latent Variables



● calibration set, ▲ validation set

# CONCLUSIONS

Predict something present in the blood using the milk spectrum is challenging -> **INDIRECT PREDICTION**

- Low-moderate  $R^2$  were as expected
- Phenotypes predicted from the milk spectra can be valuable for genetic investigations at the population level and for design of breeding programs



# CONCLUSIONS

Predict something present in the blood using the milk spectrum is challenging -> **INDIRECT PREDICTION**

→ Low-moderate  $R^2$  were how expected

→ Phenotypes predicted from the milk spectra can be valuable for **genetic investigations** at the population level and **for design of breeding programs**

Concentration of several blood traits (urea, TIBL, NEFA, creatinine, and triglyceride) can be considered as good enough for:

- ✓ POPULATION SCREENING: monitoring metabolic health of buffaloes in dairy herd
- ✓ CARRYING OUT SELECTIVE BREEDING



# CONCLUSIONS

Predict something present in the blood using the milk spectrum is challenging -> **INDIRECT PREDICTION**

- Low-moderate R<sup>2</sup> were how expected
- Phenotypes predicted from the milk spectra can be valuable for **genetic investigations** at the population level and **for design of breeding programs**

Concentration of several blood traits (urea, TIBL, NEFA, creatinine, and triglyceride) can be considered as good enough for:

- ✓ POPULATION SCREENING: monitoring metabolic health of buffaloes in dairy herd
- ✓ CARRYING OUT SELECTIVE BREEDING

Further efforts should be made to understand if MIR spectra coupled with **machine learning algorithms** can result in an improvement of the models' accuracy.





Thank you for  
attending

**Ph.D.** ANIMAL  
**COURSE** & FOOD SCIENCE  
UNIVERSITY OF PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA  
**DAFNAE**  
Department of Agronomy Food Natural  
resources Animals and Environment



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

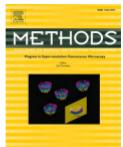
**BIG**  
ANASB • BUFALA MEDITERRANEA ITALIANA  
TECNOLOGIE INNOVATIVE PER IL MIGLIORAMENTO GENETICO

Dipartimento  
Medicina Veterinaria  
Produzioni Animali

UNIVERSITÀ  
DEGLI STUDI DI NAPOLI  
**FEDERICO II**



[silvia.magro.1@phd.unipd.it](mailto:silvia.magro.1@phd.unipd.it)



Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions

C. Grelet <sup>a</sup>, P. Dardenne <sup>a</sup>, H. Soyeurt <sup>b</sup>, J.A. Fernandez <sup>a</sup>, A. Vanlierde <sup>a</sup>, F. Stevens <sup>a</sup>, N. Gengler <sup>b</sup>, F. Dehareng <sup>a,\*</sup>

**Table 4**

Characteristics of the 7 K-mean clusters resulting from the classification of 57 milk MIR models following their mean-centered cross-validation RPD, relative RMSE and  $R^2$ .

Cluster	RPDcv	Relative RMSEcv	$R^2_{cv}$	Interpretation for application
1	> 6	<5%	> 0.97	Any application
2	4.2–6	<10%	0.94–0.97	Quality control
3	3–4.2	<10%	0.89–0.94	Quantitative screening
4	2–3	<25%	0.74–0.89	Rough screening
5	1.5–2	<25%	0.55–0.74	Allows to compare groups, discriminate high or low values
6	1.5–2	>25%	0.55–0.74	Highly imprecise, can be used to detect extreme values
7	< 1.5	–	< 0.55	Not recommended

## Near-infrared technology: getting the best out of light: a short course in the practical implementation of near-infrared spectroscopy for the user

P. Williams - 2004 - PDK Projects, Incorporated

- $R^2$  between 0.50 and 0.65 -> detection of extreme values or comparison of groups.
- $R^2$  between 0.66–0.81 -> approximate screening
- $R^2$  between 0.82-0.90 -> good quantitative screening
- $R^2 \geq 0.91$  -> punctual prediction