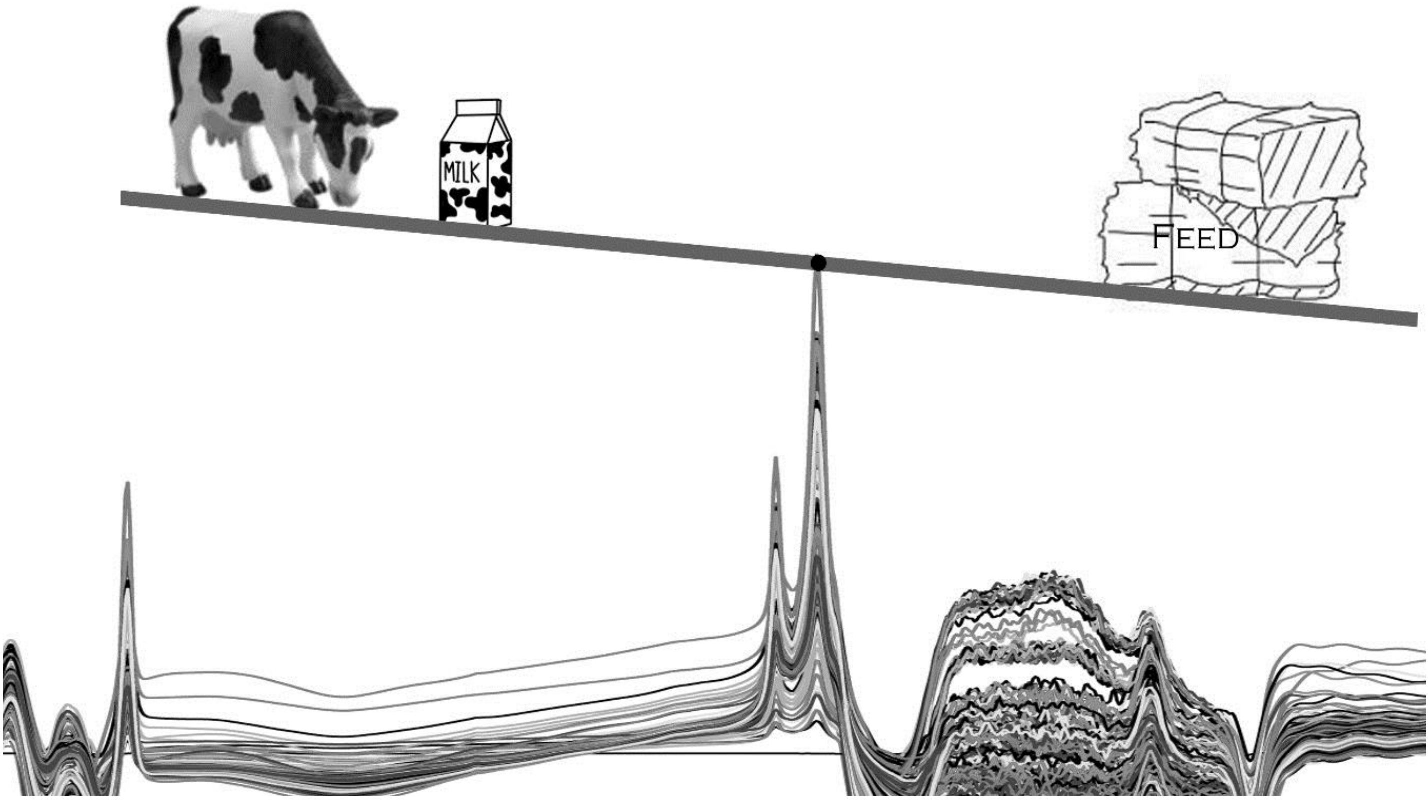
LIÈGE université
Gembloux
Agro-Bio Tech

Prediction of feed efficiency based on test-day liveweight of dairy cows estimated using animal characteristics and milk mid-infrared spectra



Lei Zhang

COMMUNAUTÉ FRANÇAISE DE BELGIQUE

UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

Prediction of feed efficiency based on test-day liveweight of dairy cows estimated using animal characteristics and milk mid-infrared spectra

Lei Zhang

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteur: Prof. Hélène Soyeurt

Année civile: 2021

This dissertation is dedicated to
my grandfather, the late Yi Zhang (张义).

You had ever told me: "Never to be too late to get a start, once you have settled your mind."
Since young, I have dreamt of collecting Cigar from the world for you, and now I can, but I lost you.
I will keep you and your words in my mind forever!

Abstract

Due to the increasing world population, the consumption of milk and dairy products is raising. Optimizing the production of milk under the economic and environmental constraints is a challenge. In this context, knowing the feed efficiency (FE) of dairy cow is a key issue. Unfortunately, the acquisition of such records at individual and large scales is unfeasible. So, the current thesis aims to fill this gap by developing a FE predictive tool for dairy sector available at large and individual scales. To achieve this objective, the FE was assessed by calculating the ratio of fat and protein corrected milk (FPCM) to dry matter intake (DMI). Although FPCM is already routinely available, this is not the case for DMI. However, a literature equation exists to predict DMI from FCM, live bodyweight (BW) and the number of weeks of lactation. Recently, a methodology using the milk mid-infrared (MIR) spectrometry combined with the animal characteristics was developed to predict test-day BW records, leading to open new perspectives about the use of FE related traits for management and breeding purposes. So, the final achievement of this thesis is the development and the implementation of a FE prediction tool from traits easily recorded by dairy herd improvement (DHI) organizations, including namely the BW estimation. To achieve this objective, 3 main research activities were conducted in this thesis.

The first research aimed to increase the calibration set and to apply feature selection algorithms during the modeling in order to improve the BW models' robustness and accuracy. Indeed, the presence of less informative variables in a prediction equation could impact negatively its robustness. Three feature selection algorithms were applied on 280 predictors to select the most informative ones from a dataset containing 5,920 records: partial least squares regression (PLS) combined with sum of ranking difference (SRD), PLS combined with uninformative variables elimination (UVE), and the output of Elastic net regression (EN). Parity, days in milk (DIM), milk yield (MY), and two MIR spectral points were selected as the most relevant variables to predict BW. Validation root mean square errors (RMSEp) of 60 kg were obtained for both PLS and EN regressions employing these 5 predictors, suggesting a better robustness of these models compared to the ones without MIR or using all 277 MIR variables. The RMSE values of validation set coming from another brand of spectrometer were around 64 kg.

The second research work focused on the implementation of a BW equation in practice. Indeed, some poor quality BW predictions can be obtained using poor quality spectral data or by applying the model on samples for which the variability was not included in the calibration set. So, the objective of this work was to develop data cleaning methods easy to implement by DHI to ensure the quality of BW predictions. So, 3 data cleaning procedures and their combinations were tested on a DHI dataset containing 346,818 records: the deletion of 1% of extreme high and low predicted values (M1), the deletion of records when the global-H (GH) distance was greater than 5 (M2), and the deletion of records if the absolute fat residual value was higher than 0.30 g/dL of milk (M3). The interest of those procedures was assessed by estimating the root mean square differences (RMSD) between fat, protein, and fatty acids traits predicted by the MIR spectrometry internally and externally. All methods allowed to decrease RMSD, the gain ranged from 0.32% to 41.39%. Based on the obtained results, the “M1 and M2” combination should be preferred to be more parsimonious in the data loss as it had the higher ratio of RMSD gain to data loss. However, to ensure the lowest RMSD, the “M2 or M3” combination was the most relevant.

Based on these 2 first works, FE records were easily obtained at large and individual scales and cleaned appropriately. Then, in order to assess the relevancy of the FE tool, the final work of this thesis consisted to study the behaviors of BW and FE predictions obtained from the 5 developed equations within and between lactations as well as per test month and compared them to the ones observed in the literature using reference values. Subsets of Hebei (N=288,607) and Walloon (N=379,472) DHI datasets were used. Even if the BW equations did not differ a lot based on the prediction performances, differences were observed on the DHI datasets and clustered the equations within 2 groups: the ones including 5 predictors and the one using the parity, DIM, MY, and the full MIR spectral data. The final one depicted a more expected evolution within lactation with a drop of BW around 30-50 days in milk and then an increase. This was not observed for the other equations suggesting a lack of MIR information to take into account sufficiently the individual variability of BW. The annual trend observed for this equation was also more expected with a BW drop during the grazing period. However, the differences in FE predictions obtained using BW estimated from different equations were less marked, suggesting a low sensibility of FE predictions to a moderate variation of BW. The validation RMSE reached 0.05 suggesting a good accuracy for this FE indicator.

In conclusion, the FE tool developed in this thesis can be implemented by DHI organizations based on the BW equation including milk, parity, DIM, and the 277 MIR spectral points. However, some additional investigations are still needed before the use of a such tool by DHI. Indeed, even if preliminary results obtained in this thesis suggested a moderate heritability of FE trait, a study with the newly developed BW equations could be done as we have observed that the past BW equation tended to overestimate BW. Moreover, knowing the relationships between FE predictions with other traits having economic interest is required before any use. A reflection about the best way to communicate the results to the farmers must be also started. Finally, the FE tool was based on an equation predicting DMI. For the future, it could be also of interest to measure the relevancy of this equation by using DMI reference data.

Key words: Milk; Mid-infrared; Body weight; Feed efficiency; Accuracy

Résumé

A cause de l'accroissement de la population mondiale, la consommation de lait et ses dérivés est en augmentation. L'optimisation de la production laitière avec les contraintes économiques et environnementales actuelles représente un challenge. Ainsi, connaitre l'efficience alimentaire (EA) de la vache est un élément clé. Malheureusement, l'acquisition de telles données à large échelle est infaisable. La présente thèse vise à combler ce manque en développant un outil de prédiction de l'EA d'une vache utilisable à large échelle. Pour ce faire, l'EA a été évalué via le calcul du rapport entre le lait corrigé par le taux en matières grasses et en protéines (LCFP) et la quantité de matières sèches ingérées (MSI). Bien que le LCFP soit déjà disponible en routine, ce n'est pas le cas pour MSI. Or, une équation publiée dans la littérature propose d'estimer le MSI d'une vache à partir de son LCFP, son nombre de semaines en lactation et son poids. Une méthodologie récente utilisant la spectrométrie moyen-infrarouge du lait et les caractéristiques de la vache permet de prédire son poids à un jour de contrôle donné. Cette trouvaille a ouvert la voie à l'emploi de caractères reliés à l'EA à des fins de gestion et de sélection des troupeaux. Ainsi, l'objectif général de cette thèse consiste à développer et implémenter un outil de prédiction de l'EA à partir de caractères facilement enregistrables par les organismes de contrôle laitier (CL), ceci incluant l'estimation du poids. Pour ce faire, le travail de recherche a été divisé en 3 parties.

La première visait à augmenter le jeu de calibrage et à appliquer des algorithmes de sélection de variables pendant la modélisation afin d'améliorer la robustesse et la justesse du modèle prédisant le poids. En effet, la présence de variables non informatives dans un modèle prédictif peut impacter négativement sa robustesse. A partir d'un jeu de données contenant 5 920 enregistrements, 3 algorithmes de sélection de variables ont été appliqués sur 280 prédicteurs: une régression aux moindres carrés partiels (PLS) combinée à la somme des différences de classement, une PLS combinée à une élimination des variables non-informatives, et un algorithme basé sur la sortie d'une régression de type ElasticNet (EN). La parité, le jour en lactation (JEL), le rendement laitier (RL) et 2 points spectraux ont ainsi été considérés comme les prédicteurs du poids les plus informatifs. Des erreurs quadratiques moyennes (EQM) de 60 kg ont été obtenues pour PLS et EN incluant les 5 prédicteurs, suggérant ainsi une bonne robustesse de ces modèles comparés à ceux n'utilisant pas l'information MIR ou ceux incluant 277 points spectraux. Les EQM du jeu de validation issu d'une autre marque de spectromètre étaient de l'ordre de 64 kg.

La seconde partie s'est focalisée sur l'implémentation de la prédiction du poids en routine. Des prédictions de poids erronées peuvent être obtenues en utilisant des spectres de mauvaise qualité ou en effectuant une prédiction pour des échantillons dont la variabilité n'a pas été considérée dans le jeu de calibrage. Ainsi, le travail mené dans cette partie visait à développer des procédures de nettoyage des données facile à implémenter par les CL pour assurer la qualité des prédictions obtenues. Trois procédures de nettoyage de données ainsi que leurs combinaisons ont été testées sur un jeu du CL incluant 346 818 enregistrements: la suppression des 1% de prédictions hautes et basses (M1), l'élimination des données sur base d'une distance global-H (GH) supérieure à 5 (M2), et finalement la suppression des enregistrements sur base d'un résidu de la matière grasse exprimé en valeur absolue supérieur à 0.30 g/dl de lait (M3). L'intérêt de ces procédures a été évalué en estimant l'erreur quadratique moyenne des différence (EQMD) entre la matière grasse, la protéine et les acides gras prédits par la spectrométrie MIR en interne de la machine ou en externe. Toutes les méthodes ont permis de diminuer le EQMD; le gain étant compris entre 0.32% et 41.39%. A partir de ces résultats, la combinaison « M1 & M2 » a été préférée afin de conserver un maximum de données alors que la combinaison « M2 ou M3 » a été plébiscitée pour assurer le plus faible EQMD.

A partir de là, il a été facile d'obtenir des données relatives à l'EA des vaches laitières à large échelle et de les nettoyer de manière appropriée. Ensuite, afin d'évaluer la pertinence de l'outil de prédiction de l'EA, le travail final a consisté à étudier les comportements, entre et au sein des lactations ainsi que par mois de test, des prédictions de poids et de EA obtenus en utilisant les 5 équations de poids développées et de les comparer aux comportements attendus sur base de la littérature à partir de données de référence. Des sets de données du CL de Hebei (N=288 607) et de Wallonie (N=379 472) ont été utilisés. Même si les équations de poids ne se différenciaient pas beaucoup sur base de leurs performances prédictives, des différences ont été observées à partir des sets des CL séparant en 2 groupes les équations : les équations incluant 5 prédicteurs et celle incluant le JEL, la parité, le RL et les 277 points spectraux. Cette dernière équation reflétait mieux l'évolution attendue au sein d'une lactation du poids avec une baisse de ce dernier aux environs du 30ème-50ème JEL et puis une augmentation. Cela n'a pas été observé pour les autres équations suggérant un manque de données MIR pour prendre en compte suffisamment la variabilité individuelle du poids. L'évolution annuelle observée était également plus en adéquation pour l'équation incluant tous les points spectraux avec une baisse de poids observée pendant la période de pâturage. Cependant, les différences observées entre les prédictions d'EA utilisant les estimations de poids obtenues au travers des 5 équations étaient moins marquées, suggérant une faible sensibilité de la prédiction de l'EA à un changement modéré du poids. L'EQM de validation obtenu pour l'EA était de 0.05 montrant ainsi une bonne justesse dans la prédiction de cet indicateur de l'EA.

En conclusion, l'outil prédictif de l'EA développé dans cette thèse peut être implémenté dans les CL à partir de l'équation du poids incluant le RL, la parité, le JEL et les 277 points spectraux. Cependant, des investigations doivent encore être menées avant que le CL n'utilise l'outil. En effet, même si les résultats préliminaires obtenus dans cette thèse suggèrent une héritabilité modérée du caractère EA, une étude utilisant la nouvelle équation de poids doit être réalisée car il a été montré que l'ancienne équation avait tendance à sur-estimer le poids. De plus, la connaissance des relations entre les prédictions de l'EA et d'autres caractères d'intérêt économique est nécessaire. Une réflexion à propos de la meilleure manière de communiquer les résultats aux producteurs doit également être envisagée. Finalement, l'outil développé utilise une équation de prédiction du MSI. Il serait plus que pertinent dans le futur de valider cette équation avec des données de MSI de référence.

Mots clés : Lait; Moyen infrarouge; Poids; Efficacité alimentaire; Précision

Acknowledgments

This PhD thesis was completed under a grant given by the Chinese Scholarship Council (CSC). So, first of all, I would like to express my sincere gratitude to my country for this funding, which allowed me to complete most of my PhD studies. Also, I would like to thank the University of Liège for the financial support which allowed me to finalize my PhD thesis in Belgium.

I would like to thank my supervisor Prof. Helene Soyeurt. You brought me into this MIR field and made me feel endless fun. I sincerely thank you for your meticulous help point by point, and sometimes even sacrifice your rest time. I appreciate that very much. I would like also to acknowledge the members of my thesis committee for giving me a lot of help: thanks to Mr. Eric Froidmont (CRA-W) for helping me to collect data and find information about the dairy industry in Europe; thanks to Mr. Frédéric Dehareng (CRA-W) for helping, encouraging and guiding me to gain more knowledge and to better communicate the results of my research; thanks to Prof. Nicolas Gengler for suggestions on my PhD study and the moments of discussion; thanks to Miss Amélie Vanlierde (CRA-W) for your kindly tips about the tele-conference and suggestions on my thesis; finally, thanks also to Mrs. Chunfang Li and the staff from Hebei DHI lab for the collaboration which making this thesis more interesting. I also appreciate the co-authors and contributors in my publications for their help. A sincere thank of mine to Prof. Tom Fearn from University College London and Prof. John Kalivas from Idaho State University for sharing knowledge with me.

I would like also to thank the members of our research team (SIMA) for your help: many thanks to Madam Dominique Marchal for her meticulous help in the office affairs so much; great thanks to Prof. Yves Brostaux for your help in statistics and software; many thanks to Mr. Lionel Luxen for your kind help with computer equipments and all aspects; I acknowledge Miss Anne-Catherine Dalcq for her help in collecting Belgian dairy farming data on the web; I am also grateful to my colleagues Mr. Anthony Tedde and Miss Pauline Delhez for the help in my study. A special thank to my colleague Miss Mireille Degraeuwe for the tremendous help and happiness; Many thanks to Mr. Clement Grelet (CRA-W) for helping me in the study. I also deeply appreciate the other members in our team for the kind help in all my problems.

Last but not least, I must thank my family for waiting for me and the understanding of my grandmother and parents. Special thanks to my wife, you encouraged me to be more brave. Without you, I cannot go even one step further.

Lei ZHANG

Table of Contents

Abstract i

Résumé i

Acknowledgments iii

Table of Contents iv

List of Tables vii

List of Figures ix

List of Abbreviations xi

Chapter I : General introduction 1

- 1. Milk consumption and production 4
- 2. Interest of feed efficiency and its measurement 7
- 3. Prediction of feed efficiency at large and individual scales 12
- 4. Improvement of feed efficiency related predicted phenotypes at large scale 14
- 5. Research framework and outline 16
- References 18

Chapter II : Mining data from milk mid-infrared spectroscopy and animal characteristics to improve the prediction of dairy cow's liveweight using feature selection algorithms based on partial least squares and Elastic net regressions 27

- Abstract 30
- 1. Introduction 31
- 2. Materials and Methods 33
  - 2.1 Modeling dataset 33
  - 2.2 Prediction models 34
  - 2.3 Feature selections 35
  - 2.4 External validation 37
  - 2.5 Computation 38
- 3. Results 38
  - 3.1 Descriptive statistics 38
  - 3.2 Predictive models and informative variable selection 39
  - 3.3 The robust model and variable importance 45
- 4. Discussion 47
- 5. Conclusion 50
- 6. Acknowledgments 51
- References 52

Chapter III : Improving the quality of milk mid-infrared predictions recorded by dairy herd improvement organizations using data quality checks 56

Abstract 59

1. Introduction................ 60

2. Materials and Methods 62

    2.1 Data 62

    2.2 Data cleaning techniques 63

3. Results 65

    3.1 Comparison Between Manufacturer’s and Externally Predicted Phenotypes................ 65

    3.2 Deletion of Extreme Predicted Phenotypes (Method 1) 66

    3.3 GH-based Data Cleaning (Method 2)................ 67

    3.4 Data Cleaning Based on the Absolute Fat Residual Limit (Method 3) 69

    3.5 Comparison of Three Tested Data Cleaning Methods................ 70

4. Discussion................ 71

5. Conclusions................ 76

6. Acknowledgments 77

References 78

Chapter IV : Can we observe expected behaviours at large and individual scales for feed efficiency-related traits predicted partly from milk mid-infrared spectra? 81

Abstract 84

1. Introduction................ 85

2. Materials and Methods 86

3. Results 87

    3.1 Parity effect................ 87

    3.2 Effect of Lactation Stage 88

    3.3 Correlations Between Studied Traits................ 89

    3.4 Test month influence 90

    3.5 Heritability................ 91

4. Discussion................ 91

    4.1 Parity effect................ 91

    4.2 Effect of Lactation Stage 92

    4.3 Correlations Between Studied Traits................ 93

    4.4 Test month influence 94

    4.5 Heritability................ 94

5. Conclusions................ 95

6. Acknowledgments 95
References 96
Chapter V : General discussion.. 101
1. Context 104
2. Body weight predictive models and derived feed efficiency tool 104
3. Predictions of bodyweight from easily recorded features.. 107
3.1 Cleaning procedures... 107
3.2 Evolutions within and between lactations of BW predictions from reference datasets. 110
3.3 Bodyweight predictions from dairy herd improvement datasets 113
4. Predictions of feed efficiency at large and individual scales 119
4.1 Prediction of feed efficiency from reference datasets 119
4.2 Prediction of feed efficiency from dairy herd improvement datasets 125
Conclusions and perspectives 133
References 137

List of Tables

Table 2-1 Descriptive statistics of cleaned data sets. 39

Table 2-2 Calibration and validation performances obtained for the 16 developed models predicting test-day body weight (kg) of dairy cows. 40

Table 2-3 Mean and standard deviation of cross-validation and validation root mean squared errors (RMSE) per kind of models as well as the absolute difference between cross-validation RMSE (RMSEcv) and validation RMSE (RMSEp, RMSEv). 41

Table 2-4 Parameters of the variable selection algorithms based on PLS-SRD, PLS-UVE and Elastic Net. 42

Table 2-5 The relative importance of selected variables of the Fin-PLS models using different calibration sets. 47

Table 3-1 Prediction performances of the external equations used to predict the contents of five milk components 63

Table 3-2 Descriptive statistics of manufacturer's and externally predicted phenotypes (g/dL of milk) and their relationship 66

Table 3-3 Characteristics of the dataset obtained after a cleaning using as thresholds, the externally predicted values observed at the 1% and 99% quartiles (Method 1) 67

Table 3-4 Descriptive statistics of global-H (GH) distances and their correlation coefficient (r) with the squared differences between manufacturer's and externally predicted traits (e^2) 67

Table 3-5 Characteristics of the dataset obtained after a cleaning based on a global-H (GH) distance set at 5 (Method 2). 68

Table 3-6 Characteristics of the dataset obtained after a cleaning based on an absolute fat residual limitation set at 0.30 g/dL of milk (Method 3) 69

Table 3-7 Interest of using the three studied methods and their combinations to clean the raw dataset 71

Table 4-1 Descriptive statistics of the studied datasets. 87

Table 4-2 Correlation coefficients observed between studied traits in first (below the diagonal) and second lactation (above the diagonal). 89

Table 4-3 Heritability (h^2) of studied traits for the first two lactations. 91

Table 5-1 The prediction performance of the 5 developed models predicting test-day bodyweight (kg/day) presented in Chapter 2 and the equation obtained by Soyeurt et al. (2019). 105

Table 5-2 Descriptive statistics of the assumed reference and predicted feed efficiency and their corresponding root mean square error (RMSE). 107

Table 5-3 The prediction improvement observed on the Australian dataset (N=3,989) using all proposed cleaning methods and their combinations. 109

Table 5-4 The descriptive statistics of the feed efficiency related traits of calibration/test and validation sets. 110

Table 5-5 Descriptive statistics of the cleaned dairy herd improvement datasets. 114

Table 5-6 The Pearson correlation coefficients among predicted body weight from the Walloon dataset (above the diagonal) and from the Hebei dataset (below the diagonal). 114

Table 5-7 Descriptive statistics of the feed efficiency related traits predicted on two dairy herd improvement datasets. 115

Table 5-8 The Pearson correlation coefficients of the feed efficiency related traits ($p < 0.001$). 131

List of Figures

Figure 1-1. The evolution of milk production and consumption in the world. 4

Figure 1-2. The scheme of nutrients intake and allocation of dairy cows. 7

Figure 1-3. The processes of methane producing under the background of nutrients digestion of dairy cows (Morgavi et al., 2010, Ungerfeld, 2020). 8

Figure 1-4. The evolution of feed efficiency (FE) by different variations. 12

Figure 1-5. Mid-infrared spectrometry (©Bentley instrument, 2019)...14

Figure 1-6. The process of development and application of MIR prediction model. 15

Figure 1-7. The overview of contents in this thesis. 16

Figure 2-1. The technical flow chart of the different modeling approaches used in this study. 33

Figure 2-2. Workflow of sum of ranking differences (SRD) algorithm 36

Figure 2-3. Distribution of parity (A), days in milk (B), and body weight (C), and the body weight evolution by parity (D) from the cleaned modeling data sets. 39

Figure 2-4. The trend of RRM value across the model number for the 4 tested calibration sets (A=first dataset until D=fourth dataset; the red arrow shows the optimal model)...43

Figure 2-5. Intersection of groups of selected variables among the PLS-SRD, PLS-UVE and EN feature selections and the four calibration sets. 44

Figure 2-6. The observed and predicted body weight of the calibration and validation herds (herd7) based on the EN regression using the final 5 selected variables (A); The validation root mean square error distribution across different body weight stage (interval=60 kg; B)..46

Figure 4-1. Evolution of predicted body weight (A), milk yield (B), predicted dry matter intake (C), and predicted consumption index (D) throughout lactation. 88

Figure 4-2. The evolution of the averaged herd least squares means (LSMEANS) for all studied traits across test months. 90

Figure 5-1. The least squares means evolution per class of 15 days in milk and for the first three lactation for the predicted test-day bodyweight. A, C, and E represent the first, second and third parity for the validation set (Australian set), respectively; B, D, and F are the first, second and third parity for the calibration/test set, respectively. BW1 to BW5 are the bodyweight predicting using Mod1 to Mod5; BWref is the observed bodyweight. 112

Figure 5-2. Evolution of least squares means of body weights predicted using the 5 available models between and within lactations; A, C, E and B, D, F denote the first, second, and third parity of cows from Hebei and Walloon dataset respectively; Mod0 is the body weight predicted using the equation proposed by Soyeurt et al. (2019). 117

Figure 5-3. The evolution of the least squares means of different predicted body weights by test month. A and B represent the Hebei and Walloon datasets, respectively. 118

Figure 5-4. The distribution of assumed reference feed efficiency (FE) and predicted FE for calibration/test, validations sets and DHI data sets. 120

Figure 5-5. The least squares means evolution of the predicted feed efficiency within lactation. A, C, and E represent the first, second and third parity for the validation set (Australia dataset), respectively; B, D, and F are the first, second and third parity for the calibration set, respectively. FE1 to FE5 are the corresponding feed efficiency predicted using bodyweight records predicted from Mod1 to Mod5; FEref is the feed efficiency calculated using the observed body weight. 122

Figure 5-6. The least squares means evolution of the fat- and protein- corrected milk production within and between lactation across the lactation stage. A represents the validation set (Australia dataset); B respects the calibration/test set...123

Figure 5-7. The least squares means evolution of the predicted dry matter intake (DMI) within lactation by the lactation stage. A, C, and E represent the first, second and third parity for the validation set (Australia dataset), respectively; B, D, and F are the first, second and third parity for the calibration set, respectively. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMIref is the dry matter intake calculated using the observed body weight...124

Figure 5-8. Evolution of least squares means of feed efficiency derived from the body weight predicted using the models Mod0 to Mod5 within and between lactations; A, C, E and B, D, F denote the first, second, and third parity of cows from Hebei and Walloon dataset, respectively. 127

Figure 5-9. Evolution of least squares means of feed efficiency derived from the body weight predicted using the models (Mod0 to Mod5) across test month; A and B represent Hebei and Walloon datasets, respectively. 128

Figure 5-10. The least squares means evolution of the fat- and protein- corrected milk production (FPCM) within and between lactation across the lactation stage (A, B) as well as the least squares means evolution of FPCM by month of test (C, D). A and C represent the Hebei set; B and D respect the Walloon set. 129

Figure 5-11. The least squares means evolution of the predicted dry matter intake (DMI) within lactation by the lactation stage. A, C, and E represent the first, second and third parity for the Hebei set, respectively; B, D, and F are the first, second and third parity for the Walloon set, respectively. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMI0 is the dry matter intake calculated using the predicted body weight of Mod0. 130

Figure 5-12. The least squares means evolution of predicted dry matter intake by month of test. A represents the Hebei set; B respects the Walloon set. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMI0 is the dry matter intake calculated using the predicted body weight of Mod0..131

Figure 5-13. The FE evolution by dry matter intake from the calibration/test and validation sets. 132

List of Abbreviations

%FAT: Fat contents
%PROT: Protein contents
BCS: Body condition score
BLUP: Best linear unbiased prediction
BW: Live body weight
C value: Reliability of predictor
DHI: dairy herd improvement
DIM: Days in milk
ECM: Energy corrected milk yield
EN: Elastic net regression
EN-FS: EN regression using variables based on feature selection
FCM: 4%Fat corrected milk yield
FPCM: Fat and protein corrected milk yield
GH: Global H distance
HTD: Herd test-day model
ICAR: International Committee for Animal Recording
LSMEANS: Least squares means
LV: Latent variables
MIR: Mid-infrared spectrum
MY: Milk yield
pBW: Predicted body weight
PC: Principal components
pDMI: Predicted dry matter intake
pIC: Predicted consumption index
PLS: Partial least squares regression
PLS-SRD: Partial least squares regression (PLS) combined with sum of ranking difference
PLS-UVE: PLS combined with uninformative variables elimination
PLS-NoMIR: PLS regression without MIR spectrum
PLS-All: PLS regression using animal characteristics and MIR spectrum
PLS-FS: PLS regression using variables based on feature selection
RMSE: Root mean square error
RMSEPv: Root mean square error of validation
RPD: Ratio of performance to deviation
RRM: Rank reliability measure
SD: Standard deviation
SRD: Sum of ranking difference
THI: Temperature humidity index
UVE: Uninformative variables elimination
WHO: World Health Organization

1

Chapter I : General introduction

Simple summary

This chapter summarized the current interests and limitations regarding the feed efficiency related traits in the dairy farming sector and the utility of mid-infrared technique in this context. The outline of this thesis study is also presented in this chapter.

1. Milk consumption and production

Milk and dairy products are considered as one of the most complete nature's food since it provides essential and various nutrients to human beings including fat (around 3.70g per dL of milk), protein (around 3.30g per dL of milk), and micronutrients like vitamins (ranged from 0.04 to 0.16 mg per dL of milk) and minerals (between 0.004 and 151 mg per dL of milk) (Park et al., 2013, USDA, 2018). Those nutrients participate in the metabolism and physiology process in the human body. For instance, the high-quality protein provided by milk is an essential structural component of cells and enzymes in the human body (Singhal et al., 2017); the milk lipids serve as the energy source of the body and provide saturated fatty acids for human brain's development (Kratz et al., 2013); the milk mineral trace-elements (e.g. Copper, Iron, Zinc, Selenium, etc.) play an essential role in the body physiological processes such as the cell metabolism and the enzyme synthesis (WHO, 1996); and the mineral macro-elements (e.g. Calcium, Phosphorus) are relevant for body bones and teeth health (Cashman, 2006). From a global perspective, milk provides around 5%, 10%, and 9% of the global energy, protein, and fat supply capita for humans per day on average if we consider a worldwide normal diet (Muehlhoff et al., 2013).

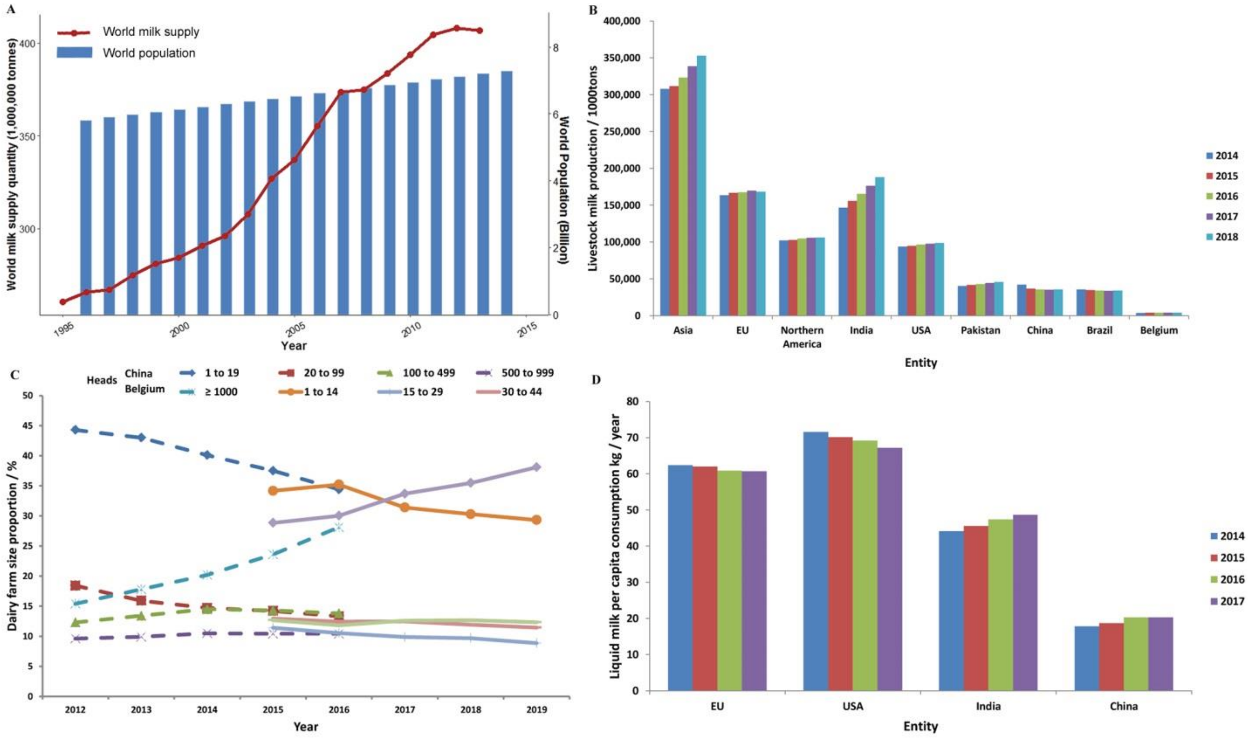


Figure 1-1. The evolution of milk production and consumption in the world.

The world supply quantity of whole milk and population across the year (A, FAO, 2020); the milk production amount of different regions and countries (B, FAO, 2019); The evolution of farm size by year of Belgium and China (C, Belgium, STATBEL, 2019; China, Koning et al, 2016); the liquid milk per capita consumption of different countries (D, China dairy statistical summary, 2019).

On account of the crucial role of dairy products in human daily life, the consumption of milk and dairy products has increased rapidly for decades (Figure 1-1A) in the world as reported by Food and Agriculture Organization of the United Nations (FAO, 2020). At the nations' level, in 2018, India, United States (USA), Pakistan, China, and Brazil are the top five milk producing countries according to the entire livestock milk production quantity (1000 tonnes per year, Figure 1-1B). According to a report from World Health Organization (WHO), the increase of consumption of milk and dairy products could be related to the following reasons: the population increase (Figure1-1A) in developing countries such as India, Pakistan, and China; and the worldwide economic expansion which is usually followed by an improvement of a country’s food supply, thereby improving the overall nutritional status of their people which leads to an increase of milk and dairy products consumption (WHO, 2003). The level of economic development was also highlighted by Muehlhoff et al. (2013) as a key factor explaining the increase in milk consumption. Moreover, following the report written by Rong et al.(2011), the dietary habit is another important factor. This last statement could explain why, at the per capita level (Figure 1-2D), the developed countries like the United States of America (USA) and the European Union (EU) have the most individual consumption (around 60−65 kg liquid milk per person per year). However, the dietary habit also explains why developing countries like China (around 20 kg liquid milk per person per year) and India (around 50 kg liquid milk per person per year) are lower consumers even if they have growing population and economic (Figure 1-2D). But, the milk consumption in these countries tend to increase (China dairy statistical summary, 2019). So, the world per capita consumption of fresh dairy products is projected to increase by 1.0% per annum over the coming decade which will be mainly driven by the consumption demand from the developing countries (FAO, 2020). So, there is a need to produce sufficient milk and dairy products to fulfill the requirement given by the increasing economics and population expansion globally, particularly in developing countries.

Driven by the increasing demand for milk, the average herd size has increased nearly 2-fold in most countries: from 200 to 400 heads for New Zealand, from 130 to 270 heads for Australia, from 75 to 175 heads for the US, from 27 to 53 heads for Germany approximately between 1996 and 2014 (Barkema et al., 2015). So, over the last few years, dairy farming has shifted towards larger herds and greater annual milk yield per individual cow to be productive and profitable (Park et al., 2013). In Belgium (Figure 1-1C), the proportion of dairy farms containing 1 to 14 heads of cows reduced from 34.16% in 2015 to 29.32% in 2019; however, the proportion of dairy farms having more than 60 animals increased from 28.84% in 2015 to 38.09% in 2019 (STATBEL, 2019). Similarly, in China (Figure 1-1C), the percentage of dairy farms having 1 to 19 heads cows decreased from 44.30% in 2012 to 34.40% in 2016; whereas the rate of the dairy farms containing more than 1,000 heads of cows increased from 15.40% in 2012 to 28.20% in 2016 (Koning et al., 2016). Accompanying the increasing size of herds, intensive and precise care to the individual animal are required to maintain high quality of animal well-being (Song et al., 2018) and, consequently, to ensure a good profitability. Regarding cow health, such care at individual scale is more needed in larger size herds because the higher risk is encountered by these herds compared to the ones with a smaller size. Indeed, a higher risk of lameness exists when a parlor milking system is used leading to a lack of cow comfort and an increased standing time (Westin et al., 2016). An intensive in-door condition mainly encountered in large herds involves also a higher risk factor for hock lesions, mastitis, and injuries of animals and contributes to higher mortality of cattle (Crump et al., 2019). Moreover, the rapid intensification of dairy farm considerably affects the animals' physical and mental well-being since the lack of the natural behaviors of cows than it is in grazing condition, leading to animals initiatively seeking more comfort (Clay et al., 2020). So, nowadays, dairy farming requires higher skill labors and enhanced management experiences farmers to provide more professional and timely management to cows at an individual level (Berckmans, 2006). Indeed, there is no doubt about the benefits for the profitability of farms of the early and quick detection of the dairy cow having abnormal behaviors or physiological imbalance (e.g. energy imbalance) in order to provide in time the required treatment to those cows and, therefore, to improve the herd health status, production, and farm profitability (Moyes et al., 2013). However, it is becoming more and more difficult for producers to conduct cow-based management due to the higher amount of cows in the herd and the limited labor resources which results from the changing of family structure (less family workers on the farm) (Barkema et al., 2015). To cope with this issue, the adoption of new technologies (e.g., automated calf feeders, cow activity monitors, and automated milking systems) is accelerating to pursue smart dairy farming (Barkema et al., 2015).

Besides the management problems brought by the growing dairy herd size, the milk production has an inevitable impact on the environment (Cheng et al., 2020). Indeed, the global dairy sector is the second larger contributor of greenhouse gas (GHG; 5.99 million tonnes CO2 -eq per annum) in the human-induced emissions (Gerber et al., 2013, FAOSTAT, 2017). This is partly due to the enteric fermentation of animals which naturally produces methane (CH_4) contributing to around 20% of the anthropogenic methane emissions (Wuebbles and Hayhoe, 2002, Yusuf et al., 2012). This emission leads also to a loss of energy available for the cow (2-12% gross energy intake) (Johnson et al., 1994). An increase of the CH_4 quantity in the atmosphere (+1081 ppb) was observed between 1750 and 2011; the average level of CH_4 in the atmosphere observed in 2011 (1,803 ± 2 ppb) is unprecedented in at least 650,000 years (Myhre et al., 2014). Moreover, even if the radiative forcing of CH_4 is lower than the one observed for CO_2 (Myhre et al., 2014), methane was estimated to have a Global warming potential (GWP) of 28-36 (relative to CO_2) over the last century (EPA, 2020). These facts reinforce the interest of reducing methane emission.

It is interesting to notice that mitigating the quantity of methane eructed by dairy cows has a beneficial effect on the environment but also on the cost of production. This is why developing strategies to mitigate the emission of methane eructed by dairy cows is a current hot topic for many dairy herd improvement (DHI) organizations, breeding companies and research institutions. However, the obtaining of individual methane measurements at cow scale is unfeasible as such quantification is done using respiration chambers, sulfur hexafluoride (SF6) tracer technique, or GreenFeed system (Hristov et al., 2016) which are not always available on commercial farms. Some alternatives were developed through the past years using proxies to quantify it (Negussie et al., 2017) such as the milk MIR (Dehareng et al., 2012), rumen meta-genome profiles (Ross et al., 2013), or milk fatty acid profiles (Chilliard et al., 2009, Dijkstra et al., 2016). Although the accuracy differs widely between methods, the accuracy of those proxies can be considered as moderate and suggests the interest of considering additional traits related to this topic.

2. Interest of feed efficiency and its measurement

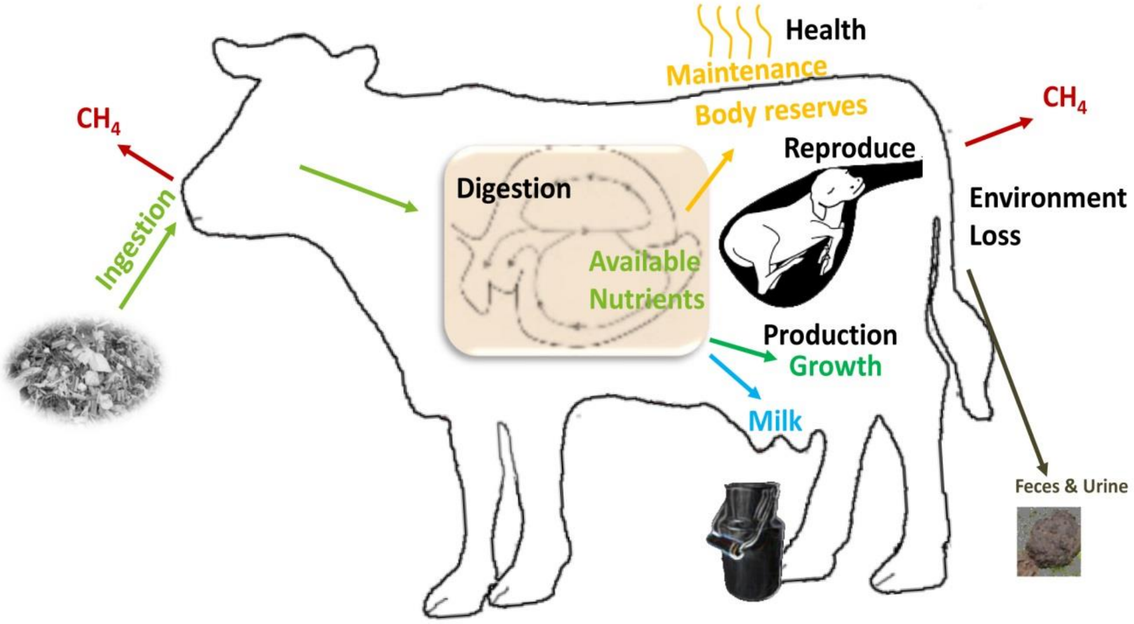


Figure 1-2. The scheme of nutrients intake and allocation of dairy cows.

The carbohydrates from feedstuffs mainly include cellulose, hemicelluloses, starch, and water-soluble carbohydrates. Firstly, the carbohydrates were breakdown into simple sugars. Secondly, during the microbial fermentation, the simple sugars are degraded into pyruvic acid (McDonald. et al., 2011). There are mainly four pathways of the pyruvatic acid to be transformed into 4 different volatile fatty acids (VFA): formic acid, acetic, propionic, and butyric acids. CO_2 and methane are also released in the rumen during this process (Figure 1-3). The rumen fermentation of carbohydrates can be resumed as (Wolin, 1960):

$$57.5\ C_6H_{12}O_6 \rightarrow 65\ \text{Acetate} + 20\ \text{Propionate} + 15\ \text{Butyrate} + 60\ CO_2 + 35\ CH_4 + 25\ H_2O$$

The produced VFAs are then mainly absorbed by the wall of the rumen and provide nearly 70% of the dairy cow energy supply (McDonald. et al., 2011). Meanwhile, the pectin from the carbohydrates is partly breakdown into methanol (Ungerfeld, 2020) which is one of the methane origins in the rumen. Methane can also be produced from CO_2 and H_2 thanks to the hydrogenation effect managed by the methanogens (Giuburunca et al., 2014). Moreover, methane could also be generated from the degradation of formic acid or acetate during the digestion process. So, there are mainly four pathways of methane producing in the cow rumen. To prevent the harm to the rumen due to the produced gas, a total of 95% of methane is eructed by the post-digestive systems of the cow; the remaining 5% is emitted through the fore-digestion system (Doyle et al., 2019). Methane was reported to account for 30-40% of the entire rumen gas proportion (McDonald. et al., 2011). The remaining part of CO_2 is released into the environment by the cow. Microbial cells as well as the remained feed components pass to the post-digestive system including the abomasum and small intestine to be further digested by enzymes. Lastly, the undigested feed components are excreted into the environment through feces (McDonald. et al., 2011).

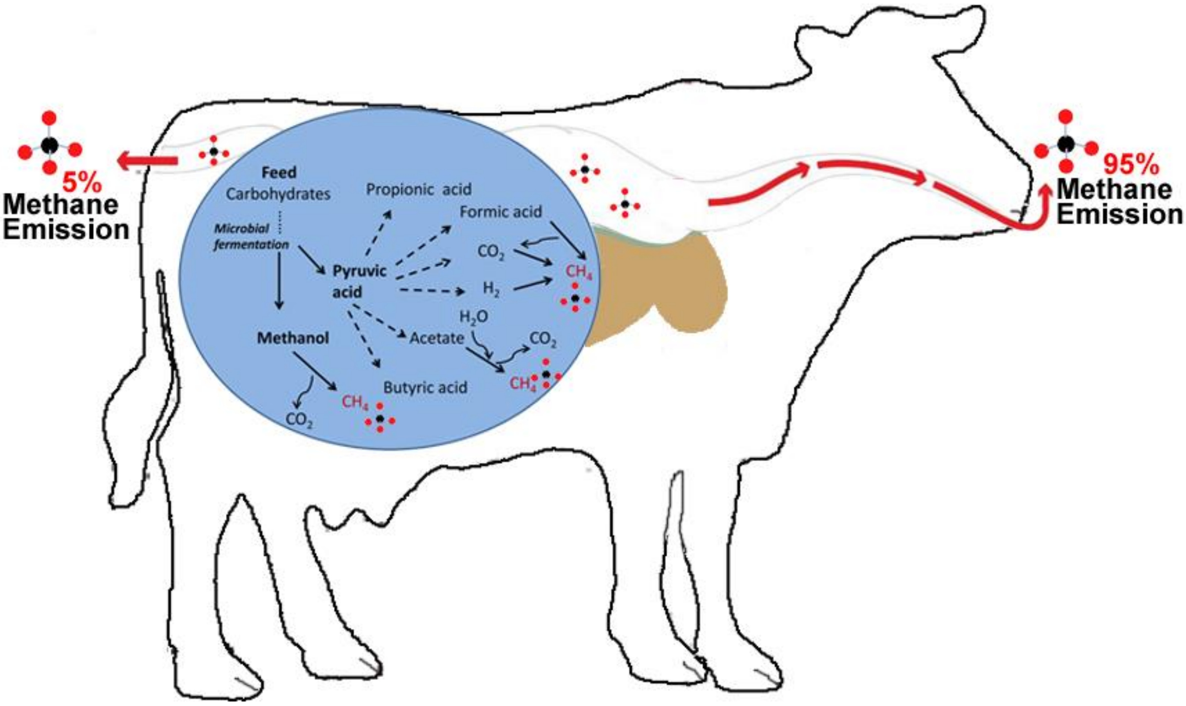


Figure 1-3. The processes of methane producing under the background of nutrients digestion of dairy cows (Morgavi et al., 2010, Ungerfeld, 2020).

The absorbed VFA like acetate contributes up to 80% of adipose tissue acetyl-groups lipogenesis, up to 30% depots of muscle (Smith and Crouse, 1984). Importantly, acetate was the main precursor for milk fat synthesis (Bauman et al., 2006, Urrutia and Harvatine, 2017). During the digestion of carbohydrates by dairy cows, the part of energy loss to produce methane represents up to 12% of the gross energy from the feed (Johnson et al., 1994). The feed proteins are hydrolysed to peptides and amino acids firstly by the rumen microorganisms, and then further into the organic acids, ammonia, as well as CO_2. Some of the ammonia and peptides can be used to synthesize the microbial proteins in rumen and some of them pass to the post-digestive systems to be digested and finally absorbed. Finally, some of the unabsorbed nutrients like ammonia and urea are excreted into the feces and urine (Figure 1-2). So, the digested protein in form of amino acids are carried by the blood to different body tissues and are used mainly for protein synthesis (e.g. milk protein) except the part metabolized to glucose which is used for energy and fat synthesis (Schwab and Broderick, 2017). The level of lipids in dairy cow diet is low (normally < 5%) to keep a balance environment for the rumen microbes (McDonald. et al., 2011). Although low, the digestion of fat in rumen is crucial for the dairy cows. Generally, in ruminants, the dietary lipids are nearly totally hydrolyzed in rumen first and mainly is the hydrogenation of the UFA; secondly, the generated fatty acids are mainly absorbed by cows in the small intestine (Doreau and Ferlay, 1994). The metabolized fatty acids are more saturated FA, which particularly give properties to the synthesis of meat and milk (Doreau and Ferlay, 1994, Lock et al., 2006). Therefore, after the digestion of the carbohydrates, protein, and lipids, the metabolized/degraded nutrients such as VFAs, amino acids, microbial protein, and peptide are then available for the animal to support maintenance, production, and reproduction (McDonald. et al., 2011). The conversion efficiency of the nutrients from feedstuffs into animal products (e.g. milk production from unit feed intake) is relevant to know to optimize the production cost of a farm.

Feed efficiency represents this dairy efficiency and provides indices of how closely the diets meet a cow's specific nutrient requirements including demands of nutrients used for maintenance and milk production (Hall, 2003). So, FE puts a high emphasis on whether the available energy (i.e., net energy) is used for milk production or for maintenance. Consequently, the FE of a dairy cow is improved by allocating more nutrients to the lactation, allowing to optimize the production of milk per consumption of unit feedstuff (Korver, 1988).

The metabolic priority of milk production is high in early lactation corresponding to a higher FE (Coleman et al., 2010). However, especially during the first two months of the lactation period (Puillet et al., 2016), this higher FE can be accompanied by an excessive body weight loss resulting in a poor fertility performance (Roche et al., 2007) and health problems (Berry et al., 2007). Therefore, the monitoring of FE of a cow would benefit to follow the health status and fertility problems (De Vries et al., 1999). This is expected as FE is related to the energy balance of cows directly linked to the metabolism of body reserves which is a key factor in the health and fertility of cows (Hansen, 2000, Banos et al., 2006). Moreover, FE of a cow is related to the economic efficiency (Mosheim and Lovell, 2009, VandeHaar et al., 2016) as the feedstuff represents a major expense of a farm. Indeed, the feed cost represents around 50% of dairy farmer's cost (Beever and Doyle, 2007). Therefore, the selection of cows with high feed efficiency (FE) is favorable as they can produce more milk per unit of feedstuff. Moreover, the elimination of cows with low FE could reduce the investment at herd level because, in a herd with a typical cow population structure, more than 20% of the dry matter intake (DMI) is consumed by the non-productive cows (Michel, 2011). Besides reducing the nutrient waste (Connor, 2015), an improved FE reduces the loss of feed energy leading to reduce the carbon footprint emitted by cows and thus decreases the overall environmental impact of dairy production (Capper et al., 2009). Indeed, the variation of digestion ability and energy usage are important factors affecting FE: low digestibility of nutrients and great methane yield will lead to a low FE (de Assis Lage et al., 2019). Indeed, there is a negative relationship between the metabolic energy utilization and the methane energy output (Yan et al., 2010). Moreover, beef cattle with high FE (expressed as RFI) were reported to have lower methane yield. On another side, FE is related to body size as it is a function of milk yield which is indirectly related to the body size (Blake and Custodio, 1984). The available energy is allocated to the usage of maintenance and milk production under the regulation of the hormones (e.g. growth hormone, insulin) (Blake and Custodio, 1984). Based on these regulatory mechanisms, the cows with larger body weight require more energy for maintenance and accordingly resulted in a relatively lower FE or under a risk of negative energy balance (Hofstetter et al., 2011). Considerable interest has been taken, for decades, in improving the feed utilization (i.e., FE) of dairy cows by limiting the energy loss (e.g., methane) to the environment (Difford et al., 2020). FE is relevant for biological, economic, and environmental reasons (Gravert, 1985, Korver, 1988, Connor, 2015). However, although FE are widely studied in poultry, swine, or beef cattle, a lag exists in evaluating dairy cows' FE in scientific literature (Connor, 2015). The reason could be due to the hinder to follow the FE across the milking cycle exactly under the situation of lacking sufficient data and the paradox of improving FE of dairy cow without affecting the other aspects (Pryce et al., 2014).

The definition of FE could be categorized into two classes: (1) residual or regression traits and (2) ratio traits but, in both cases, DMI records are needed (Pryce et al., 2014, Hurley et al., 2016). For the residual traits, the residual feed intake (RFI) is the most known one. This trait was firstly used in beef cattle and then for growing animals and gradually being used in lactating dairy populations (Hurley et al., 2016). The RFI calculation can be resumed as :

$$RFI = DMI_{actual} - DMI_{pred}$$

where the DMI_{pred} is the linear regression results encompassing the factors of interest. RFI reflects the real net feed intake needed by the cows (Koch et al., 1963, Pryce et al., 2014) as it takes into account the variation in activity, protein turnover, and even the animals' digestibility (Mayes et al., 1986). However, RFI is easy to acquire for growing and non-lactation cows but is more difficult to obtain for lactating cows due to the more complicated energy allocation during the lactating stage (Pryce et al., 2014). Moreover, the FE expressed as RFI is more complicated to be determined in practice because people need to obtain both the real and expected DMI of cows to calculate their difference (Connor, 2015). For the ratio traits, FE is defined as a "ratio" between output (i.e., milk production) and input (i.e., the quantity of feed intake) (Korver, 1988) which was also deemed to gross feed efficiency. Ratio trait is a more traditional definition of FE (Hurley et al., 2016). In this case, the milk production is taken into account using different forms such as the fat- and protein-corrected milk production (FPCM) (Seymour et al., 2020), the milk yield (Barros et al., 2017), as well as the energy corrected milk production (ECM) (Ben Meir et al., 2018). So, FE can be calculated using the ratio FPCM to DMI. This ratio was reported to be related mainly to the body weight loss and body condition score (BCS) throughout the lactation (Vallimont et al., 2011).

The FE phenotypes were used in different studies to better understand its variability. So, the physiological stage of a cow influences its FE (Bauman et al., 1985, Phuong et al., 2013) due to the maintenance requirements or the nutrients partitions of cows. In specific, DIM was reported to have a negative correlation (around −0.50) with FE (Britt et al., 2003). This is expected as the cows tend to mobilize more their body reserves at the early stage of lactation resulting in a high FE value (Connor et al., 2012). Indeed, BW and DIM were claimed to be significantly related to FE based on a meta-analysis (Phuong et al., 2013). Besides, the feed composition and feeding level were reported also to be related to the FE by affecting the energy partitioning variables (Ramin and Huhtanen, 2013). Indeed, the forage level in the diet influences the rumen environment (Stone et al., 1960, Pino et al., 2018). Furthermore, the temperature was also reported to affect the FE of cows. For instance, the FE seems to be around 7% lower during the warm season (> 21 ℃) compared to the cool season (< 21 ℃) (Britt et al., 2003). Additionally, parity is another significant factor impact the FE. For example, using 26 cows, Ben Meir et al.(2019) found that the cows at second parity produced more milk and consumed more feed than cows in first parity and resulted in a significant ($p <$

0.05) higher FE (1.51 v.s. 1.40 kg of milk/kg DM). Similarly, multiparous cows were reported to have higher FE than primiparous cows (Figure 1-4A) (DeVries et al., 2011, Løvendahl et al., 2018, Ben Meir et al., 2019). More than that, FE depends also on the genetic ability of a cow to transform feed intake into milk production. In specific, the FE from different breeds of cows were estimated to be different (DIM < 60 days, Figure 1-4B): the Norwegian Red cows and Holstein cows presented a higher FE than the Normande cows in the early lactation (Buckley et al., 2006). Moreover, the FE was reported presented a decline trend across the stage of lactation for both the cows with high and low efficiency (Figure 1-4C) (St-Pierre, 2019). In addition, the numerator and the denominator used to assess the FE (e.g. Milk yield and DMI) are also factors contributing to the FE variation (Phuong et al., 2013); as shown in Figure 1-4D, the FE was nearly present a linear relationship with milk yield or DMI (St-Pierre, 2019). Finally, although the genetic differences between cows for available energy utilization was reported not significant ($P > 0.05$) (Moe, 1981, Blake and Custodio, 1984), the genetic parameters (e.g., heritability which varied between 0.01 and 0.52) depend on the population used but their values confirm that FE can be improved according to an appropriate genetic selection (Brito et al., 2020). Because the FE is directly related to the digestive and metabolic efficiency which are both affected by the genetic variation (Arthur et al., 2001), it was expected that FE will be also affected by animal genetics. As reported, the gross feed efficiency (i.e., the "ratio") was more heritable than RFI (Veerkamp and Emmans, 1995, Vallimont et al., 2011).

Knowing FE of the herd helps to optimize the production of dairy cows for a given amount of feedstuff. Due to the diversity of FE definitions, there is a various recommended optimal range of FE. For instance, defining FE as the ratio of 3.5% FCM to DMI, the acceptable limit of FE is 0.95 to 1.9 kg of milk/kg DM following the DIM and parity (DAIREXNET, 2019). The optimal range of FE defined as ECM divided by DMI is comprised between 1.3 and 1.8 kg of milk/kg DM (Jud Heinrichs and Ishler, 2016). Using the same ratio as the current study (FE=FPCM/DMI), the range of FE must be between 0.90 and 2.70 kg of milk/kg DM based on studies conducted from Holsteins cows in various stage of lactation, parities, and fed with different kinds of diet (Aguerre et al., 2016, Zucali et al., 2016, Barros et al., 2017, Seymour et al., 2020).

In conclusion, as FE is crucial in dairy cow husbandry regarding cow nutrient allocation, cow health status, farm profit, and environment, selecting high FE dairy cows is the long-standing objective for animal breeding. But the most tough task faced by dairy sector consists of collecting enough FE information at individual and large scales (Seymour et al., 2019).

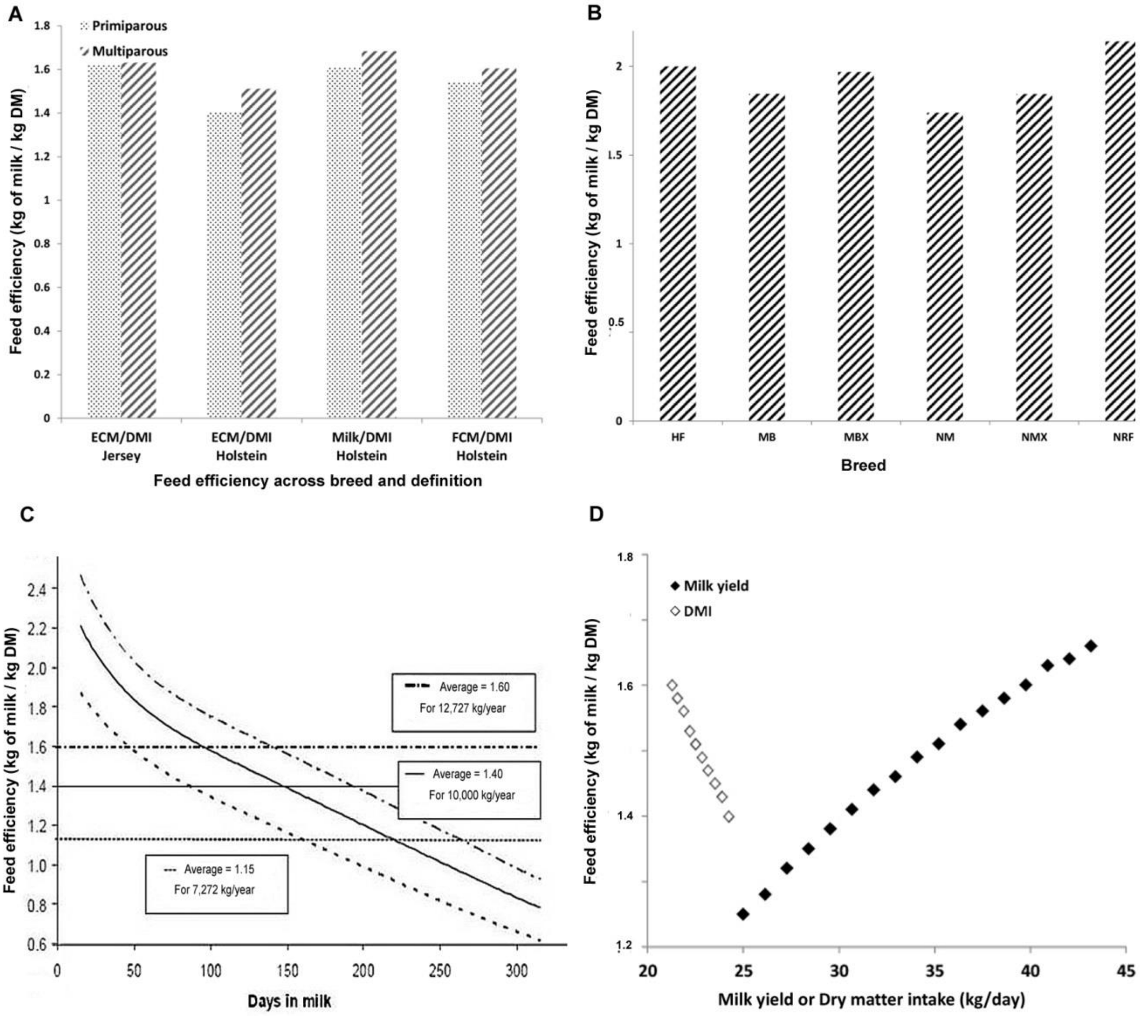


Figure 1-4. The evolution of feed efficiency (FE) by different variations.

A:FE evolution by parity (DeVries et al., 2011, Løvendahl et al., 2018, Ben Meir et al., 2019); B:FE evolution by breed (HF=Holstein, MB=Montbeliarde, MBX=Holstein×Montbeliarde, NM=Normande, NMX=Holstein×Normande, NRF=Norwegian Red) (Buckley et al., 2006); C:FE evolution by days in milk (DIM) for high and low FE cows (St-Pierre, 2019); D:FE evolution by milk yield and DMI (St-Pierre, 2019).

3. Prediction of feed efficiency at large and individual scales

According to the definition of FE mentioned above, the DMI is a key factor to calculate FE of a cow (Berry and Crowley, 2013). Two steps are needed to measure the DMI ingested by a cow: the first is to measure the feed intake and the second is to estimate the dry matter content in feed consumed by cow. Unfortunately, it is difficult for the commercial farms to be well equipped to perform this measurement or have limited time to measure feed intake at large and individual scales. Even though the DMI can be measured in some experimental groups under the indoor condition, evaluating the DMI of cows in the pasture condition is always a tough work (Méndez et al., 2019). For instance, to acquire individual animal DMI at pasture conditions, the n-alkane technique was used (Mayes et al., 1986). However, this method based on internal markers is restricted when used on a large scale. On another hand, to assess the DMI at pasture conditions, knowing the biomass at a unit time is necessary (Méndez et al., 2019). The photogrammetry is a useful method to assess the pasture biomass in order to assess the DMI of a group of animals but not for individuals (Wigley et al., 2019). So, knowing the quantity of ingested feed is a traditional hinder (Seymour et al., 2019). To cope with this issue, several models were built and tested. For instance, the equations using animal characteristics and MIR spectrum obtained an accuracy ranging from 1.59 to 2.23 kg/day (Lahart et al., 2019). The equation from Agricultural Research Council (ARC, 1980), employing live body weight (BW), FCM, and lactation stage as predictors, and the one from Cornell Net Carbohydrate and Protein System (CNCPS., 2004), employing BW, FCM, and lactation stage adjusted factors as predictors were tested using 777 data points in the study of (Ellis et al., 2006). The obtained RMSE were around 1.96 - 2.08 kg/day and 2.39 - 2.49 kg/day for ARC and CNCPS equations, respectively. Additionally, with an accuracy of 1.82 kg/day, the National Research Council (NRC) equation to predict the DMI using the live body weight (BW, kg), 4% Fat-corrected milk (FCM, kg/d), and the number of weeks of lactation (wol) (NRC, 2001, Lahart et al., 2019). This equation can be resumed as:

$$DMI = (0.372 * FCM + 0.0968 * BW^{0.75}) * (1 - e^{(-0.192*(wol+3.67))})$$

The FCM and the week of lactation can be easily measured at individual cow scale using traits routinely recorded by dairy herd improvement (DHI) organizations during the milk recording. However, this is not the case for the live BW. The gold-standard method to evaluate BW is using a weighing scale. However, due to the drawback of cost and labor intensive, alternatives are needful (Soyeurt et al., 2019). For example, the estimation of BW using morphological traits (Heinrichs et al., 1992, Enevoldsen and Kristensen, 1997), and camera (Song et al., 2018) are alternatives to the gold-standard method. However, the cost or the frequency of data acquisition limits their use on a routine basis. Owing to the potential of milk mid-infrared (MIR) spectra routinely recorded in the framework of DHI, predicting the BW from this information and animal characteristics is a promising perspective allowing an accuracy around 50 kg (Soyeurt et al., 2019). The spectrum provided by the MIR spectrometry of milk sample corresponds to the absorbance of the infrared ray by specific chemical bonds in the matter and represents therefore the interactions between propagating light and matter (McDermott et al., 2016, Türker-Kaya and Huck, 2017). When the MIR light passes through the sample, some of the energy of this ray is absorbed by specific chemical bonds (Sun, 2009). This absorption is then captured by the detector (Figure 1-5). According to the varied energy of the MIR ray (about thousand times following the spectrometer used), the sample spectrum is obtained finally. Works using similar approach were conducted to predict DMI. Indeed, predicting dry matter intake (DMI) at the individual and large scale using MIR technique was reported by researchers as well, such as (Lahart et al., 2019) obtained a root mean squared error (RMSE) of predicted DMI at 1.68 kg, and (Shetty et al., 2017) revealed an RMSE of RFI at 1.49 kg.

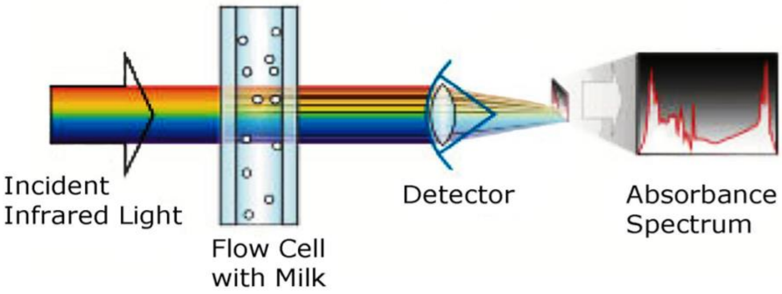


Figure 1-5. Mid-infrared spectrometry (©Bentley instrument, 2019).

Predicting DMI, RFI, or BW from animal characteristics and milk MIR spectrum opens new perspectives in the inclusion of FE concept for management and breeding perspective for the dairy sector as those predictors are provided at low cost, at the individual scale, and on a monthly basis (4-6 weeks) through the milk recording managed by DHI organizations. In other words, feed efficiency provided by such approach can be considered as "Big Data" because it fulfilled the requirement of data volume, data velocity and heterogeneity. This large amount of observations opens the perspective to conduct quantitative genetic studies as well as genomic studies. Indeed, more phenotypes allow enhancing the genomic selection (Hayes et al., 2009) leading potentially a selection of animals with high FE at a younger stage.

4. Improvement of feed efficiency related predicted phenotypes at large scale

In recent years, several phenotypes prediction equations using the MIR technology have been developed related to a wide aspect of dairy farming such as cow energy status (McParland et al., 2011), BW (Soyeurt et al., 2019), feed intake or DMI (Shetty et al., 2017, Lahart et al., 2019), feed composition (Klaffenböck et al., 2017), methane emission (Dehareng et al., 2012, Vanlierde et al., 2015), nitrogen efficiency (Grelet et al., 2020b), fertility status (Ho et al., 2019), lameness biomarkers (Mineur et al., 2020), milk lactoferrin content (Soyeurt et al., 2012), milk fatty acids content (Soyeurt et al., 2006), fat globule size (Fleming et al., 2017), cheese making properties (Karoui et al., 2006), pregnancy status (Delhez, 2020), and so on. However, limitations related to the use of these MIR equations exist, since most of them were developed from specific populations and the performance of these models towards new samples from other conditions need to be validated. So, although the use of MIR spectroscopy provides fast, cost-effective quantification or qualification phenotypes to farmers or their advisors, there are still some issues related to the model development and implementation (Figure 1-6) to be concerned and settled such as creating accurate and robust models using limited data sets; the detection of samples to be predicted out of the model capacity (i.e., outliers); the model transfer between different instruments/brand and even among different countries.

So, the elementary issue that must be achieved to improve the prediction ability of a prediction equation is to enlarge the number of records used for developing the models in order to gain a sufficient variability (Grelet et al., 2020a). Indeed, as the prediction performance is highly affected by the data structure and variability of the calibration sets used to develop the models; the more variability covered by the calibration set, the more powerful the model will be (Edward V. Thomas, 2000). In other words, if the models are used towards the new samples not covered by the variability of the calibration sets, the accuracy will be impaired (Tomas Naes, 2002). Regarding the traits of interest in the context of this thesis, this means that the variability of BW, RFI, or DMI must be represented in the calibration set used. Comparing to RFI and DMI reference data acquisition, the obtaining of BW records is easiest and therefore enlarges the possibility of building a large scale dataset to build the predictive model leading to the confection of a dataset more representative of the dairy cow population. So, it is why BW must be considered as a key trait to develop a large scale feed efficiency tool. However, the dataset used by Soyeurt et al.(2019) was limited because it considered only 735 records collected from 270 cows belonging to six herds. So, there is a potentiality of improvement for this equation by increasing the variability of the dataset.

On other hand, the number of the variables employed to build the predictive model also impacts the prediction quality (Centner et al., 1996). So, feature selection approaches before final modeling is essential and can be adopted to put the most informative variables into calibration models without the interference of noise (Saeys et al., 2008). However, even if a gross selection of pin numbers has been done by experience, Soyeurt et al.(2019) used nearly 300 MIR spectral points to develop the BW prediction models.

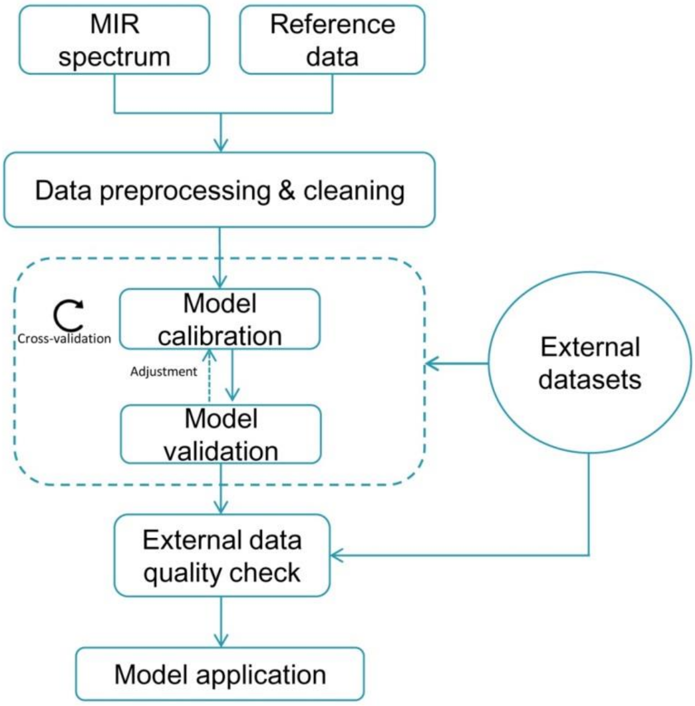


Figure 1-6. The process of development and application of MIR prediction model.

Even if researchers always pursue to cover as more as possible the variability of the trait to predict in the training sets, this is often unfeasible for financial and practical reasons to have enough samples to be representative of the studied cow population. This situation leads to provide a prediction, which is the fruit of a spectral extrapolation, for a milk sample of which the variability was not taken into account in the calibration set (Grelet et al., 2020a). Unfortunately, until now, no quality procedure is developed by milk laboratories or DHI organizations to detect those samples. However, a quality procedure named "ring test" exists for fat and protein to fix a correction slope and bias for the MIR-based predictions. Following the protocol proposed by the International Dairy Federation (FIL-IDF; ISO9622/IDF 141), this ring test is performed by analyzing a series of samples which have a wide range of fat and protein contents certified using reference chemical measurement methods by different laboratories. Unfortunately, it is not possible to adapt this procedure for the MIR predictions of BW and DMI as the reference measurement is not the fruit of a chemical analysis. Some novel cleaning procedures must be found. In this context, the use of Mahalanobis distance (Mahalanobis, 1936), Global H distance (GH) (Shenk and Westerhaus, 1991, De Maesschalck et al., 2000), threshold between the corrected fat percentage and the one predicted directly from the spectra as proposed by Dale et al.(2019), or other methods to detect extreme predictions could be relevant to highlight and discard those potentially erroneous predictions.

5. Research framework and outline

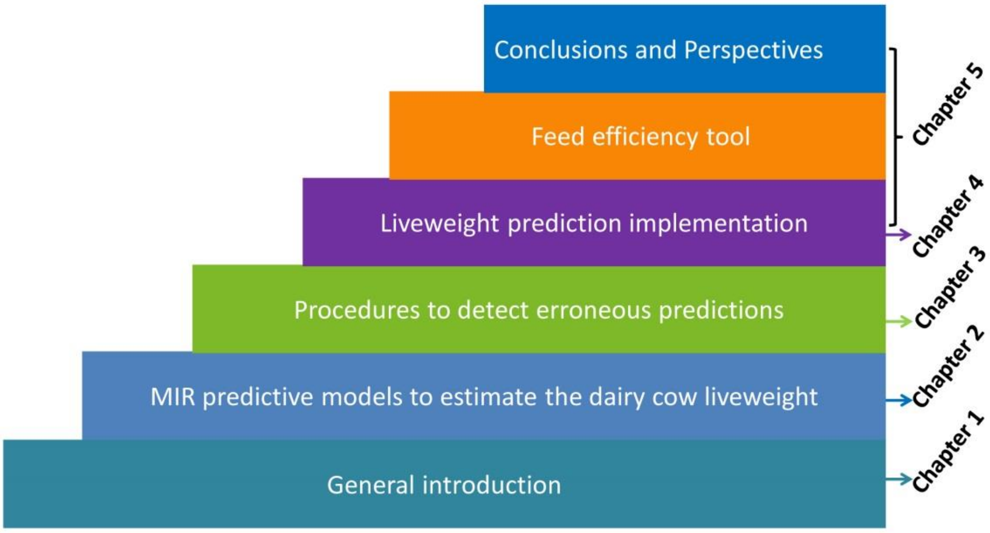


Figure 1-7. The overview of contents in this thesis.

The development of a feed efficiency tool available at low cost, large scale and at cow basis is of interest in the current dairy farming context where the profitability of the farm must be ensured as well as its low environmental impact. This PhD thesis aims to contribute to this global subject under different aspects. To achieve this global objective, the FE trait must be accessible at large and individual scale. To fulfill this requirement, the FE traits will be estimated in this thesis using the ratio of FPCM to DMI with a DMI phenotype predicted using the NRC equation and the BW predicted from animal characteristics and milk MIR spectrum. Even if an equation is already available (Soyeurt et al., 2019), the work conducted in this thesis will try to improve the accuracy of this equation by enlarging the calibration set and by testing feature selection approaches. Then, this thesis aims also to contribute to the definition of quality protocol easily to be implemented by DHI organizations in order to ensure a high quality of prediction of traits, for which the reference chemical analysis of milk is not the gold standard. Finally, this research will also study the behaviors of different traits related to FE like methane, DMI, BW and a consumption index (IC) predicted using entirely or partially the milk MIR information when they are predicted on large scale DHI databases. All of this work will contribute to the development of a FE tool available for dairy farmers and breeders.

This thesis is a compilation of three published or submitted scientific papers. The overview of the structure of this thesis work is shown in Figure 1-7. After a general introduction presented in this Chapter (Chapter 1), the improvement of BW MIR prediction model by using a large calibration set as well as an appropriate feature selection will be exposed in Chapter 2. Even if the predictive model is robust, the lack of variability for the studied trait in the calibration set as well as the use of poor quality spectral data from the new samples can lead to obtaining erroneous predictions. So, to avoid this issue, the development of appropriate data quality check procedures which can be easily implemented by DHI or milk laboratories will be investigated in Chapter 3. Then, a large scale study of the behaviors of different traits related directly or indirectly to FE and predicted fully or partially from milk MIR spectral data will be conducted from the use of herd test-day models (Chapter 4). Finally, a discussion related to all aspects concerning the development of a FE tool by DHI organizations will be done in Chapter 5 and will finish by the formulation of conclusions and perspectives related to the current PhD topic.

References

Aguerre, M. J., M. C. Capozzolo, P. Lencioni, C. Cabral, and M. A. Wattiaux. 2016. Effect of quebracho-chestnut tannin extracts at 2 dietary crude protein levels on performance, rumen fermentation, and nitrogen partitioning in dairy cows. Journal of dairy science 99(6):4476-4486.

ARC. Agricultural Research Council., 1980. The nutrient requirements of ruminant livestock: technical review. CAB International.

Arthur, P. F., J. A. Archer, D. J. Johnston, R. M. Herd, E. C. Richardson, and P. F. Parnell. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. Journal of Animal Science 79(11):2805-2811.

Banos, G., M. P. Coffey, E. Wall, and S. Brotherstone. 2006. Genetic relationship between first-lactation body energy and later-life udder health in dairy cattle. Journal of dairy science 89(6):2222-2232.

Barkema, H. W., M. A. von Keyserlingk, J. P. Kastelic, T. J. Lam, C. Luby, J. P. Roy, S. J. LeBlanc, G. P. Keefe, and D. F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. Journal of dairy science 98(11):7426-7445.

Barros, T., M. A. Quaassdorff, M. J. Aguerre, J. J. O. Colmenero, S. J. Bertics, P. M. Crump, and M. A. Wattiaux. 2017. Effects of dietary crude protein concentration on late-lactation dairy cow performance and indicators of nitrogen utilization. Journal of dairy science 100(7):5434-5448.

Bauman, D. E., I. H. Mather, R. J. Wall, and A. L. Lock. 2006. Major advances associated with the biosynthesis of milk. Journal of dairy science 89(4):1235-1243.

Bauman, D. E., S. N. McCutcheon, W. D. Steinhour, P. J. Eppard, and S. J. Sechen. 1985. Sources of Variation and Prospects for Improvement of Productive Efficiency in the Dairy Cow: A Review. Journal of Animal Science 60(2):583-592.

Beever, D. E. and P. T. Doyle. 2007. Feed conversion efficiency as a key determinant of dairy herd performance: a review. Australian Journal of Experimental Agriculture 47(6):645-657.

Ben Meir, Y. A., M. Nikbachat, Y. Fortnik, S. Jacoby, H. Levit, G. Adin, M. Cohen Zinder, A. Shabtay, E. Gershon, M. Zachut, S. J. Mabjeesh, I. Halachmi, and J. Miron. 2018. Eating behavior, milk production, rumination, and digestibility characteristics of high- and low-efficiency lactating cows fed a low-roughage diet. Journal of dairy science 101(12):10973-10984.

Ben Meir, Y. A., M. Nikbachat, S. Jacoby, Y. Portnik, H. Levit, A. K. Elazary, E. Gershon, G. Adin, M. Zinder-Cohen, A. Shabtay, M. Zachut, S. J. Mabjeesh, I. Halachmi, and J. Miron. 2019. Effect of lactation trimester and parity on eating behavior, milk production and efficiency traits of dairy cows. Animal : an international journal of animal bioscience 13(8):1736-1743.

Berckmans, D. 2006. Automatic on-line monitoring of animals by precision livestock farming. Livestock production and society 287.

Berry, D. P. and J. J. Crowley. 2013. CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle1. Journal of Animal Science 91(4):1594-1613.

Berry, D. P., J. M. Lee, K. A. Macdonald, and J. R. Roche. 2007. Body condition score and body weight effects on dystocia and stillbirths and consequent effects on postcalving performance. Journal of dairy science 90(9):4201-4211.

Blake, R. W. and A. A. Custodio. 1984. Feed efficiency: a composite trait of dairy cattle. Journal of dairy science 67(9):2075-2083.

Brito, L. F., H. R. Oliveira, K. Houlahan, P. A. S. Fonseca, S. Lam, A. M. Butty, D. J. Seymour, G. Vargas, T. C. S. Chud, F. F. Silva, C. F. Baes, A. Cánovas, F. Miglior, and F. S. Schenkel. 2020. Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle. Canadian Journal of Animal Science 100(4):587-604.

Britt, J. S., R. C. Thomas, N. C. Speer, and M. B. Hall. 2003. Efficiency of Converting Nutrient Dry Matter to Milk in Holstein Herds1. Journal of dairy science 86(11):3796-3801.

Buckley, F., S. W. Walsh, and P. Dillon. 2006. Comparison of breed of dairy cow under grass-based spring milk production systems. Teagasc.

Capper, J. L., R. A. Cady, and D. E. Bauman. 2009. The environmental impact of dairy production: 1944 compared with 2007. Journal of Animal Science 87(6):2160-2167.

Cashman, K. D. 2006. Milk minerals (including trace elements) and bone health. International Dairy Journal 16(11):1389-1398.

Centner, V., D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, and C. Sterna. 1996. Elimination of uninformative variables for multivariate calibration. Analytical chemistry 68(21):3851-3858.

Cheng, D., Y. Liu, H. H. Ngo, W. Guo, S. W. Chang, D. D. Nguyen, S. Zhang, G. Luo, and Y. Liu. 2020. A review on application of enzymatic bioprocesses in animal wastewater and manure treatment. Bioresource Technology 313:123683.

Chilliard, Y., C. Martin, J. Rouel, and M. Doreau. 2009. Milk fatty acids in dairy cows fed whole crude linseed, extruded linseed, or linseed oil, and their relationship with methane output¹. Journal of dairy science 92(10):5199-5211.

China dairy statistical summary. 2019. Dairy association of China, Beijing.

Clay, N., T. Garnett, and J. Lorimer. 2020. Dairy intensification: Drivers, impacts and alternatives. Ambio 49(1):35-48.

Coleman, J., D. P. Berry, K. M. Pierce, A. Brennan, and B. Horan. 2010. Dry matter intake and feed efficiency profiles of 3 genotypes of Holstein-Friesian within pasture-based systems of milk production. Journal of dairy science 93(9):4318-4331.

Connor, E. E. 2015. Invited review: improving feed efficiency in dairy production: challenges and possibilities. Animal : an international journal of animal bioscience 9(3):395-408.

Connor, E. E., J. L. Hutchison, K. M. Olson, and H. D. Norman. 2012. Triennial lactation symposium: Opportunities for improving milk production efficiency in dairy cattle. Journal of Animal Science 90(5):1687-1694.

Crump, A., K. Jenkins, E. J. Bethell, C. P. Ferris, and G. Arnott. 2019. Pasture Access Affects Behavioral Indicators of Wellbeing in Dairy Cows. Animals : an open access journal from MDPI 9(11):902.

DAIREXNET. 2019. Practical Approaches To Feed Efficiency and Applications on the Farm. Vol. 2020.

Dale, L. M., A. Werner, H. Spiekers, E. Stamer, M. Au, and F. Onken. . 2019. Prediction of evaluated energy balance (NEL and ME) in dairy cows by milk mid- infrared (MIR) spectra. Page 137 in Proc. ICAR, Prague.

de Assis Lage, C. F., S. Gesteira Coelho, H. d. C. Diniz Neto, V. M. Rocha Malacco, J. P. Pacheco Rodrigues, J. P. Sacramento, F. Samarini Machado, L. G. Ribeiro Pereira, T. Ribeiro Tomich, and M. Magalhães Campos. 2019. Relationship between feed efficiency indexes and performance, body measurements, digestibility, energy partitioning, and nitrogen partitioning in pre-weaning dairy heifers. PLOS ONE 14(10):e0223368.

De Maesschalck, R., D. Jouan-Rimbaud, and D. L. Massart. 2000. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems 50(1):1-18.

De Vries, M. J., S. Van Der Beek, L. Kaal-Lansbergen, W. Ouweltjes, and J. B. M. Wilmink. 1999. Modeling of energy balance in early lactation and the effect of energy deficits in early lactation on first detected estrus postpartum in dairy cows. Journal of dairy science 82(9):1927-1934.

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal : an international journal of animal bioscience 6(10):1694-1701.

Delhez, P. 2020. Developing decision-support tools with economic and environmental interests for the dairy sector using milk mid-infrared spectrometry. in Gembloux Agro-Bio Tech. University of Liege.

DeVries, T. J., L. Holtshausen, M. Oba, and K. A. Beauchemin. 2011. Effect of parity and stage of lactation on feed sorting behavior of lactating dairy cows. Journal of dairy science 94(8):4039-4045.

Difford, G. F., P. Løvendahl, R. F. Veerkamp, H. Bovenhuis, M. H. P. W. Visker, J. Lassen, and Y. de Haas. 2020. Can greenhouse gases in breath be used to genetically improve feed efficiency of dairy cows? Journal of dairy science 103(3):2442-2459.

Dijkstra, J., S. van Gastelen, E. C. Antunes-Fernandes, D. Warner, B. Hatew, G. Klop, S. C. Podesta, H. J. van Lingen, K. A. Hettinga, and A. Bannink. 2016. Relationships between milk fatty acid profiles and enteric methane production in dairy cattle fed grass- or grass silage-based diets. Animal Production Science 56(3):541-548.

Doreau, M. and A. Ferlay. 1994. Digestion and utilisation of fatty acids by ruminants. Animal Feed Science and Technology 45(3):379-396.

Doyle, N., P. Mbandlwa, W. J. Kelly, G. Attwood, Y. Li, R. P. Ross, C. Stanton, and S. Leahy. 2019. Use of Lactic Acid Bacteria to Reduce Methane Production in Ruminants, a Critical Review. Frontiers in Microbiology 10:2207.

Edward V. Thomas, N. G. 2000. Development of Robust Multivariate Calibration Models. Technometrics (42):168-177.

Ellis, J. L., F. Qiao, and J. P. Cant. 2006. Prediction of Dry Matter Intake Throughout Lactation in a Dynamic Model of Dairy Cow Performance. Journal of dairy science 89(5):1558-1570.

Enevoldsen, C. and T. Kristensen. 1997. Estimation of Body Weight from Body Size Measurements and Body Condition Scores in Dairy Cows. Journal of dairy science 80(9):1988-1995.

EPA. 2020. Understanding Global Warming Potentials.

FAO. 2020. OECD-FAO Agricultural Outlook 2020-2029, OECD Publishing, Paris/FAO, Rome.

FAOSTAT. 2017. Food and agriculture organization of the world united nations.

Fleming, A., F. S. Schenkel, J. Chen, F. Malchiodi, R. A. Ali, B. Mallard, M. Sargolzaei, M. Corredig, and F. Miglior. 2017. Variation in fat globule size in bovine milk and its prediction using mid-infrared spectroscopy. Journal of dairy science 100(3):1640-1649.

Fox, D. G., L. O. Tedeschi, T. P. Tylutki, J. B. Russell, M. E. Van Amburgh, L. E. Chase, A. N. Pell, and T. R. Overton. 2004. The Cornell Net Carbohydrate and Protein System model for evaluating herd nutrition and nutrient excretion. Animal Feed Science and Technology 112(1-4):29-78.

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. 2013. Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities. Food and Agriculture Organization of the United Nations (FAO).

Giuburunca, M., A. Criste, D. Cocan, R. Constantinescu, C. Răducu, and V. Mireșan. 2014. Methane production in the rumen and its influence on global warming. ProEnvironment/ProMediu 7(18).

Gravert, H. O. 1985. Genetic factors controlling feed efficiency in dairy cows. Livestock Production Science 13(2):87-99.

Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Steevens, N. Gengler, and F. Dehareng. 2020a. Large-scale phenotyping in dairy sector using milk MIR spectra: key factors affecting the quality of predictions. Methods.

Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C. P. Ferris, K. L. Ingvartsen, M. A. Crowe, M. T. Sorensen, J. A. Fernandez Pierna, A. Vanlierde, N. Gengler, and F. Dehareng. 2020b. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. Journal of dairy science 103(5):4435-4445.

Hall, M. B. 2003. What you feed vs. what you get: feed efficiency as an evaluation tool. Pages 25-31.

Hansen, L. B. 2000. Consequences of selection for milk yield from a geneticist's viewpoint. Journal of dairy science 83(5):1145-1150.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of dairy science 92(2):433-443.

Heinrichs, A. J., G. W. Rogers, and J. B. Cooper. 1992. Predicting Body Weight and Wither Height in Holstein Heifers Using Body Measurements. Journal of dairy science 75(12):3576-3581.

Ho, P. N., V. Bonfatti, T. D. W. Luke, and J. E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. Journal of dairy science 102(11):10460-10470.

Hofstetter, P., M. Steiger Burgos, R. Petermann, A. Münger, J. W. Blum, P. Thomet, H. Menzi, S. Kohler, and P. Kunz. 2011. Does body size of dairy cows, at constant ratio of maintenance to production requirements, affect productivity in a pasture-based production system? J Anim Physiol Anim Nutr (Berl) 95(6):717-729.

Hristov, A. N., J. Oh, F. Giallongo, T. Frederick, M. T. Harper, H. Weeks, A. F. Branco, W. J. Price, P. J. Moate, and M. H. Deighton. 2016. Comparison of the GreenFeed system with the sulfur hexafluoride tracer technique for measuring enteric methane emissions from dairy cows. Journal of dairy science 99(7):5461-5465.

Hurley, A. M., N. López-Villalobos, S. McParland, E. Kennedy, E. Lewis, M. O'Donovan, J. L. Burke, and D. P. Berry. 2016. Inter-relationships among alternative definitions of feed efficiency in grazing lactating dairy cows. Journal of dairy science 99(1):468-479.

Johnson, K., M. Huyler, H. Westberg, B. Lamb, and P. Zimmerman. 1994. Measurement of methane emissions from ruminant livestock using a sulfur hexafluoride tracer technique. Environmental Science & Technology 28(2):359-362.

Jud Heinrichs and V. A. Ishler. 2016. Feed Efficiency in Lactating Cows and Relationship to Income Over Feed Costs. Vol. 2020.

Karoui, R., A. M. Mouazen, É. Dufour, L. Pillonel, E. Schaller, D. Picque, J. De Baerdemaeker, and J.-O. Bosset. 2006. A comparison and joint use of NIR and MIR spectroscopic methods for the determination of some parameters in European Emmental cheese. European Food Research and Technology 223(1):44-50.

Klaffenböck, M., A. Steinwidder, C. Fasching, G. Terler, L. Gruber, G. Mészáros, and J. Sölkner. 2017. The use of mid-infrared spectrometry to estimate the ration composition of lactating dairy cows. Journal of dairy science 100(7):5411-5421.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of Feed Use in Beef Cattle. Journal of Animal Science 22(2):486-494.

Koning, C. J. A. M. d., A. C. G. Beldman, C. H. G. Daatselaar, J. B. v. d. Fels, R. W. Melse, F. E. d. Buisonjé, R. A. Jongeneel, L. Shengli, L. Shengyu, L. Yan, L. Changquan, L. Xiao, L. Chang, W. Yajing, X. Jianming, Y. Yang, Y. Kun, and Z. Liebing. 2016. White paper on China dairy 2016.

Korver, S. 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: a review. Livestock Production Science 20(1):1-13.

Kratz, M., T. Baars, and S. Guyenet. 2013. The relationship between high-fat dairy consumption and obesity, cardiovascular, and metabolic disease. European Journal of Nutrition 52(1):1-24.

Løvendahl, P., G. F. Difford, B. Li, M. G. G. Chagunda, P. Huhtanen, M. H. Lidauer, J. Lassen, and P. Lund. 2018. Review: Selecting for improved feed efficiency and reduced methane emissions in dairy cattle. Animal : an international journal of animal bioscience 12(s2):s336-s349.

Lahart, B., S. McParland, E. Kennedy, T. M. Boland, T. Condon, M. Williams, N. Galvin, B. McCarthy, and F. Buckley. 2019. Predicting the dry matter intake of grazing dairy cows using infrared reflectance spectroscopy analysis. Journal of dairy science 102(10):8907-8918.

Lock, A. L., K. J. Harvatine, J. K. Drackley, and D. E. Bauman. 2006. Concepts in fat and fatty acid digestion in ruminants. Pages 85-100.

Méndez, M. N., P. Chilibroste, and M. Aguerre. 2019. Pasture dry matter intake per cow in intensive dairy production systems: effects of grazing and feeding management. Animal : an international journal of animal bioscience 14(4):846-853.

Mahalanobis, P. C. 1936. On the Generalized Distance in Statistics. Pages 49-55 in Proc. Proceedings of the National Institute of Science of India, India.

Mayes, R. W., C. S. Lamb, and P. M. Colgrove. 1986. The use of dosed and herbage n-alkanes as markers for the determination of herbage intake. The Journal of Agricultural Science 107(1):161-170.

McDermott, A., G. Visentin, M. De Marchi, D. P. Berry, M. A. Fenelon, P. M. O'Connor, O. A. Kenny, and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. Journal of dairy science 99(4):3171-3182.

McDonald., P., R. A. Edwards., J.F.D.Greenhalgh., C.A.Morgan., L.A.Sinclair., and R.G.Wilkinson. 2011. Animal nutrition. Pearson United Kingdom, United Kingdom.

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows1. Journal of dairy science 94(7):3651-3661.

Michel, W. A. 2011. Heifer raising – weaning to calving, Chapter 33 Feeding and housing. University of Wisconsin, Madison, WI, USA.

Mineur, A., H. Hammami, C. Grelet, C. Egger-Danner, J. Sölkner, and N. Gengler. 2020. Short communication: Investigation of the temporal relationships between milk mid-infrared predicted biomarkers and lameness events in later lactation. Journal of dairy science 103(5):4475-4482.

Moe, P. W. 1981. Energy Metabolism of Dairy Cattle. Journal of dairy science 64(6):1120-1139.

Morgavi, D. P., E. Forano, C. Martin, and C. J. Newbold. 2010. Microbial ecosystem and methanogenesis in ruminants. Animal : an international journal of animal bioscience 4(7):1024-1036.

Mosheim, R. and C. A. K. Lovell. 2009. Scale Economies and Inefficiency of U.S. Dairy Farms. American Journal of Agricultural Economics 91(3):777-794.

Moyes, K. M., T. Larsen, and K. L. Ingvartsen. 2013. Generation of an index for physiological imbalance and its use as a predictor of primary disease in dairy cows during early lactation. Journal of dairy science 96(4):2161-2170.

Muehlhoff, E., A. Bennett, and D. McMahon. 2013. Milk and dairy products in human nutrition. Food and Agriculture Organization of the United Nations (FAO).

Myhre, G., D. Shindell, and J. Pongratz. 2014. Anthropogenic and natural radiative forcing.

Negussie, E., Y. de Haas, F. Dehareng, R. J. Dewhurst, J. Dijkstra, N. Gengler, D. P. Morgavi, H. Soyeurt, S. van Gastelen, T. Yan, and F. Biscarini. 2017. Invited review: Large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions. Journal of dairy science 100(4):2433-2453.

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th ed. National Academies Press, Washington, DC.

Park, Y. W., G. Haenlein, and D. S. Ag. 2013. Milk and dairy products in human nutrition. Wilet-Blackwell. A John Wiley & Sons, Ltd., Publication 700.

Phuong, H. N., N. C. Friggens, I. J. M. de Boer, and P. Schmidely. 2013. Factors affecting energy and nitrogen efficiency of dairy cows: A meta-analysis. Journal of dairy science 96(11):7245-7259.

Pino, F., L. K. Mitchell, C. M. Jones, and A. J. Heinrichs. 2018. Comparison of diet digestibility, rumen fermentation, rumen rate of passage, and feed efficiency in dairy heifers fed ad-libitum versus precision diets with low and high quality forages. Journal of Applied Animal Research 46(1):1296-1306.

Pryce, J. E., W. J. Wales, Y. de Haas, R. F. Veerkamp, and B. J. Hayes. 2014. Genomic selection for feed efficiency in dairy cattle. Animal : an international journal of animal bioscience 8(1):1-10.

Puillet, L., D. Réale, and N. C. Friggens. 2016. Disentangling the relative roles of resource acquisition and allocation on animal feed efficiency: insights from a dairy cow model. Genetics Selection Evolution 48(1):72.

Ramin, M. and P. Huhtanen. 2013. Development of equations for predicting methane emissions from ruminants. Journal of dairy science 96(4):2476-2493.

Roche, J. R., K. A. Macdonald, C. R. Burke, J. M. Lee, and D. P. Berry. 2007. Associations among body condition score, body weight, and reproductive performance in seasonal-calving dairy cattle. Journal of dairy science 90(1):376-391.

Rong, Q., C. Huang, D. U. HuiZhang, Z. Guo, L. I. Ling, and Y. E. Sheng. 2011. Milk consumption and lactose intolerance in adults. Biomedical and Environmental Sciences 24(5):512-517.

Ross, E. M., P. J. Moate, L. C. Marett, B. G. Cocks, and B. J. Hayes. 2013. Metagenomic

Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle. PLOS ONE 8(9):e73056.

Saeys, Y., T. Abeel, and Y. Van de Peer. 2008. Robust Feature Selection Using Ensemble Feature Selection Techniques. Pages 313-325 in Proc. Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, Berlin, Heidelberg.

Schwab, C. G. and G. A. Broderick. 2017. A 100-Year Review: Protein and amino acid nutrition in dairy cows. Journal of dairy science 100(12):10094-10112.

Seymour, D. J., A. Cánovas, C. F. Baes, T. C. S. Chud, V. R. Osborne, J. P. Cant, L. F. Brito, B. Gredler-Grandl, R. Finocchiaro, R. F. Veerkamp, Y. de Haas, and F. Miglior. 2019. Invited review: Determination of large-scale individual dry matter intake phenotypes in dairy cattle. Journal of dairy science 102(9):7655-7663.

Seymour, D. J., A. Cánovas, T. C. S. Chud, J. P. Cant, V. R. Osborne, C. F. Baes, F. S. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. Journal of dairy science 103(2):1528-1540.

Shenk, J. S. and M. O. Westerhaus. 1991. Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy. Crop Science 31(2):cropsci1991.0011183X003100020049x.

Shetty, N., P. Løvendahl, M. S. Lund, and A. J. Buitenhuis. 2017. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. Journal of dairy science 100(1):253-264.

Singhal, S., R. D. Baker, and S. S. Baker. 2017. A Comparison of the Nutritional Value of Cow's Milk and Nondairy Beverages. Journal of Pediatric Gastroenterology and Nutrition 64(5):799-805.

Smith, S. B. and J. D. Crouse. 1984. Relative contributions of acetate, lactate and glucose to lipogenesis in bovine intramuscular and subcutaneous adipose tissue. The Journal of nutrition 114(4):792-800.

Song, X., E. A. M. Bokkers, P. P. J. van der Tol, P. W. G. Groot Koerkamp, and S. van Mourik. 2018. Automated body weight prediction of dairy cows using 3-dimensional vision. Journal of dairy science 101(5):4448-4459.

Soyeurt, H., C. Bastin, F. G. Colinet, V. M. R. Arnould, D. Berry, E. Wall, F. Dehareng, H. N. Nguyen, P. Dardenne, and J. Schefers. 2012. Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. Journal of dairy science 89(9):3690-3695.

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F. G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livestock Science 227:82-89.

St-Pierre, N. 2019. Managing Measures of Feed Costs: Benchmarking Physical and Economic Feed Efficiency. Vol. 2021. The Ohio State University.

STATBEL. 2019. Agricultural data. STATBEL.

Stone, J. B., G. W. Trimberger, C. E. Henderson, J. T. Reid, K. L. Turk, and J. K. Loosli. 1960. Forage Intake and Efficiency of Feed Utilization in Dairy Cattle1. Journal of dairy science 43(9):1275-1281.

Sun, D.-W. 2009. Infrared spectroscopy for food quality analysis and control. Academic press.

Türker-Kaya, S. and C. W. Huck. 2017. A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. Molecules 22(1):168.

Tomas Naes, T. I., Tom Fearn and Tony Davies. 2002. A user-friendly guide to multivariate calibration and classification Journal of Chemometrics 17(10):571-572.

Ungerfeld, E. M. 2020. Metabolic Hydrogen Flows in Rumen Fermentation: Principles and Possibilities of Interventions. Frontiers in microbiology 11:589-589.

Urrutia, N. L. and K. J. Harvatine. 2017. Acetate dose-dependently stimulates milk fat synthesis in lactating dairy cows. The Journal of Nutrition 147(5):763-769.

USDA. 2018. Milk, producer, fluid, 3.7% milk fat (SR LEGACY, 171266) Vol. Access date Oct. 01, 2020.

Vallimont, J. E., C. D. Dechow, J. M. Daubert, M. W. Dekleva, J. W. Blum, C. M. Barlieb, W. Liu, G. A. Varga, A. J. Heinrichs, and C. R. Baumrucker. 2011. Short communication: Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. Journal of dairy science 94(4):2108-2113.

VandeHaar, M. J., L. E. Armentano, K. Weigel, D. M. Spurlock, R. J. Tempelman, and R. Veerkamp. 2016. Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency1. Journal of dairy science 99(6):4941-4954.

Vanlierde, A., M. L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M. H. Deighton, F. Grandl, M. Kreuzer, B. Gredler, P. Dardenne, and N. Gengler. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. Journal of dairy science 98(8):5740-5747.

Veerkamp, R. F. and G. C. Emmans. 1995. Sources of genetic variation in energetic efficiency of dairy cows. Livestock Production Science 44(2):87-97.

Westin, R., A. Vaughan, A. M. de Passille, T. J. DeVries, E. A. Pajor, D. Pellerin, J. M. Siegford, A. Witaifi, E. Vasseur, and J. Rushen. 2016. Cow- and farm-level risk factors for lameness on dairy farms with automated milking systems. Journal of dairy science 99(5):3732-3743.

WHO. 1996. Trace elements in human nutrition and health. Public Health and Environment, World Health Organization. , Geneva, Switzerland.

WHO. 2003. Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation. Vol. 916. World Health Organization.

Wigley, K., J. L. Owens, M. Westerschulte, P. Riding, J. Fourie, and A. Werner. 2019. Photogrammetry for assessment of pasture biomass. Journal of New Zealand Grasslands 81:33-40.

Wolin, M. J. 1960. A theoretical rumen fermentation balance. Journal of dairy science 43(10):1452-1459.

Wuebbles, D. J. and K. Hayhoe. 2002. Atmospheric methane and global change. Earth-Science Reviews 57(3):177-210.

Yan, T., C. S. Mayne, F. G. Gordon, M. G. Porter, R. E. Agnew, D. C. Patterson, C. P. Ferris, and D. J. Kilpatrick. 2010. Mitigation of enteric methane emissions through improving efficiency of energy utilization and productivity in lactating dairy cows. Journal of dairy science 93(6):2630-2638.

Yusuf, R. O., Z. Z. Noor, A. H. Abba, M. A. A. Hassan, and M. F. M. Din. 2012. Methane emission by sectors: A comprehensive review of emission sources and mitigation methods. Renewable and Sustainable Energy Reviews 16(7):5059-5070.

Zucali, M., G. Battelli, M. Battini, L. Bava, M. Decimo, S. Mattiello, M. Povolo, and M. Brasca.

2016. Multi-dimensional assessment and scoring system for dairy farms. Italian Journal of Animal Science 15(3):492-503.Zucali, M., G. Battelli, M. Battini, L. Bava, M. Decimo, S. Mattiello, M. Povolo, and M. Brasca. 2016. Multi-dimensional assessment and scoring system for dairy farms. Italian Journal of Animal Science 15(3):492-503.

2

Chapter II : Mining data from milk mid-infrared spectroscopy and animal characteristics to improve the prediction of dairy cow's liveweight using feature selection algorithms based on partial least squares and Elastic net regressions

Adapted from: Lei Zhang, Anthony Tedde, Phuong Ho, Clément Grelet, Frédéric Dehareng, Eric Froidmont, Nicolas Gengler, Yves Brostaux, Dagnachew Hailemariam, Jennie Pryce, GplusE Consortium, Hélène Soyeurt. Mining data from milk mid-infrared spectroscopy and animal characteristics to improve the prediction of dairy cow’s liveweight using feature selection algorithms based on partial least squares and Elastic net regressions. Computers and Electronics in Agriculture: Under review.

Simple summary

In this Chapter, different strategies such as enlarging the data size of the calibration set and eliminating the uninformative variables were explored to improve the accuracy and the robustness of the past MIR based BW predictive model.

Abstract

Body weight (BW) of dairy cows is relevant for breeding programs and farm management to assess the maintenance requirements, reproduction performance, or health status of cow. Currently, it is still difficult to follow BW changes of individual cows routinely in large herds. Combined with animal characteristics, milk mid-infrared (MIR) spectrum was proposed as an additional source of information to predict BW under the framework of dairy herd improvement (DHI) programs. However, the presence of less informative variables in the prediction equation could impact negatively its robustness. This research aims to improve the robustness of BW regression models by applying a feature selection before modeling. A total of 5,920 BW records composed of animal characteristics and milk MIR spectrum were collected from Holstein cows. Three feature selection algorithms were applied to select the most informative variables: partial least squares regression (PLS) combined with sum of ranking difference (PLS-SRD), PLS combined with uninformative variables elimination (PLS-UVE), and the output of Elastic net regression (EN). Four herd independent validation sets and the corresponding remained calibration datasets having on average 163 and 1,708 records, respectively, were used to develop models using PLS or EN approaches. Ten-fold cross-validation was conducted to parametrize each model. Parity, days in milk (DIM), milk yield (MY), and two MIR spectral points were selected as relevant variables to predict BW. PLS (root mean square error of validation, RMSEp=60 kg) and EN (RMSEp=60 kg) regressions employing these 5 predictors were more robust than the models developed without MIR or using MIR without feature selection. The EN models had a cross-validation root mean square error of around 53 kg. The 2 MIR points explained up to 4.20% variation in predicting BW. The RMSE of validation sets using another brand of spectrometer were around 64 kg. This study confirms the possibility to predict an indicator of BW from animal characteristics and MIR variables. The variable selection procedures improved the model's robustness and transferability. The accuracy of BW prediction seems to be sufficient to provide useful information for breeding program and farm management decisions under a DHI framework.

Keywords: Dairy Cow Liveweight; Mid-Infrared; Feature Selection; Bias and Robustness; Modeling

1. Introduction

Phenotypes related directly or indirectly to feed efficiency have to be considered in animal breeding and farm management (Miglior et al., 2017; Zhang et al., 2020). The body weight (BW) is one of the indicators of importance to monitor the energy balance which is related to the cow health, reproduction, and feed efficiency (Thorup et al., 2012; Zhang et al., 2020). For instance, approximating feed intake in dairy cattle is important in order to accurately compare feed efficiency of dairy cattle. Unfortunately, measuring individual feed intake is difficult and expensive. However, BW can be used to approximate maintenance requirements (Haile-Mariam et al., 2014). For breeding purposes, a limited number of BW measurements can be sufficient. However, for management purpose, the estimation of BW changes on a routine basis is more relevant and this requires repeated estimates (Mäntysaari and Mäntysaari, 2015).

For decades, strategies for the use of BW as a management tool have been explored in both production and research conditions (Bewley and Schutz, 2008; Maltz et al., 1997). Moreover, assessing information at the individual scale is required to improve the accuracy of genetic parameters estimated for this trait (Korver, 1988). Thus, efforts have been made to monitor dairy cow BW at large and individual levels (Song et al., 2018; Hélène Soyeurt et al., 2019). Several methods exist to quantify or estimate the BW of dairy cows such as the electronic weighing scale, the 3-D camera (accuracy around 43 kg) (Kuzuhara et al., 2015), the prediction of BW from body condition score (BCS) (accuracy around 47kg) (Haile-Mariam et al., 2014), the BW prediction using animal conformation traits (accuracy varied between 37 kg and 110 kg) (Banos and Coffey, 2012; Haile-Mariam et al., 2014; Vanrobays et al., 2015). However, the first and second methods are not always available on commercial farms due to its cost. The third and fourth ones are laborious and time-consuming (Soyeurt et al., 2019; Yan et al., 2009) and subjected to the assessor's ability (Edmonson et al., 1989). Moreover, the last one is often made only one time during the cow productive life in most countries (Soyeurt et al., 2019). However, dairy herd improvement (DHI) programs can provide beneficial information to estimate BW at individual and large scales. For instance, BW can be predicted from the milk yield, number of lactation and parity (Enevoldsen and Kristensen, 1997; Kuzuhara et al., 2015; Song et al., 2018). Moreover, a larger part of the individual variability of BW can be assessed by adding the test-day milk mid-infrared (MIR) spectrum (Soyeurt et al., 2019) in the predictive model.

However, some issues could exist which indicate that these preliminary BW prediction equations could be improved. First, expanding the size of the calibration set would be meaningful to cover a larger variability of BW. Second, because the accuracy and the robustness of a predictive model evaluated from the same calibration set partly depend on the predictors used (Mehmood et al., 2020), the elimination of uninformative variables from the BW model should be relevant to avoid over-fitting.

Therefore, the accuracy and robustness of a calibration model could be improved by optimizing the variable selection before modeling (Cai et al., 2008). To address this issue, several informative variable selection algorithms have been developed such as the sum of ranking differences (SRD) procedure (Kalivas et al., 2015; Tencate et al., 2016), the uninformative variables elimination (UVE) (Centner et al., 1996), and the use of the output of elastic net (EN) penalized regressions. In brief, compared to the most usual way to parametrize a model that is only based on the root mean squared error of cross-validation (RMSEcv), SRD procedure selects the most relevant predictors based on the computation of considered merits derived from statistical parameters denoting the model complexity, variance and bias (e.g., RMSEcv, reliability of regression coefficients, L2 norms). This is conducted without being affected by parameters' weight allocation problems (Nie et al., 2019). The UVE combined with Partial Least Squares (PLS) is another method allowing to select the most relevant subset of predictors. The hypothesis of this method is that the most informative variables have a higher importance in the regression compared to the artificial noisy variables. The EN method was proposed to fix the problems in the presence of high correlations among variables; it select groups of correlated variables and conducts continuous shrinkage, and it can also simultaneously conducts variable selection; the output of this algorithm can be considered as selected variables (Zou and Hastie, 2005).

In conclusion, the objective of this research was to improve the robustness of BW prediction models, based on cow’s characteristics and milk MIR spectrum, by selecting the most informative variables based on the feature selection algorithms mentioned above.

2. Materials and Methods

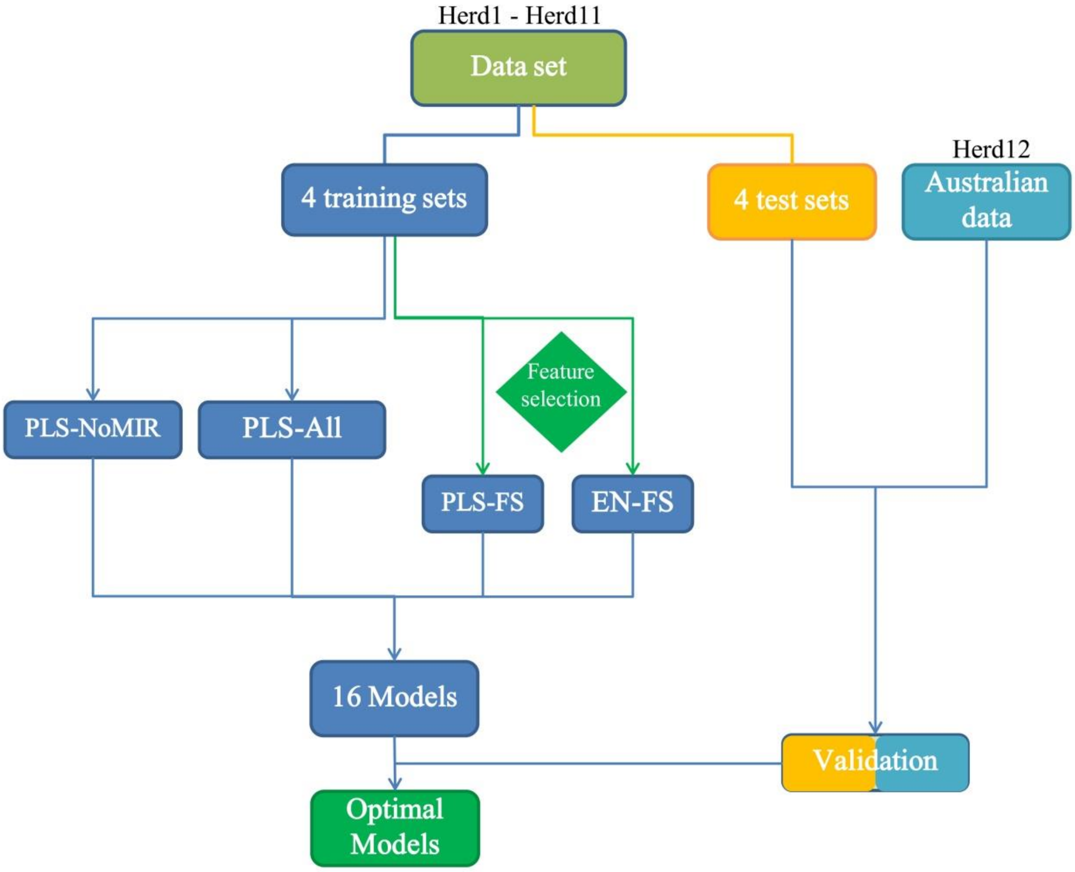


Figure 2-1. The technical flow chart of the different modeling approaches used in this study.

PLS-NoMIR=Partial least squares regression included milk yield, parity, and days in milk; PLS-All=PLS using milk yield, parity, days in milk and 277 MIR data; PLS-FS=PLS using the 5 variables selected by feature selection; EN-FS=Elastic Net regression using the 5 variables selected by feature selection.

2.1 Modeling dataset

For the data from herd1 to herd11 (Figure 2-1), a total of 1,915 records were collected from 363 Holstein cows during 2007 to 2016 from the herd of following institutions: University of Liège (Herd 5, Liège, Belgium), Walloon Agricultural Research Centre (Herd 3, Herd 4 and Herd 9,Gembloux, Belgium), University of Alberta (Herd 6, Alberta, Canada), Aarhus University (Herd 8, Tjele, Denmark), Agri-Food and Biosciences Institute (Herd 7, Northern Ireland), University College Dublin (Herd 11, Dublin, Ireland), Walloon Breeding Association (Herd 1 and Herd 2, Ciney, Belgium), Leibniz Institute for Farm Animal Biology (Herd 10, Dummerstorf, Germany). Part of those data (N=754) was initially used by Soyeurt et al. (Soyeurt et al., 2019) and part of those data were from Genotype Plus Environment (GplusE) project (http://www.gpluse.eu). The spectrum from Bentley spectrometer was standardized based on the procedures proposed by Grelet et al. (2017). Parity ranged from 1 to 11 and the number of days in milk (DIM) ranged from 1 to 512 days. Milk yield (MY) was recorded on daily basis and fat and protein contents were measured locally using Foss MilkoScan FT+ or FT6000 spectrometers (Foss, Hillerod, Denmark) or Standard Lactoscope FT-MIR automatic (PerkinElmer, Waltham, United States) in Walloon Agricultural Research

Center (Gembloux, Belgium). The spectrum was also recorded during this infrared analysis and standardized. Then, a first-derivation was applied to these data with a window of 5 spectral points to correct the baseline drift. The weighing scales used to measure the BW were manufactured by Fullwood (Shropshire, UK), Gallagher (Canley, UK) or Griffiths Elder (Bury St Edmunds, UK) instruments. The dataset was cleaned using the following editing. First, the records were cleaned according to the International Committee for Animal Recording (ICAR) standard (ICAR, 2017): milk yield between 3.0 to 99.9 kg/d, fat content between 1.5 g/dL and 9.0 g/dL, and protein content between 1.0 g/dL and 7.0 g/dL. Before modeling, the DIM records were split into 35 classes of 15 days. Due to the low number of records beyond 4th parity, parity records were split into five classes (i.e., 1, 2, 3, 4, and 5 or more). After removing the noisy MIR regions, which are well known to be related to water absorbance as described by Soyeurt et al. (2019), the first derived absorbance values located in the range of 950-1,600 cm–1, 1,750-1,800 cm–1, and 2,600-3,000 cm–1 (i.e., 277 spectral variables) were combined to DIM classes, parity classes, MY and BW to constitute the modeling dataset. So, this contained 1,871 records composed of 280 predictors collecting from 355 cows (i.e., 5.27 records per cow on average).

2.2 Prediction models

As presented in Figure 2-1, four herd independent validation datasets (herd5, herd7, herd10 or herd11) having approximately the same number of records in each set were used in this study to assess the performance of the developed models. Consequently, the calibration sets using the remaining samples had also a similar number of records between them. For each calibration set, a principal component analysis was performed on the spectral data to calculate the Global H distance (GH) of each record to detect potential spectral outliers. The number of principal components used explained 99% of the spectral variability. Records with a GH distance higher than 5 were discarded.

A common modeling procedure was applied on each calibration set. First, the predictors were scaled and centered. Second, a 10-fold cross-validation was used to choose the best parametrization of models (i.e, optimum number of latent variables (LV) for PLS and the best penalty for EN regression). The best parametrization was the one for which the next value for the considered parameters (LV or penalty) did not contribute to a substantial reduction in the RMSEcv. Third, the potential BW outliers were discarded based the prediction residuals. If the prediction residual was higher than the mean + 3SD or lower than the mean – 3SD, the BW record was discarded. Fourth, PLS models were rebuilt using this second cleaned data set and the obtained regression was validated using the corresponding validation set. Finally, the model performance was assessed by calculating the calibration (RMSEc), cross-validation (RMSEcv) and validation RMSE as well as the ratio of performance to deviation (RPD) which was the ratio of the standard deviation of BW variable to the RMSEcv. The relative error was also quantified by dividing RMSEp by BW mean calculated from the validation set. The robustness of the models was assessed according to the difference between RMSEcv and RMSEp, since in robust calibration models, the RMSEp is equal or nearly equal to the RMSEcv (Müllertz et al., 2016; Wang et al., 2018).

Four different modeling approaches were tested and compared: PLS regression including only cow's characteristics (parity + MY + DIM) named PLS-NoMIR; PLS regression using cow's characteristics and MIR spectrum (parity + MY + DIM + 277 MIR variables) as done by Soyeurt et al., (2019) and named as PLS-All; PLS and EN regressions employing the variables selected in common by the 3 used feature selection algorithms, named as PLS-FS and EN-FS, respectively. The feature selection algorithms used are explained further in this section. Consequently, a total of 16 models were built to predict test-day BW of dairy cows (Figure 2-1). Then, Pearson correlation coefficients between predictions were calculated to assess the prediction consistency between models. The optimal model will be the one having low RMSEp and RMSEcv combined with a small difference between RMSEcv and RMSEp. The performance of this optimal model was evaluated on different 60kg BW intervals through the estimation of RMSEp per class of BW. The weighted absolute regression coefficients were finally used to assess the contribution of each predictor to the BW prediction.

2.3 Feature selections

Based on the working hypothesis that the most informative variables must be selected by different feature selection algorithms from different training sets, three feature selection algorithms were performed on the 280 variables included in the 4 different training sets. Then, from those 12 selections, the common selected variables were grouped as they were assumed to be the most informative variables.

The first variable selection method was based on SRD algorithm proposed by Kalivas et al. (Kalivas et al., 2015). A 10-fold cross-validation was firstly used to predict BW from the 280 scaled and mean-centered predictors. Consequently, 10 PLS models were built, providing 10 regression coefficients per predictor. For each predictor, the mean and standard deviation (SD) of those coefficients were calculated to compute the reliability (C) (i.e., mean divided by SD (Centner et al., 1996)). The higher the value of C, the more informative and reproducible is the predictor. In a second step, following a decreasing value of C, predictors were included into new PLS models one by one. The tested number of LV was the one considered as optimal in PLS-All and the ones around it having a similar mean and SD of RMSEcv. Then, for each number of LV, different models were tested considered a different number of predictors ranging from LV + 1 to 280. For each tested model, 8 merits were calculated in order to assess the model relevancy. Merit 1 was the jaggedness (J) denoting the model complexity and the degree of over-fitting of a regression model (eq 1) (Gowen et al., 2011).

$$J = \sum_{i=2}^{m} \sqrt{\left(b_i - b_{i-1}\right)^2} \qquad \text{(eq 2-1)}$$

where the b_i denotes the regression coefficient (b) of i^{th} predictor of a model; m is the number of predictors in this model. Merit 2 (called B2) corresponds to the L2 norm of a model (eq 2) and reflects the variance of the model.

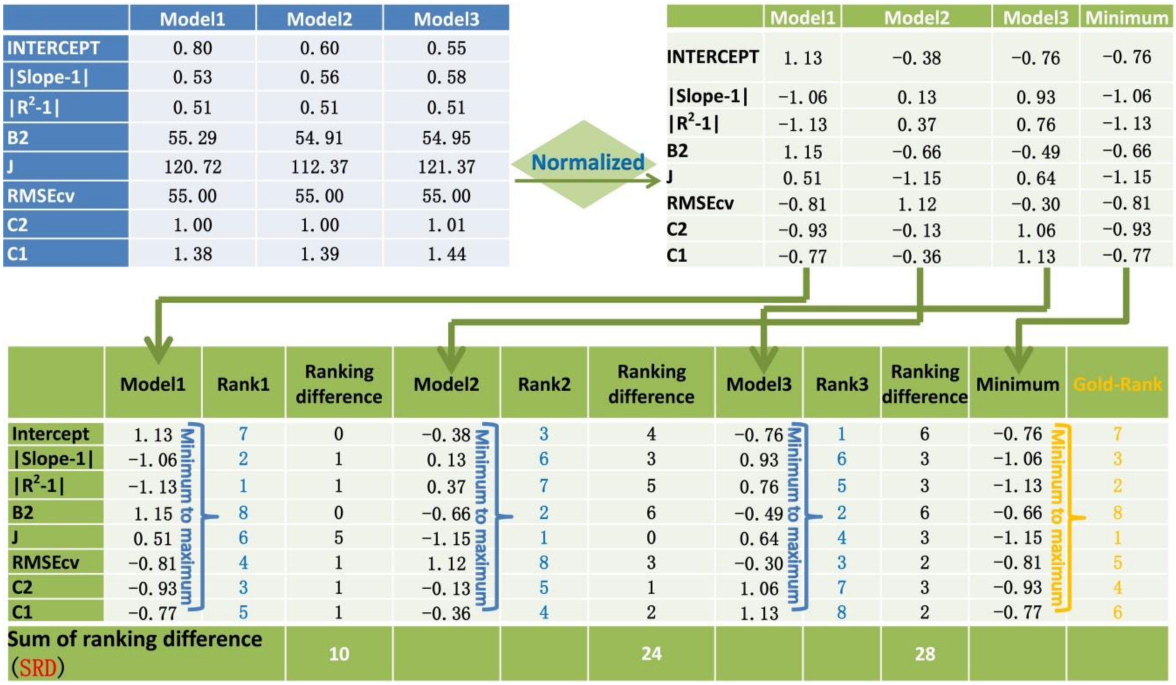
$$\mathrm{B2} = \sqrt{\sum_{i=1}^{m} {b_i}^2} \qquad \text{(eq 2-2)}$$

where b_i is the regression coefficient (b) of the i^{th} predictor in a model; m is the number of predictors in this model. Merit 3 was the RMSEcv denoting the model accuracy. Merit 4 (C1, eq 3) and Merit 5 (C2, eq 4) reflect the model bias and variance.

$$\mathrm{C1} = \left(\frac{\|B2\|-\|B2\|min}{\|B2\|\mathrm{max}-\|B2\|min}\right) + \left(\frac{\mathrm{RMSECV} - \mathrm{RMSECV}min}{RMSE\mathrm{CVmax}-RMSE\mathrm{CV}min}\right) \quad \text{(eq 2-3)}$$

$$\mathrm{C2} = \left(\frac{\|J\|-\|J\|min}{\|J\|\mathrm{max}-\|J\|min}\right) + \left(\frac{\mathrm{RMSECV} - \mathrm{RMSECV}min}{RMSE\mathrm{CVmax}-RMSE\mathrm{CV}min}\right) \quad \text{(eq 2-4)}$$

where the max and min represent the maximum and minimum values of B2 observed among the models. Merit 6 to 8 were defined after the calibration process and correspond to the absolute value of $R^2 - 1$ ($|R^2-1|$) where R is the calibration correlation coefficient, the absolute value of the regression slope - 1 (|slope -1|), and the absolute value of the estimated intercept (|intercept|) calculated from the linear regression fitted from the measured and predicted BW records. The lower the values of the 8 merits, the better the PLS model is fitted.



	Model1	Model2	Model3
INTERCEPT	0.80	0.60	0.55
|Slope-1|	0.53	0.56	0.58
$\|R^2-1\|$	0.51	0.51	0.51
B2	55.29	54.91	54.95
J	120.72	112.37	121.37
RMSEcv	55.00	55.00	55.00
C2	1.00	1.00	1.01
C1	1.38	1.39	1.44

	Model1	Model2	Model3	Minimum
INTERCEPT	1.13	−0.38	−0.76	−0.76
|Slope-1|	−1.06	0.13	0.93	−1.06
$\|R^2-1\|$	−1.13	0.37	0.76	−1.13
B2	1.15	−0.66	−0.49	−0.66
J	0.51	−1.15	0.64	−1.15
RMSEcv	−0.81	1.12	−0.30	−0.81
C2	−0.93	−0.13	1.06	−0.93
C1	−0.77	−0.36	1.13	−0.77

	Model1	Rank1	Ranking difference	Model2	Rank2	Ranking difference	Model3	Rank3	Ranking difference	Minimum	Gold-Rank
Intercept	1.13	7	0	−0.38	3	4	−0.76	1	6	−0.76	7
|Slope-1|	−1.06	2	1	0.13	6	3	0.93	6	3	−1.06	3
$\|R^2-1\|$	−1.13	1	1	0.37	7	5	0.76	5	3	−1.13	2
B2	1.15	8	0	−0.66	2	6	−0.49	2	6	−0.66	8
J	0.51	6	5	−1.15	1	0	0.64	4	3	−1.15	1
RMSEcv	−0.81	4	1	1.12	8	3	−0.30	3	2	−0.81	5
C2	−0.93	3	1	−0.13	5	1	1.06	7	3	−0.93	4
C1	−0.77	5	1	−0.36	4	2	1.13	8	2	−0.77	6
Sum of ranking difference (SRD)			10			24			28		

Figure 2-2. Workflow of sum of ranking differences (SRD) algorithm

The calculation of SRD value was based on a matrix, which was constituted by 8 rows and up to 822 columns. The workflow is shown in Figure 2-2. First, to make the merits comparable at a same amplitude, all merits calculated for each tested model were normalized by models. Second, the gold-rank was defined using the minimum normalized merit value obtained from column (i.e., among the models) and ranked according to an increasing trend by row (i.e., among the merits). Third, for each model (i.e., column), a model ranking based on an increasing trend of their normalized merits value were defined. Fourth, the ranking difference of each model was calculated using the gold-rank and the corresponding rank of each merit of each model. Finally, the sum of ranking difference of each model was calculated by accumulating all ranking differences obtained for the considered modeling. Thus, from the example illustrated in Figure 2-2, Model 1 has the best prediction performances among these 3 models due to its minimum SRD value. Furthermore, to assess the consistency of the ranking result, the rank reliability measure (RRM) (eq 5) (Tencate et al., 2016) was calculated from the SRD values based on the 8 merits and used to choose the optimal model and the corresponding predictors. The optimal model was the one with the lowest RRM value. For all the tested models, the RRM was calculated as follows:

$$RRM_i = \left(\frac{\sigma_{FR_i} - \sigma_{FR_{min}}}{\sigma_{FR_{max}} - \sigma_{FR_{min}}}\right) + \left(\frac{\overline{FR}_i - \overline{FR}_{min}}{\overline{FR}_{max} - \overline{FR}_{min}}\right) \qquad \text{(eq 2-5)}$$

where σ_{FR_i} and $\overline{FR}_i$ are the standard deviation and mean of SRD scores of tested models and scaled to range at 0 and 1 using the minimum and maximum values of all the models. Finally, the predictors included in the optimal model having the lowest RRM value were considered as the most interesting combination of predictors. If equal RRM were obtained for different models, the one with the lowest RMSEcv was considered as having a better fitting.

The second method used for the selection of informative variables combined PLS and UVE which was proposed by Vitezslav et al. (Centner et al., 1996). First, the predictors were scaled and centered. Second, a matrix of artificial noise variables having the same number of rows and columns than the calibration set was randomly created and added into the predictor matrix column by column. The minimum and maximum value of the noise was set according to the range of the MIR spectra used (–0.35 and 0.36 in this case) in order to have the same magnitude than the MIR signal. A 10-fold cross-validation was applied on the new predictor matrix. A Monte Carlo process (Li et al., 2009) was simulated 8 times during the PLS procedure in order to acquire a stability in the estimation of the regression coefficient from both the MIR and artificial noise variables. Finally, 10 PLS models were generated. Similar to the SRD method, the C value was calculated for each predictor. The information amplitude of the noise must be determined in advance to estimate the suitable cutoff. The amplitude was set at 10^{-11} and an arbitrary value k was set as proposed by Cai et al. (Cai et al., 2008). The k value was adopted based on the average of C value estimated for each predictor from the artificial noise matrix. Finally, the predictors having higher C rank than the artificial noises were selected as the most informative variables.

The third feature selection method was based on the output of the penalized EN regression including the whole scaled and centered predictors. The alpha parameter was set at 0.5 to be between Ridge and Lasso penalized regressions. The tested range for the penalty (called lambda) varied between 0.5 and 1. The optimal lambda was fixed according to the obtained RMSEcv estimated from a 10-fold cross-validation. Finally, the variables with a regression coefficient different than 0 were considered as informative variables.

2.4 External validation

To assess the model transferability, the 16 developed regressions were furtherly validated using a completely independent data set coming from another country (i.e., herd12, validation RMSE expressed as RMSEv) and MIR spectral data obtained from another brand of spectrometer. This validation set totalized 4,005 records collected between 2015 and 2017 from 231 cows belonging to the research farm of the Victorian Department of Jobs, Precincts and Regions (Melbourne, Australia). MIR spectra as well as fat and protein contents were provided by a Bentley spectrometer (Model 2000, Chaska, MN, USA). The BW was measured using walkover scales (DeLaval, Tumba, Sweden) as described by HO et al. (Ho et al., 2019). As the BW equation was initially developed from milk Foss MIR spectra, the Bentley spectra were standardized (Grelet et al., 2017) to be expressed on the Foss basis. The RMSE estimated between the fat and protein contents predicted externally from the standardized spectra and the ones provided by the Bentley spectrometer were 0.11 and 0.27 g/dL of milk, respectively. To ensure the quality of the standardized spectral data, the absolute difference between the fat content predictions (standardized vs. Bentley) was calculated for each record. Records having an absolute fat difference above 0.30g/dL of milk were discarded. Then, the cleaning procedure and data editing were the same as the ones used for the modeling dataset. The final validation dataset contained 3,956 records. The average parity was 2.05 with a range comprised from 1 to 8. The DIM records ranged from 37 to 161 days with a mean of 104 days.

2.5 Computation

All computations and modeling were performed using R software (version 3.6.2) (R Core Team, 2019). The descriptive statistics were analyzed using base package (R Core Team, 2019). The PLS were implemented using the pls package (Mevik et al., 2019) and the caret package (Max et al., 2019)) as also for EN regression. The PLS-SRD was computed using a homemade script using R software (version 3.6.2). The PLS-UVE procedure was performed partly from the plsVarSel package (Mehmood et al., 2012)).

3. Results

3.1 Descriptive statistics

The descriptive statistic of the cleaned modeling dataset and for the Australian data (herd12) is shown in Table 2-1. From the modeling dataset (i.e., herd1 until herd11), the average milk yield was 30.51 ± 11.99 kg/d with a range comprised between 4.30 and 60.50 kg/d. The average fat and protein contents were 3.96 and 3.29 g/dL of milk, respectively. The BW varied between 309 and 984 kg with an average of 619 ± 79 kg. From Australian data (i.e., herd12), the average BW was 550 kg and ranged between 340 kg and 770 kg with a standard deviation of 65 kg.

As shown in Figure 2-3, for herds used in the modeling sets (i.e., from herd 1 to herd 11), the mode of studied parity was 2 (Figure 2-3A). The DIM records covered the 35 DIM classes (classes over 25 were shown as ⩾25) and most records were distributed in the 1 to 4 DIM classes (Figure 2-3B). BW records were normally distributed (Figure 2-3C) and globally, BW increased with the parity to a stable trend above the third parity (Figure 2-3D). The average BW was 540, 607, 662, 666 and 694 kg for first, second, third, fourth and fifth or more parity respectively.

Table 2-1 Descriptive statistics of cleaned data sets.

		BW		DIM		Parity		Milk yield		Fat		Protein	
Herd	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	32	680	71.	220.8	142.2	2.66	1.21	21.83	8.07	3.91	0.52	3.53	0.47
2	41	680	79	217.3	112.5	3.00	2.44	28.67	9.44	4.27	0.54	3.54	0.36
3	60	603	38	166.8	101.6	2.10	1.00	21.68	6.28	4.03	0.68	3.27	0.45
4	70	656	79	159.3	85.5	3.54	1.05	21.20	6.73	4.18	0.73	3.34	0.39
5	149	618	72	185.3	112.2	2.56	1.57	19.68	7.04	3.97	1.01	3.55	0.51
6	358	628	69	125.4	77.6	2.36	1.35	19.39	5.20	3.16	1.13	3.11	0.31
7	188	612	84	28.0	13.1	2.80	1.57	30.29	11.07	4.15	0.61	3.19	0.33
8	635	599	86	28.5	12.6	2.40	1.55	37.69	10.47	4.03	0.86	3.39	0.34
9	23	573	51	26.0	11.2	2.48	1.53	34.48	9.36	3.95	0.66	3.15	0.29
10	180	612	69	28.1	12.4	2.44	0.64	41.30	6.66	4.58	1.42	3.18	0.35
11	135	654	57	30.0	12.0	3.78	1.44	34.76	7.37	4.47	0.88	3.10	0.33
12	3,956	550	65	104.0	18.7	2.05	1.06	25.77	4.95	4.18	0.96	3.17	0.34

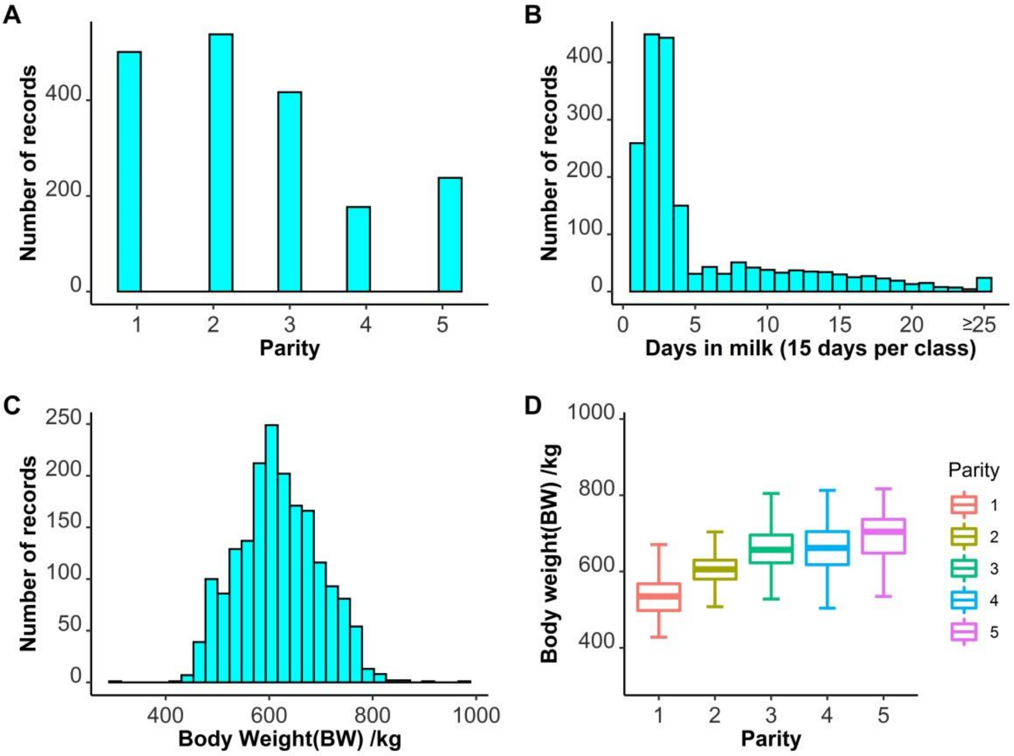


Figure 2-3. Distribution of parity (A), days in milk (B), and body weight (C), and the body weight evolution by parity (D) from the cleaned modeling data sets.

3.2 Predictive models and informative variable selection

Table 2-2 summarizes the calibration and validation performances observed for the 16 developed models. As R^2 of a model depends highly on the data structure like the distribution and the range of the data (Davies and Fearn, 2006), RMSE is the most relevant statistical parameter to assess the performance of a model (Grelet et al., 2020). So, only validation RMSE are mentioned in Table 2-2. Discarding of potential BW outliers from residual analysis lead to a slightly different number of records between the calibration sets. Therefore, the

10-fold cross-validation performances cannot be fully compared between models using the same dataset. However, we can compare the validation performances between methods and assess the variability of prediction accuracy for one method thanks to the use of several validation datasets. So, the means and SD of RMSEcv, RMSEp and RMSEv as well as the absolute differences between RMSEp and RMSEcv estimated for each kind of models are mentioned in Table 2-3.

Table 2-2 Calibration and validation performances obtained for the 16 developed models predicting test-day body weight (kg) of dairy cows.

Method	**Validation Herd**	**Herd5 (N=149)**	**Herd7 (N=188)**	**Herd10 (N=180)**	**Herd11 (N=135)**
	Model number	1	2	3	4
PLS-NoMIR (Npredictors=3; Parity, milk yield, and days in milk)	N samples	1699	1656	1662	1714
	N latent variables	2	2	2	2
	RMSEcv (kg)	53	53	52	53
	R^2cv	0.52	0.51	0.55	0.52
	RMSEp (kg)	60	55	68	58
	RMSEv (kg)	71	70	70	69
PLS-All (Npredictors=280; Parity, milk yield, days in milk, and 277 MIR spectral points)	N samples	1,699	1,658	1,663	1,716
	N latent variables	5	5	6	5
	RMSEcv (kg)	50	51	48	48
	R^2cv	0.59	0.55	0.63	0.62
	RMSEp (kg)	59	59	64	59
	RMSEv (kg)	64	63	93	126
PLS-FS (Npredictors=5; Parity, milk yield, days in milk, and 2 MIR spectral points)	N samples	1696	1655	1661	1713
	N latent variables	2	2	2	2
	RMSEcv (kg)	53	53	52	53
	R^2cv	0.54	0.52	0.55	0.53
	RMSEp (kg)	61	54	67	57
	RMSEv (kg)	64	64	64	63
EN-FS (Npredictors=5; Parity, milk yield, days in milk, and 2 MIR spectral points)	N samples	1696	1655	1661	1713
	Lambda	0.5	0.45	0.25	0.15
	RMSEcv (kg)	53	53	52	53
	R^2cv	0.54	0.52	0.55	0.53
	RMSEp (kg)	60	54	67	57
	RMSEv (kg)	65	64	64	63

PLS-NoMIR=Partial least squares regression included milk yield, parity, and days in milk; PLS-All=PLS using milk yield, parity, days in milk and 277 MIR data; PLS-FS=PLS using the 5 variables selected by feature selection; EN-FS=Elastic Net regression using the 5 variables selected by feature selection; R2cv=10-fold cross-validation coefficient of determination; RMSEp=root mean squared error of validation using herd5, herd7, herd10 or herd11; RMSEv=root mean squared error estimated from herd12 (Australian data).

The averaged RMSEp for PLS-NoMIR was 60 ± 5.57 kg (Table 2-3). The validation relative prediction error and RPD varied from 8.90% to 13.14%, and 1.4 to 1.5, respectively. The absolute difference between averaged RMSEcv and RMSEp was 8 kg. Higher RMSEp were observed for herd5 and herd10 but more constant error was observed between models using herd12 (RMSEv) (Table2-2).

The PLS-All models had LV number comprised between 5 and 6. The inclusion of 277 spectral data decreased the reproducibility of the models. Indeed, the SD of RMSEcv for PLS-All was more than 3 times higher than the one observed for PLS-NoMIR (Table 2-3). The same phenomenon was observed for RMSEv. Only the SD of RMSEp decreased (Table

2-3). The model calibrated using a training set leaving herd10 and herd11 out had the worse prediction accuracy (Table 2-2). The RPD ranged from 1.48 to 1.63. The relative error of the RMSEp ranged between 9.04% and 15.46%. The average absolute difference between RMSEcv and RMSEp was 11 kg (Table 2-4) which was higher than the models from PLS-NoMIR approach. This high difference and the high observed SD of RMSE suggested a low robustness of PLS-All models.

Table 2-3 Mean and standard deviation of cross-validation and validation root mean squared errors (RMSE) per kind of models as well as the absolute difference between cross-validation RMSE (RMSEcv) and validation RMSE (RMSEp, RMSEv).

Models	**RMSEcv**		**RMSEp**		**Diff_abs**		**RMSEv**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
PLS-NoMIR	53	0.50	60	5.57	8	6.03	70	0.82
PLS-All	49	1.57	60	2.65	11	3.88	86	29.80
PLS-FS	53	0.50	60	5.62	7	6.05	64	0.80
EN-FS	53	0.50	60	5.57	7	6.02	64	0.69

RMSEcv=10-fold cross-validation root mean squared error; RMSEp=RMSE estimated from herd5, herd7, herd10, or herd11; Diff_abs denotes the absolute difference between RMSEcv and RMSEp; PLS-FS=PLS using the 5 variables selected by feature selection; EN-FS=Elastic Net regression using the 5 variables selected by feature selection.

The instability of PLS-All models could be related to the presence of uninformative or non-reproducible in time or among instruments variables in the model. Indeed, the presence of such variables can impact negatively the robustness of the model (Centner et al., 1996). This confirms the relevance of testing different feature selection algorithms to select the best combination of predictors. Table 2-4 summarizes the results of the feature selections realized in this study.

Multiple models were created during the PLS-SRD selection due to the inclusion of different LV number and predictors ranked based on the reliability of their regression coefficients. As mentioned in the materials and methods section, the number of LVs tested for PLS-SRD feature selection was the one selected for PLS-All and the ones around it having similar mean and SD of RMSEcv. Therefore, the range of tested LV were 5 to 7, 4 to 5, 5 to 6 and 4 to 5, respectively for the first, second, third and fourth calibration datasets. Based on the number of tested LV and the number of predictors possible for each model (ranged from LV+1 to 280), 822 models were built from the first calibration set (275 for LV=5, 274 for LV=6, and 273 for LV=7). The number of models tested for the 3 other calibration sets were 551, 549 and 551, respectively (Table 2-4). Globally, RRM value obtained for each model fluctuated for all calibration sets along with the change of LV and the number of predictors included in the model (Figure 2-4). The range was comprised from 0 to 1.69. The optimal model is the one having the lowest RRM. So, for the first dataset, the 79th, 109th and 121st models had the lowest RRM value (around 0.02; Figure 2-4A) revealing a similar robustness. However, 109th model presented the lowest RMSEcv. So, the 114 predictors including in this model into 5 LVs were considered as the most informative based on the first calibration set (i.e., leaving herd5 out). The same procedure was repeated for all other datasets and the results were summarized in Table 2-4. The best models are illustrated for each dataset in Figure 2-4 by the red arrow. From PLS-SRD results, the number of selected variables ranged from 61 to 138 predictors (Table 2-4) with an average of 105 predictors. In other words, the decrease of features number varied between 51% and 78% of the whole predictors.

The PLS-UVE method selected 106, 120, 69 and 116 informative variables based on the 4 datasets used (Table 2-4). As for the PLS-SRD method, a strongly lower number of predictors was selected using the third dataset. Except for the third and fourth datasets, PLS-UVE selected a lower number of predictors compared to PLS-SRD. The number of LVs considered in PLS-UVE models were similar to the ones used by PLS-SRD models and turned around 4 to 5.

Using EN feature selection, the number of selected variables ranged from 62 to 73 (Table 2-4). The decrease of variable number was stronger here compared to PLS-SRD and PLS-UVE except for the third dataset. The number of discarded samples after the residual analysis was around 27 records which was similar to the ones obtained from PLS-SRD and PLS-UVE methods. The lambda penalty fixed based on the 10-fold RMSE was the same for all datasets and set to 0.5.

Table 2-4 Parameters of the variable selection algorithms based on PLS-SRD, PLS-UVE and Elastic Net.

Method[1]	**Validation Herd**	**Herd5 (N=149)**	**Herd7 (N=188)**	**Herd10 (N=180)**	**Herd11 (N=135)**
	Number of Models	822	551	549	551
PLS-SRD	N samples	1,689	1,654	1,643	1,707
	N predictors	114	138	61	106
	N latent variables	5	5	6	5
PLS-UVE	N samples	1,696	1,657	1,661	1,712
	N predictors	106	120	69	116
	N latent variables	5	5	4	5
EN	N samples	1,696	1,656	1,663	1,711
	N predictors	67	62	73	62
	Penalty parameter	λ=0.5	λ=0. 5	λ=0.5	λ=0.5

[1] *PLS=Partial least squares regression; SRD=Sum of Ranking Difference; UVE=Uninformative Variable Elimination; EN=Elastic Net.*

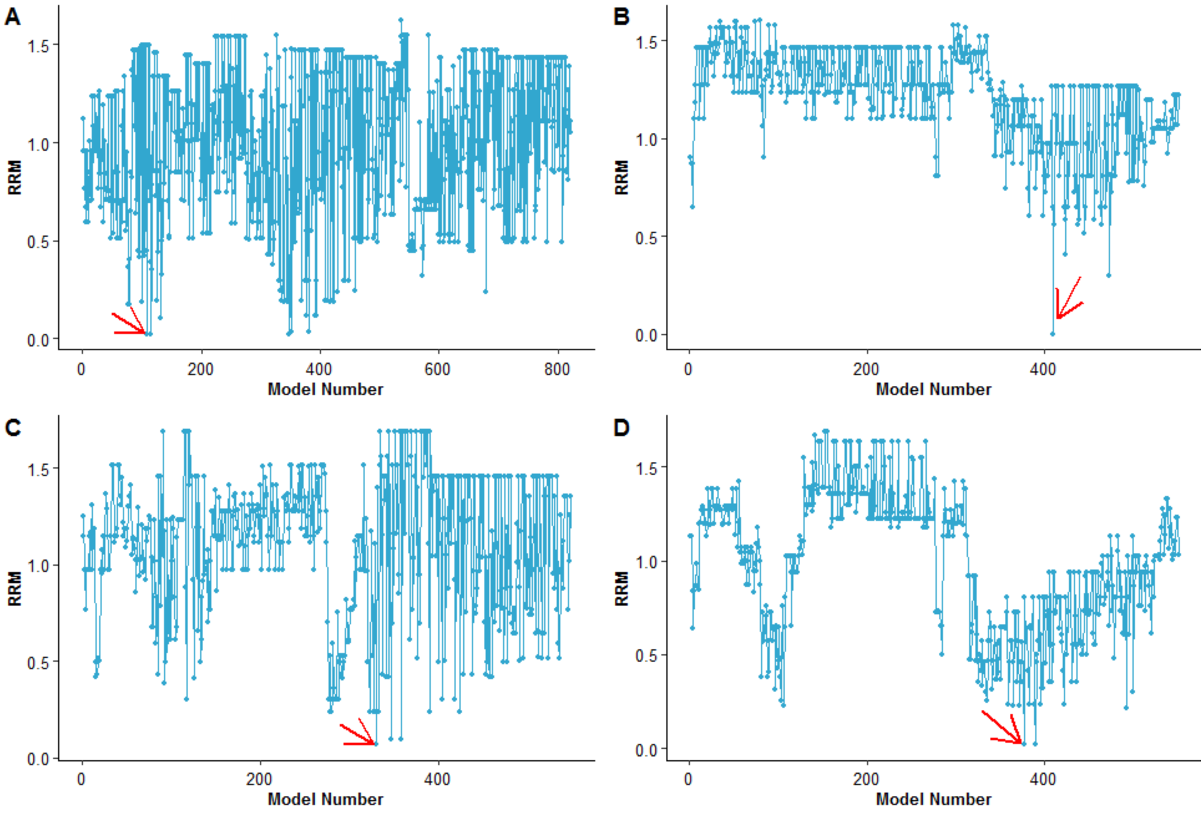


Figure 2-4. The trend of RRM value across the model number for the 4 tested calibration sets (A=first dataset until D=fourth dataset; the red arrow shows the optimal model)

Based on the hypothesis that the most informative variables must be selected by different feature selection algorithms from different training sets, the intersection of selected features was defined (Figure 2-5). Therefore, from those 12 selections (i.e., 4 training sets and 3 feature selection algorithms), only 5 predictors were in common. They were then included into the third (PLS-FS) and fourth (EN-FS) modeling approaches to develop BW prediction models.

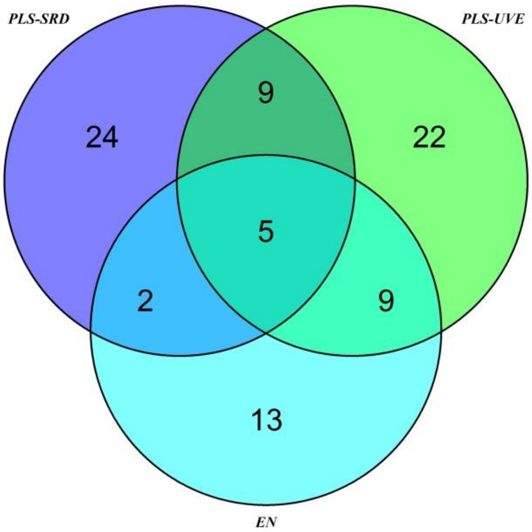


Figure 2-5. Intersection of groups of selected variables among the PLS-SRD, PLS-UVE and EN feature selections and the four calibration sets.

The RMSEcv obtained for PLS-FS was around 53 kg (Table 2-2). The RPD ranged from 1.4 to 1.5. The relative predictive error ranged from 8.72% to 11.24% among the validation sets. The RMSEp estimated from those four validation sets was 60 ±5.62 kg (Table 2-3). The trend of validation performance between datasets was similar to the one observed with PLS-NoMIR (Table 2-3). The absolute difference between RMSEcv and RMSEp was 7 kg (Table 2-3) which was lower than the one observed from models of PLS-NoMIR and PLS-All. Similar to PLS-All and PLS-NoMIR, the worse validation results were observed for herd10 (Table 2-2). The external validation on Australian data (herd12) had a RMSE of 64 ±0.80 kg. The standard deviation is both lower than PLS-NoMIR and PLS-All. These same five selected predictors were also included in an EN regression. RMSEcv was similar for all four data sets and ranged between 52 and 53 kg (Table 2-2). The range of RPD and the relative error were from 1.4 to 1.5, and 8.72% to 11.50%, respectively. Validation performances were similar to PLS-FS but the SD of RMSEp and RMSEv were lower (Table 2-3). Therefore, these models were the optimal ones: low absolute difference between RMSEcv and RMSEp, low mean and SD of RMSEcv and RMSEPp and low mean and SD of RMSEv.

Besides the calibration and validation results, it is also interesting to assess the prediction consistency between the developed models. So, the 16 models were applied to the 4 validation datasets. The correlation coefficients among predictions using the 16 models applied on the four validation sets were high and ranged from 0.73 to 1.00. For all calibration sets, the predictions of PLS-All models using the whole 280 predictors showed lower correlation coefficients (ranged from 0.73 to 0.96) with predictions obtained from PLS-NoMIR, PLS-FS and EN-FS models. Besides the lower RMSEp, RMSEv and their corresponding lower SD of models developed using EN-FS approach, the prediction correlation among EN-FS models ranged between 0.99 to 1 also revealing the robustness of this kind of modeling.

3.3 The robust model and variable importance

All EN-FS models presented a similar regression trend. Figure 2-6A illustrates the relationships between predicted and real measured BW records (when herd7 as validation set). The slope of fitted regression line was 0.51, which was moderately close to 1; the bias was 302.94 kg. For other calibration sets, the slope ranged between 0.51 and 0.53; the bias varied from 288.74 to 303.39kg. No obvious herds presented a special distribution in the population cloud (Figure 2-6A). The validation prediction errors were higher for the high and low BW (Figure 2-6A). There was no obvious region revealed herds had extremely poor prediction at the BW range between 520 and 700 kg (Figure 2-6A). To be more specific, the RMSEp and RMSEv were also estimated for 8 different 60kg BW intervals (Figure 2-6B). Clearly, the cows with BW lower than 400 kg or higher than 760 kg showed a relatively lower prediction accuracy. However, BW between 460 to 700 kg revealed a lower RMSE.

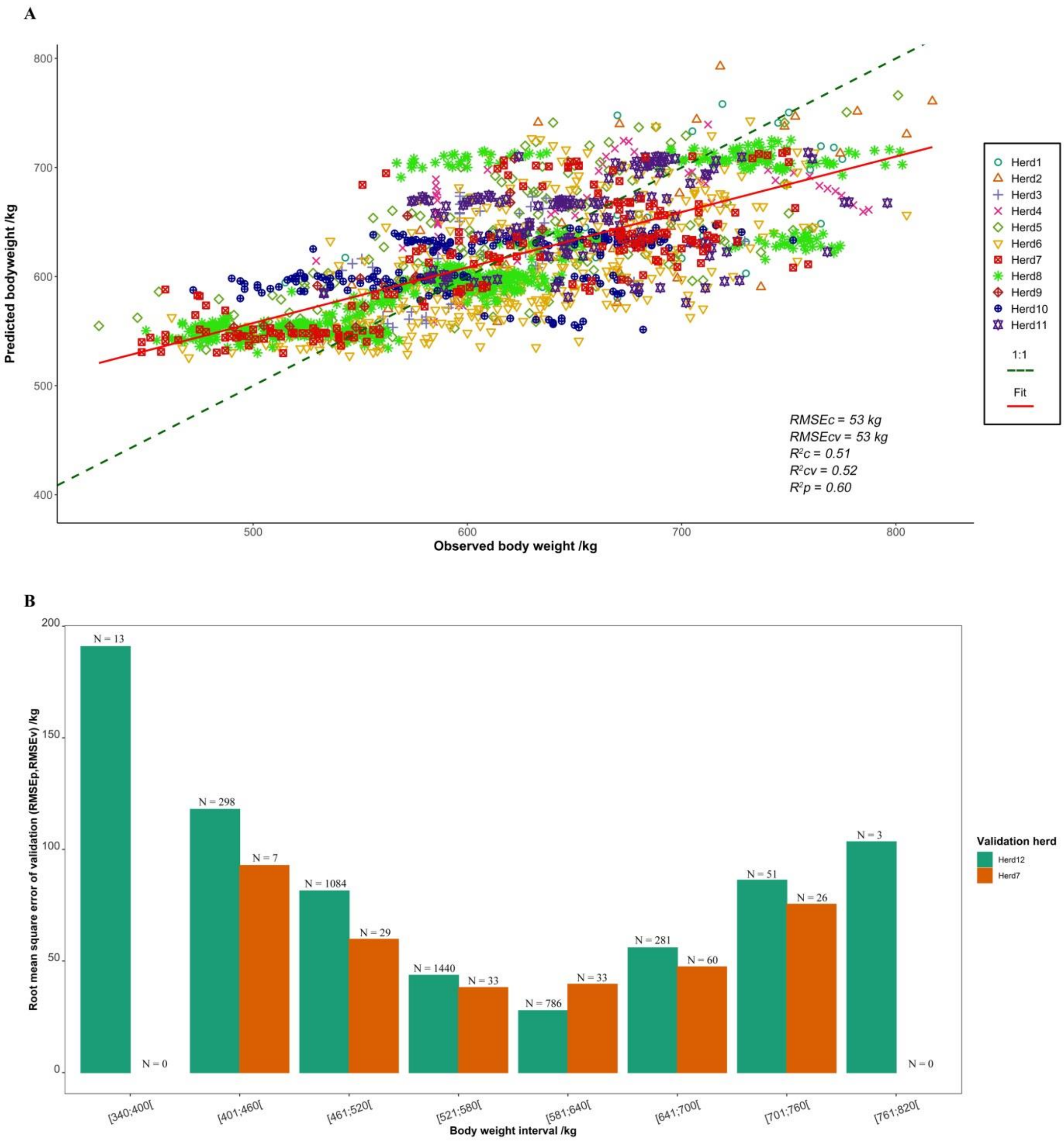


Figure 2-6. The observed and predicted body weight of the calibration and validation herds (herd7) based on the EN regression using the final 5 selected variables (A); The validation root mean square error distribution across different body weight stage (interval=60 kg; B).

The variable relative importance estimated for the 5 selected variables included in the EN-FS model from the 4 datasets are given in Table 2-5. Based on the C value, the sequence of important variables was the parity, MIR0123, DIM, milk yield, and MIR0094. Those spectral points represent the MIR region of 1,396.21 cm^{-1} and 1,284.36 cm^{-1}. The variability of variable importance among models were low except for milk yield and MIR0094 (Table 2-5). Finally, according to the R^2 of the cross-validation step during the EN-FS regression based on the final 5 selected variables, the adding of MIR contributed on average 2.13 %

(R^2cv increased from around 0.53 to around 0.54) and up to 4.20% information to predict BW globally.

Table 2-5 The relative importance of selected variables of the Fin-PLS models using different calibration sets.

Variables relative importance (%)	**Herd5**	**Herd7**	**Herd10**	**Herd11**	**Mean**	**SD**	**CV**	**C value (Mean/SD)**	**Importance ranking**
Parity	100.00	100.00	100.00	100.00	100.00	0.00	0.00	∞	1
MIR0123	12.11	12.49	11.15	11.67	11.86	0.57	4.84	2066.72	2
Days in milk (days)	40.88	41.38	43.74	45.74	42.93	2.25	5.24	1908.92	3
Milk yield (kg/day)	19.51	17.01	23.58	25.56	21.41	3.87	18.07	553.34	4
MIR0094	5.76	0.18	1.53	3.68	2.79	2.45	87.98	113.66	5

4. Discussion

To ensure the development of a robust model, the modeling dataset must be representative of the variability existing for the trait of interest (i.e., BW in this study) but also for the predictors. According to the mean and SD (Table 2-1), the average coefficient of variation for milk yield, fat, and protein contents were of 29.27%, 20.45% and 11.32%. This variability of the main milk components allowed indirectly to ensure a good milk spectral variability. The variability coefficient of BW varied between 6.30% and 14.36% in this study which was similar to the range of 8.17% to 16.47% reported in literature (Koenen et al., 1999; Kuzuhara et al., 2015; McParland et al., 2015; Song et al., 2018; Yan et al., 2009, 2006). The datasets used can be therefore considered as representative of the cow population.

Due to the interest of knowing the test-day BW at an individual scale for management and breeding purposes, there is a need to try to predict this trait with predictors that are easily available on field by DHI organizations. So, it is why this study focused on the use of milk yield, DIM and parity as well as the milk MIR spectrum as predictors. These cow's characteristics related predictors were not chosen randomly. Indeed, these traits routinely recorded by DHI at cow level were positively related to BW. The correlation coefficients with measured BW were 0.12 for milk yield, 0.67 with parity, and 0.20 for DIM. Similar correlation values were already reported in the literature. For instance, Song et al. reported correlation coefficients of BW with parity and DIM equal to 0.81 and 0.18, respectively (Song et al., 2018). Although the correlation value between DIM and BW is not as high as the one observed with parity, this trait is valuable to be included in a prediction model as it is indirectly related to the pregnancy status of a cow (Song et al., 2018; Zhang et al., 2020). Moreover, the energy reallocation occurring after new calving cows (i.e., first days in milk) leads to a negative energy balance leading to a mobilization of body reserves (Soyeurt et al., 2019) influencing BW. Parity with its high correlation with BW is incontestably a relevant predictor. BW increased until the fourth parity and then stayed stable (Figure 2-3D). This trend was expected as the cow's body keep growing until the fourth or fifth parity (Artegoitia et al., 2013; Mellado et al., 2011; Ray et al., 1992). Other authors reported the interest of using DIM and parity to predict BW as well (Enevoldsen and Kristensen, 1997; Kuzuhara et al., 2015). Along with the growth of cow's body, the increase of mammary volume from primiparous to multiparous of a healthy cow always result in an increase of milk yield (Davis and Hughson, 1988), explaining the positive relationship observed between milk yield and measured BW. The regression including those predictors (i.e., PLS-NoMIR) had a RMSEcv (53kg; Table 2-3) similar to the value of 56 kg reported by Soyeurt et al. (2019) based on a model including parity, milk yield, DIM and test month. Even if these models seem to present satisfactory prediction performances on cross-validation and herd validation, the application of them on Australian data provided worse results suggesting a poor model transferability.

It is known that the milk composition can be related to the energy balance of the cow which is also related to the BW changes of a dairy cows (Friggens et al., 2007). Therefore, using the milk MIR spectral data reflecting the global milk composition as additional predictors is relevant to predict test-day BW. The modeling already proposed by Soyeurt et al. (2019) was not robust enough. Indeed, the absolute differences between RMSEp and RMSEcv were globally higher for PLS-All compared to the PLS-NoMIR models (Table 2-3). Moreover, the SD of RMSEcv and RMSEv were more than 3 times higher than the ones observed for PLS-NoMIR suggesting the absence of robustness for those models. This could be related to the presence of uninformative variables in the model (Gottardo et al., 2016).

Kalivas and Palmer (Kalivas and Palmer, 2014) reported that a harmonious model requires a suitable balance of bias and variance to gain an optimal performance. However, using too few informative variables will lead to under-fit and the model is therefore not complex enough to capture the variability in the samples; whereas, adopting too many variables will result in an over-fitted model which is too specific to the calibration set used and presents therefore a poor predictive ability (Lavine, 2003). So, an optimum compromise must be found. In this context, a total of 3 different feature selection algorithms were employed in this study as the working hypothesis was that if a predictor is selected by different feature selection algorithms, the probability to highlight an informative variable is high. By using 4 different calibration sets, we have also reinforced this idea because the variable must be also selected from all datasets. By this way, only 5 predictors were in common from the 12 selection procedures done in this study. There were then included in PLS and EN regressions. Both models (PLS-FS and EN-FS) had similar cross-validation and herd validation performances compared to PLS-NoMIR but a largely better validation performance on the Australian data (Table 2-3) suggesting a good transferability of those models. Moreover, as SD of RMSEv was slightly lower (while the SD of RMSEp was equal to PLS-NoMIR) for EN-FS, this model was assumed to be the most optimal model to predict test-day BW for dairy cows. In conclusion, the elimination of uninformative variables allows to keep the most informative variables and reduce the effect from noisy variables (Cai et al., 2008; Centner et al., 1996) leading to a better transferability of those models (i.e., mean and SD of RMSEv; Table 2-3). The lowest correlation found between predictions made using PLS-All approach confirmed the presence of noisy information.

It is interesting to notice that the number of variables were different following the feature selections used and ranged from 61 to 138. At the end, by realizing the intersection, we have finally only 5 variables (Parity, DIM, milk yield, and two MIR spectral points). This means that using only one feature selection algorithm and one dataset could not be sufficient and lead to over-fit the model. Normally PLS algorithm should deal with that (Wold et al., 2001). Indeed, by fixing its LVs by taking into account the variability of predictors and the trait to be predicted, PLS must give a low absolute regression coefficient to uninformative variable. This was not observed in this study. This could be related to the size of the dataset. Maybe by increasing the number of records, the PLS should improve its ability to recognize the most informative variables. This is why the use of a repeatability file (such as predicted BW as an indicator) composed of spectra without reference value could be interesting (Soyeurt et al., 2019; Vanrobays et al., 2015) and to test the selection of the whole MIR spectral points as well.

The relationship between the observed BW and the ones predicted using EN-FS records revealed the goodness of the calibration model (Figure 2-6A). No obvious herd effect was visible on this graph. This can be explained by the high robustness of the EN-FS revealing by the small difference between RMSEp and RMSEcv. Furthermore, the information from the calibration herds have a good representativeness to the validation herd, especially in the models when leaving out herd5 and herd7 (lower RMSEp and RMSEv). Meanwhile, as we can observe from the fitted line of the predicted BW versus the real measured BW in Figure 2-6A, the cows having extreme BW had a relative lower prediction performance. As reported by Thomas and Ge (2000), the representativeness and structure of a calibration set affect the accuracy of a predictive model. So, this phenomenon could be explained by the BW variability of the calibration set as there were a low amount of high and low BW in the calibration set (Figure 2-3C). But the cows with a moderate BW had a better prediction accuracy.

The five selected variables were the parity, MIR0123, DIM, milk yield, and MIR0094. The mean and SD of relative variable importance values, and C value calculated from the EN-FS models showed that those first two predictors were the most important in all models. By averaging the relative variable importance values calculated for DIM, this predictor was the third more important and it was very close to MIR0123 which is related to the MIR region at 1,396 cm^{-1}. Not like the first three predictors, the variability of the milk yield relative variable importance values and the C value suggested that the importance of this trait can fluctuate between models. Similar trend was observed for the last predictor, MIR0094, which is related to the MIR region at 1284.36 cm^{-1}. These two MIR spectral data were located in the MIR spectral region related to the carbohydrates and organic acids (Bittante and Cecchinato, 2013; Picque et al., 1993). This could be related to the fact that the milk carbohydrate-related contents or fatty acids reflect partly the status of energy metabolism of body which influences the BW changes (Vacek and Beran, 2012; Yan et al., 2006). The highlighting of parity, DIM and milk yield as important predictors of BW is in agreement with the findings mentioned in the first part of this discussion about the traits significantly related to BW (Enevoldsen and Kristensen, 1997; Kuzuhara et al., 2015; Song et al., 2018). The interest of using MIR information as additional predictors was confirmed in this study. In the past, McParland et al. (McParland et al., 2015) and Soyeurt et al. (Soyeurt et al., 2019) reported also the potential interest of using MIR spectra to predict BW changes or BW. However, in this study, the part of the explained BW variability brought by MIR data up to 4.20% was lower than the 11.67% reported by Soyeurt et al (2019). Moreover, the current study revealed also the interest of adding MIR data to enhance the model robustness and transferability.

The RPD values for the EN-FS models averaged at 1.5, suggesting that the predicted BW is more an indicator of this trait. Indeed, a RPD between 1.5 and 2.0 representing a discrimination of low and high value of dependent variance (Saeys et al., 2005). However, although the value of RPD between 1.5 and 2.0 showed limited prediction accuracy, the prediction results could be accepted in a breeding view point (Chen et al., 2011). So, the RPD obtained in the current study around an average value of 1.5 indicating the usefulness of the models developed in this study to predict the body weight as least to distinguish the light and the heavier ones, and could be useful especially in the breeding programs. For instance, this predicted BW could be used in a breeding goal to measure the maintenance requirements (Koenen et al., 1999).

The RMSEp around 60 kg for the EN-FS representing 8.72% to 11.46% of the real BW for different validation sets. This level of accuracy is acceptable, since, not like a certain research experiment, the measurement of BW in commercial farms, is also affected by many aspects like the measurement period (i.e., before or after milking of a cow, before or after feeding etc.) and the calibration of the weighing scale. In this study, no protocol was fixed to collect the BW records. Therefore, a certain uncertainty exists around the reference BW values used. However, it is important to notice that the accuracy obtained from 3-D camera or BCS regression methods ranged around 40kg to 50kg regarding the cross-validation (Haile-Mariam et al., 2014; Kuzuhara et al., 2015). Although the prediction error isn't as small as the daily body weight changes (around 1 kg) resulted from body reserves mobilize (Picque et al. (1993), whereas this extremely small changes appears normally during the early lactation stage (Bittante et al., 2013). So, in the current study, the accuracy of predicted BW is at least meaningful at global herd level. In this view, the mean live BW is often used in field to determine a global decision such as the nutritional requirement of a herd related to the goodness management of maintenance. So, from the breeding and herd level nutrient determination view, the residual of BW prediction, in the current study, using the selected models met the usage requirement practically.

5. Conclusion

This research confirmed the feasibility of predicting an indicator of test-day BW at an individual cow level using traits routinely recorded by DHI such as parity, days in milk, milk yield, and even animal conformation types. The use of the intersection of groups of variables using PLS-SRD, PLS-UVE and EN feature selection algorithms allowed to improve the model robustness and its transferability to another brand of spectrometers. The selection of parity, DIM, milk yield, MIR0094 and MIR0123 were logical based on the relationships between measured BW and those traits found in the literature. Even if PLS algorithm is normally able to deal with uninformative variables, this study showed the interest of realizing an elimination of uninformative variables before the final modeling. Moreover, in the context of this study, we demonstrated that the use of one feature selection algorithm and one data set was not enough to isolate the most informative variables. The EN-FS models developed in this study could be easily implemented into a routine DHI framework allowing to the creation of a large scale database of BW indicator. However, the herd specific BW evolution features could be considered in the future studies.

6. Acknowledgments

The China Scholarship Council (CSC) is gratefully acknowledged for funding Mr. Zhang's PhD grant.

The members of GplusE Consortium which including: Mark Crowe, Alan Fahey, Fiona Carter, Elizabeth Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O'Flaherty, Sinead Hallinan, Claire Wathes, Mazdak Salavati, Zhangrui Cheng, Ali Fouladi, Geoff Pollott, Dirk Werling, Beatriz Sanz Bernardo, Conrad Ferris, Alistair Wylie, Matt Bell, Mieke Vaneetvelde, Kristof Hermans, Miel Hostens, Geert Opsomer, Sander Moerman, Jenne De Koster, Hannes Bogaert, Jan Vandepitte, Leila Vandevelde, Bonny Vanranst, Klaus Ingvartsen, Martin Tang Sorensen, Johanna Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Leslie Foldager, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Torben Larsen, Victor H. Silva de Oliveira, Cinzia Marchitelli, Federica Signorelli, Francesco Napolitano, Bianca Moioli, Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney, Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen, Huanchun Chen, Junlong Zhao, Liguo Yang, Guohua Hua, Chen Tan, Guiqiang Wang, Michel Bonneau, Marlène Sciarretta, Armin Pearn, Arnold Evertson, Linda Kosten, Anders Fogh, Thomas Andersen, Matthew Lucy, Chris Elsik, Gavin Conant, Jerry Taylor, Deborah Triant, Nicolas Gengler, Michel Georges, Frederic Colinet, Marilou Ramos Pamplona, Hedi Hammami, Catherine Bastin, Haruko Takeda, Aurelie Laine, Anne-Sophie Van Laere, Rodrigo Mota, Saied Naderi Darbagshahi, Frederic Dehareng, Clement Grelet, Amelie Vanlierde, Eric Froidmont, Frank Becker, Martin Schulze, and Sergio Palma Vera are deeply acknowledged in this study.

The authors gratefully thank the farmers and staff of milk recording organizations for the collection of data used in this study: Ellinbank (Australia), University of liege (Liege, Belgium), Walloon Agricultural Research Centre (Gembloux, Belgium), University of Alberta (Alberta, Canada), Aarhus University (Tjele, Denmark), Agri-Food and Biosciences Institute (Northern Ireland), University College Dublin (Dublin, Ireland), Walloon Breeding Association (Ciney, Belgium), Leibniz Institute for Farm Animal Biology (Dummerstorf, Germany).

References

Artegoitia, V., Meikle, A., Olazabal, L., Damián, J.P., Adrien, M.L., Mattiauda, D.A., Bermudez, J., Torre, A., Carriquiry, M., 2013. Milk casein and fatty acid fractions in early lactation are affected by nutritional regulation of body condition score at the beginning of the transition period in primiparous and multiparous cows under grazing conditions. J. Anim. Physiol. Anim. Nutr. (Berl). 97, 919–932.

Banos, G., Coffey, M.P., 2012. Technical note: Prediction of liveweight from linear conformation traits in dairy cattle. J. Dairy Sci. 95, 2170–2175. https://doi.org/10.3168/jds.2011-4838

Bewley, J.M., Schutz, M.M., 2008. An Interdisciplinary Review of Body Condition Scoring for Dairy Cattle. Prof. Anim. Sci. 24, 507–529. https://doi.org/10.15232/s1080-7446(15)30901-3

Bittante, G., Cecchinato, A., 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J. Dairy Sci. 96, 5991–6006. https://doi.org/10.3168/jds.2013-6583

Bjørn-Helge Mevik, R.W. and K.H.L., 2019. Partial Least Squares and Principal Component Regression. R package version 2.7-2.

Cai, W., Li, Y., Shao, X., 2008. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. Chemom. Intell. Lab. Syst. 90, 188–194. https://doi.org/https://doi.org/10.1016/j.chemolab.2007.10.001

Centner, V., Massart, D.-L., de Noord, O.E., de Jong, S., Vandeginste, B.M., Sterna, C., 1996. Elimination of Uninformative Variables for Multivariate Calibration. Anal. Chem. 68, 3851–3858. https://doi.org/10.1021/ac960321m

Chen, G.L., Zhang, B., Wu, J.G., Shi, C.H., 2011. Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy. Anim. Feed Sci. Technol. 165, 111–119. https://doi.org/https://doi.org/10.1016/j.anifeedsci.2011.02.004

Davies, A.M.C., Fearn, T., 2006. Back to basics: calibration statistics. Spectrosc. Eur. 18, 31–32.

Davis, S.R., Hughson, G.A., 1988. Measurement of functional udder capacity in lactating Jersey cows. Aust. J. Agric. Res. 39, 1163–1168.

Edmonson, A.J., Lean, I.J., Weaver, L.D., Farver, T., Webster, G., 1989. A Body Condition Scoring Chart for Holstein Dairy Cows. J. Dairy Sci. 72, 68–78. https://doi.org/10.3168/jds.S0022-0302(89)79081-0

Enevoldsen, C., Kristensen, T., 1997. Estimation of body weight from body size measurements and body condition scores in dairy cows. J. Dairy Sci. 80, 1988–1995.

Friggens, N.C., Ridder, C., Løvendahl, P., 2007. On the use of milk composition measures to predict the energy balance of dairy cows. J. Dairy Sci. 90, 5453–5467. https://doi.org/10.3168/jds.2006-821

Gibb, M.J., Ivings, W.E., Dhanoa, M.S., Sutton, J.D., 1992. Changes in body components of autumn-calving Holstein-Friesian cows over the first 29 weeks of lactation. Anim. Sci. 55, 339–360.

Gottardo, P., Penasa, M., Lopez-Villalobos, N., De Marchi, M., 2016. Variable selection procedures before partial least squares regression enhance the accuracy of milk fatty acid composition predicted by mid-infrared spectroscopy. J. Dairy Sci. 99, 7782–7790.

Gowen, A.A., Downey, G., Esquerre, C., O'Donnell, C.P., 2011. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. J. Chemom. 25, 375–381. https://doi.org/10.1002/cem.1349

Grelet, C., Froidmont, E., Foldager, L., Salavati, M., Hostens, M., Ferris, C.P., Ingvartsen, K.L., Crowe, M.A., Sorensen, M.T., Fernandez Pierna, J.A., Vanlierde, A., Gengler, N., Dehareng, F., 2020. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. J. Dairy Sci. https://doi.org/10.3168/jds.2019-17910

Grelet, C., Pierna, J.A.F., Dardenne, P., Soyeurt, H., Vanlierde, A., Colinet, F., Bastin, C., Gengler, N., Baeten, V., Dehareng, F., 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100, 7910–7921.

Haile-Mariam, M., Gonzalez-Recio, O., Pryce, J.E., 2014. Prediction of liveweight of cows from type traits and its relationship with production and fitness traits. J. Dairy Sci. 97, 3173–3189. https://doi.org/10.3168/jds.2013-7516

Ho, P.N., Marett, L.C., Wales, W.J., Axford, M., Oakes, E.M., Pryce, J.E., 2019. Predicting milk fatty acids and energy balance of dairy cows in Australia using milk mid-infrared spectroscopy. Anim. Prod. Sci. 60, 164–168. https://doi.org/10.1071/AN18532

ICAR, 2017. Section 2 - Guidelines for Dairy Cattle Milk Recording, in: Guidelines. p. 27.

Jensen, C., Østergaard, S., Bertilsson, J., Weisbjerg, M.R., 2015. Responses in live weight change to net energy intake in dairy cows. Livest. Sci. 181, 163–170. https://doi.org/10.1016/j.livsci.2015.09.016

Kalivas, J.H., Héberger, K., Andries, E., 2015. Sum of ranking differences (SRD) to ensemble multivariate calibration model merits for tuning parameter selection and comparing calibration methods. Anal. Chim. Acta 869, 21–33.

Kalivas, J.H., Palmer, J., 2014. Characterizing multivariate calibration tradeoffs (bias, variance, selectivity, and sensitivity) to select model tuning parameters. J. Chemom. 28, 347–357. https://doi.org/10.1002/cem.2555

Koenen, E.P.C., Groen, A.F., Gengler, N., 1999. Phenotypic variation in live weight and live-weight changes of lactating Holstein-Friesian cows. Anim. Sci. 68, 109–114. https://doi.org/10.1017/s135772980005013x

Korver, S., 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: a review. Livest. Prod. Sci. 20, 1–13.

Kuzuhara, Y., Kawamura, K., Yoshitoshi, R., Tamaki, T., Sugai, S., Ikegami, M., Kurokawa, Y., Obitsu, T., Okita, M., Sugino, T., 2015. A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system. Comput. Electron. Agric. 111, 186–193.

Lavine, B., 2003. A user-friendly guide to multivariate calibration and classification, Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002. J. Chemom. 17, 571–572. https://doi.org/10.1002/cem.815

Li, H., Liang, Y., Xu, Q., Cao, D., 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta 648, 77–84.

Maltz, E., Devir, S., Metz, J.H.M., Hogeveen, H., 1997. The body weight of the dairy cow: I. Introductory study into body weight changes in dairy cows as a management aid. Livest. Prod. Sci. 48, 175–186. https://doi.org/10.1016/S0301-6226(97)00024-9

Mäntysaari, P., Mäntysaari, E.A., 2015. Modeling of daily body weights and body weight changes of Nordic Red cows. J. Dairy Sci. 98, 6992–7002. https://doi.org/10.3168/jds.2015-9541

Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, C.C. and T.H., 2019. caret: Classification and Regression Training.

McParland, S., Kennedy, E., Lewis, E., Moore, S.G., McCarthy, B., O'Donovan, M., Berry, D.P., 2015. Genetic parameters of dairy cow energy intake and body energy status predicted using mid-infrared spectrometry of milk. J. Dairy Sci. 98, 1310–1320. https://doi.org/10.3168/jds.2014-8892

Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in Partial Least Squares Regression. Chemom. Intell. Lab. Syst. 118, 62–69.

Mehmood, T., Sæbø, S., Liland, K.H., 2020. Comparison of variable selection methods in partial least squares regression. J. Chemom. n/a, e3226. https://doi.org/10.1002/cem.3226

Mellado, M., Antonio-Chirino, E., Meza-Herrera, C., Veliz, F.G., Arevalo, J.R., Mellado, J., De Santiago, A., 2011. Effect of lactation number, year, and season of initiation of lactation on milk yield of cows hormonally induced into lactation and treated with recombinant bovine somatotropin. J. Dairy Sci. 94, 4524–4530.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100, 10251–10271.

Müllertz, A., Perrie, Y., Rades, T., 2016. Analytical techniques in the pharmaceutical sciences. Springer.

Nie, M., Meng, L., Chen, X., Hu, X., Li, L., Yuan, L., Shi, W., 2019. Tuning parameter identification for variable selection algorithm using the sum of ranking differences algorithm. J. Chemom. 33, e3113. https://doi.org/10.1002/cem.3113

Picque, D., Lefier, D., Grappin, R., Corrieu, G., 1993. Monitoring of fermentation by infrared spectrometry: Alcoholic and lactic fermentations. Anal. Chim. Acta 279, 67–72. https://doi.org/https://doi.org/10.1016/0003-2670(93)85067-T

R Core Team, 2019. A language and environment for statistical computing. R Foundation for Statistical Computing.

Ray, D.E., Halbach, T.J., Armstrong, D. V, 1992. Season and lactation number effects on milk production and reproduction of dairy cattle in Arizona. J. Dairy Sci. 75, 2976–2983. https://doi.org/10.3168/jds.S0022-0302(92)78061-8

Saeys, W., Mouazen, A.M., Ramon, H., 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosyst. Eng. 91, 393–402. https://doi.org/10.1016/j.biosystemseng.2005.05.001

Song, X., Bokkers, E.A.M., van der Tol, P.P.J., Koerkamp, P.W.G.G., van Mourik, S., 2018. Automated body weight prediction of dairy cows using 3-dimensional vision. J. Dairy Sci. 101, 4448–4459.

Soyeurt, Hélène, Froidmont, E., Dufrasne, I., Hailemariam, D., Wang, Z., Bertozzi, C., Colinet, F.G., Dehareng, F., Gengler, N., 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227, 82–89.

Soyeurt, H, Froidmont, E., Dufrasne, I., Hailemariam, D., Wang, Z., Bertozzi, C., Colinet, F.G., Dehareng, F., Gengler, N., 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227, 82–89.

Tencate, A.J., Kalivas, J.H., White, A.J., 2016. Fusion strategies for selecting multiple tuning parameters for multivariate calibration and other penalty based processes: A model updating application for pharmaceutical analysis. Anal. Chim. Acta 921, 28–37. https://doi.org/https://doi.org/10.1016/j.aca.2016.03.046

Thomas, E. V, Ge, N., 2000. Development of Robust Multivariate Calibration Models. Technometrics 42, 168–177. https://doi.org/10.1080/00401706.2000.10485996

Thorup, V.M., Edwards, D., Friggens, N.C., 2012. On-farm estimation of energy balance in dairy cows using only frequent body weight measurements and body condition score. J. Dairy Sci. 95, 1784–1793. https://doi.org/10.3168/jds.2011-4631

Ducháček, J., Vacek, M., Stádník, L., Beran, J., Okrouhlá, M., 2012. Changes in milk fatty acid composition in relation to indicators of energy balance in Holstein cows. Acta Univ. Agric. Silvic. Mendelianae Brun. 60, 29–38. https://doi.org/10.11118/actaun201260010029

Vanrobays, M.L., Vandenplas, J., Hammami, H., Froidmont, E., Gengler, N., 2015. Short communication: Novel method to predict body weight of primiparous dairy cows throughout the lactation. J. Dairy Sci. 98, 692–697. https://doi.org/10.3168/jds.2014-8504

Wang, Y., Jiang, F., Gupta, B.B., Rho, S., Liu, Q., Hou, H., Jing, D., Shen, W., 2018. Variable Selection and Optimization in Rapid Detection of Soybean Straw Biomass Based on CARS. IEEE Access 6, 5290–5299. https://doi.org/10.1109/ACCESS.2017.2763596

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130. https://doi.org/https://doi.org/10.1016/S0169-7439(01)00155-1

Yan, T., Mayne, C.S., Keady, T.W.J., Agnew, R.E., 2006. Effects of Dairy Cow Genotype with Two Planes of Nutrition on Energy Partitioning Between Milk and Body Tissue. J. Dairy Sci. 89, 1031–1042. https://doi.org/10.3168/jds.S0022-0302(06)72170-1

Yan, T., Mayne, C.S., Patterson, D.C., Agnew, R.E., 2009. Prediction of body weight and empty body composition using body size measurements in lactating dairy cows. Livest. Sci. 124, 233–241. https://doi.org/10.1016/j.livsci.2009.02.003

Zhang, L., Gengler, N., Dehareng, F., Colinet, F., Froidmont, E., Soyeurt, H., 2020. Can we observe expected behaviors at large and individual scales for feed efficiency-related traits predicted partly from milk mid-infrared spectra? Animals 10, 873. https://doi.org/10.3390/ani10050873

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B (Statistical Methodol). 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

3

Chapter III : Improving the quality of milk mid-infrared predictions recorded by dairy herd improvement organizations using data quality checks

Adapted from: Lei Zhang, C.F. Li, F. Dehareng, C. Grelet, F. Colinet, N. Gengler, Y. Brostaux, H. Soyeurt. Improving the quality of milk mid-infrared predictions recorded by dairy herd improvement organizations using data quality checks. Animals: Revised.

Simple summary

There is a growing interest of using milk mid-infrared (MIR) spectrometry to obtain new phenotypes useful in the complex management of dairy farms. Those predictions can be erroneous for many reasons even if the predictive models used are accurate. In this Chapter, three different statistical methods were proposed to detect abnormal predictions and limit the spectral extrapolation in order to improve the quality of the MIR-based predictions.

Abstract

The use of abnormal milk MIR spectrum strongly affects the prediction quality even the predictive model used is accurate. So this record must be detected after or before the prediction process to avoid erroneous spectral extrapolation or the use of poor quality spectral data by dairy herd improvement (DHI) organizations. For financial or practical reasons, adapting the quality protocol used currently to improve the accuracy of fat and protein contents is unfeasible. So, this study proposed three different statistical methods easy to implement by DHI organizations to solve this issue: the deletion of 1% of extreme high and low predicted values (M1), the deletion of records based on the global-H (GH) distance (M2), and the deletion of records based on the absolute fat residual value (M3). Additionally, the combinations of these three methods were investigated. A total of 346,818 milk samples were analyzed by MIR spectrometry to predict the contents of fat, protein, and fatty acids. Then, the same traits were also predicted externally using their corresponded standardized MIR spectra. The interest of cleaning procedures was assessed by estimating the root mean square differences (RMSD) between those internal and external predicted phenotypes. All methods allowed to decrease RMSD, the gain ranged from 0.32% to 41.39%. Based on the obtained results, the “M1 and M2” combination should be preferred to be more parsimonious in the data loss as it had the higher ratio of RMSD gain to data loss. This method deleted the records based on the 2% extreme predictions and a GH threshold set at 5. However, to ensure the lowest RMSD, the “M2 or M3” combination, considering as GH threshold of 5 and an absolute fat residual difference set at 0.30 g/dL of milk was the most relevant. Both combinations involved M2 confirming the high interest of calculating the GH distance for all samples to predict. However, if it is impossible to estimate the GH distance due to a lack of relevant information to compute this statistical parameter, the obtained results recommended the use of M1 combined with M3. The limitation used in M3 must be adapted by the DHI as this will depend on the spectral data and the equation used. The methodology proposed in this study can be applied to find the best threshold for the considered database and could be implemented on any MIR-based phenotypes.

Keywords: Milk Component Prediction; Mid-infrared Spectrum; Mahalanobis Distance; Quality Assurance System; Holstein cow

1. Introduction

Since the 1980s, the milk contents like fat and protein are predicted using milk mid-infrared (MIR) spectrometry. Thanks to the high accuracy of the predictive models used, those phenotypes were included in the equation defining the milk price given to the farmers for their production. Due to the direct economic link between those MIR predictions and the farm profitability, most of milk laboratories follow a quality insurance procedure proposed by the Internal Dairy Federation during the most commonly called “ring test”. During this test, identical sets of modified milk samples (i.e., often 10 milk samples having a large variation of fat content) are measured by the MIR spectrometers in the different laboratories. The predictions are then compared to the contents given by certified reference chemical measurement methods. This allows an inter-laboratories comparison, an intra-laboratory comparison (i.e., machine differences or operators) as well as the definition of the slope and the bias corrections used to calibrate more accurately the fat and protein predictions given by the manufacturer’s equation after the prediction process. Moreover, if the bias and slope corrections for the considering trait are too strong, the ring test allows alerting also the laboratory that an adjustment is needed or, a maintenance of the spectrometer is required. A procedure is also implemented to check a daily (short-term) stability of the instrument between two ring tests. Therefore, the predictions from the spectrometer could be assumed close to the real values due to the bias and slope corrections done during the ring test procedure. However, the ring test has a cost due to the creation of milk samples, the sending of them, the reference chemical analysis as well as the laboratory comparison. But this cost is the price to ensure the best precision and accuracy for the fat and protein phenotypes predicted by MIR. In parallel, as the milk laboratories involved in the milk analysis for the milk payment and the routine performance recording of dairy cows are the same, the ring test is a great opportunity for the dairy herd improvement (DHI) organizations to ensure a high quality for the fat and protein phenotypes used in their breeding selection programs.

Farms require more and more highly skilled labor and techniques to handle complex cow management strategies (Gargiulo et al., 2018). Fast and low-cost phenotypes related to milk production, animal husbandry, animal health, and reproduction are therefore required currently for dairy farming (Boichard and Brochard, 2012; Gengler et al., 2016; Lamb, 1978). This explains why, increasingly, these phenotypes were predicted with a fluctuating accuracy by milk MIR spectrometry (Gengler et al., 2016): the milk fatty acids composition (De Marchi et al., 2011; Soyeurt et al., 2006), the protein composition (De Marchi et al., 2011), the mineral contents (Soyeurt et al., 2009), the milk acidity (De Marchi et al., 2009), the energy status of cow (McParland et al., 2011), the nitrogen utility efficiency (Grelet et al., 2020b), the cow fertility (Delhez et al., 2020; Ho et al., 2019), the methane emissions (Vanlierde et al., 2018), and the detection of milk adulteration (Hansen and Holroyd, 2019). So, this means that all of those phenotypes are derived from the same raw data: the MIR spectrum. For this reason, more and more DHI organizations record the milk MIR spectrum of each tested cow in their database in order to perform a specific prediction externally using their own predictive models. For instance, the Walloon Breeding Association (Awé, Ciney, Belgium) was the first to realize this recording. However, compared to fat and protein predicted contents, no quality protocol is currently implemented in the milk laboratory (except the alert for the machine maintenance) or in the DHI organizations for these new MIR traits. The improvement of the ring test procedure to take into account these traits is sometimes unfeasible as it is impossible to quantify them using chemical analysis on milk (i.e., methane, energy balance...). Moreover, for traits measurable directly on milk, the high cost of the reference chemical analysis will increase dramatically the final cost of the ring test impacting potentially its large use by milk laboratories or DHI organizations, especially because no direct economic return for the farmer exists for those traits currently. So, an alternative low-cost and quick quality protocol must be defined.

To achieve this objective, it is important to highlight the different factors influencing the final accuracy of a MIR prediction. First, the predictive model itself can provide an inaccurate prediction due to the variability of the dataset used to build it. From a dataset containing reference data and their corresponding milk MIR spectra, predictive models can be developed and easily applied to new milk spectra to predict the phenotypes of interest. However, the performance of this modelling can fluctuate regarding the conditions of application. The stability of the predictive model, more often called robustness, is worth considering. For decades, there is no standard definition of robustness (Zeaiter et al., 2004). It is certain that a high accuracy (i.e., low root mean square error (RMSE)) coupled with a low RMSE variability under various conditions contribute to robust predictive models. Grelet et al. (2020a) reported that the inclusion of the variability related to milk spectra, breeds, herds, and diets can notably improve the model robustness. Thomas and Ge (2000) also concluded that the validity of a predictive model depends partly on the representativeness and the structure of the calibration set used. However, for practical and financial reasons, it is impossible to include all samples as needed to cover the entire variability of field data (Melfsen et al., 2013). This could potentially affect the prediction accuracy for new samples. This means that even if we have a high quality model, the obtained predictions can be erroneous due to the spectral extrapolation as the sample variability is not taken into account in the calibration set. Consequently, knowing the calibration set used to develop a predictive model is of interest to limit the prediction of samples which are out of the variability covered by this set. Second, a poor quality prediction can be obtained due to the use of poor quality spectral data. Unfortunately, there is no consensus about the best strategies to be implemented routinely to address these both issues. As mentioned previously, a strategy based on the analysis of traits using reference chemical analysis is not feasible for financial or practical reasons. However, a quality protocol based on statistical procedures applied directly or indirectly on the spectral data massively available by DHI organizations could be relevant and cost efficient.

Indeed, multivariate distance can be calculated between the spectrum to be predicted and those used in the calibration set in order to limit spectral extrapolation. Mahalanobis distance (Mahalanobis, 1936) (also called H distance) is preferable to the Euclidean distance as the H distance takes the variability and the internal correlations of samples into consideration (Dokas et al., 2002; Mark, 1986). However, the Global H (GH) distance is preferred in practice because it is not affected by the number of variables used to calculate this distance. The GH distance is defined as the ratio of squared H distance to the number of variables used (De Maesschalck et al., 2000; Shenk and Westerhaus, 1991). A GH value equal to 3 was proposed to eliminate potential outliers (Shenk and Westerhaus, 1991). However, in the case of multivariate normal model, less than 99% of the population would fall within the boundary of 3 and the optimal limit of the GH value depends actually on the size of the calibration set, the number of variables used, and the alpha risk limit adopted (Whitfield et al., 1987). The high quality of information that can be provided by the calculation of such a distance could be implemented in milk laboratories or DHI organizations in the future.

To identify potential outliers for a particular trait, experimenters often use the mean ± 3 times the standard deviation if the trait is normally distributed (alpha=0.3%). If the GH distance is not available and as the MIR traits are often normally distributed and due to the high quantity of information available by DHI organizations or milk laboratories, a cleaning based on these easily to compute position statistical parameters is relevant. Moreover, these potential outliers can be related to a spectral extrapolation during the prediction process or the use of poor quality spectral data. This last case could also be detected using the content of fat given directly by the spectrometer as this prediction is routinely adjusted for bias and slope by the milk laboratories during the ring test. Due to this correction and the high accuracy of the predictive model used, this predicted content can therefore be assumed to be a control value. The estimation of a large error between this control value and the value predicted using an external calibration model (i.e., without bias and slope corrections) could highlight the presence of abnormal spectra due to analytical issues or a wrong association between sample and data. Recently, Dale et al. (2019) proposed to clean a MIR prediction dataset using a threshold of 2% of the relative fat error.

In conclusion, an increasing number of phenotypes is predicted from milk MIR spectra externally of the spectrometers, a definition of a quality procedure is needed to ensure the reliability of these phenotypes. To be on a realistic context, this study deals with DHI data collected during the routine milk recording. Milk fatty acids and protein contents will be used in this study to illustrate the interest of the cleaning statistical approaches previously detailed. However, these proposed procedures can be generalized to any other MIR phenotypes. So, three data quality checks were tested: the deletion of 1% of extreme high and low predicted values (Method 1), the deletion based on the spectral GH distance (Method 2), and the deletion of records based on the absolute residual value between the predicted and control milk fat contents (Method 3). Additionally, the combinations of these three methods were investigated.

2. Materials and Methods

2.1 Data

A total of 397,131 milk records were collected from 49,522 Holstein dairy cows belonging to 279 herds during the year 2018 to 2019 in Shijiazhuang (Hebei province, China). All milk samples were analyzed on four Bentley FTS instruments to predict the contents (g/dL of milk) of fat and protein that were then corrected using the slope and bias estimated from ring tests. The spectrometer predicted also the contents of monounsaturated fatty acid (MFA), unsaturated FA (UFA), and saturated FA (SFA) (g/dL of milk) from the generated MIR spectral data using predictive models established by the manufacturer.

All MIR spectra were standardized following the procedure detailed by Grelet et al. (2017) based on piecewise regressions. Briefly, spectral standardization consists in comparing the spectral data obtained by different milk laboratories and instruments using the same milk samples in order to estimate standardization coefficients needed to ensure a high reproducibility of spectral data between laboratories. For this study, the standardization procedure was performed two times in December 2018 and February 2019. The standardization coefficients which were obtained closest to the test date were applied to the recorded spectral data. The average interval between the spectral measurement and the standardization was 117 days. Based on the standard requirements proposed by the International Committee for Animal Recording (ICAR) (International Organization for Standardization, 2017), records out of the range of 1.5%-9.0% for fat and 1.0%-7.0% for protein were deleted. Only records having a day in milk between 5 and 365 were kept. The average lactation number ranged from 1 to 13 with a mean value of 2. Finally, the cleaned dataset contained 346,818 records collected from 49,522 Holstein cows. The average number of records per cow was 7 ± 11.

The contents of fat, protein, MFA, UFA, and SFA predicted by the spectrometer were supposed as control values and were compared to the values predicted from recorded standardized spectra using external prediction equations, for which basic information is listed in Table 3-1.

Table 3-1 Prediction performances of the external equations used to predict the contents of five milk components

			Cross-validation	
Trait (g/dL of milk)	**N**[1]	**Mean ± SD**[2]	**RMSE***	**R^2**
Fat	1,799	3.93 ± 1.00	0.0086	0.9999
Protein	4,305	3.36 ± 0.41	0.0200	0.9976
Monounsaturated fatty acids	1,793	1.08 ± 0.34	0.0581	0.9705
Saturated fatty acids	1,790	2.69 ± 0.74	0.0719	0.9904
Unsaturated fatty acids	1,788	1.24 ± 0.37	0.0648	0.9698

**RMSE=root mean square error;*
[1] N is the number of samples used in the calibration set;
[2] SD is the standard deviation.

2.2 Data cleaning techniques

A total of three different methods were tested in this study to clean the MIR-based predicted phenotypes. Their relevancy was assessed by estimating the improvement of the relationship between manufacturer's and predicted values, measured by calculating the RMSD using the following formula:

$$RMSD = \sqrt{\sum_{i=1}^{n} (EXT_{trait} - INT_{trait})^2 / n} \quad \text{(eq 3-1)}$$

where n is the total number of observations of the considered trait; EXT_{trait} is the content predicted externally using the prediction equation; and INT_{trait} is the manufacturer's prediction.

As the MIR predictions were nearly normally distributed based on their estimated skewness and kurtosis values (Table 3-2) and as the dataset was large and representative of the studied dairy population (i.e., more than 300,000 records), the first data cleaning method consisted in removing extreme MIR external predicted values based on the observed 1% and 99% quartile values (**Method 1**). The fixed thresholds used to clean the dataset for all predicted traits are given in Table 3-3.

The second data cleaning technique tested in this study was based on the calculation of GH distance between a considered sample spectrum and the average spectrum calculated from the calibration set used to build the external prediction equation (**Method 2**). This was done to observe if this spectral record was not too far away from the calibration set. This distance was calculated for all spectral records by first calculating the H distance and then by standardizing this distance using the number of variables used (Williams, 2003). As many spectral data points are highly correlated between each other, a reduction of the spectral dimensionality was needed to inverse the (co)variance matrix required to calculate the H distance. So, a principal component analysis was performed on the first-derived standardized spectra in the calibration set. A first derivation using a gap of 5 was applied to the standardized spectra in order to correct the baseline drift. Then, the obtained eigenvectors were applied to the recorded spectral data in order to resume their spectral information into new variables called principal components (PCs). The number of PCs was fixed to cover 95% of the calibration spectral variability. The GH value between a recorded sample spectrum and the calibration average spectrum was calculated using the following formula:

$$GH = ((\bar{x} - \bar{\mu})^T S^{-1}(\bar{x} - \bar{\mu})) / nPC \qquad \text{(eq 3-2)}$$

where $\bar{x}$ is the PC scores of the spectrum to be predicted; $\bar{\mu}$ is the mean of PC scores estimated from the calibration set; S corresponds to the (co)variance matrix between PC scores estimated from the calibration spectra; and nPC is the number of principal components used. Other formulas exist to calculate the GH distance by introducing factors related to the number of samples used in the calibration set (e.g., n/n-1 (Garrido-Varo et al., 2019)) but we decided to use a simplified formula as the number of samples used in the calibration set is large. Therefore, the additional factor is close to 1. As the squared differences between internally and externally predicted values were not normally distributed, Kruskal-Wallis non-parametric tests and pairwise two-sample Wilcoxon comparisons were used to assess the significance of observed squared differences from GH thresholds ranging from 1 to 10. Moreover, the correlation coefficient between GH distances and those squared differences was also estimated in order to observe the strength of their linear relationship.

The third method tested in this study was derived from the procedure proposed by Dale et al. (2019) (**Method 3**). As explained in the introduction section, milk samples with known contents of fat and protein are analyzed using MIR spectrometry by many milk laboratories during the ring tests. This allows the estimation of bias and slope needed to correct the prediction of fat and protein contents done using milk MIR spectrometry. Therefore, those corrected predictions can be assumed to be similar to the reference values measured by a reference chemical analysis. This means that the calculation of the absolute difference between this assumed control fat content and the content predicted externally using a prediction equation could be relevant for cleaning the prediction dataset. As the fat content predicted externally is not corrected by applying the slope and bias, this quantification is directly influenced by the quality of the spectral data. So, the working hypothesis of Method 3 is that the absolute fat residual difference could be used to detect poor quality spectra due to instrumental issue, analytical issue or wrong association between sample and data. Four absolute fat residual thresholds (0.08, 0.10, 0.30, and 0.40 g/dL of milk) were tested and the obtained squared differences between manufacturer's and externally predicted contents were compared using Kruskal-Wallis non-parametric tests and pairwise two-sample Wilcoxon comparisons.

Finally, using a similar statistical approach, the interest of combining these cleaning methods was also studied.

3. Results

3.1 Comparison Between Manufacturer's and Externally Predicted Phenotypes

The average values of externally predicted fat, protein, and fatty acids contents were higher than those obtained directly from the Bentley spectrometers (Table 3-2). The maximum values of externally predicted phenotypes were also higher for all traits except for SFA. However, the coefficient of variation (CV) stayed within the same range for all studied traits as well as the standard deviation (SD) values. Moreover, the relationships between the studied traits were good as high correlation coefficients were obtained between the manufacturer's and externally predicted phenotypes (i.e., $\geqslant$ 0.94; Table 3-2) even if they were inferior to the square root of the cross-validation R^2 listed in Table 3-1. However, even if the error was higher, the prediction performance was better for the traits predicted with a higher accuracy by MIR spectrometry like fat, protein and SFA.

Table 3-2 Descriptive statistics of manufacturer's and externally predicted phenotypes (g/dL of milk) and their relationship

N=346,818		**Fat**	**Protein**	**MFA**[1]	**SFA**[1]	**UFA**[1]
Manufacturer's values	Mean	3.93	3.41	0.85	2.58	0.92
	SD[2]	1.10	0.42	0.32	0.76	0.38
	CV[3]	27.99	12.32	37.65	29.46	41.30
	Minimum	1.01	1.01	0.04	0.10	0.01
	Maximum	8.99	6.99	4.61	8.00	5.19
	Skewness	0.65	0.40	1.59	0.79	1.44
	Kurtosis	1.52	2.62	6.00	2.12	5.07
Externally predicted values	Mean	3.94	3.53	1.13	2.62	1.28
	SD	1.10	0.47	0.42	0.75	0.46
	CV	27.92	13.31	37.17	28.63	35.94
	Minimum	0.72	0.52	–5.62	0.13	–6.44
	Maximum	9.92	7.10	5.26	7.32	5.66
	Skewness	0.69	0.28	1.22	0.72	1.12
	Kurtosis	1.63	1.88	5.56	1.80	5.39
Prediction relationship[4]	RMSD	0.173	0.187	0.327	0.199	0.397
	r	0.99	0.95	0.94	0.97	0.94

[1] *MFA=monounsaturated fatty acids; SFA=saturated fatty acids; UFA=unsaturated fatty acids;*

[2] *SD=standard deviation;*

[3] *CV=coefficient of variation;*

[4] *Root mean square difference (RMSD) and correlation value (r) between manufacturer's and externally predicted values.*

3.2 Deletion of Extreme Predicted Phenotypes (Method 1)

Method 1 consisted in cleaning the raw dataset by deleting records that were out of the range estimated using as thresholds the externally predicted values observed at the 1% and 99% quartiles. The impact of this deletion on the dataset shown in Table 3-3. Obviously, the loss of data in each trait was always around 2% and the SD of externally predicted traits decreased due to the deletion of extreme values. Method 1 allowed a RMSE decrease around 1% for fat and protein traits and around 2% for fatty acids traits (Table 3-3). This was expected for UFA and MFA as these traits presented a large range of variation (Table 3-2).

Table 3-3 Characteristics of the dataset obtained after a cleaning using as thresholds, the externally predicted values observed at the 1% and 99% quartiles (Method 1)

Traits	**Threshold**		**N**	**Data loss (%)**	**g/dL of milk**			**Gain$_{\text{RMSD}}$[2] (%)**
	1%	**99%**			**Mean**	**SD**	**RMSD[1]**	
Fat	1.59	7.35	339,909	1.99	3.93	0.99	0.17	1.29
Protein	2.49	4.78	339,941	1.98	3.52	0.42	0.185	0.97
MFA[3]	0.35	2.51	339,708	2.05	1.12	0.36	0.319	2.62
SFA[3]	1.06	4.92	339,852	2.00	2.61	0.68	0.195	1.95
UFA[3]	0.4	2.78	339,892	2.00	1.27	0.40	0.389	2.16

[1] *RMSD=root mean square difference between manufacturer's and externally predicted phenotypes;*
[2] *Gain=difference expressed in % between RMSD estimated before and after cleaning with Method 1;*
[3] *MFA=monounsaturated fatty acids; SFA=saturated fatty acids; UFA=unsaturated fatty acids.*

3.3 GH-based Data Cleaning (Method 2)

The estimated GH distances varied between traits (Table 3-4) due to the use of different calibration sets to build the external prediction models (Table 3-1). These calibration sets differed based on their size but also on their spectral representativeness. The eigenvectors estimated using the PC analysis were different between traits. Consequently, the PC scores used to estimate the GH distances were also different. A large majority of samples (i.e., more than 79%) had a GH lower or equal to 3. A higher proportion of samples had a GH greater than 5 for MFA and UFA. This means that the models used to predict those fatty acids were less adapted for some Chinese samples even if on average the GH distances stayed below 3 (Table 3-4). In other words, the presence of extreme high GH distances highlights the presence in the dataset of samples having a spectrum largely different than the ones used in the calibration set.

Table 3-4 Descriptive statistics of global-H (GH) distances and their correlation coefficient (r) with the squared differences between manufacturer's and externally predicted traits (e^2)

	GH distance			**GH≤3**	**3<GH≤5**	**GH>5**	**r**
Traits	**Mean**	**SD**	**Max**	**(%)**	**(%)**	**(%)**	**(GH,e^2)**
Fat	1.84	2.66	182.5	86.61	8.27	5.12	0.35
Protein	1.90	4.11	182.6	86.36	7.99	5.65	0.07
Monounsaturated FA[1]	2.37	3.23	211.3	79.73	11.42	8.85	0.48
Saturated FA	1.78	2.48	170.3	87.52	7.84	4.64	0.47
Unsaturated FA	2.37	3.23	209.7	79.70	11.46	8.85	0.47

[1] *FA=fatty acids.*

The correlation values between GH distance and the squared differences between manufacturer's and externally predicted traits were moderately positive and ranged between

0.35 and 0.48 (Table 3-4), except for protein for which a correlation coefficient of 0.07 was obtained. This low correlation could signify that the prediction of this milk component is less affected by the spectral extrapolation.

As there is no agreement about the GH threshold, we have decided to study the prediction relationship between manufacturer's and externally predicted traits after applying different GH thresholds ranging from 1 to 10. Data loss for the different studied traits ranged from around 1% when a GH threshold was set at 10, to around 72% at 1. When the GH threshold was lower than 4, the data loss for each trait was higher than 10%. As the squared differences between manufacturer's and externally predicted traits were not normally distributed, Kruskal-Wallis non-parametric tests and pairwise two-sample Wilcoxon comparisons were performed to compare those values. No differences ($P > 0.05$) of squared differences for all studied traits were observed between groups based on a GH limitation comprised between 7 and 8, and between 9 and 10 (except MFA and UFA). Globally, the average squared differences decreased significantly ($P < 0.05$) from GH limitation at 7 to GH limitation at 1 for all studied traits.

Table 3-5 Characteristics of the dataset obtained after a cleaning based on a global-H (GH) distance set at 5 (Method 2).

Traits	**N**	**Data loss (%)**	**g/dL of milk**			**Gain$_{RMSD}$[2] (%)**
			Mean	**SD**	**RMSD[1]**	
Fat	329,064	5.12	3.89	1.01	0.159	8.88
Protein	327,230	5.65	3.52	0.42	0.185	1.29
Monounsaturated fatty acids	316,128	8.85	1.1	0.37	0.303	7.79
Saturated fatty acids	330,739	4.64	2.59	0.70	0.188	5.97
Unsaturated fatty acids	316,128	8.85	1.25	0.40	0.376	5.72

[1] *RMSD=root mean square difference between manufacturer's and externally predicted values;*
[2] *Gain=difference expressed in % between RMSD estimated before and after cleaning with Method 2.*

If we applied a GH threshold equal to 5 which corresponds to an acceptable data loss (less than 10%), a significant decrease of the root mean square difference, and a value recognized in the literature as an appropriate threshold to detect outliers, the decrease of RMSE ranged from 1.29% for protein to 8.88% for fat (Table 3-5). The low gain observed for the protein content is related to the low correlation existing between GH distance and the squared differences (Table 3-4). Indeed, there is a correlation value of 0.75 between the gain and the correlation between GH distance and the squared differences. However, the number of discarded data had also an influence but its intensity was weak (correlation of 0.15 between the gain and the percentage of data loss). The highest RMSE was observed for the fat and fatty acids contents. It is also interesting to notice that the SD values for externally predicted traits were now closer to those estimated from the manufacturer's traits shown in Table 3-2. By applying a GH threshold set at 5, the RMSD gain estimated for all studied traits were higher than the one observed by applying Method 1.

3.4 Data Cleaning Based on the Absolute Fat Residual Limit (Method 3)

The strategy in Method 3 consisted in applying a limitation based on the absolute value of the difference between the fat content externally predicted and the one quantified by the Bentley spectrometer and then corrected using the slope and bias estimated after ring tests (i.e., assumed control fat content). This strategy should allow detecting spectra with poor quality. Indeed, the squared differences for fat contents were positively correlated with all studied traits (i.e., 0.13 for protein and 0.40 for the remaining studied traits). This moderate relationship could confirm the potential interest to use the absolute fat differences to detect abnormal spectra. Compared to the study of Dale et al. (2019), where data loss was equal to 30% due to the threshold (⩽2% relative error fat value), the current study proposed to use the absolute residual between predicted and assumed control fat contents to clean the dataset. Several limitations were tested and ranged from 0.08 to 0.40 g of fat per dL of milk. The data loss was equal to 55.93%, 47.31%, 8.88% and 3.47% for the thresholds set at 0.08, 0.10, 0.30 and 0.40 g/dL of milk, respectively. The significance of the squared differences between internally and externally predicted phenotypes were assessed based on Kruskal-Wallis non-parametric tests and pairwise two-sample Wilcoxon comparisons. From these results, we can conclude that the stricter the absolute fat residual limitation was, the greater was the data loss and the better was the RMSD gain, except for protein. More specifically, in most cases, the squared differences decreased ($P < 0.05$) from the group cleaned using a limit set to 0.40 g/dL of milk compared to the group with a limitation of 0.30 g/dL of milk. The loss of data using this last fat residual limitation was less than 10% (Table 3-6).

When a threshold of 0.30 g/dL of milk was applied to clean the dataset, the RMSD gain ranged from 1.49% for protein to 38.82% for fat. The higher improvement observed for fat is obvious because this trait was directly involved in the cleaning process proposed in Method 3. However, even if high correlation values exist between fat and fatty acids, the RMSD gain for fatty acids was not as high as the one observed for fat. The lowest gain was reached by the protein content. However, even if low, this improvement was better than the ones observed based on Method 1 and Method 2. This is also confirmed for the other studied traits except for MFA for which the method 2 provided a slightly better improvement.

Table 3-6 Characteristics of the dataset obtained after a cleaning based on an absolute fat residual limitation set at 0.30 g/dL of milk (Method 3)

N=316,025	**g/dL of milk**			**Gain$_{RMSD}$^{2} (%)**
	Mean	**SD**	**RMSD^1**	
Fat	3.88	1.06	0.125	38.82
Protein	3.53	0.46	0.184	1.49
Monounsaturated fatty acids	1.1	0.38	0.304	7.52
Saturated fatty acids	2.58	0.73	0.185	7.64
Unsaturated fatty acids	1.24	0.42	0.373	6.35

[1] *RMSD=root mean square difference between manufacturer's and externally predicted values;*
[2] *Gain=difference expressed in % between RMSD estimated before and after cleaning with Method 2.*

3.5 Comparison of Three Tested Data Cleaning Methods

In this study, we have tested three different techniques to clean a MIR-predicted dataset and we have concluded that all of them allow decreasing the squared differences between manufacturer's and externally predicted phenotypes. Method 3 had the highest RMSD gain except for MFA. However, these methods can be combined during a quality procedure. Highlighting the best combination of methods will depend on the aim. In this study, we propose two different aims. The first one could be to highlight the method offering the highest RMSD gain. The second aim is based on a compromise by selecting the model offering a high RMSD gain as well as a low data loss. To achieve this issue, we can calculate the ratio between the gain and the data loss expressed both in percentage. All results about method combinations are listed in Table 3-7.

The method combinations offering the highest RMSD gain were: "M2 or M3" for fat and UFA and "M1 or M2 or M3" for the remaining traits. So, the combination of method M2 and M3 seems to be crucial. So, this combination allowed detecting samples for which the prediction could be hazardous due to a spectral extrapolation as well as the use of spectral data with poor quality. By comparison with "M2 & M3" or "M1 & M2 & M3", it seems important to discard not only a record which fulfil the condition of M2 and M3 as the RMSD gain was higher.

The method combinations offering the best compromise between the number of samples discarded (i.e., N loss in Table 3-7) and the RMSD gain were: "M1 & M2" (i.e. higher gain:loss). For protein, the combination "M1 & M2 & M3" provided a slightly better result. So, the combination "M1 & M2" seems to be important. This means that to achieve the best gain:loss ratio, we need to conserve samples for which the prediction values were not extreme and the GH distance was lower than 5 to avoid spectral extrapolation.

Table 3-7 Interest of using the three studied methods and their combinations to clean the raw dataset

		M1[1]	M2[1]	M3[1]	M1 & M2	M1 & M3	M2 & M3	M1 & M2 & M3	M1 or M2	M1 or M3	M2 or M3	M1 or M2 or M3
Fat	Gain$_{RMSD}$[2] (%)	1.29	8.88	38.82	38.82	1.70	8.39	1.40	8.88	38.77	41.39	41.08
	N loss[3] (%)	1.99	5.12	8.88	0.76	0.33	2.30	0.26	6.35	10.54	11.69	12.85
	Gain:loss[4]	0.65	1.73	4.37	51.23	5.09	3.64	5.48	1.40	3.68	3.54	3.20
Protein	Gain$_{RMSD}$ (%)	0.97	1.29	1.49	1.49	0.35	0.66	0.32	1.60	2.21	2.22	2.54
	N loss (%)	1.98	5.65	8.88	1.41	0.36	1.42	0.29	6.22	10.50	13.11	13.61
	Gain:loss	0.49	0.23	0.17	1.06	0.97	0.47	1.12	0.26	0.21	0.17	0.19
MFA	Gain$_{RMSD}$ (%)	2.62	7.79	7.52	7.52	1.74	4.69	1.62	7.96	8.77	11.36	11.42
	N loss (%)	2.05	8.85	8.88	0.93	0.57	2.98	0.48	9.97	10.36	14.75	15.78
	Gain:loss	1.28	0.88	0.85	8.07	3.07	1.58	3.38	0.80	0.85	0.77	0.72
SFA	Gain$_{RMSD}$ (%)	1.95	5.97	7.64	7.64	2.27	3.34	0.65	6.63	9.29	10.99	11.60
	N loss (%)	2.01	4.64	8.88	0.58	1.02	2.13	0.15	6.07	10.63	11.38	12.71
	Gain:loss	0.97	1.29	0.86	13.22	2.22	1.57	4.20	1.09	0.87	0.97	0.91
UFA	Gain$_{RMSD}$ (%)	2.16	5.72	6.35	6.35	1.57	3.82	1.48	5.74	7.18	8.74	8.67
	N loss (%)	2.00	8.85	8.88	0.92	0.56	2.95	0.47	9.92	10.32	14.77	15.76
	Gain:loss	1.08	0.65	0.72	6.87	2.82	1.29	3.17	0.58	0.70	0.59	0.55

[1] *M1=Method 1; M2=Method 2 with a GH limitation set at 5; M3=Method 3 with an absolute residual fat difference set at 0.30 g/dL of milk;*

[2] *Gain=difference expressed in % between the root mean square difference between manufacturer's and externally predicted values estimated before and after cleaning with the considered method;*

[3] *Nloss=the percentage of samples discarded using the tested cleaning procedure;*

[4] *Gain:loss=the ratio of the GainRMSD to Nloss.*

4. Discussion

As mentioned in the introduction section, more and more phenotypes are predicted from milk MIR spectrometry to develop management and breeding tools for dairy farmers. Unfortunately, for financial and practical issues, it is not possible for milk laboratories or DHI organizations to analyze routinely a set of reference samples. Therefore, imagining to adapt the ring tests, developed to ensure the quality of the MIR predictions of fat and protein contents, is not feasible. So, it is why this study proposed to test different statistical approaches easily implementable on practice at low cost and at large scale to clean the DHI database related to MIR-based predictions. To achieve this objective, three methods were tested and combined. By predicting the MIR phenotypes on a large scale, the obtained dataset can be assumed as representative of the studied dairy cow population and, therefore, useful to identify extreme values. As the majority of MIR-predicted traits are normally distributed, the thresholds to consider a record as an outlier can be based on the 1% and 99% quartiles. This was the principle applied in Method 1 to clean the DHI dataset. Even if this method presents the advantage to be easy to implement and robust thanks to the definition of the threshold on a large database based on individual cow measurements, this method forces automatically the deletion of around 2% of extreme records without knowing the reasons and leading potentially to decrease the existing variability for the studied trait. However, knowing the origin of the obtaining of extreme values could be interesting to confirm the outlier status. So, wrong MIR predictions can be obtained due a spectral extrapolation related to a prediction process applied on samples for which the spectral variability was not taken into account in the calibration set used to build the predictive model. To solve this issue, the calculation of the GH distance between the average calibration spectrum and the spectrum of the sample to predict is useful. This is the basics of Method 2. The greater the GH distance, the more the prediction can be the result of spectral extrapolation. However, this method has the inconvenient of requiring the (co)variance matrix of the calibration set and are not able to identify predictions made on poor quality spectrum (i.e., sample for which an important slope and bias correction was applied on fat and protein predictions). So, a third method was proposed in this study to achieve this objective and was based on the absolute residual fat difference. If this difference estimated between the corrected fat prediction content and the one predicted directly from the spectral data is high, the probability that the spectrum has a poor quality is high. In conclusion, this means that Method 1 is a rough method to clean a database compared to Method 2 and Method 3 which focus more on a potential origin of prediction uncertainty.

Ideally, to compare those methods, we need a set of reference values. However, for financial and practical reasons, this was unrealistic. Moreover, when the milk analysis is available, their analysis cost limit the size of the dataset which is often no more representative of the studied dairy cow population. For instance, this is the case for the calibration set used to build MIR predictive models. Indeed, in order to increase the variability needed in the calibration set to develop a robust model, the samples are not selected randomly but their selection are controlled to ensure a good covering of the existing variability of the studied trait. Consequently, in this study, we have decided to use directly the records provided by the spectrometers: the contents of fat, protein and the major groups of milk fatty acids. Fat and protein predictions were assumed to be control values as these values were corrected based on the slope and bias estimated by the milk laboratory after ring tests. Moreover, it is known that the prediction models to quantify these traits are very accurate (Table 3-1). For fatty acids, the accuracy of the prediction models used in the spectrometers was unknown. However, it is known that this accuracy is higher for the equations predicting SFA compared to the ones quantifying MFA or UFA (Table 3-1). Fortunately, the FA predictions provided by the used Bentley spectrometers were within the expected range. Indeed, although few previous Chinese investigations have been conducted into the analysis of milk fat composition from Holstein cows based on a large scale MIR spectral database, the contents obtained in the current study were within the expected range for Chinese Holstein cows (Cao et al., 2010; Yang et al., 2013). Using the same technology to predict the fatty acid contents, Bastin et al. (2012) reported average contents of MIR-predicted MFA, SFA, and UFA equal to 1.13, 2.79, and 1.31 g/dL of milk produced by Holstein cows. Those contents were slightly higher than the values obtained in the current study. On other hand, the contents of MFA, SFA, and UFA obtained in the current study were lower than those reported by Soyeurt et al. (2006) (i.e., 1.44, 2.95, and 1.65 g/dL of milk) and Fleming et al. (2017) (1.00, 2.97, 1.16 g/dL of milk from Holstein, Brown Swiss, and Jersey cows).

The fat, protein, SFA, MFA, and UFA were predicted both by the Bentley spectrometers and using external prediction equations (Table 3-1). Even if the prediction accuracy of Bentley spectrometer to predict the contents of those traits were unknown, good relationships between those predictions were observed with correlation coefficients higher or equal to 0.94 (Table 3-2). The higher squared differences between those predictions compared the cross-validation RMSE, a traditional parameter to measure the accuracy of a prediction model (De Vries and Feleke, 2008), can be attributed to several reasons. First, the calibration dataset used to build the external prediction models did not contain samples collected on the Chinese studied dairy cow population. Therefore, the spectral variability of some samples was not taken into account leading an uncertainty in the prediction due to the spectral extrapolation. Second, the spectra were standardized to be expressed on Foss basis in order to apply the available external prediction equations. This could bring an additional uncertainty. Third, even if the contents of fat and protein were assumed to be a control value, all records were not analyzed using reference chemical analysis. The bias and slope correction cannot correct perfectly all predictions. However, the aim of this study was not to validate the prediction equations used in the spectrometer or applied externally. Therefore, the error shift was not so problematic as this study focused on the changes of RMSD value following the cleaning procedures. Even if the RMSD values listed in Table 3-2 were higher than the RMSE values mentioned in Table 3-1, the high correlations between the predicted traits as well as their similar CV (Table 3-2) are sufficient to confirm the interest of using those data to conduct the proposed research and take the profit of their representativeness due to the large individual data acquisition. The higher SD observed for the traits predicted externally compared to the ones provided by the spectrometer were due to the presence of outliers in the dataset. Indeed, Naes et al. (2002) noted that if the new sample to be predicted was beyond the capacity of calibration models, the accuracy was impaired accordingly.

For all traits, the cleaning methods decreased RMSD and allowed having SD values for the externally predicted phenotypes closer to the ones estimated from manufacturer's predictions (Tables 3-3, 3-4, and 3-6). These SD decreases confirmed that a part of the variability observed in the raw data was related to several outlier records. However, all proposed cleaning methods did not discard the same records. For instance, less than 1% of records were considered as outliers for the 3 methods used. More or less 1% of records had M1 and M2 in common or M1 and M3. Close to 3% of records had M2 and M3 in common (Table 3-7). Method 1 is a rough cleaning procedure deleting the 2% of extreme predictions. Therefore, this method can keep outliers in the dataset if more than 2% of records are outliers. Following the considered traits and from the results listed in Table 3-7, Method 2 had between 25% and 46% of its discarded samples in common with Method 3. This was less than 25% observed for Method 1.

This moderate relationship between Method 2 and Method 3 can be easily explained. Method 2 discards the samples out of the spectral range present in the calibration set. A GH limit of 3 was proposed by Mark and Tunnell (1985), as three standard deviations from the center of a group was considered to be the boundary. In the case of a multivariate model, the threshold of 3 may be unjustifiable for the database set and the number of wavelengths used (Whitfield et al., 1987). A threshold around 4 was proposed by Ritchie et al. (2003) based on the distribution of samples. In our study, this GH threshold was set at 5 to limit the data losses (i.e., lower than 6% except for MFA and UFA; Table 3-3) and to decrease significantly RMSD gain. Method 3 detects spectra data with poor quality. Therefore, samples with poor spectral data must have a high GH distance. So, the samples discarded by Method 2 and Method 3 had certainly a very poor spectral quality. The existence of these abnormal or extreme spectra can be related to different factors which can be associated, for instance, to instrument failure, incorrect milk sampling, an analytical issue or poor milk conservation between the sampling and analysis period. The remaining samples discarded by Method 3 were related to samples for which the spectral data was only far from the one present in the calibration but the spectral data did not seem to have a poor quality. For instance, these abnormal or extreme spectra could be related to specific cows showing a unique spectral pattern and therefore special milk composition. This could have a potential negative impact on a breeding program as those cows could potentially be interesting. However, this was not observed in this study. The average loss of records for a specific cow was equal to 0.91 (i.e., less than 1 record per cow). The most probable reason could be related to a different farm management and context. Indeed, it is known that the milk composition can be influenced by many factors related to the animal or to the farming system. For this last issue, the feeding system has a major impact on the milk fat composition. The external prediction models used in this study were not developed using Chinese samples. Therefore, the milk spectral data could be not fully considered. However, a large part of the spectra used in this study had a GH lower than 3 (Table 3-3), this means that those spectra were within the spectral range of the calibration set even if these studied records were not included in the prediction models. This also indirectly shows that globally the milk composition of studied samples was similar to the samples included in the calibration set which were collected in Europe. However, even if only 5 to 9% of samples were discarded based on a GH threshold set at 5 (Table 3-3), it remained samples in the dataset with an uncovered spectral variability. It was especially true for the models predicted the contents of MFA and UFA which are known to be influenced by the feeding but also the heat stress and the body condition score (Bastin and Gengler, 2013; Hammami et al., 2015; Liu et al., 2016). This finding suggests therefore the interest to collect specifically this kind of samples to improve the robustness of the fatty acids prediction models. However, the finding cannot mention if the accuracy could be improved as the gas chromatographic fatty acid profiles were not available for these samples having a GH distance out of the desired range. The percentage of samples discarded using a GH threshold were similar to previous results. For instance, based on a drugs samples, Morozova et al. (2013) found 77.3% of spectra with a H distance (also GH distance as GH is obtained by dividing the squared H distance by the number of variables) less than 3, 18.2% of spectra with a H distance between 3 and 4, and 4.5% of spectra greater than 4. Based on a GH threshold equal to 3, Soyeurt et al. (2011) deleted 3.9% of the raw data. This is lower than the data loss observed in this study using the same threshold, which ranged from 12.48% to 20.27%. This could be related to the fact that the calibration sets used to build the prediction models did not consider the records included in the current study, leading to a potentially higher number of records considered as outliers but related to a spectral milk variability which was not considered in the prediction models. However, as mentioned previously, the main part of the data had a GH lower than 3.

Concerning the samples discarded by Method 2, only between 25% and 45% of those samples were in common with Method 3 (results derived from Table 3-7). This means that the majority of the samples discarded by Method 3 were kept by Method 2. As the RMSD gain was the highest for nearly all traits (Table 3-7), we can assume that this cleaning even if stricter provided beneficial effect. These samples presented spectral data considered as poor quality but their spectral profile was not extreme as the GH distance was lower than 5. The average GH distance for those samples was equal to 2.20 with minimum and maximum values of 0 and 5.00, respectively. Consequently, as Method 2 and Method 3 reflect different outlier origins, this was logical to observe an additional improvement of RMSD gain by combining these methods with the “or” logical disjunction (Table 3-7).

For protein, SFA and MFA, a slight improvement was obtained by adding the Method 1 using an “or” logical disjunction to Method 2 and Method 3 (Table 3-7). This means that Method 1 and Method 3 were not able to detect fully all outliers. However, the improvement was low compared to “M2 or M3” combination leading to highlight a poor contribution of Method 1. However, the cleaning procedures having the highest gain:loss ratios were method combinations involving the Methods 1 and 2 using an “and” logical disjunction (Table 3-7). Only for the protein content, adding Method 3 in this combination improved slightly the gain:loss ratio. Method 1 and Method 2 dealt both with extreme values: the first one based on the predictions and the second one based on spectral data. The selection of a cleaning procedure based on this ratio represents an optimisation between the data loss and the RMSD gain. The “M1 & M2” combination allows detecting samples with an extreme predictions and a GH distance higher than 5. This means that samples out of the spectral range covered by the set used to build the prediction models but having a reasonable prediction were kept in the dataset. Indeed, the samples kept in Method 1 but not in Method 2 had an average GH ranging from 7.02 to 9.08 with a minimum value of 5.10 and a maximum value of 113.00 for all studied traits. In other words, this implies that we accept the spectral extrapolation during the prediction process if the prediction is between a realistic range. This working hypothesis is acceptable as the correlation values between GH distance and the squared residuals were moderate for fat and fatty acids and low for the protein (Table 3-3). The positive correlation found in this study is in agreement with the findings of Whitfield et al. (1987) who reported that the H distance was directly proportional to the absolute residuals between predicted and reference values. The GH-based cleaning deleted less than 1% of raw records which is largely lower than the “M2 or M3” combination discarding from 11 to 15% (Table 3-7).

Depending the aim behind the deletion of potential outlier samples (lower RMSD gain or higher gain:loss ratio), the selected cleaning method is not the same. However, Method 2 is the common denominator. This reveals the high interest of measuring the GH distance for all samples to predict. Practically, this means that simultaneously providing the GH distance of the spectrum to be predicted with the prediction could be relevant to indirectly inform customers about the potential accuracy of the given phenotype. However, the GH calculation requires the mean PC spectrum, the eigenvectors, and the (co)variance matrix between calibration PC scores. This information could be provided by the manufacturer or the creator of the equation. Currently, they are reluctant to provide this information because they do not want to communicate the calibration set. However, the milk laboratories and DHI organizations did not need the dataset but need only the average PC spectrum, the eigenvectors and the covariance matrix which should not be confidential. Some international institutions like the International Committee for Animal recording (ICAR) or the International Dairy Federation (IDF) could make a pressure on the model’s providers to obtain this information. However, if it is impossible to find, a cleaning based on the deletion of 1% of extreme high and low predicted values estimated from a large spectral database is still of interest combined with Method 3 (Table 3-7).

The protein content was the trait which was the less affected by the data cleaning. Indeed, the RMSD gain ranged from 0.32 to 2.54 (Table 3-7). This low improvement could be explained by the fact that this trait was less impacted by the spectral extrapolation. Of course, the correlation value between the GH distance and the squared differences between manufacturer’s and externally predicted phenotypes were low (Table 3-3). Moreover, this trait presents also a low variability compared to the other studied traits (i.e., CV of 12% vs. 30-40% for the other traits; Table 3-2). It could be supposed that Method 3 as based on fat content which is correlated with fatty acids could be more profitable for these traits compared to the protein. But Method 3 considered lonely offered the best improvement also for the protein content highlighting the interest of this cleaning process also for trait which are less correlated with the fat content.

The choice of the fat content to develop the cleaning criteria of Method 3 was related to the fact that this trait is always adjusted for the slope and bias by the milk laboratory conducting routine ring tests and presents a natural large variability (Table 3-2). So, in the DHI database, this trait can be assumed to be closer to the real reference value. By comparing this assumed control value to the one predicted using the external calibration model, it was possible to detect spectral abnormal values which can lead to erroneous MIR predictions. So, from the results obtained in this study, a fat limitation of 0.30 g/dL of milk seems to be preferable in order to increase the chance to gain accurate predictions without losing much data. However, the limitation must be adapted to the spectral database used as this will depend on the spectral data and the equation used.

5. Conclusions

This study confirms that spectral outliers are generated within the framework of DHI, as the current corrections for the slope and bias used in the milk laboratories to ensure high prediction quality for fat and protein contents had no effect on the spectral data. Moreover, some spectral outliers being different to the spectra constituting the calibration dataset were also identified. So, there is a need to define a quality procedure for MIR phenotypes, especially as the amount of traits predicted by this technology is growing rapidly. Three methods were tested and combined in this study. To be more parsimonious in the data loss, the results recommend to prefer the “M1 and M2” combination: deletion of the 2% extreme predictions + samples with a GH higher than 5. However, to ensure the lowest squared differences between manufacturer’s and externally predicted phenotypes, the “M2 or M3” combination, where M3 consisted of deleting samples with an absolute fat residual higher than 0.30 g/dL of milk, must be applied on the DHI dataset. Both of those combinations involved Method 2 confirming the high interest of calculating the GH distance for all samples to predict. However, if it is impossible to estimate the GH distance due to a lack of relevant information to compute this statistical parameter, we recommend the use of Method 1 combined with Method 3 (i.e., M1 or M3). The limitation used in Method 3, which was set to 0.3 g/dL of milk, must be adapted to the spectral database used as this will depend on the spectral data and the equation used. The methodology proposed in this study can be applied to find the best threshold for the considered database and could be implemented on any MIR-based phenotypes.

6. Acknowledgments

The China Scholarship Council (CSC) is gratefully acknowledged for funding Mr. Zhang’s PhD grant. The authors wish to thank the staff of the DHI lab station in China for arranging the analysis of milk samples. The authors also would like to thank the European Milk Recording team for their help in milk spectral standardization and the University of Liège - Gembloux Agro-Bio Tech for providing computing resources.

References

Bastin, C., Berry, D.P., Soyeurt, H., Gengler, N., 2012. Genetic correlations of days open with production traits and contents in milk of major fatty acids predicted by mid-infrared spectrometry. J. Dairy Sci. 95, 6113–6121.

Bastin, C., Gengler, N., 2013. Genetics of body condition score as an indicator of dairy cattle fertility. A review. Biotechnol. Agron. Société Environ. 17, 64–75.

Boichard, D., Brochard, M., 2012. New phenotypes for new breeding goals in dairy cattle. Animal 6, 544–550.

Cao, Z., Huang, W., Wang, T., Wang, Y., Wen, W., Ma, M., Li, S., 2010. Effects of parity, days in milk, milk production and milk components on milk urea nitrogen in Chinese Holstein. J. Anim. Vet. Adv. 9, 688–695.

Dale, L.M., Werner, A., Spiekers, H., Hertel-Böhnke, P., Stamer, E., Gollé-Leidreiter, F., Au, M., Onken, F., 2019. Prediction of evaluated energy balance (NEL and ME) in dairy cows by milk mid-infrared (MIR) spectra, in: ICAR Technical Series. pp. 137–141.

De Maesschalck, R., Jouan-Rimbaud, D., Massart, D.L., 2000. The Mahalanobis distance. Chemom. Intell. Lab. Syst. 50, 1–18.

De Marchi, M., Fagan, C.C., O’Donnell, C.P., Cecchinato, A., Dal Zotto, R., Cassandro, M., Penasa, M., Bittante, G., 2009. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. J. Dairy Sci. 92, 423–432.

De Marchi, M., Penasa, M., Cecchinato, A., Mele, M., Secchiari, P., Bittante, G., 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. animal 5, 1653–1658.

De Vries, A., Feleke, S., 2008. Prediction of future uniform milk prices in Florida Federal Milk Marketing Order 6 from milk futures markets. J. Dairy Sci. 91, 4871–4880.

Delhez, P., Ho, P.N., Gengler, N., Soyeurt, H., Pryce, J.E., 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? J. Dairy Sci. 103, 3264–3274.

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., Tan, P.-N., 2002. Data mining for network intrusion detection, in: Proceedings of the Proc. NSF Workshop on Next Generation Data Mining. pp. 21–30.

Fleming, A., Schenkel, F.S., Chen, J., Malchiodi, F., Bonfatti, V., Ali, R.A., Mallard, B., Corredig, M., Miglior, F., 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. J. Dairy Sci. 100, 5073–5081.

Gargiulo, J.I., Eastwood, C.R., Garcia, S.C., Lyons, N.A., 2018. Dairy farmers with larger herd sizes adopt more precision dairy technologies. J. Dairy Sci. 101, 5466–5473.

Garrido-Varo, A., Garcia-Olmo, J., Fearn, T., 2019. A note on Mahalanobis and related distance measures in WinISI and The Unscrambler. J. Near Infrared Spectrosc. 27, 253–258.

Gengler, N., Soyeurt, H., Dehareng, F., Bastin, C., Colinet, F., Hammami, H., Vanrobays, M.-L., Lainé, A., Vanderick, S., Grelet, C., Vanlierde, A., Froidmont, E., Dardenne, P., 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99, 4071–4079.

Grelet, C., Dardenne, P., Soyeurt, H., Fernandez, J.A., Vanlierde, A., Stevens, F., Gengler, N., Dehareng, F., 2020a. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods Under pres.

Grelet, C., Froidmont, E., Foldager, L., Salavati, M., Hostens, M., Ferris, C.P., Ingvartsen, K.L., Crowe, M.A., Sorensen, M.T., Fernandez Pierna, J.A., Vanlierde, A., Gengler, N., Dehareng, F., 2020b. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. J. Dairy Sci. 103, 4435–4445.

Grelet, C., Pierna, J.A.F., Dardenne, P., Soyeurt, H., Vanlierde, A., Colinet, F., Bastin, C., Gengler, N., Baeten, V., Dehareng, F., 2017. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J. Dairy Sci. 100, 7910–7921.

Hammami, H., Vandenplas, J., Vanrobays, M.-L., Rekik, B., Bastin, C., Gengler, N., 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. J. Dairy Sci. 98, 4956–4968.

Hansen, P.W., Holroyd, S.E., 2019. Development and application of Fourier transform infrared spectroscopy for detection of milk adulteration in practice. Int. J. Dairy Technol. 72, 1471–0307.12592.

Ho, P.N., Bonfatti, V., Luke, T.D.W., Pryce, J.E., 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102, 10460–10470.

International Organization for Standardization, 2017. ICAR Section 2 - Guidelines for Dairy Cattle Milk Recording, in: Proceedings of the Guidelines. p. 27.

Lamb, R.C., 1978. Improving Large Dairy Herd Management Practices: Review of Regional Project NC-119. J. Dairy Sci. 61, 1284–1291.

Liu, S., Zhang, R., Kang, R., Meng, J., Ao, C., 2016. Milk fatty acids profiles and milk production from dairy cows fed different forage quality diets. Anim. Nutr. 2, 329–333.

Mahalanobis, P.C., 1936. On the generalized distance in statistics. National Institute of Science of India.

Mark, H., 1986. Normalized distances for qualitative near-infrared reflectance analysis. Anal. Chem. 58, 379–384.

Mark, H.L., Tunnell, D., 1985. Qualitative near-infrared reflectance analysis using Mahalanobis distances. Anal. Chem. 57, 1449–1456.

McParland, S., Banos, G., Wall, E., Coffey, M.P., Soyeurt, H., Veerkamp, R.F., Berry, D.P., 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94, 3651–3661.

Melfsen, A., Hartung, E., Haeussermann, A., 2013. Robustness of near-infrared calibration models for the prediction of milk constituents during the milking process. J. Dairy Res. 80, 103–112.

Morozova, M., Elizarova, T., Pleteneva, T., 2013. Discriminant analysis and Mahalanobis distance (NIR diffuse reflectance spectra) in the assessment of drug's batch-to-batch dispersion and quality threshold establishment. Eur. Sci. J. 9, 8–25.

Naes, T., Isaksson, T., Fearn, T., Davies, T., 2002. A user friendly guide to multivariate calibration and classification. NIR publications.

Ritchie, G.E., Mark, H., Ciurczak, E.W., 2003. Evaluation of the Conformity Index and the Mahalanobis distance as a tool for process analysis: A technical note. AAPS PharmSciTech 4, 109–118.

Shenk, J.S., Westerhaus, M.O., 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci. 31, 469–474.

Soyeurt, H., Bruwier, D., Romnee, J.-M., Gengler, N., Bertozzi, C., Veselko, D., Dardenne, P., 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92, 2444–2454.

Soyeurt, H., Dardenne, P., Dehareng, F., Lognay, G., Veselko, D., Marlier, M., Bertozzi, C., Mayeres, P., Gengler, N., 2006. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. J. Dairy Sci. 89, 3690–3695.

Soyeurt, H., Dehareng, F., Gengler, N., McParland, S., Wall, E., Berry, D.P., Coffey, M., Dardenne, P., 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94, 1657–1667.

Thomas, E. V., Ge, N., 2000. Development of robust multivariate calibration models. Technometrics 42, 168–177.

Vanlierde, A., Soyeurt, H., Gengler, N., Colinet, F.G., Froidmont, E., Kreuzer, M., Grandl, F., Bell, M., Lund, P., Olijhoek, D.W., Eugène, M., Martin, C., Kuhla, B., Dehareng, F., 2018. Short communication: Development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J. Dairy Sci. 101, 7618–7624.

Whitfield, R.G., Gerger, M.E., Sharp, R.L., 1987. Near-infrared spectrum qualification via Mahalanobis distance determination. Appl. Spectrosc. 41, 1204–1213.

Williams, P., 2003. Near-infrared technology: getting the best out of light. PDK Grain, Nanaimo, Canada.

Yang, L., Yang, Q., Yi, M., Pang, Z.H., Xiong, B.H., 2013. Effects of seasonal change and parity on raw milk composition and related indices in Chinese Holstein cows in northern China. J. Dairy Sci. 96, 6863–6869.

Zeaiter, M., Roger, J.-M., Bellon-Maurel, V., Rutledge, D.N., 2004. Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. TrAC Trends Anal. Chem. 23, 157–170.

4

Chapter IV : Can we observe expected behaviours at large and individual scales for feed efficiency-related traits predicted partly from milk mid-infrared spectra?

Adapted from: Lei Zhang, Nicolas Gengler, Frédéric Dehareng, Frédéric Colinet, Eric Froidmont, Hélène Soyeurt. Can We Observe Expected Behaviors at Large and Individual Scales for Feed Efficiency-Related Traits Predicted Partly from Milk Mid-Infrared Spectra? Animals, 10(5): 873. https://doi.org/10.3390/ani10050873

Simple summary

Feed efficiency is a key point to ensure the sustainability of dairy farming. Knowing feed efficiency at large scale would be beneficial for the cow breeding and the management on farm, but it is unrealistic in the current context due to practical and financial issues. In this Chapter, we have used the Walloon milk recording database to predict, using namely a BW predictive model, traits related to feed efficiency at large and individual scales. This aimed to compare their behaviors with the expected ones published in the literature.

Abstract

Phenotypes related to feed efficiency were predicted from records easily acquired by breeding organizations. A total of 461,036 and 354,148 records were collected from first and second parity Holstein cows. Equations were applied to the milk mid-infrared spectra to predict the main milk components and coupled with animal characteristics to predict the body weight (pBW). Dry matter intake (pDMI) was predicted from pBW using the NRC equation. The consumption index (pIC) was estimated from pDMI and fat and protein corrected milk. All traits were modelled using single trait test-day models. Descriptive statistics were within the expected range. Milk yield, pDMI, and pBW were phenotypically positively related (r ranged from 0.08 to 0.64). As expected, pIC was phenotypically negatively correlated with milk yield (-0.77 and -0.80 for first and second lactation) and slightly positively correlated with pBW (0.16 and 0.07 for first and second lactation). Later parity cows seemed to have a better feed efficiency as they had a lower pIC. Although the prediction accuracy was moderate, the observed behaviours of studied traits by year, stage of lactation and parity were in agreement with literature. Moreover, as a genetic component was highlighted (heritability around 0.18), it would be interesting to realize a genetic evaluation of these traits and compare the obtained breeding values with the ones estimated for sires having daughters with reference feed efficiency records.

Keywords: Mid-infrared spectra; milk; feed efficiency; phenotypes

1. Introduction

The theory that life is a chemical process corresponding to metabolism and respiration was first proposed in 1783 (Underwood E., et al., 1944). The body has the ability to use nutrients to synthesize tissues (Carpenter et al., 1997). Appropriate nutrient level and management allows the genetic potential of a specific animal to be maximized and the dairy farm sustainability to be improved. Sustainability interrelates different aspects of milk production, from economics (i.e., maximization of milk production per feed quantity), well-being (i.e., knowing the feed efficiency of a cow allows farmers to better feed it and ensures a healthy status), to the environment (such as a lower amount of feed for the same level of milk production leads to less methane eructed by dairy cows) (Olesen et al., 1999). The improvement of feed efficiency is an opportunity to solve issues facing the dairy industry resulting from production and environmental aspects (Connor, 2015). Evaluating several traits, like body weight (BW), dry matter intake (DMI), milk production and composition, and consumption index (IC) reflects the feed efficiency of individuals in a certain way (Yusuf et al., 2012) and is of interest. This is why most are considered to be important selection traits in animal breeding (Miglior et al., 2017).

Feed efficiency can be assessed, for instance, by using the IC defined as the ratio of DMI to fat and protein corrected milk (FPCM) (Phocas et al., 2014). Although the IC denominator is easily estimated at large and individual scales thanks to routine milk recording, the numerator is harder to record. There have been many attempts to collect this trait, using for instance an automated feed monitor, but this recording is limited to a small population of cattle and suffers from inherent difficulty when monitoring pasture-based herds (Olesen et al., 1999). To solve this problem, the literature proposes an indirect methodology to predict DMI based on more easily recorded traits such as fat corrected milk (FCM), BW, and week of lactation (Council, 2001). FCM and lactation stage are currently documented by all milk recording organizations, but BW is not so frequently recorded. BW can be measured using a weighing scale, predicted using modern technology such as cameras (Song et al., 2018), or based on the cow’s morphological traits (Enevoldsen and Kristensen, 1997; Heinrichs et al., 1992; Tasdemir et al., 2011). For this last approach, the classification of a cow is unfortunately often realized only once in its life; and due their cost the first and second techniques are mainly used in research herds and in some high level commercial herds (Vanrobays et al., 2015). Recently, Soyeurt et al., (2019) developed an equation to predict an indicator of BW from the milk mid-infrared spectrum (MIR), lactation stage, parity, month of test, and milk yield. The obtained external validation root mean squared error (RMSEPv) ranged from 39 to 64 kg. In conclusion, based on the existing common dairy production traits and using the prediction equations existing for BW and DMI, it is possible to predict several traits related directly or indirectly to feed efficiency at a large scale. This research studies the behaviour of the predicted traits with their corresponding reference values published in the literature for several sources of variation such as herd, test day, stage of lactation, and parity. Heritability is also estimated as this genetic parameter reflects the part of the variability transmitted from generations to generations. If the predicted traits available at a large scale depict the expected behaviour, those traits could be used in the future to optimize feed efficiency at herd and breeding levels.

2. Materials and Methods

This work was carried out in accordance with the EU Directive 2010/63/EU for animal experiments. Milk samples collected for this study were part of the regular milk sampling work undertaken in the Walloon region of Belgium by the Walloon Breeding Association (AWÉ, Ciney, Belgium). Samples were collected from January 2007 until December 2017 from 47,176 first parity Holstein cows in 1,204 herds (i.e., 460,765 records, after cleaning) and from 36,208 second parity Holstein cows in 1,184 herds (i.e., 353,936 records, after cleaning). Fat (%FAT) and protein (%PROT) contents (g/dL of milk) were determined using FOSS Milkoscan FT6000 spectrometers (Hillerød, Denmark) at the milk laboratory "Comité du Lait" (Battice, Belgium). The generated MIR spectra were also recorded. BW records were predicted (pBW; kg) from those spectra and animal characteristics using the equation developed by Soyeurt et al. (Soyeurt et al., 2019). Fat and protein corrected milk (FPCM; kg/d) was calculated using the following formula (FAO, 2010):

$$\text{FPCM} = \text{Milk yield} * (0.337 + 0.116 * \%\text{FAT} + 0.06 * \%\text{PROT}) \qquad \text{(eq 4-1)}$$

Then, DMI records at specific test days were predicted (pDMI; kg/day) using pBW, FCM, and the week of lactation using the NRC equation. This equation has a root mean square error of 1.82 kg/day (Council, 2001) and can be summarized as follows:

$$\text{pDMI} = (0.372 * \text{FCM} + 0.0968 * \text{pBW}^{0.75}) * (1 - \exp^{(-0.192 * (\text{wol} + 3.67))}) \qquad \text{(eq 4-2)}$$

where wol is week of lactation. Fat-corrected milk (FCM; kg/d) was calculated according to the formula (Council, 2001; Gaines and Davidson, 1923):

$$4\%\ \text{FCM} = \text{Milk yield} * 0.4 + 15 * \text{FAT} \qquad \text{(eq 4-3)}$$

Finally, the consumption index (pIC) was predicted as the ratio of pDMI to FPCM. All records fulfilled the following requirements: records from 5 to 365 days in milk (DIM), pBW between 400 kg and 900 kg (Soyeurt et al., 2019), milk yield less than 71 kg/d, fat content within the range of 1.5% to 9.0% and protein content ranged from 1.0% to 7.0% (International Organization for Standardization, 2017). Significance of differences for original predicted traits between parities was assessed using the TTEST procedure in SAS (Version 9.4, SAS institute, Inc, Cary, NC, USA), as the records were normally distributed and the SD differences for all traits between two lactations did not exceed 4 times. Correlation values per lactation were also estimated between studied traits to assess their relationships and for comparisons with the literature. The pedigree related to the first and second parity datasets contained 191,400 animal (180,136 dams and 11,264 sires) and 165,032 animal (154,585 dams and 10,447 sires) born after 1990.

The studied traits were modelled per lactation using 6 univariate test-day models summarized as follows:

$$y = X\beta + \text{CDIM} + \text{QG} + \text{Zpe} + e \qquad \text{(eq 4-4)}$$

where y was the vector of observations for each trait (i.e., milk yield, %FAT, %PROT, pBW, pDMI, and pIC); β was the vector of herd*test-day (HTD; i.e., 51,334 levels for parity 1 and 50,161 levels for parity 2); DIM was the vector related to the stage of lactation (i.e., 25 classes of 15 DIM); G was the vector of genetic effect (i.e., 191,401 levels for parity 1 and

165,033 levels for parity 2); pe was the vector of permanent environment effect (i.e., 47,175 levels for parity 1 and 36,207 levels for parity 2); e was the vector of residuals. X, C, Q, and Z were the corresponding incidence matrices assigning observations to effects.

The variance components were estimated by the restricted maximum likelihood method (REML) using the REMLF90 program developed by Misztal (Misztal, 1994). Heritability for each studied trait was calculated as the ratio of the genetic variance to the total variance (i.e., the sum of genetic, permanent environmental, and residual variances). Solutions for each effect were estimated using the best linear unbiased prediction (BLUP) method using the BLUPF90 program developed by (Misztal, 2011). In order to have a cleaned trend, the DIM tendency was assessed using the least squared means (LSMEANS) calculated from the solutions of DIM effect obtained from the used test-day models. Similarly, the test month effect was studied by averaging per month LMEANS estimated from the HTD solutions.

3. Results

3.1 Parity effect

Table 4-1 Descriptive statistics of the studied datasets.

	Mean ± SD^{1}			
Traits	**Lactation 1**	**Lactation 2**	**Minimum**	**Maximum**
Milk yield (g/day)	23.54 ± 6.00^{B}	26.63 ± 8.19^{A}	3.10	70.60
Fat content (g/dL of milk)	3.98 ± 0.69^{B}	4.07 ± 0.73^{A}	1.50	9.00
Protein content (g/dL of milk)	3.38 ± 0.36^{B}	3.46 ± 0.39^{A}	1.19	7.00
Predicted body weight (kg)	597 ± 35^{B}	639 ± 31^{A}	421	820
Predicted dry matter intake (kg/day)	19.45 ± 2.71^{B}	21.16 ± 2.92^{A}	7.50	39.21
Predicted consumption index(kg of DM/kg milk)	0.87 ± 0.18^{A}	0.84 ± 0.22^{B}	0.33	4.98

^{1}*A,B Means within a row with different superscripts differ significantly from each other (P < 0.001).*

The pBW of the second lactation was significantly higher than the first lactation, and the difference was 42 kg (Table 4-1). A significant increase occurred for pDMI between first and second lactations (19.45 to 21.16 kg/day, Table 4-1). The global mean of pDMI for first and second lactation was 20.30 kg/day. A positive difference of 3.09 kg/day was observed for milk yield between first and second lactation. The contents of fat and protein in first lactation were 3.98 and 3.38 g/dL of milk, which were lower than the ones observed in the second lactation (4.07 and 3.46 g/dL of milk). The pIC decreased significantly from the first lactation (0.87 kg of DM/kg milk) to the second lactation (0.84 kg of DM/kg milk).

3.2 Effect of Lactation Stage

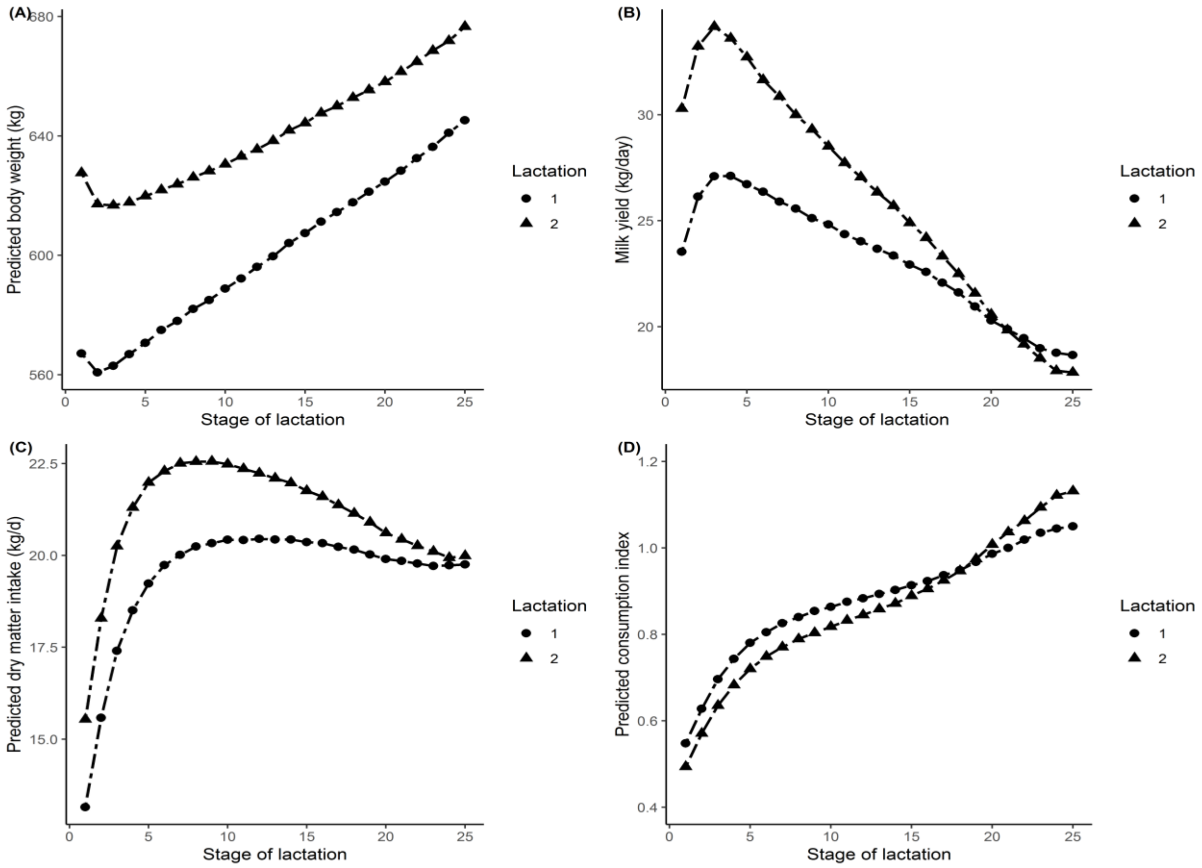


Figure 4-1. Evolution of predicted body weight (A), milk yield (B), predicted dry matter intake (C), and predicted consumption index (D) throughout lactation.

The evolution of initial predicted traits and milk yield are shown in Figure 4-1. The interest of this paragraph is to ensure that the predicted traits used in this study followed the expected trends. The pBW dropped to a lower point after calving and then increased slightly. The pDMI got a top peak around two months after calving and then decreased slowly. The peak of pDMI appeared later than the appearance of milk peak. In brief, milk yield increased sharply compared to pDMI at the beginning of lactation, involving a negative energy balance and a loss of weight (i.e., pBW in this study). A smooth slope appears in the pIC curve, which means that the feed efficiency turned worse gently at that time.

3.3 Correlations Between Studied Traits

Table 4-2 Correlation coefficients observed between studied traits in first (below the diagonal) and second lactation (above the diagonal).

Trait*	**Milk**	**%FAT**	**%PROT**	**pBW**	**pDMI**	**pIC**
Milk yield (kg/day)		–0.37	–0.50	0.08	0.64	–0.80
Fat content (g/dL of milk; %FAT)	–0.35		0.54	0.13	–0.06	0.11
Protein content (g/dL of milk; %PROT)	–0.35	0.49		0.51	–0.12	0.37
Predicted body weight (kg; pBW)	0.09	0.16	0.61		0.47	0.08
Predicted dry matter intake (kg/day; pDMI)	0.59	0.01	0.15	0.62		–0.38
Predicted consumption index (kg of DM/kg milk; pIC)	–0.77	0.08	0.31	0.16	–0.17	

* The correlations for all traits were significant ($P < 0.001$).

As shown in Table 4-2, the first and second parity Holstein cows showed different correlation values estimated between all studied traits. The pBW was slightly positively correlated with milk yield for both lactations (r=0.09 or 0.08). The pDMI was also positively correlated with milk yield and pBW. There were positive relationships between pIC and fat or protein contents in milk (Table 4-2). A similar correlation value was observed between milk yield and pIC (Table 4-2). In this study, we found a negative correlation between pIC and pDMI (r=–0.17 and –0.38) but this relationship was weaker compared to the one observed with milk yield (r=–0.77 and –0.80). Slightly positive relationships were found between pBW and pIC (0.16 and 0.08 for parity 1 and 2).

3.4 Test month influence

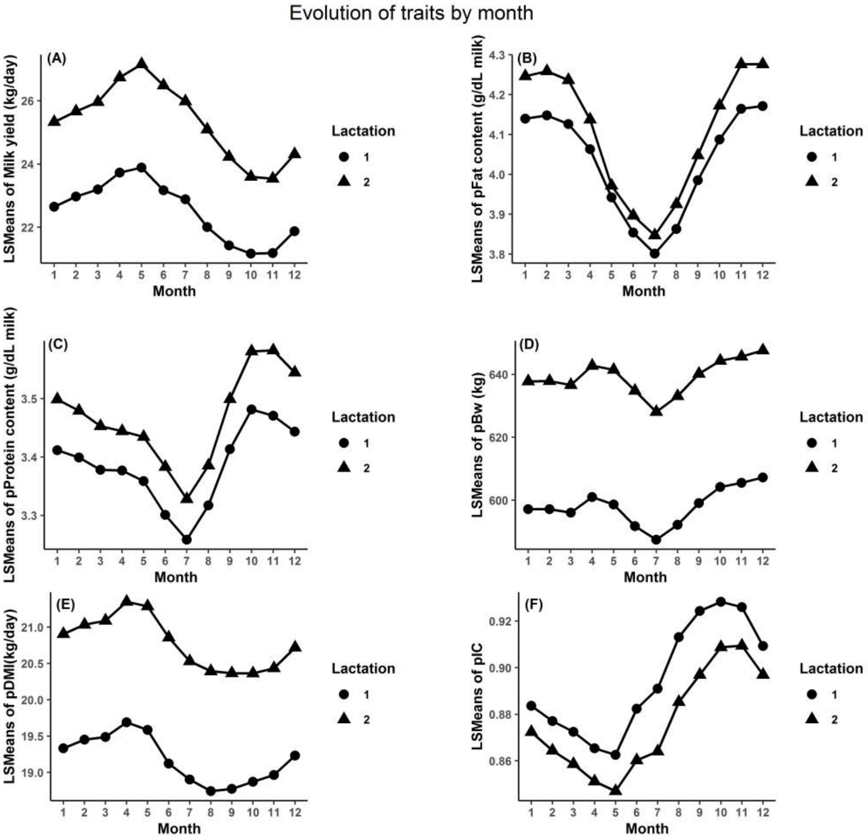


Figure 4-2. The evolution of the averaged herd least squares means (LSMEANS) for all studied traits across test months.

The averaged herd LSMEANS varied throughout the year for all studied traits and for both lactations studied (Figure 4-2). As observed previously from raw data, the second lactation showed a higher milk yield, fat and protein content, pBW, and pDMI compared to the first lactation. There are 3 kinds of seasonal trends for the studied traits: 1) milk yield and pDMI; 2) %PROT, %FAT, and pBW; 3) pIC. There was a slight increase of milk yield and pDMI until May, then a smooth decrease, and another increase in August and November for pDMI and milk, respectively. The %PROT, %FAT, and pBW showed a similar annual behaviour. Specifically, pBW was mostly stable over time, showing a slight drop during the summer time and then an increase. The %PROT was lower in July (i.e., 3.26 g/dL ± 0.20 g/dL of milk for first lactation and 3.33% ± 0.21% for second lactation, respectively), and reached a peak in October (Figure 4-2). This tendency was analogous to the trend in %FAT, except that this trait had a peak in November and December which was approximately one month later compared to %PROT. The evolution of pIC for both studied lactations dropped smoothly until May, and then peaked in November, which was quite contrary to the annual trend observed for milk yield. The best feed efficiency did not occur at the point of the highest milk yield or the lowest DMI.

3.5 Heritability

The estimated heritability of pDMI was 0.14 and 0.11 for first and second lactation. The heritability estimates for pBW were also moderate and equal to 0.18 and 0.17 for first and second lactation. Moderate heritability values were also estimated for pIC (0.14 and 0.09 for first and second lactation).

Table 4-3 Heritability (h^2) of studied traits for the first two lactations.

Traits	Heritability	
	Lactation1	Lactation2
Milk yield	0.20	0.16
%FAT	0.37	0.41
%PROT	0.41	0.40
pBW	0.18	0.17
pDMI	0.14	0.11
pIC	0.14	0.09

4. Discussion

4.1 Parity effect

The growth of a cow directly impacts its morphology and therefore the body weight. So, a positive BW difference is expected between primiparous and multiparous dairy cows. Although BW was predicted using MIR spectrometry and animal characteristics, the trend observed for this predicted phenotype (Table 4-1) is in agreement with the behaviour observed for the reference measured trait. Indeed, the same trend was reported by Ferris et al. (2017), who found reference BW for Holstein-Friesian cows for first lactation to subsequent lactations ranged from 588 kg to 702 kg.

The BW is not the only morphological trait influenced by growth; the whole body of the cow is impacted. These changes of morphological traits induce an increase in the capacity for feed ingestion by the cow. So, more concretely, a significant positive difference between first and second parity was therefore expected for DMI and also for the predicted trait related to DMI, as it is the case in this study. The global mean of pDMI was similar to the DMI value reported by De Boever et al. (2017). They observed an average DMI of 20.87 kg/day for Holstein-Friesian cows at an average lactation of 2.3.

Moreover, growth and related morphological changes also indirectly impact the quantity and the composition of milk. Less mature cows tend to use their limited nutrition intake to complete their own physical development rather than producing more milk (Dado and Allen, 1994; Wathes et al., 2007). This happens alongside growth until the fourth or fifth lactation (Artegoitia et al., 2013; Mellado et al., 2011; Ray et al., 1992). Moreover, the volume of the mammary gland increases with growth and this is positively correlated to the milk yield (Davis and Hughson, 1988). So, all of these things can explain the positive difference of 3.09 kg/day observed for milk yield in this study (similar to the finding of 3.0 kg/d made by Davis and Hughson (1988)) as well as the lower amount of fat and protein produced in it. Younger cows need more nutrients to fulfill body growth and therefore produce milk with a lower content of fat and protein (Oltner et al., 1985). Such differences of milk fat and protein contents between first and second parities were also observed by other authors. For instance, Craninx et al. (2008) observed 3.83% and 4.30% fat from primiparous and multiparous cows. Morales-Almaráz et al. (2018) measured 3.66% and 3.93% of fat in milk produced by primiparous and multiparous non-grazing cows. So, there is a link between morphological traits and dairy production phenotypes. This is why the calculation of the consumption index expressed as the ratio of an indirectly related morphological trait, like DMI, to the milk production trait (FPCM) is of interest to assess feed efficiency. Theoretically, the most productive cows will be the healthy one which can be able to maximize the production of milk for a given amount of feed. As a basis, the cows' genetic background of the feed utilization must be taken into consideration. In field, farmers also need to balance the feed efficiency and the veterinarian cost of a cow to enhance the sustainability of a farm.

The difference of pIC between lactations means that the cows in the second lactation seemed to be more efficient than the cows in first lactation. This finding is biologically relevant as primiparous cows tend to use nutrients to grow rather than to produce milk (Wathes et al., 2007). However, Oldenbroek (1989) studied different breeds of cow and found there was no significant effect of parity on the ratio of milk energy production to the net energy intake, but the cows in second parity were more efficient than the cows in first parity. This conclusion was similar to the trend observed in the present study.

4.2 Effect of Lactation Stage

We know that growth, as well as the gestation of the cow, can impact the body weight and also the quantity and the composition of milk. This can be observed by checking the evolution of traits throughout lactation (Figure 4-1). It is interesting to note that the curves for pBW and pDMI were similar to those presented in the studies (van Knegsel et al., 2007a, 2007b) conducted on Holstein-Friesian cows for BW and DMI up to 9 weeks after calving. Another study (Berry et al., 2014) also concluded a similar curve of DMI within lactation from dairy cows from nine different countries. However, their DMI curve decreased more gently compared to the pDMI trend observed in the present study. Logically, the observed trends can be explained again by the concept of nutrient allocation. A study reported that new calving dairy cows cannot maintain a positive dietary energy balance and must mobilize body reserves to compensate the negative energy balance (Coppock et al., 1972). By contrast, pDMI increased from calving until the mid-lactation, therefore increasing the quantity of available energy and leading to an increase of BW. So, pBW increased after the slight drop in early lactation. From reference measurements, Ntallaris et al. (2017) reported a significant increase of DMI during the first 120 days postpartum in Holstein cows. A research reported that energy equilibrium for first and older lactating cows was reached at week 13 and week 18 of lactation, respectively (Braund and Steele, 1972). This could explain why pDMI had a peak at early lactation. The peak delay of pDMI compared to milk peak is consistent with the results reported by Hristov et al. (2005), supporting the theory that the feed intake is driven by milk production (Council, 2001).

For the smooth slope of the pIC curve, it can be explained by the fact that the feed efficiency related traits like FPCM and DMI were not constant across the whole lactation period (Manzanilla Pech et al., 2014).

4.3 Correlations Between Studied Traits

A positive relationship between pDMI with milk yield was observed in the current study (Table 4-2). A similar positive relationship between these traits (r=0.57 for DMI and milk yield) was observed from reference measurements (Ben Meir et al., 2018). Similarly, DMI increased with milk yield and BW in Holstein-Friesian cows (Garnsworthy et al., 2012). Although there were not many reports on correlations between BW and daily milk yield, the absence of a relationship between BW and milk yield was observed from Holstein (N=1,584) and Jersey (N=679) cows and the reason may be related to the health status of the cows (Badinga et al., 1985). A positive correlation (r=0.40) between metabolic BW and DMI was observed (Ben Meir et al., 2018). Asher et al. (2014) reported a correlation coefficient equal to 0.26 for Israeli Holstein cows between BW and DMI but this value was not significant. Those authors concluded that the positive correlation is due to the biological increase of feed digestibility. However, the lack of significance (*P*=0.14) observed by those authors may be related to the size of the dataset (N=35); a larger cohort could produce a different result.

Positive relationships between pIC and fat or protein contents were observed in this study. Similarly, another study (Ben Meir et al., 2018) reported that high feed efficiency cows produced lower fat and protein contents in milk compared to low feed efficiency cows. This is related to the fact that a cow with high feed efficiency has higher milk production, which explains why a negative correlation between milk yield and pIC is expected. The observed similar correlation value between milk yield and pIC is in agreement with the conclusion of Korver (1988) who stated that high producing cows had a better feed efficiency. Ben Meir et al. (2018) reported that the ratio of ECM to DMI (i.e., another expression of feed efficiency) in Holstein cows was positively correlated with milk yield but negatively correlated with DMI. Similarly, a positive correlation (r=0.18) between feed efficiency (defined as FCM divided by energy intake) and energy intake was reported suggesting that the variation in feed efficiency depends more on the milk production than the feed consumption (Hooven et al., 1968).

The negative correlation between pIC and pDMI is a consequence related to the positive relationships existing between pDMI and milk yield. Moreover, as the animal satisfies first its maintenance (and growth) requirements essentially. Afterwards, it uses nutrients better for milk production and this is known that above a certain limit, the marginal use of nutrients for milk production decrease. Therefore, there may be a curvilinear relation between pIC and pDMI. A part of this positive correlation is also related to the fact that pDMI includes in its formula the information about milk yield.

The slightly positive relationships between pBW and pIC means that heavier cows would tend to present lower feed efficiency. This statement was also expressed in the past (Yerex et al., 1988). A negative correlation (r=–0.20) was reported between BW and feed conversion efficiency (DMIE, defined as 305-d FCM divided by 305-d DMI) (Vallimont et al., 2011).

This study confirmed the positive value observed in the current research as we have used an inverse indicator of feed efficiency. The low value of correlation was expected based on the findings (Badinga et al., 1985) which observed a quadratic relationship between feed efficiency with the BW in the first lactation of Holstein cows, and the correlation coefficient may also change.

4.4 Test month influence

The trends of %PROT (Figure 4-2) is in agreement with the findings from previous researches (Bittante et al., 2015). Those authors also found that the %FAT and %PROT dropped during the summer time and then slightly increased. The gentle increase in pBW observed in the current study, during the period around winter when concentrate supplement is given to cows, was consistent with results reported previously (Schöbitz et al., 2013).

Holstein cows showed lower feed efficiency in warm seasons than cold seasons (Britt et al., 2003). In the present study, the findings suggest that the dairy cows had higher feed efficiency around the spring time when the temperature was comfortable and concentrate supplementation was available.

Different reasons can explain the three trends observed. The first reason and the main one is the change in diet given to the cows throughout the year. Feed rations richer in concentrates are currently given to cows during the non-grazing time (i.e., winter season) leading to increased DMI, milk yield, and %PROT (Bargo et al., 2003; Schöbitz et al., 2013). Some authors have found that milk fat has an increasing trend when a higher amount of concentrate is given to the cow (70% vs. 50% DM) (Murphy et al., 2000; Sutton et al., 1986) but this effect depends heavily on the composition of the concentrate which is related to changes in the ratio of acetate to propionate in the rumen (Woodford and Murphy, 1988). In the current study, higher milk fat contents were observed during the winter period (i.e., when concentrate supplements were given to the cows) and the milk yield was relatively low. The second reason, coupled to the first one, is the physical activity of the cows. During the grazing period, the cows will have higher physical activity which impacts the energy partitioning and may lead to a decrease of milk production and BW. The third reason for annual changes observed during the summer is that high temperatures relate to discomfort of the cow living in this environment (Council, 1988). During this time, the cow will eat less (i.e., lower DMI) due to heat stress, negatively impacting milk production (Ominski et al., 2002) and potentially also BW (Rees, 1964).

4.5 Heritability

The estimated heritability of pDMI (Table 4-3) was similar to the value of 0.17 reported before for DMI (Vallimont et al., 2011). This is also within the range of heritability for DMI proposed by Hardie et al. (2017) for dairy cows (from 0.02 to 0.52). The heritability of pBW is similar to the value of 0.19 (Clark and Touchberry, 1962) and within the range proposed in a research for the BW of Holstein cows (from 0.20 to 0.80) (Hooven et al., 1968). The heritability of pIC was moderate. Generally, the heritability of gross feed efficiency ranges from 0.12 to 0.63 for dairy cows (Korver, 1988). These values were in agreement with those estimated by other authors for other traits related to feed efficiency. A value of 0.18 for feed efficiency defined as 305-d FCM divided by net-energy intake for 970 dairy cows was reported (Vallimont et al., 2011). So, heritability estimates of pBW, pDMI, and pIC were moderate and slightly lower for the second lactation than first lactation (Table 4-3) . A similar trend was found (Song, 2010) which the 305-d DMI of cows in the third parity was 0.08, lower than the 0.12 for cows in first parity, and concluded this may be due to the ongoing breeding selection of cows, leading to lower genetic variance for mature cows.

5. Conclusions

Due to practical issues, measuring traits related directly or indirectly to feed efficiency, such as BW, IC, and DMI at large and individual scales, is difficult if not impossible. However, predicting those traits using animal production characteristics and MIR spectrometry is feasible. However, as it was not possible to conduct a large scale validation of these predictions, this study aimed to predict these traits using the Walloon milk recording database in order to observe the behaviour of the traits in real conditions, and compared them to findings reported in the literature. The results revealed that behaviour throughout the year, the stage of lactation, and the parity for these predicted traits fully agreed with the expected values suggesting that relevant information was provided by the traits, even if their prediction accuracy was moderate. Additionally, these traits presented moderate genetic variability. All of these aspects suggest the potential interest of using predicted traits to develop management and breeding tools. To confirm this interest, a large scale study may be conducted to assess the accuracy of predictions. For that, it will be relevant to combine the efforts made internationally in many research farms where the routine acquisition of DMI and BW is possible. This could allow comparisons of the prediction with real reference values, collected in many different conditions, in order to establish variability that is representative of the dairy cow population. Moreover, as a genetic component exists, it could be interesting to realize a genetic evaluation of the feed efficiency related traits and compare the breeding values estimated for sires having daughters with predicted and reference records. If the ranking of bulls is similar, this will open new perspectives on the use of these traits to develop future selection programs aiming to improve the feed efficiency of dairy cows.

6. Acknowledgments

The China Scholarship Council (CSC) is gratefully acknowledged for funding Mr. Zhang’s PhD grant. The authors acknowledge the technical support provided by the milk laboratory (CECI, AWé, Comité du Lait, Belgium) and thank the Walloon Breeding Association (Ciney, Belgium) for providing access to milk-recording data.

References

Artegoitia, V., Meikle, A., Olazabal, L., Damián, J.P., Adrien, M.L., Mattiauda, D.A., Bermudez, J., Torre, A., Carriquiry, M., 2013. Milk casein and fatty acid fractions in early lactation are affected by nutritional regulation of body condition score at the beginning of the transition period in primiparous and multiparous cows under grazing conditions. J. Anim. Physiol. Anim. Nutr. (Berl). 97, 919–932.

Asher, A., Shabtay, A., Haim, A., Aharoni, Y., Miron, J., Adin, G., Tamir, A., Arieli, A., Halachmi, I., Moallem, U., Orlov, A., Brosh, A., 2014. Time required to determine performance variables and production efficiency of lactating dairy cows. J. Dairy Sci. 97, 4340–4353.

Badinga, L., Collier, R.J., Wilcox, C.J., Thatcher, W.W., 1985. Interrelationships of milk yield, body weight, and reproductive performance. J. Dairy Sci. 68, 1828–1831.

Bargo, F., Muller, L.D., Kolver, E.S., Delahoy, J.E., 2003. Invited Review: Production and Digestion of Supplemented Dairy Cows on Pasture. J. Dairy Sci. 86, 1–42.

Ben Meir, Y.A., Nikbachat, M., Fortnik, Y., Jacoby, S., Levit, H., Adin, G., Cohen Zinder, M., Shabtay, A., Gershon, E., Zachut, M., Mabjeesh, S.J., Halachmi, I., Miron, J., 2018. Eating behavior, milk production, rumination, and digestibility characteristics of high- and low-efficiency lactating cows fed a low-roughage diet. J. Dairy Sci. 101, 10973–10984.

Berry, D.P., Coffey, M.P., Pryce, J.E., de Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J.J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K., Veerkamp, R.F., 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. J. Dairy Sci. 97, 3894–3905.

Bittante, G., Cipolat-Gotet, C., Malchiodi, F., Sturaro, E., Tagliapietra, F., Schiavon, S., Cecchinato, A., 2015. Effect of dairy farming system, herd, season, parity, and days in milk on modeling of the coagulation, curd firming, and syneresis of bovine milk. J. Dairy Sci. 98, 2759–2774.

Braund, D.G., Steele, R.., 1972. Performance of cows individually fed total mixed rations ad libitum, Coop. Res. Farms. Trial CF2--269, Charlottesville, NY.

Britt, J.S., Thomas, R.C., Speer, N.C., Hall, M.B., 2003. Efficiency of converting nutrient dry matter to milk in Holstein herds. J. Dairy Sci. 86, 3796–3801.

Carpenter, K.J., Harper, A.E., Olson, R.E., 1997. Experiments That Changed Nutritional Thinking. J. Nutr. 127, 1017S-1053S.

Clark, R.D., Touchberry, R.W., 1962. Effect of body weight and age at calving on milk production in Holstein cattle. J. Dairy Sci. 45, 1500–1510.

Connor, E.E., 2015. Improving feed efficiency in dairy production: challenges and possibilities. Animal 9, 395–408.

Coppock, C.E., Noller, C.H., Wolfe, S.A., Callahan, C.J., Baker, J.S., 1972. Effect of forage-concentrate ratio in complete feeds fed ad libitum on feed intake prepartum and the occurrence of abomasal displacement in dairy cows. J. Dairy Sci. 55, 783–789.

Council, N.R., 2001. Nutrient Requirements of Dairy Cattle, Seventh Revised Addition 2001. Natl. Res. Counc.

Council, N.R., 1988. Factors affecting the composition of milk from dairy cows, in: Designing Foods: Animal Product Options in the Marketplace. National Academies Press, US.

Craninx, M., Steen, A., Van Laar, H., Van Nespen, T., Martín-Tereso, J., de Baets, B., Fievez, V., 2008. Effect of lactation stage on the odd-and branched-chain milk fatty acids of dairy cattle under grazing and indoor conditions. J. Dairy Sci. 91, 2662–2677.

Dado, R.G., Allen, M.S., 1994. Variation in and relationships among feeding, chewing, and drinking variables for lactating dairy cows. J. Dairy Sci. 77, 132–144.

Davis, S., Hughson, G., 1988. Measurement of functional udder capacity in lactating Jersey cows. Aust. J. Agric. Res. 39, 1163–1168.

De Boever, J.L., Goossens, K., Peiren, N., Swanckaert, J., Ampe, B., Reheul, D., De Brabander, D.L., De Campeneere, S., Vandaele, L., 2017. The effect of maize silage type on the performances and methane emission of dairy cattle. J. Anim. Physiol. Anim. Nutr. (Berl). 101, e246–e256.

Enevoldsen, C., Kristensen, T., 1997. Estimation of body weight from body size measurements and body condition scores in dairy cows. J. Dairy Sci. 80, 1988–1995.

FAO, 2010. Food and Agriculture Organization of the United Nations (FAO). Greenhouse Gas Emissions from the Dairy Sector - A Life Cycle Assessment [WWW Document]. FAO, Anim. Prod. Heal. Div. URL http://www.fao.org/3/k7930e/k7930e00.pdf (accessed 1.13.19).

Ferris, C.P., Jiao, H., Murray, S., Gordon, A.W., Carson, A.F., 2017. Methane emissions from non-lactating pregnant dairy cows while grazing. Livest. Sci. 206, 121–124.

Gaines, W.L., Davidson, F.A., 1928. Relation between percentage fat content and yield of milk.

Garnsworthy, P.C., Craigon, J., Hernandez-Medrano, J.H., Saunders, N., 2012. Variation among individual dairy cows in methane measurements made on farm during milking. J. Dairy Sci. 95, 3181–3189.

Hardie, L.C., VandeHaar, M.J., Tempelman, R.J., Weigel, K.A., Armentano, L.E., Wiggans, G.R., Veerkamp, R.F., de Haas, Y., Coffey, M.P., Connor, E.E., Hanigan, M.D., Staples, C., Wang, Z., Dekkers, J.C.M., Spurlock, D.M., 2017. The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. J. Dairy Sci. 100, 9061–9075.

Heinrichs, A.J., Rogers, G.W., Cooper, J.B., 1992. Predicting body weight and wither height in Holstein heifers using body measurements. J. Dairy Sci. 75, 3576–3581.

Hooven, N.W., Miller, R.H., Plowman, R.D., 1968. Genetic and environmental relationships among efficiency, yield, consumption and weight of Holstein cows. J. Dairy Sci. 51, 1409–1419.

Hristov, A.N., Price, W.J., Shafii, B., 2005. A meta-analysis on the relationship between intake of nutrients and body weight with milk volume and milk protein yield in dairy cows. J. Dairy Sci. 88, 2860–2869.

International Organization for Standardization, 2017. ICAR Guidelines, Section 2 - Guidelines for Dairy Cattle Milk Recording [WWW Document]. URL https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf (accessed 6.6.19).

Korver, S., 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: A review. Livest. Prod. Sci. 20, 1–13.

Manzanilla Pech, C.I.V., Veerkamp, R.F., Calus, M.P.L., Zom, R., van Knegsel, A., Pryce, J.E., De Haas, Y., 2014. Genetic parameters across lactation for feed intake, fat- and protein-corrected milk, and liveweight in first-parity Holstein cattle. J. Dairy Sci. 97, 5851–5862.

Mellado, M., Antonio-Chirino, E., Meza-Herrera, C., Veliz, F.G., Arevalo, J.R., Mellado, J., de Santiago, A., 2011. Effect of lactation number, year, and season of initiation of lactation on milk yield of cows hormonally induced into lactation and treated with recombinant bovine somatotropin. J. Dairy Sci. 94, 4524–4530.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle.

J. Dairy Sci. 100, 10251–10271.

Misztal, I., 2011. BLUPF90 family of programs [WWW Document]. URL http://nce.ads.uga.edu/~ignacy/numpub/blupf90/ (accessed 1.2.07).

Misztal, I., 1994. MTCAFS (MTC)—Multitrait REML estimation of variance components program by canonical transformation, with support for multiple random effects [WWW Document]. URL http://nce.ads.uga.edu/wiki/doku.php?id=application_programs (accessed 10.2.11).

Morales-Almaráz, E., de la Roza-Delgado, B., Soldado, A., Martínez-Fernández, A., González, A., Domínguez-Vara, I.A., Vicente, F., 2018. Parity and grazing-time effects on milk fatty acid profile in dairy cows. Anim. Prod. Sci. 58, 1233–1238.

Murphy, M., Åkerlind, M., Holtenius, K., 2000. Rumen fermentation in lactating cows selected for milk fat content fed two forage to concentrate ratios with hay or silage. J. Dairy Sci. 83, 756–764.

Ntallaris, T., Humblot, P., Båge, R., Sjunnesson, Y., Dupont, J., Berglund, B., 2017. Effect of energy balance profiles on metabolic and reproductive response in Holstein and Swedish Red cows. Theriogenology 90, 276–283.

Oldenbroek, J.K., 1989. Parity effects on feed intake and feed efficiency in four dairy breeds fed ad libitum two different diets. Livest. Prod. Sci. 21, 115–129.

Olesen, I., Gjerde, B., Groen, A.F., 1999. Methodology for deriving non-market trait values in animal breeding goals for sustainable production systems. Interbull Bull. 13.

Oltner, R., Emanuelson, M., Wiktorsson, H., 1985. Urea concentrations in milk in relation to milk yield, live weight, lactation number and amount and composition of feed given to dairy cows. Livest. Prod. Sci. 12, 47–57.

Ominski, K.H., Kennedy, A.D., Wittenberg, K.M., Moshtaghi Nia, S.A., 2002. Physiological and production responses to feeding schedule in lactating dairy cows exposed to short-term, moderate heat stress. J. Dairy Sci. 85, 730–737.

Phocas, F., Agabriel, J., Dupont-Nivet, M., GEURDEN, J., Médale, F., Mignon-Grasteau, S., Gilbert, H., Dourmad, J.-Y., 2014. Le phénotypage de l’efficacité alimentaire et de ses composantes, une nécessité pour accroître l’efficience des productions animales. INRAE Prod. Anim. 27, 235–248.

Ray, D.E., Halbach, T.J., Armstrong, D.V., 1992. Season and lactation number effects on milk production and reproduction of dairy cattle in Arizona. J. Dairy Sci. 75, 2976–2983.

Rees, H.V., 1964. Heat stress in dairy cattle: physiological responses and variations in milk composition and equilibrium. Master Thesis, University of Tasmania, Australia.

Schöbitz, J., Ruiz-Albarrán, M., Balocchi, O.A., Wittwer, F., Noro, M., Pulido, R.G., 2013. Effect of increasing pasture allowance and concentrate supplementation on animal performance and microbial protein synthesis in dairy cows. Arch. Med. Vet. 45, 247–258.

Song, J., 2010. Estimation of heritability of feed intake in Canadian Holsteins. Master Thesis, McGill University, Montreal.

Song, X., Bokkers, E.A.M., van der Tol, P.P.J., Koerkamp, P.W.G.G., van Mourik, S., 2018. Automated body weight prediction of dairy cows using 3-dimensional vision. J. Dairy Sci. 101, 4448–4459.

Soyeurt, H., Froidmont, E., Dufrasne, I., Hailemariam, D., Wang, Z., Bertozzi, C., Colinet, F.G., Dehareng, F., Gengler, N., 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livest. Sci. 227, 82–89.

Sutton, J.D., Hart, I.C., Brosters, W.H., Elliott, R.J., Schuller, E., 1986. Feeding frequency for lactating cows: effects on rumen fermentation and blood metabolites and hormones. Br. J. Nutr. 56, 181–192.

Tasdemir, S., Urkmez, A., Inal, S., 2011. Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis. Comput. Electron. Agric. 76, 189–197.

Underwood, E.A.; M.A.. M.D.. D.P.H., 1944. Lavoisier and the history of respiration, in: Proceedings of the Royal Society of Medicine. November 3. Longmans, Green and Co, London.

Vallimont, J.E., Dechow, C.D., Daubert, J.M., Dekleva, M.W., Blum, J.W., Barlieb, C.M., Liu, W., Varga, G.A., Heinrichs, A.J., Baumrucker, C.R., 2011. Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. J. Dairy Sci. 94, 2108–2113.

van Knegsel, A.T.M., van den Brand, H., Dijkstra, J., van Straalen, W.M., Heetkamp, M.J.W., Tamminga, S., Kemp, B., 2007a. Dietary energy source in dairy cows in early lactation: energy partitioning and milk composition. J. Dairy Sci. 90, 1467–1476.

van Knegsel, A.T.M., van den Brand, H., Graat, E.A.M., Dijkstra, J., Jorritsma, R., Decuypere, E., Tamminga, S., Kemp, B., 2007b. Dietary energy source in dairy cows in early lactation: metabolites and metabolic hormones. J. Dairy Sci. 90, 1477–1485.

Vanrobays, M.L., Vandenplas, J., Hammami, H., Froidmont, E., Gengler, N., 2015. Short communication: Novel method to predict body weight of primiparous dairy cows throughout the lactation. J. Dairy Sci. 98, 692–697.

Wathes, D.C., Cheng, Z., Bourne, N., Taylor, V.J., Coffey, M.P., Brotherstone, S., 2007. Differences between primiparous and multiparous dairy cows in the inter-relationships between metabolic traits, milk yield and body condition score in the periparturient period. Domest. Anim. Endocrinol. 33, 203–225.

Woodford, S.T., Murphy, M.R., 1988. Effect of forage physical form on chewing activity, dry matter intake, and rumen function of dairy cows in early lactation. J. Dairy Sci. 71, 674–686.

Yerex, R.P., Young, C.W., Donker, J.D., Marx, G.D., 1988. Effects of selection for body size on feed efficiency and size of Holsteins. J. Dairy Sci. 71, 1355–1360.

Yusuf, R.O., Noor, Z.Z., Abba, A.H., Hassan, M.A.A., Din, M.F.M., 2012. Methane emission by sectors: A comprehensive review of emission sources and mitigation methods. Renew. Sustain. Energy Rev. 16, 5059–5070.

5

Chapter V : General discussion

Simple summary

This chapter uses the findings mentioned in the past chapters to propose a prediction tool of the cow feed efficiency using traits easily recorded by DHI organizations. The robustness, the accuracy, and the relevancy of this tool will be assessed in this chapter by using 2 DHI datasets and 2 datasets including reference BW measurements.

1. Context

For decades, scientists from multiple disciplines such as animal nutrition, animal genetics and breeding, and microbiology seek a better understanding and an improvement of the feed efficiency (FE) of dairy cows in order to enhance sustainable dairy production (Lovendahl et al., 2018). The aim of this is to optimize the milk production of cows under the finite natural resources and to reduce the negative impact of it on the environment like GHG emissions (Nam et al., 2020). Besides its direct relationships with nutrition, production, and environment, FE is also related to animal well-being, reproductive ability, and disease resilience (Gengler et al., 2018). In conclusion, knowing the dairy cow feed efficiency at a large scale is a hot topic for dairy farming.

The FE concerns both the profit aspect related to the optimization of milk production, and the biological aspect linked to the efficiency of the nutrient usage (de Ondarza and Tricarico, 2017). Consequently, the assessment of FE through the calculation of the ratio of fat- and protein- corrected milk yield (FPCM) to dry matter intake (DMI) (or it's reciprocal) is relevant (Korver, 1988, de Ondarza and Tricarico, 2017). Due to the cost and labor intensive, the collection of individual DMI information at a large scale is unfeasible. However, the National Research Council (NRC) proposed in 2001 an equation to predict this trait from more easily recorded features which are the milk production, the body weight, and the number of lactation weeks (NRC, 2001). So, as the FPCM calculation is easy to compute using records collected routinely by dairy herd improvement (DHI) organizations, calculating an individual indicator of feed efficiency is feasible when the BW information of a cow is available.

2. Body weight predictive models and derived feed efficiency tool

Even if the acquisition of cow BW is technically easier than DMI, knowing this information on a routine basis is impossible in most farms due to the equipment cost. Therefore, techniques allowing to assess the BW using traits easily recorded by DHI were of interest for a long time. First, methods using the measurements of specific morphological traits to predict the cow BW were revealed efficient. The accuracy ranged from 37 kg to 110 kg (Banos and Coffey, 2012, Haile-Mariam et al., 2014, Vanrobays et al., 2015). Unfortunately, this acquisition of morphological features cannot be realized on a daily or monthly basis. Indeed, for financial reasons, this process is mainly done one time during the cow lifetime by the DHI organizations leading to limit its use to drive daily decisions. More recently, Soyeurt et al.(2019) highlighted the possibility to use the animal characteristics and milk MIR spectrum to compute a BW indicator. The cow cross-validation root mean square error (RMSE) ranged from 37 to 64 kg. As those features are collected on average every 4-6 weeks from all productive dairy cows in herds involved in the routine milk recording, this BW prediction could be useful. However, this study was performed on a limited dataset and proposed to use nearly 300 spectral features. However, the inclusion of massive features in a predictive model might lead to its over-fitting (Centner et al., 1996). It is why variable selection features were tested in Chapter 2 in order to improve the robustness of BW predictive model. Finally, a total of 5 BW predictive models were built: one equation was developed using ElasticNet method adopting 5 selected predictors using calibration set when leaving out herd7 (Mod1，Chapter 2); one equation was built from ElasticNet method adopting 5 selected predictors using calibration set when leaving out herd10 (Mod2，Chapter 2); one equation was developed using PLS method adopting 5 selected predictors using calibration set when leaving out herd7 (Mod3，Chapter 2); one equation was built from PLS method adopting 5 selected predictors using calibration set when leaving out herd10 (Mod4，Chapter 2); and the fifth equation was developed using PLS method adopting 280 predictors without variable selection using calibration set when leaving out herd7 (Mod5，Chapter 2). More details are given in Chapter 2.

As shown in Table 5-1, the validation prediction error ranged from 54 to 65 kg. Globally, the models developed using the predictors after a feature selection presented a similar validation performance (i.e., RMSEv) than the models developed using the initial 280 variables. As explained previously, RMSEcv and RMSEp were assessed using different sets. However, the values were within the same range for all models. Smaller absolute difference between RMSEcv and RMSEp whereas observed for Mod1, Mod3, and Mod5. The lower values could be attributed to the fact that the variability of the test set was better taken into account in the calibration set leading to the obtaining of a more robust predictive model. Moreover, from the BW calibration sets, the values obtained for the RPD, defining as the ratio of RMSEcv to SD, ranged between 1.4 and 1.5 (Table 5-1). If RPD is comprised between 1.5 and 2, this means the possibility to apart the low and high values (Saeys et al., 2005). Moreover, the interpretation of RPD value depends on the context and purpose of the predictions used (Fearn, 2002). Some authors reported that the RPD no less than 1.4 was also considered as acceptable (Camacho-Tamayo et al., 2014). So, these RPD values confirmed the relevance of the BW predictions to detect light and heavy cows.

Table 5-1 The prediction performance of the 5 developed models predicting test-day bodyweight (kg/day) presented in Chapter 2 and the equation obtained by Soyeurt et al. (2019).

Model	**Method***	**Predictors**	**RMSEcv**	**RMSEp**	**Diff_abs**	**RMSEv**	**RPD**
Mod1	ElasticNet	5	53	60	7	65	1.4
Mod2	ElasticNet	5	52	67	15	64	1.5
Mod3	PLS	5	53	54	1	64	1.4
Mod4	PLS	5	52	67	15	64	1.5
Mod5	PLS	280	50	59	9	64	1.5
Soyeurt et al. (2019)	PLS	281	40	37-64	3-24	62	1.4-1.6

* *PLS=Partial Least Square regression; RMSEcv=Root mean square error of cross-validation; RMSEp=Root mean square error of validation; Diff_abs=Absolute difference between RMSEcv and RMSEp; RMSEv=Root mean square error of validation based on Australian dataset; RPD=Ratio of performance to deviation.*

The equation published by Soyeurt et al. (2019) including nearly 300 MIR spectra had a large range of RMSEp. These values represented the accuracy for the prediction obtained for each cow. Therefore, it is not directly comparable with the results obtained in Chapter 2. However, the RMSEv is comparable between BW models as the value was computed using the same validation set. Based on RMSEv mentioned in Table 5-1, there are no strong differences among models, highlighting a similar robustness. This could be due to the variable selections. Indeed, all models presented in Table 5-1 can be considered as the results of variable selection features. Indeed, the model developed by Soyeurt et al.(2019) or Mod 5 used 277 instead of 1,060 spectral data. Those spectral variables were selected by the experience gained through the development of other MIR models. Mod1-Mod4 used 2 spectral variables. Soyeurt et al.(2019) suggested that the inclusion of MIR information inside the BW equation allows to consider the individual BW variation. So, by using only 2 spectral variables, there is a risk to under-fit the individual variation. Indeed, an excessive exclusion of informative predictors could lead to building an under-fitted model (Khaire and Dhanalakshmi, 2019). So, to observe the impact of this BW individual variation, we need to have repeated measurements of the required features on the same cow. This was not the case for the cows included in the calibration sets. Consequently, this highlights the interest of predicting BW on DHI datasets which contain many records from the same cow. The prediction on a trait on a large scale repeated database is always a source of relevant information. For instance, for the model predicting the daily quantity of methane eructed by dairy cows, observing the prediction of this trait on a DHI database allowed to improve the modeling. Indeed, Vanlierde et al.(2015) observed that the methane predictions had a behavior within lactation different than the one known for reference methane measurements. To solve this issue, the authors introduced Legendre Polynomials in the modeling, leading to the obtaining of a pattern with a more expected shape (Vanlierde et al., 2015).

Even if we cannot select easily the best BW predictive model, from the results mentioned in Table 5-1, it is important to keep in mind that the global interest of this thesis is to predict cow FE at large and individual scales. Indeed, as mentioned previously (Chapter 4), we can achieve this objective by calculating the ratio of FPCM to DMI predicted from the NRC equation (2001):

$$DMI = (0.372 * FCM + 0.0968 * BW^{0.75}) * (1 - e^{(-0.192*(wol+3.67))})$$

where $BW^{0.75}$ denotes the metabolic body weight (BW, kg), FCM is 4% Fat-corrected milk (FCM, kg/d), and wol is the number of lactation weeks. As all required information is now available, the predicted FE values were estimated from the BW reference datasets. Table 5-2 provides the descriptive statistics of FE predicted using predicted and reference BW records as well as their corresponding RMSE values.

The FE RMSE values estimated from the calibration/test set were slightly lower than the ones calculated from the validation set. This was expected as the accuracy of BW predicted on the validation set was lower (Table 5-1). Globally, the RMSE accounted for 3.18% and 4.80% of the average FE values in the calibration/test and validation sets, respectively. This estimated FE error lower than 5% reveals the interest of this indicator for a potential use in the dairy sector. It is important to notice that this FE is also related to the accuracy of the

DMI equation developed by NRC (2001) (RMSE=1.82 kg/day across parity).

The correlation values between assumed reference and predicted FE between equations were 0.99 and 0.93 from the calibration/test sets and validation sets, respectively. Based on these strong relationships and the descriptive statistics mentioned in Table 5-2, we can conclude that the BW prediction variation observed between equations did not impact a lot the final FE predictions. However, as mentioned for BW, the ability of the BW equations to model the individual variation could be impacted by the lower consideration of spectral variables in Mod1 to Mod4. Therefore, this could impact also the prediction of FE. So, to assess this impact, it is important also to predict this trait on a large DHI database.

Table 5-2 Descriptive statistics of the assumed reference and predicted feed efficiency and their corresponding root mean square error (RMSE).

	Feed efficiency (kg of milk/kg DM)[*]	**Mean**	**SD**	**Min**	**Max**	**RMSE**
Calibration/test set (N=1,844)	FEref	1.57	0.51	0.34	2.90	NA
	FE1	1.57	0.51	0.37	2.92	0.05
	FE2	1.57	0.51	0.37	2.92	0.05
	FE3	1.57	0.51	0.37	2.92	0.05
	FE4	1.57	0.51	0.37	2.91	0.05
	FE5	1.57	0.51	0.37	2.92	0.05
Validation set (N=3,956)	FEref	1.28	0.12	0.69	1.91	NA
	FE1	1.25	0.13	0.67	1.83	0.06
	FE2	1.25	0.13	0.67	1.83	0.06
	FE3	1.25	0.13	0.67	1.83	0.06
	FE4	1.25	0.13	0.67	1.83	0.06
	FE5	1.25	0.13	0.67	1.83	0.06

[*] *FE1 to FE5 are related to FE values estimated from BW records predicted from Mod1 to Mod5; FEref is estimated using reference BW measurements.*

3. Predictions of bodyweight from easily recorded features

To confirm the potentiality of the predicted trait, we have observed the evolution of predicted BW records following several sources of variation such as the parity, the month of test and the stage of lactation. Four different datasets were used. First, we have merged together the calibration and test sets used to build the 5 predictive models mentioned in Table 5-1 (N=1,915). The second set contained the validation set composed of Australian records (N=4,005). Finally, the third and fourth datasets were composed of DHI records collected in Hebei region of China and in the Walloon Region of Belgium. The Hebei dataset included 302,091 records collected from 44,640 cows in 244 herds by Hebei DHI milk laboratory (Hebei, China) from April to December 2018. The Walloon dataset contained 379,751 records collected from 59,126 cows in 692 herds by the Walloon Breeding Association (Ciney, Belgium) from April 2018 to April 2019.

3.1 Cleaning procedures

The BW predictive models were developed using data collected through past research projects or from research herds by different international partners. So, this dataset cannot be assumed to be representative of the dairy cow population. Indeed, to be representative, the samples should be collected randomly in the entire population and the sample number must be function of the variability of the studied trait. However, for financial and practical reasons, this is unfeasible. The dataset had the following specificities: more records collected from cows belonging to the first three parities (Figure 2-3A, Chapter 2), and the majority of records were distributed in the early lactation (i.e., less than 75 days, Figure 2-3B, Chapter 2). However, the BW range was large with BW measurements comprised between 309 and 984 kg. Therefore, due to this under representativeness, it is possible that the variability of samples to predict was not taken into account in this dataset. This would lead to an extrapolation that could impact negatively the prediction quality. Moreover, a bad prediction quality could be also achieved using poor quality spectral data. Unfortunately, as mentioned in Chapter 3, the method used routinely to ensure high quality for the prediction of fat and protein contents cannot be implemented in the BW context. So, quality protocols based on the computation of statistical parameters estimated from large scale datasets were defined in Chapter 3. To summarize, 3 methods were tested individually or combined: the first method (M1) removed extreme MIR predicted values based on the observed 1% and 99% quartile values; the second method (M2) was based on the GH distance between the sample to predict and the average principal components calculated from the predictors and the calibration set used to build the prediction equation; finally, the third method (M3) was based on the calculation of the absolute fat residual difference. As explained in Chapter 3, those three methods did not discard the same samples as they used different statistical approaches to detect outliers. Based on the obtained results, the best quality protocol to be parsimonious with the number of discarded samples was based on the combination “M1 and M2”.

However, during the reviewing process of the article presented in Chapter 3, some reviewers argued that it will be more interesting to test the quality protocol on reference measurements instead of values provided by the spectrometers. So, in the discussion section of this thesis, we have decided to test this issue on the Australian dataset used to validate the BW predictive models. According to the cleaning procedures, Method M1 removed samples with a predicted BW value lower than 500 or higher than 700 kg; Method M2 discarded samples for which the GH distance was higher than 5; Method M3 deleted samples with an absolute fat difference greater than 0.3 g/dL. The impact of this cleaning on the accuracy is summarized in Table 5-3. All Australian records had a GH distance greater than 5 for Mod5. So, the cleaning methods related to M2 for Mod5 were not performed leading to the writing of NA values in Table 5-3. This is probably related to the inclusion of a higher number of spectral features in Mod5 which were standardized to be on FOSS basis. Indeed, the Australian spectra were provided by Bentley spectrometer. As the BW equations were built using the spectra from FOSS machines, a standardization was applied to make the conversion. Grelet et al.(2020) found a less reproducibility of some spectral points, leading to recommend the use of 212 wavenumbers. In the current work, we have used 277 spectral points. This means that Mod5 must be applied on standardized Bentley spectra with caution.

After the data cleaning, the RMSE gains were close to 0 for each of the predicted BW which might mainly due to a low data loss. Indeed, the methods allowed to discard between 0 to 5.14% of samples. Instead of turning low, the RMSE gain was sometimes negative. This is related to the fact that the error related to the equation (around 50-60 kg; Table 5-1) is higher than the error gain that we can reach by cleaning the dataset. Moreover, the dataset size is also limited and the caution done for the acquisition of sample was high. This could explain the weak presence of extreme outliers in this dataset. Therefore, this confirms the interest of using large scale datasets as done in Chapter 3 to highlight the interest of cleaning procedures other than using database having reference values with limited number of records. The validation of those cleaning procedures could be also highlighted by creating artificial noise in a dataset.

Table 5-3 The prediction improvement observed on the Australian dataset (N=3,989) using all proposed cleaning methods and their combinations.

Model		**M1**	**M2**	**M3**	**M1&M2**	**M1&M3**	**M2&M3**	**M1&M2&M3**	**M1orM2**	**M1orM3**	**M2orM3**	**M1orM2orM3**
Mod1	N_clean	3909	3883	3956	3989	3989	3982	3989	3810	3877	3857	3784
	N_loss	80	106	33	0	0	7	0	179	112	132	205
	N_loss%	2.01	2.66	0.83	0.00	0.00	0.18	0.00	4.49	2.81	3.31	5.14
	RMSE_raw	64.30	64.30	64.30	64.30	64.30	64.30	64.30	64.30	64.30	64.30	64.30
	RMSE_clean	64.20	64.62	64.45	64.30	64.30	64.33	64.30	64.54	64.33	64.73	64.65
	Gain_RMSE	0.10	-0.32	-0.15	0.00	0.00	-0.03	0.00	-0.24	-0.03	-0.43	-0.35
Mod2	N_clean	3909	3873	3956	3989	3989	3980	3989	3801	3877	3849	3777
	N_loss	80	116	33	0	0	9	0	188	112	140	212
	N_loss%	2.01	2.91	0.83	0.00	0.00	0.23	0.00	4.71	2.81	3.51	5.31
	RMSE_raw	64.01	64.01	64.01	64.01	64.01	64.01	64.01	64.01	64.01	64.01	64.01
	RMSE_clean	63.92	64.34	64.14	64.01	64.01	64.04	64.01	64.28	64.06	64.45	64.39
	Gain_RMSE	0.09	-0.33	-0.13	0.00	0.00	-0.03	0.00	-0.27	-0.05	-0.44	-0.38
Mod3	N_clean	3909	3883	3956	3989	3989	3982	3989	3812	3877	3857	3786
	N_loss	80	106	33	0	0	7	0	177	112	132	203
	N_loss%	2.01	2.66	0.83	0.00	0.00	0.18	0.00	4.44	2.81	3.31	5.09
	RMSE_raw	63.43	63.43	63.43	63.43	63.43	63.43	63.43	63.43	63.43	63.43	63.43
	RMSE_clean	63.44	63.78	63.57	63.43	63.43	63.46	63.43	63.80	63.58	63.89	63.91
	Gain_RMSE	-0.01	-0.35	-0.14	0.00	0.00	-0.03	0.00	-0.37	-0.15	-0.46	-0.48
Mod4	N_clean	3909	3873	3956	3989	3989	3980	3989	3801	3877	3849	3777
	N_loss	80	116	33	0	0	9	0	188	112	140	212
	N_loss%	2.01	2.91	0.83	0.00	0.00	0.23	0.00	4.71	2.81	3.51	5.31
	RMSE_raw	63.58	63.58	63.58	63.58	63.58	63.58	63.58	63.58	63.58	63.58	63.58
	RMSE_clean	63.48	63.92	63.71	63.58	63.58	63.61	63.58	63.85	63.62	64.02	63.95
	Gain_RMSE	0.10	-0.34	-0.13	0.00	0.00	-0.03	0.00	-0.27	-0.04	-0.44	-0.37
Mod5	N_clean	3909	NA	3956	NA	3989	NA	NA	NA	3877	NA	NA
	N_loss	80	NA	33	NA	0	NA	NA	NA	112	NA	NA
	N_loss%	2.01	NA	0.83	NA	0.00	NA	NA	NA	2.81	NA	NA
	RMSE_raw	64.30	NA	64.30	NA	64.30	NA	NA	NA	64.30	NA	NA
	RMSE_clean	64.19	NA	64.44	NA	64.30	NA	NA	NA	64.33	NA	NA
	Gain_RMSE	0.11	NA	-0.14	NA	0.00	NA	NA	NA	-0.03	NA	NA

3.2 Evolutions within and between lactations of BW predictions from reference datasets

In order to assess the relevancy of predicted BW records, we have decided to study their evolutions following known sources of variation like the parity and the stage of lactation. First, we have conducted this study using the calibration/test and validation datasets. Records with a day in milk (DIM) greater than 365 were discarded. The descriptive statistics of those datasets are given in Table 5-4.

Table 5-4 The descriptive statistics of the feed efficiency related traits of calibration/test and validation sets.

Traits[*]	Calibration/test set (N=1,823)				Validation set (N=3,956)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
DIM(day)	71.61	78.11	1.00	362.00	104.02	18.71	37.00	161.00
Parity	2.59	1.49	1.00	11.00	2.05	1.06	1.00	8.00
FPCM(kg/day)	30.55	12.42	5.26	67.68	25.97	5.27	10.19	57.76
BWref (kg)	616	76	428	817	550	65	340	770
BW1(kg)	616	53	526	761	589	28	523	687
BW2(kg)	616	54	519	763	589	29	526	686
BW3(kg)	616	53	526	766	587	28	512	691
BW4(kg)	616	54	521	769	588	29	523	687
BW5(kg)	616	53	526	761	589	28	523	687
DMIref(kg/day)	19.00	3.82	7.44	33.34	20.08	2.58	13.12	33.82
DMI1(kg/day)	19.01	3.76	7.80	32.73	20.66	2.20	14.95	32.23
DMI2(kg/day)	19.02	3.80	7.70	32.82	20.67	2.22	14.89	32.35
DMI3(kg/day)	19.01	3.75	7.85	32.69	20.63	2.18	14.97	32.07
DMI4(kg/day)	19.02	3.78	7.74	32.83	20.65	2.21	14.87	32.27
DMI5(kg/day)	19.01	3.76	7.80	32.73	20.66	2.20	14.95	32.23
FEref(kg of milk/kg DM)	1.58	0.50	0.34	2.90	1.28	0.12	0.69	1.91
FE1(kg of milk/kg DM)	1.58	0.50	0.37	2.92	1.25	0.13	0.67	1.83
FE2(kg of milk/kg DM)	1.58	0.50	0.37	2.92	1.25	0.13	0.67	1.83
FE3(kg of milk/kg DM)	1.58	0.50	0.37	2.92	1.25	0.13	0.67	1.84
FE4(kg of milk/kg DM)	1.58	0.50	0.37	2.91	1.25	0.13	0.67	1.83
FE5(kg of milk/kg DM)	1.58	0.50	0.37	2.92	1.25	0.13	0.67	1.83

** BW1 to BW5 values were obtained by applying the models Mod1 to Mod5. DMI1 to DMI5 was predicted using namely BW1 to BW5. FE1 to FE5 is calculated from DMI1 to DMI5; ref means the use of reference BW measurements.*

The evolutions of BW records within and between lactations were obtained by estimating the least squares means (LSMEANS) for the parity*class of 15 DIM (cDIM) effect using the following model:

$$y = Parity*cDIM + Herd*Test\text{-}date + e$$

where y was the vector of BW records and e was the residuals.

The trends of the predicted and reference BW within and between lactations obtained from the calibration and validation sets are shown in Figure 5-1. From both datasets, all predicted BW records depicted a similar trend. From the calibration set, although failed to have the same value as the reference BW, the BW predicted obtained from all equations showed a similar lactation trend compared to the reference BW, except for the third parity (Figure 5-1 B, D, F). From the Australian data, especially for the 2nd and 3rd lactations, the BW predictions did not follow the reference BW trend (Figure 5-1 A, C, E).

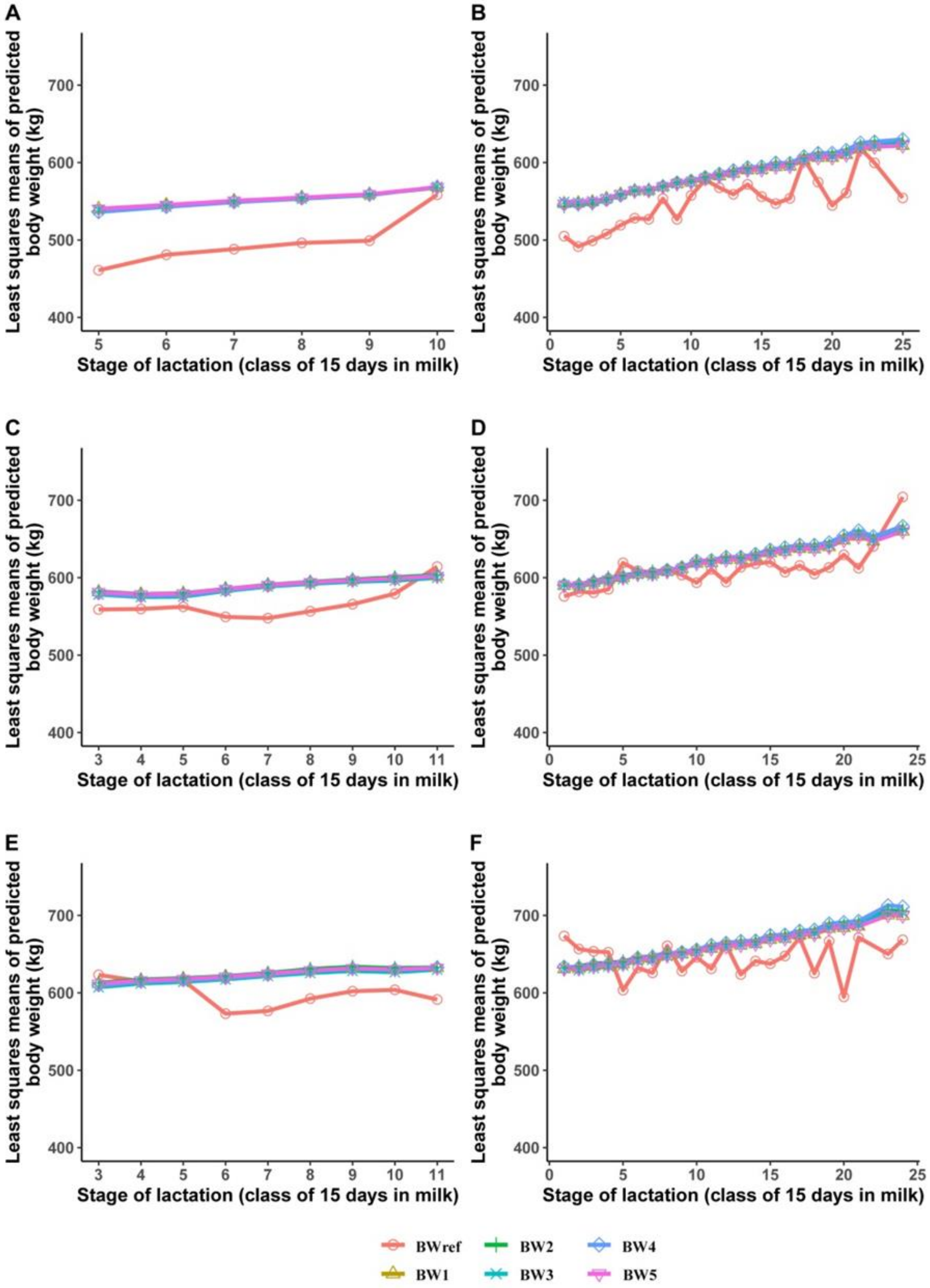


Figure 5-1. The least squares means evolution per class of 15 days in milk and for the first three lactation for the predicted test-day bodyweight. A, C, and E represent the first, second and third parity for the validation set (Australian set), respectively; B, D, and F are the first, second and third parity for the calibration/test set, respectively. BW1 to BW5 are the bodyweight predicting using Mod1 to Mod5; BWref is the observed bodyweight.

The reason for observing a poorer relationship using Australian data could be related to a lack of representativeness for the variability of the validation set in the calibration set. Indeed, if the calibration set is not sufficiently representative, the created model will have a lower accuracy for the new samples (Tomas Naes, 2002).

However, it is important to notice that the trend obtained for the reference BW within the lactation is not in agreement with the expected curve. Indeed, van Knegsel et al.(2007b) observed a BW drop at the early lactation and then a slight increase. The curves obtained using observed BW reference values showed an unsmoothed signal. This could be related to the data acquisition. Indeed, as mentioned by Soyeurt et al.(2019), all BW records were not collected under a fixed measurement protocol. Therefore, the reference BW records presented a certain quantity of noise which could be reflected in the obtained LSMEANS curves. This could explain why the signal was smoother for the validation set compared to the calibration set. Indeed, the validation set contained data coming from only one herd. However, the calibration set contained records collected in several farms and countries.

The lactation patterns observed for the predicted BW had a more expected shape. Even if we cannot observe a BW drop at the beginning of the lactation for the predicted BW traits, the increased tendency with the DIM is expected. This absence of BW drop could be related to a lack of repeated measurements for the same cows throughout their lactation. Moreover, as mentioned previously, the calibration and validation sets were not representative of the cow population leading to the obtaining of not representative DIM LSMEANS curves. In other words, LSMEANS estimated for a DIM can be more influenced by specific cows or herds. Consequently, there is a need to use a larger database to study the trends of predicted BW within and between lactations.

Based on the prediction performances on the validation set (Table 5-1) and the observed within lactation trend (Figure 5-1 A, C, E), it appeared that the variability of the calibration set to predict BW was not large enough. The addition of new samples must be a priority. However, to achieve this objective and avoid the addition of noisy spectral information, it is of interest to increase the calibration set with samples for which the spectral standardization would be done correctly with a small time period between the estimation of the standardization coefficients and the spectral correction. This was not the case for the Australian data as the estimation of standardization coefficients was done only one time. Moreover, as proposed by Grelet et al.(2020), it could be relevant to use 212 predictors instead of 277 if we want to improve the transferability of the models on Bentley instruments.

3.3 Bodyweight predictions from dairy herd improvement datasets

As the calibration/test and validation sets were not representative of the dairy cow population due to their limited size, we have decided to observe the evolution of BW predictions between and within lactation as well as per month of test using DHI datasets. In this case, we don't have the reference BW values but the size of the dataset allowed us to observe a smoother signal within and between lactations or per month of test. The Hebei and Walloon datasets were cleaned using the procedure "M1&M2" mentioned in Chapter 3 to discard potential outliers due to spectral extrapolation or poor quality spectral data. So, the predicted BW records with 1% upper/lower extreme values and with a GH value higher than 5 were supposed to be outliers and were discarded. In Hebei data set, the data loss rate was 1.77% in the step of BW cleaning using M1&M2 method; the loss rate was 2.74% using ICAR standard for milk yield and fat or protein contents. In Walloon data set, the loss rate was 0.06% in the step of BW cleaning using M1&M2 method; the loss rate was 0.01% using ICAR standard for milk yield and fat or protein contents. The descriptive statistics of the cleaned data sets are given in Table 5-5.

Table 5-5 Descriptive statistics of the cleaned dairy herd improvement datasets.

		Milk	Fat	Protein	Predicted body weight (kg)				
		kg/day	g/dL of milk		Mod1	Mod2	Mod3	Mod4	Mod5
Hebei dataset (N=288,607)	Mean	29.89	3.97	3.42	619	622	620	622	596
	SD	11.00	1.12	0.42	46	47	47	48	53
	Min	3.01	1.01	1.01	511	501	507	498	407
	Max	99.82	8.99	6.97	781	791	787	795	820
Walloon dataset (N=379,472)	Mean	27.12	4.08	3.40	634	635	635	636	635
	SD	8.40	0.69	0.37	54	55	55	55	51
	Min	3.10	1.50	1.89	525	520	523	519	438
	Max	97.5	8.99	6.85	782	790	789	795	845

The EN and PLS BW equations employing 5 predictors (i.e., Mod1 to Mod4) were strongly related as the correlation coefficients between their predictions were higher than 0.99 (Table 5-6). However, the PLS equation including 280 predictors (i.e., Mod5) showed a weaker relationship with Mod1-Mod4 (Table 5-6). The lower correlation values observed for Mod5 and Mod0 from the Hebei dataset compared to the Walloon set might be due to the spectral standardization process used. As mentioned previously, some spectral predictors are known to be less reproducible (Grelet et al., 2020). From the Walloon dataset, as expected, the highest correlation for Mod0 (i.e., equation published by Soyeurt et al.(2019)) was observed with Mod5 since the two equations sharing many common predictors.

Table 5-6 The Pearson correlation coefficients among predicted body weight from the Walloon dataset (above the diagonal) and from the Hebei dataset (below the diagonal).

	Mod0*	**Mod1**	**Mod2**	**Mod3**	**Mod4**	**Mod5**
Mod0*		0.830	0.845	0.825	0.841	0.865
Mod1	0.706		0.999	0.999	0.999	0.850
Mod2	0.733	0.997		0.997	0.999	0.851
Mod3	0.692	0.997	0.991		0.999	0.847
Mod4	0.724	0.997	0.998	0.996		0.848
Mod5	0.683	0.591	0.595	0.571	0.581	

** Mod0 represents the bodyweight equation published by Soyeurt, et al. (2019).*

Table 5-7 Descriptive statistics of the feed efficiency related traits predicted on two dairy herd improvement datasets.

Traits*	Hebei (N=271,139)				Walloon (N=356,382)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Parity	2.01	1.19	1.00	13.00	2.49	1.57	1.00	13.00
DIM(day)	168.33	93.16	2.00	365.00	168.69	95.14	5.00	365.00
Milk yield (kg/d)	30.00	10.15	3.10	90.96	27.07	7.86	6.20	67.60
FPCM(kg/day)	29.71	10.19	3.36	73.92	27.09	7.19	8.83	62.12
DMI0(kg/day)	22.85	4.31	11.12	35.36	21.79	3.03	12.96	29.50
DMI1(kg/day)	21.99	4.02	11.06	33.81	21.22	2.88	12.64	28.55
DMI2(kg/day)	22.03	4.06	10.98	33.94	21.24	2.91	12.60	28.65
DMI3(kg/day)	22.00	4.00	11.10	33.76	21.24	2.87	12.66	28.54
DMI4(kg/day)	22.03	4.05	11.01	33.92	21.26	2.90	12.62	28.63
DMI5(kg/day)	21.65	4.14	10.85	33.88	21.22	2.91	12.84	28.71
Methane yield (g/d)	415	83	170	675	401	78	158	648
Methane yield (g/kg MY)	16	10	3	180	16	6	4	69
Methane yield (g/kg DMI)	19	5	5	57-61	19	4	6	47
BW0(kg)	682	59	428	997	675	51	458	996
BW1(kg)	620	46	520	781	633	53	532	782
BW2(kg)	622	47	512	791	635	54	528	789
BW3(kg)	620	47	512	787	635	54	530	789
BW4(kg)	623	48	510	795	636	55	525	795
BW5(kg)	596	51	407	820	634	50	438	811
FE0(kg of milk/kg DM)	1.27	0.29	0.29	2.18	1.24	0.26	0.64	2.22
FE1(kg of milk/kg DM)	1.32	0.30	0.29	2.27	1.27	0.27	0.64	2.30
FE2(kg of milk/kg DM)	1.32	0.30	0.29	2.27	1.27	0.27	0.64	2.30
FE3(kg of milk/kg DM)	1.32	0.30	0.29	2.28	1.27	0.27	0.64	2.30
FE4(kg of milk/kg DM)	1.32	0.30	0.29	2.27	1.27	0.27	0.64	2.30
FE5(kg of milk/kg DM)	1.34	0.29	0.30	2.26	1.27	0.26	0.65	2.26

**BW0 denotes the BW predictions from the equation published by Soyeurt et al. (2019); BW1 to BW5 values were predicted from the models Mod1 to Mod5; DMI1 to DMI5 was predicted using namely BW1 to BW5 using NRC equation (NRC, 2001); FE1 to FE5 were calculated from DMI1 to DMI5; Methane yield results expressed as unit of DMI were averaged among different models.*

The lower correlation coefficients for BW values predicted from Mod5 with the ones predicted from Mod1-Mod4 revealed a different prediction behavior which was not visible from the calibration (Figure 5-1 B, D, F) or validation sets (Figure 5-1 A, C, E). So, different evolutions within and between lactations could be occurred. This statement could be extended to other sources of variation like herd, month of tests or genetic background.

So, to study the effects of days in milk, parity, and the month of test, we have used the same modeling as the one applied previously on calibration and validation sets. Generalized linear models were used for each dataset and BW predictions. In the next sections, we will also discuss the interest of other FE related traits like predicted FE, DMI, DIM, milk yield, and methane. So, in order to keep the same datasets throughout this discussion, the cleaning process took into account all of those traits. The quantity of methane was predicted using the equations developed by Vanlierde et al.(2020) from milk yield, parity, DIM, breed, and milk MIR spectrum. The Hebei and Walloon data sets were cleaned by deleting the 1% upper and lower extreme values of the predicted FE, DMI, and methane. The thresholds used for Hebei dataset were 11 and 34 kg/d for DMI, 0.30 and 2.30 for FE, and 170 and 675 g/d for methane. The thresholds used to clean the Walloon dataset were 12.50 kg/d and 29 kg/d for DMI, 0.60 and 2.30 for FE, and 158 and 648 g/d for methane. The descriptive statistics of these datasets are given in Table 5-7.

The LSMEANS evolutions of BW predictions within and between lactations are illustrated in Figure 5-2. For the first parity and from the Hebei dataset, the BW predictions ranged between 540 to 600 kg for the predictions obtained from Mod1 to Mod4, between 540 to 560 kg for the values predicted from Mod5, and between 590 to 680 kg for the prediction obtained from Mod0. For the first parity Walloon cows, the BW predictions ranged between 540 to 620 kg for the predictions obtained from Mod1 to Mod4, between 560 to 590 kg for the values predicted from Mod5, and between 590 to 680 kg for the prediction coming from Mod0. For the second parity the BW predictions ranged between around 580 to 700 kg for BW predictions obtained from the different models used. For the third parity, the BW predictions varied between 610 to 740 kg. In early lactation (i.e., DIM within the first 100 days), these BW ranges were consistent with the BW values reported by Poncheki et al.(2015) from Holstein dairy cows. The values found by those authors ranged between 521 to 575 kg for the first parity, between 609 and 654 kg for the 2nd parity, and between 648 and 700 kg for the 3rd parity. In the middle and late lactation (i.e., DIM from 120th to 300th days), these BW predictions were also within the range observed by Miller et al.(1969). Those authors found BW records comprised between 576 to 626 kg for the first parity, between 630 to 682 kg for the 2nd parity, and between 670 to 724 kg for the 3rd parity.

However, the BW values predicted from Mod1 to Mod4 showed a strong linear relationship with the DIM (Figure 5-2). This tendency was not observed for the BW predictions obtained from Mod0 and Mod5. From the Hebei dataset, the BW predictions obtained from these 2 models were, respectively, always higher or lower than BW values predicted from Mod1-Mod4. From the Walloon dataset, the predictions from Mod5 were first higher until around 150th DIM and then were lower. For Mod0, the BW values were always higher than other BW predictions. From both datasets, the BW predictions from Mod5 dropped at the beginning of lactation and then increased slightly with the DIM. As reported by literature, the BW always dropped at the early lactation and then recovered slightly later (van Knegsel et al., 2007a). This drop is related to the negative energy balance of the cows after calving as body reserves is mobilized to fulfill the energy gap between the energy intake and the energy required for ensuring the milk production and the maintenance (van Knegsel et al., 2007a). A research done by Řehák et al.(2012) reported the BW loss of Holstein cow at the early lactation (i.e., DIM around 50th day) can be around 40 kg and then BW moderately increases until the end of lactation. From the same report, the BW of Holstein cow recovered to their body condition just before calving at around 28 weeks after calving. In conclusion, BW5 was more in agreement with the expected BW evolution trend across DIM.

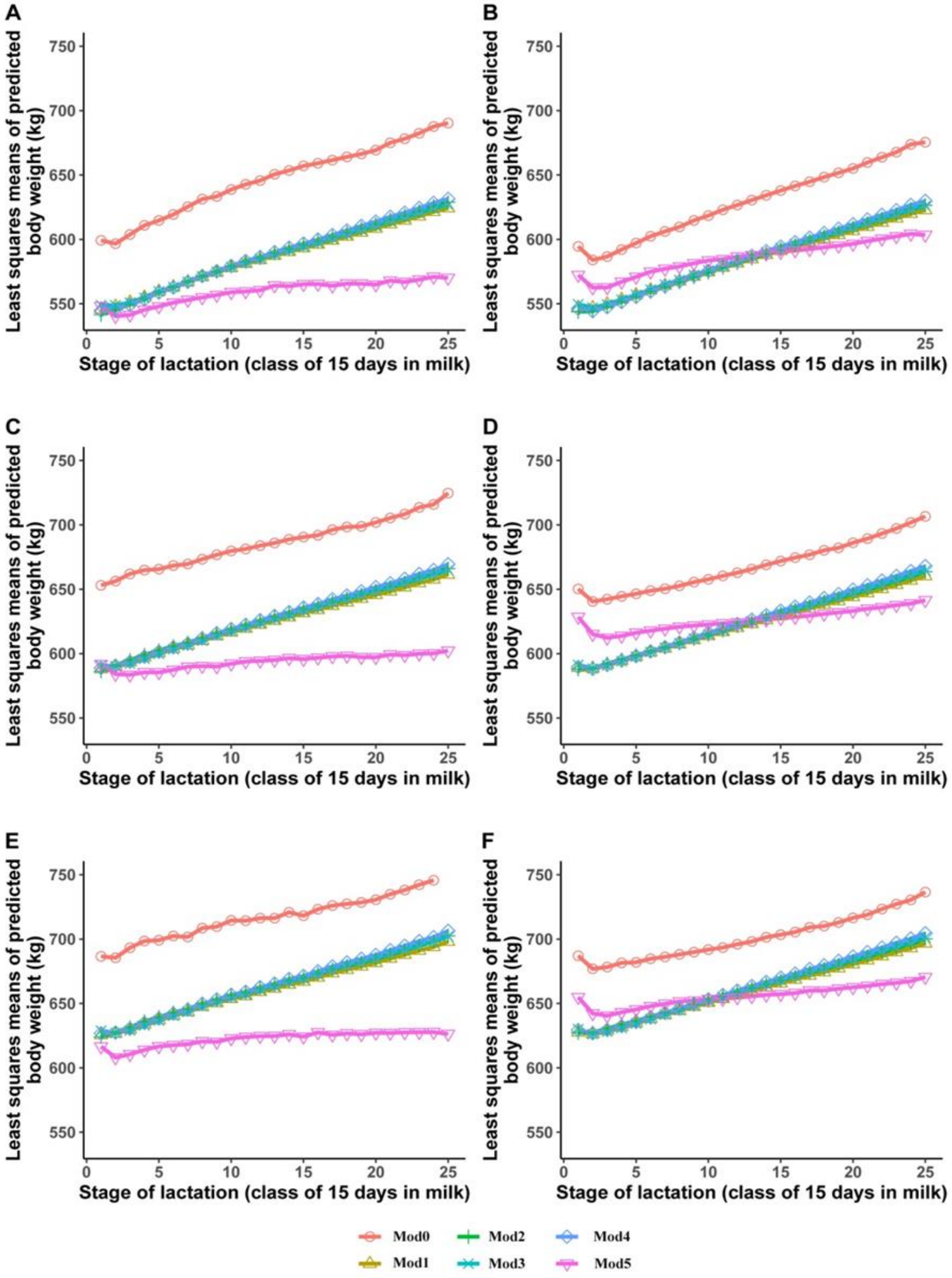


Figure 5-2. Evolution of least squares means of body weights predicted using the 5 available models between and within lactations; A, C, E and B, D, F denote the first, second, and third parity of cows from Hebei and Walloon dataset respectively; Mod0 is the body weight predicted using the equation proposed by Soyeurt et al. (2019).

The LSMEANS of BW predictions calculated for the herd*test-date fixed effect was used to depict the evolution of predicted BW following the month of test (Figure 5-3). From both datasets, the BW predictions from Mod5 were always lower than the ones obtained from other equations (i.e., Mod1 to Mod 4) while the Mod0 gave always higher values. All BW predictions dropped during the summer period and then increased slightly until November. For the Walloon dataset, the main reason is the higher mobility of cows and the feeding changes occurring at this period. Indeed, the cows are on pasture during the spring and summer (Peeters, 2010). As the LSMEANS of BW was calculated based on the test month removing the other variations like the lactation stage and the year of test, the BW loss across the test month was thereby caused by the moment of test. During this period, the cows had grass-based diet which leading to more variable of the cow. As a result, the energy allocation status changed a bit to the activity of the cows, therefore, the cows might be lighter. Moreover, same as the Hebei dairy cows, this trend can be punctually impacted by the heat stress during the summer time as the high temperatures in the summer involve a discomfort of the cow (NRC, 1988), leading to a drop of feed intake which is followed potentially by a BW loss (Rees, 1964). Based on the observed annual variation of predictions, Mod5 seems also to be the most optimal model.

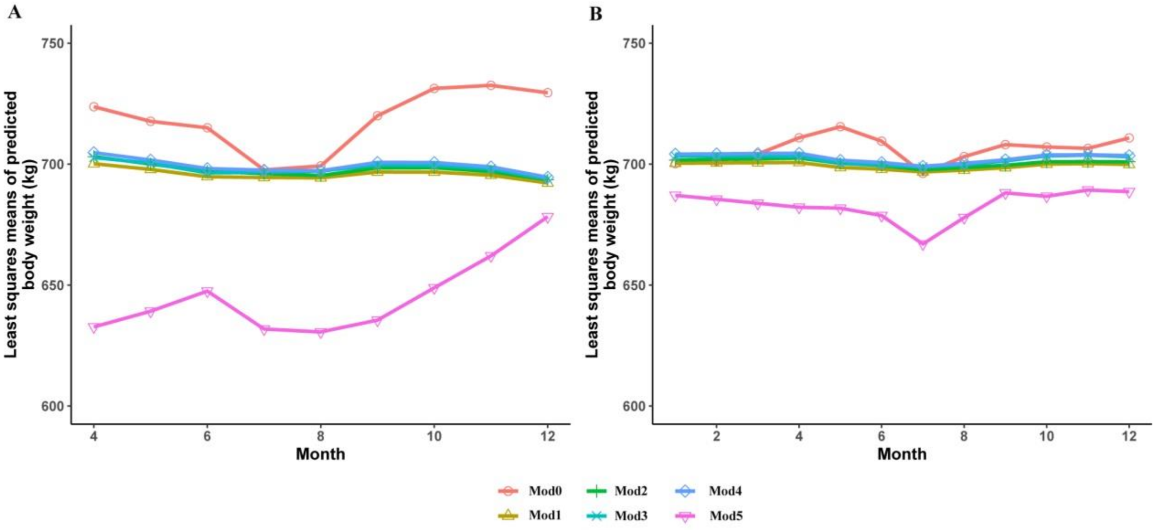


Figure 5-3. The evolution of the least squares means of different predicted body weights by test month. A and B represent the Hebei and Walloon datasets, respectively.

So, based on the obtained results from DHI datasets, 2 classes of BW equations seem to exist: Mod1-Mod4 v.s. Mod5. This splitting is directly related to the number of spectral variables included in the predictive model. The higher number of spectral features included in Mod5 allowed the predictions to have a more expected curve within and between lactations as well as per month of test. The same trend was observed for Mod0. However, this equation developed in the past on a more limited dataset seemed to over-estimate the BW.

4. Predictions of feed efficiency at large and individual scales

The main objective of this thesis is to propose a tool to predict the feed efficiency of a cow at large and individual scales. As mentioned previously, this could be done by using the NRC equation (NRC, 2001) and dairy traits routinely recorded by DHI organizations. Similar to the work done until now for BW, the relevancy of predicted FE was studied by observing its evolution following the parity, the stage of lactation, and the month of test. This was achieved by estimating the LSMEANS for cDIM*parity and herd*test-date effects from the four datasets used previously to model BW predictions.

4.1 Prediction of feed efficiency from reference datasets

The descriptive statistics of FE records used in this section are given in Table 5-4. The distribution of the assumed reference and predicted FE values are shown in Figure 5-4. For the reference data set, a mode FE value of 1.88 kg of milk/kg DM was obtained. The average FE was 1.57±0.51 kg of milk/kg DM. Similarly to the descriptive statistic (Table 5-4), the predicted FE followed well the assumed reference values. For the validation set, the assumed reference FE values presented a normal distribution (mean=1.25 kg of milk/kg DM, SD=0.13 kg of milk/kg DM) and followed also the assumed FE values. Compared to the behaviors observed for BW predictions (Figure 5-1), the trends between FE predictions within lactation were more similar to the assumed reference FE values. The main reason for that is related to the way of calculating reference FE. Indeed, the FE values assumed here to be the reference is in reality a prediction as DMI was predicted using NRC equation. Therefore, it is logical to have a smoother signal as the trait is less noisy. The closeness between assumed reference and predicted FE points can also be explained by the fact that the predicted BW is only one of the features used to predict FE. However, bigger differences within lactation were observed for the first parity from the validation set (Figure 5-5 A). This can be attributed to higher BW prediction (Figure 5-1 A), leading to having higher predicted DMI values and, therefore, lower FE estimations (Table 5-4). The similar trend of the predicted FE between BW equations indicated a good robustness of these values to a moderate change of BW. As mentioned previously, the reason might result from the formula used to calculate the FE.

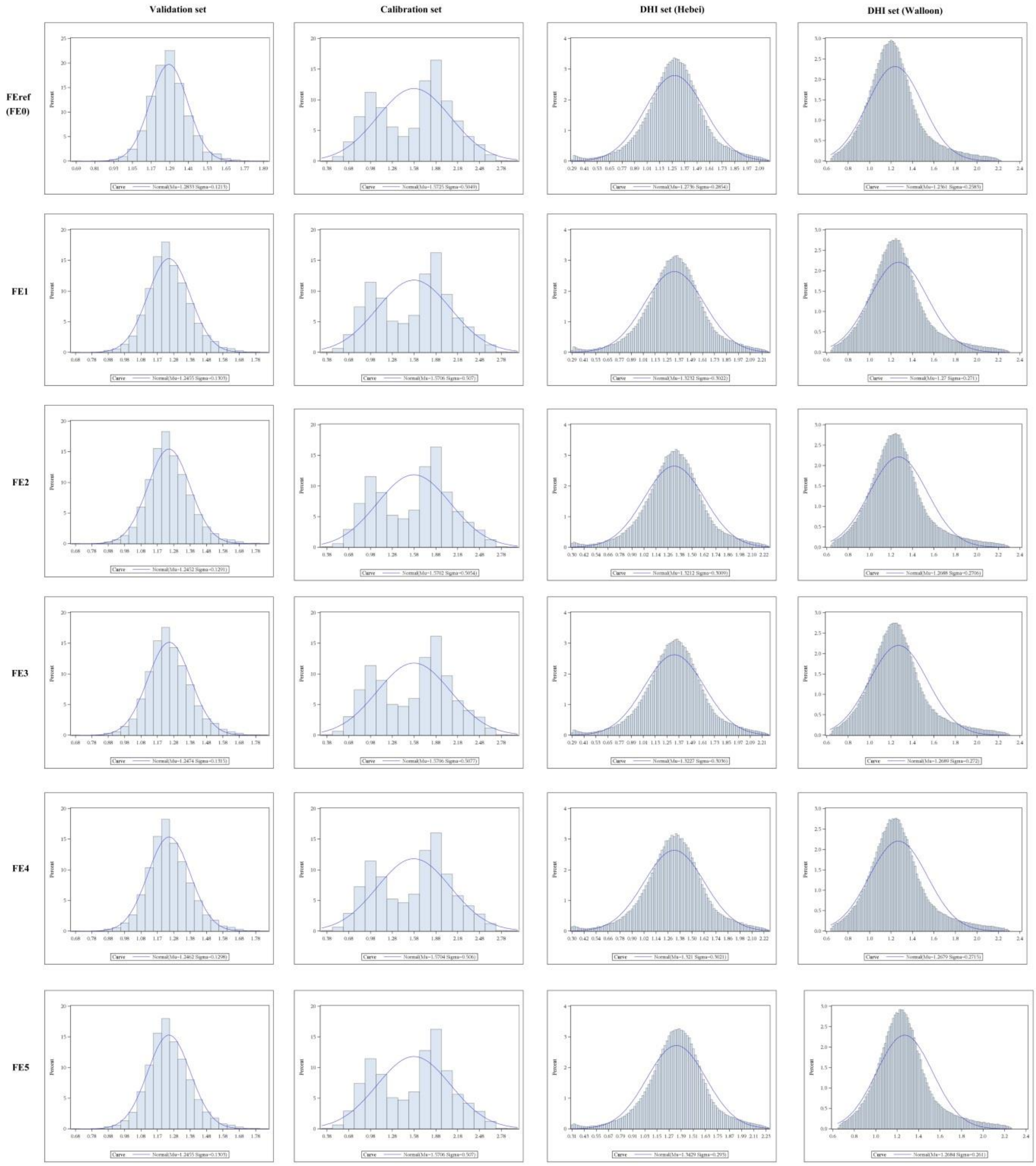


Figure 5-4. The distribution of assumed reference feed efficiency (FE) and predicted FE for calibration/test, validations sets and DHI data sets.

To explain better the FE evolution, it could be of interest to observe the FPCM evolution as this trait is the numerator used to calculate FE. As shown in Figure 5-6, the evolution of FPCM throughout the lactation showed a decreasing trend for both datasets except a flat trend observed for the first parity in the validation set (Figure 5-6 A). The reduction of FPCM degree was up to around 24% (from around 33kg/d to around 25 kg/d) and close to 35% (from around 43 kg/d to around 28 kg/d) for the validation and calibration sets, respectively. The evolution trends observed for FPCM were the ones expected. The milk yield has a peak around 45-90 days in milk (VandeHaar and St-Pierre, 2006). Similarly, the FPCM trends across DIM observed here were in agreement with the curve observed in Chapter 4 which was based on a larger Walloon dataset. In the case of the amount of milk production, the range observed was also within the recommended range formulated by ICAR. Moreover, it could be also useful to study the DMI evolution by DIM as it was the denominator of FE ratio. As presented in Figure 5-7, the predicted and assumed reference DMI also revealed an increased tendency across the lactation stage. In specific, the increase of DMI was not dramatic for the validation set, the maximum increase was around 20% (from around 15 kg/d to around 19 kg/d); the increase was up to 50% (from around 14 kg/d to around 21 kg/d) for the calibration/test sets. Therefore, it is reasonable to observe a decreasing trend of the predicted or assumed reference FE as the formula used to calculate the FE was constituted by the FPCM as numerator and the DMI as denominator. The DMI evolution trends at the first 5 classes of DIM were inconsistent with the findings reported by van Knegsel et al.(2007b). These authors found that the DMI of Holstein-Friesian cows presented an increased trend calving to around 60 days in milk. In another study, based on 224,174 records from Holstein-Friesian cows coming from 9 countries, the DMI peak was reported at 50 to 60 days in milk and then hold steady or slightly decrease until the late lactation (around 300 DIM) (Berry et al., 2014). Also, from the same study, the DMI of cows varied between 5 and 23 kg/d which is in agreement with the results observed here. Regarding the trend of the DMI evolution, the current study was inconsistent with the evolution mentioned by their study except for the first lactation of the calibration/test set.

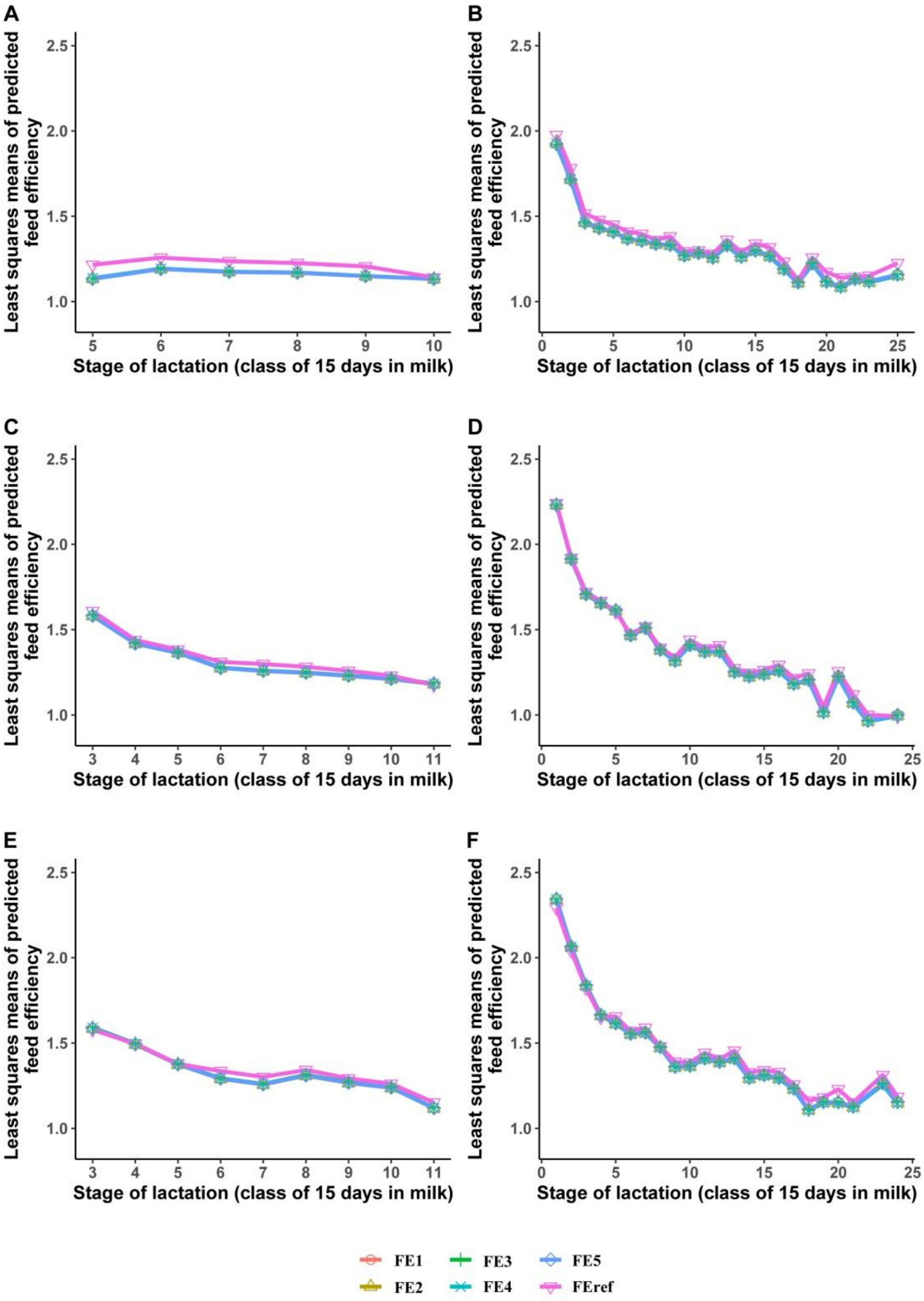


Figure 5-5. The least squares means evolution of the predicted feed efficiency within lactation. A, C, and E represent the first, second and third parity for the validation set (Australia dataset), respectively; B, D, and F are the first, second and third parity for the calibration set, respectively. FE1 to FE5 are the corresponding feed efficiency predicted using bodyweight records predicted from Mod1 to Mod5; FEref is the feed efficiency calculated using the observed body weight.

The reason could be due to the cows at first lactation were not mature and still growing even during the late lactation time (Dado and Allen, 1994). Another reason could be due to the lack of data as the current data is not representative of the dairy cow population.

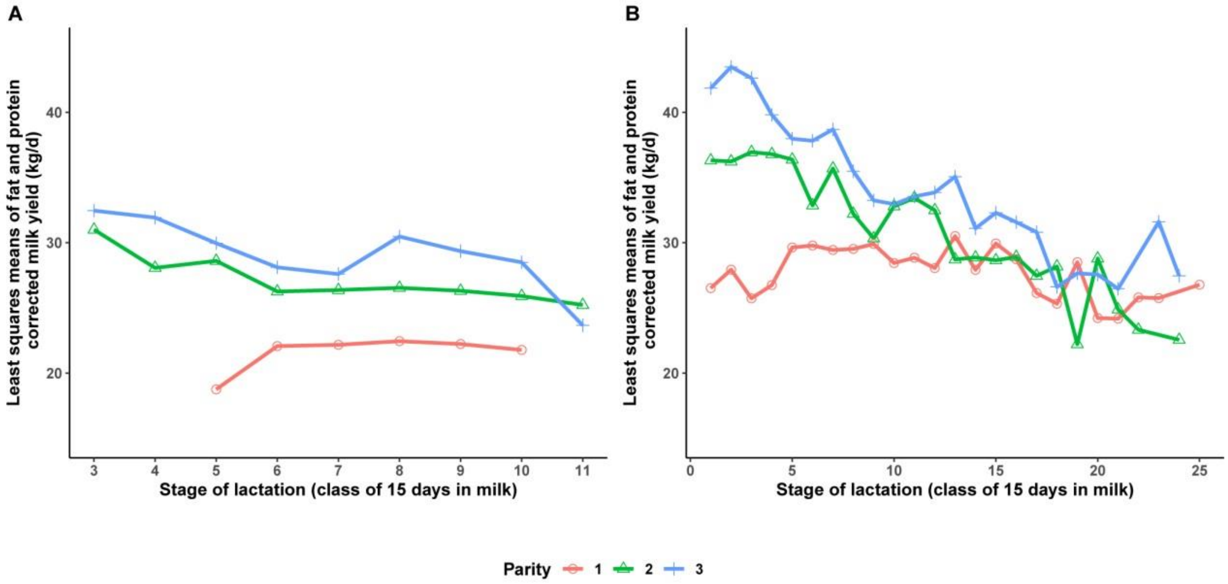


Figure 5-6. The least squares means evolution of the fat- and protein- corrected milk production within and between lactation across the lactation stage. A represents the validation set (Australia dataset); B respects the calibration/test set.

So, based on the evolution of DMI and FPCM, it is expected to observe a FE lactation trend. The comparison of FE results obtained in this study with the literature must be done with caution as there are different ratios to assess the FE: the ratio between milk yield to DMI, the ratio of the energy corrected milk to DMI, the ratio of FPCM to DMI (Chapter 1). Before comparing the lactation trend, it is interesting to observe that the descriptive statistics of FE were presented a larger variability compared to the literature. Indeed, some articles reported the following ranges of FE (i.e., FPCM/DMI): between 1.16 and 1.41 kg of milk/kg DM of the mean from 21 Holstein cows belong to different classes (Zucali et al., 2016); between 1.15 and 1.35 kg of milk/kg DM of the mean for Holstein cows consuming diets with different protein level at 224±54 days in milk (Barros et al., 2017); and between 1.44 and 1.62 kg of milk/kg DM for Holstein cows at 61±23 days in milk fed under different diet conditions (Guyader et al., 2016). But, there was another report obtained a FE (FPCM/DMI) for the Holstein cows up to 1.75 on average at the 87±36 DIM (Aguerre et al., 2016), which was closer to the results revealed in the current study in the validation data sets with reference BW values. Moreover, from the ratio of milk yield to DMI, St-Pierre.(2008) reported that the highest FE occurred at the early stage of lactation, and then the FE values decreased monotonically until the late lactation. This finding is in agreement with the trend observed in the current study. Most recently, the FE (FPCM/DMI) was reported up to 2.7 kg of milk/kg DM at the start of the lactation and gradually went to around 1.4 kg of milk/kg DM and presented a slight decrease tendency until the mid-lactation (DIM around 150 days) for 40 first parity Holstein cows (Seymour et al., 2020). This FE evolution trend and the values were in agreement with the current study too.

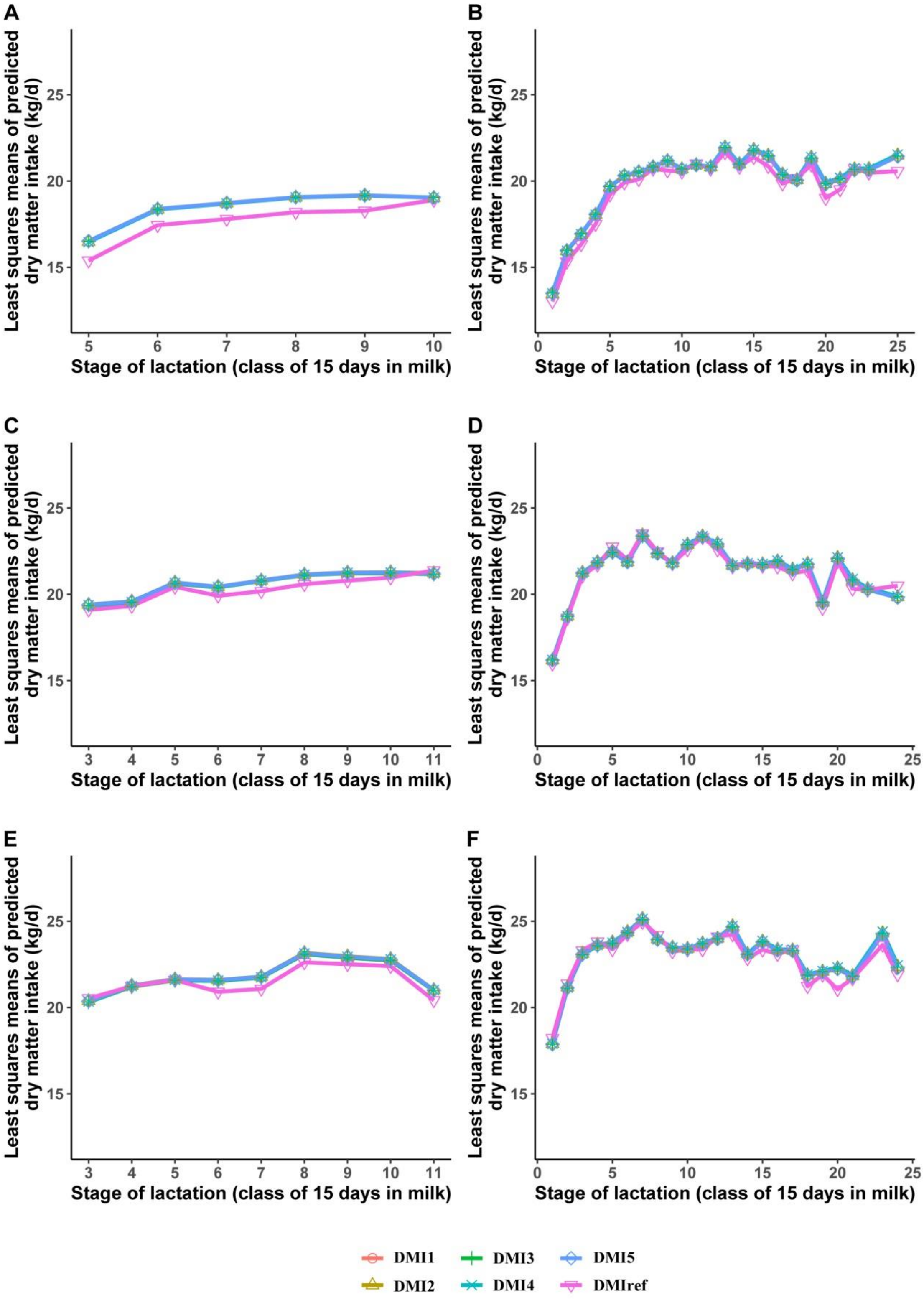


Figure 5-7. The least squares means evolution of the predicted dry matter intake (DMI) within lactation by the lactation stage. A, C, and E represent the first, second and third parity for the validation set (Australia dataset), respectively; B, D, and F are the first, second and third parity for the calibration set, respectively. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMIref is the dry matter intake calculated using the observed body weight.

However, there is no need to deeper explain the results obtained using the calibration and validation sets as those sets were not representative of a typical cow population. Therefore, there is more interest to observe the lactation LSMEANS curves based on DHI datasets.

4.2 Prediction of feed efficiency from dairy herd improvement datasets

To assess the interest of using FE, we have also studied its relationships with other FE related traits like DMI and methane. The quantity of methane was predicted using the equations developed by Vanlierde et al.(2019) from milk yield, parity, DIM, breed, and milk MIR spectrum. To avoid the presence of outliers, the Hebei and Walloon data sets were cleaned before the LSMEANS calculation by deleting the 1% upper and lower extreme values of the predicted FE, DMI, and methane. The thresholds used for Hebei dataset were 11 and 34 kg/d for DMI, 0.30 and 2.30 kg of milk/kg DM for FE, and 170 and 675 g/d for methane. The thresholds used to clean the Walloon dataset were 12.50 kg/d and 29 kg/d for DMI, 0.60 and 2.30 for FE, and 158 and 648 g/d for methane. The descriptive statistics of these datasets are given in Table 5-7. According to the literature, the possible range for DMI estimated from Holstein cows is from 5 to 43.28 kg/d (Gulay et al., 2003, Phuong et al., 2013, Berry et al., 2014); from 0.90 to 2.7 kg of milk/kg DM for FE estimated using the ratio FPCM/DMI (Donoghue et al., 2016, Zucali et al., 2016, Seymour et al., 2020); from 267 g/d (Danielsson et al., 2012) to 648 g/d (Aguerre et al., 2011) for the methane emissions. In conclusion, the thresholds we used in this cleaning step was inconsistent with the literature. Moreover, the average MIR predicted methane yield expressed as grams of methane by unit of milk yield was 16 g/kg of milk for both Hebei and Walloon data sets. This was similar to the value of 16.4 g/kg of FPCM reported by van Gastelen et al.(2017) observed from 29 Holstein dairy cows. However, the current study had a larger SD (5 g/kg of milk v.s. 1.87 g/kg of milk) which might be mainly due to the larger data size used in this study. The average MIR predicted methane yield expressed as unit of predicted DMI was of 19 g/kg of predicted DMI for both Hebei and Walloon data sets. This value was in agreement with the ones reported in literature. For instance, Charmley et al.(2016) reported in average a value of 21 ± 0.45 g of CH_4/kg of DMI from dairy cows belonging to several breeds. van Gastelen et al.(2017) observed a methane production of 24.1 ± 1.87 g/kg of DMI for Holstein dairy cows. As the datasets used in the current study were large, the bigger range of methane production in the current study might attribute to the diet variation as the enteric fermentation was impacted by the diet considerably (Olijhoek et al., 2018). However, it is also important to consider that our results coming from a prediction and cannot be considered as reference values.

Although the predicted BW varied differently among equations, as we addressed previously, the prediction of FE was more robust to a moderate change of BW. This was observable based on the comparison of the descriptive statistics observed for FE using the different BW equations (Table 5-7). The distribution of the predicted FE values for the DHI data sets are given in Figure 5-4. The predicted FE from different models presented a normal distribution for both data sets. The LSMEANS trends within and between lactations for the predicted FE are depicted in Figure 5-8. Generally, the evolution of FE from both the Hebei and Walloon DHI datasets presented a decreased trend following the DIM. The FPCM yield decreased after its peak located at around 50th DIM from both DHI datasets (Figure 5-10 A, B). The predicted DMI had a peak at around 90 to 100th DIM from the two studied dairy cow populations (Figure 5-11). As the FE was the ratio of FPCM to DMI, there was logical to observe the lower values of FE at the period around 60 to 100 days DIM. This lactation trend is similar to the ones observed in the calibration and validation sets (Figure 5-5) which was in agreement with the literature (St-Pierre, 2008, Seymour et al., 2020). These FE values are similar to the ones predicted in Chapter 4 (Table 4-1, IC=1/FE). This is expected as both studies shared identical records. Moreover, the high and low FE values observed in Table 5-7 were also within the published range 0.90 - 2.70 (Zucali et al., 2016, Barros et al., 2017, Seymour et al., 2020). However, the FE range observed in this section was lower and the trend was smoother than the ones observed in the data sets with BW reference values. The reason might be due to the larger data size allowing a better more representability.

Besides the evolution trends within and between lactations, the FE varied also in function of the month of test. The trends of predicted FE LSMEANS per month of test were similar from both Hebei and Walloon datasets (Figure 5-9). The predicted FE decreased during the summer time as also observed for the predicted BW (Figure 5-3).

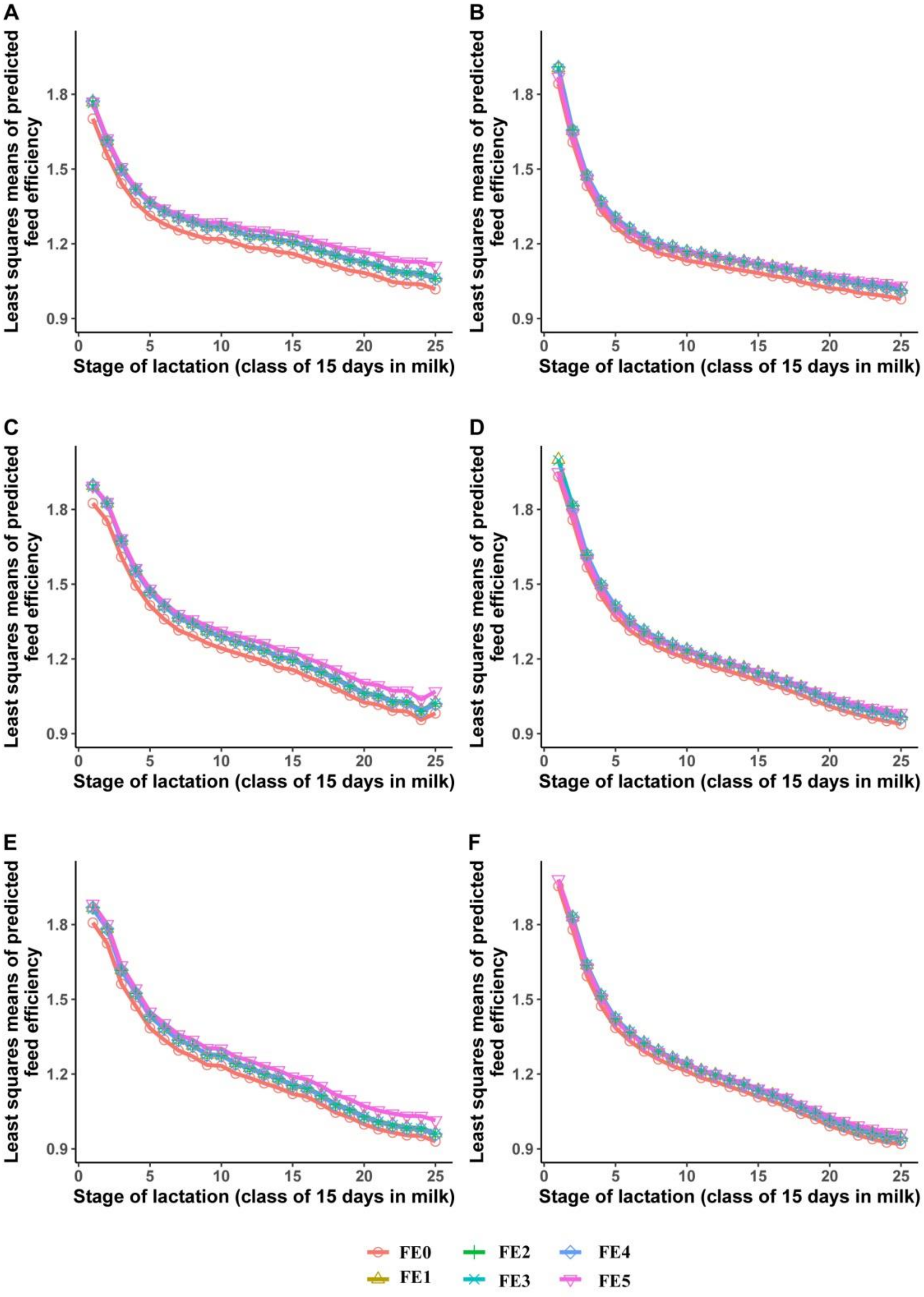


Figure 5-8. Evolution of least squares means of feed efficiency derived from the body weight predicted using the models Mod0 to Mod5 within and between lactations; A, C, E and B, D, F denote the first, second, and third parity of cows from Hebei and Walloon dataset, respectively.

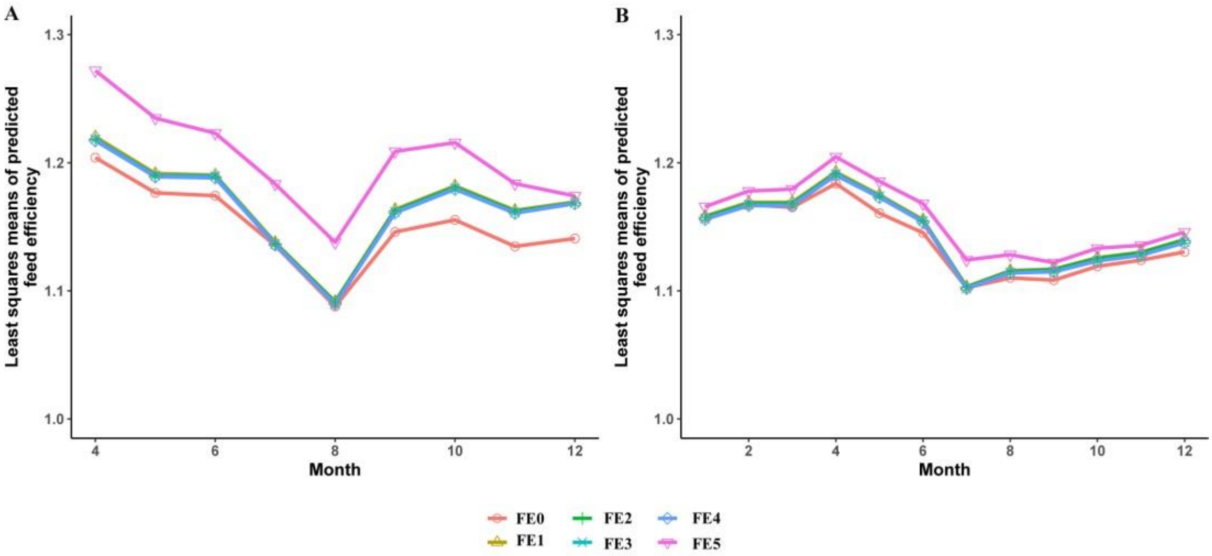


Figure 5-9. Evolution of least squares means of feed efficiency derived from the body weight predicted using the models (Mod0 to Mod5) across test month; A and B represent Hebei and Walloon datasets, respectively.

This was expected due to the changes of diet occurring at this period for most of the Walloon cows. This change impacted negatively both FPCM (Figure 5-10 D) and DMI (Figure 5-11) explaining the low FE values. For the Hebei (Figure 5-10 C) and Walloon data sets, the most possible reason could be related to the heat stress (Ranjitkar et al., 2020). Additionally, for the Walloon set, the drop might also be related to the more activity of the animals in the same pasture time.

In order to confirm the relevancy of the proposed FE prediction tool, we have also decided to calculate the correlation coefficients between FE predictions and predicted BW, DMI, and milk yield (MY) in order to compare them with the literature. The results are given in Table 5-8. The correlation between milk yield with FE ranged between 0.80 and 0.83 from both databases. This is inconsistent with the values (between 0.77 and 0.80) observed in Chapter 4 (FE expressed as IC=1/FE). Using FE expressed as ratio of FCM to DMI, (Hooven et al., 1968) reported slightly lower values which varied from 0.62 to 0.70. This high correlation was expected as the milk yield is used to calculate FPCM. Phuong et al.(2013) mentioned that the FPCM production was a significant factor to affect FE.

The correlation values of FE with the predicted BW values were slightly negative for BW predicted from Mod1-Mod4 (around −0.15) but were positive with Mod5 and Mod0 (the average was 0.16 for Hebei data set and 0.01 for Walloon data set). The correlation values for Mod5 and Mod1 were similar to the values observed in Chapter 4 (coefficient values were 0.16 and 0.08 between BW and IC for Holstein cows of first and second parity.) These low correlation values are related to the quadratic relationship existing between BW and FE (Hooven et al., 1968, Badinga et al., 1985).

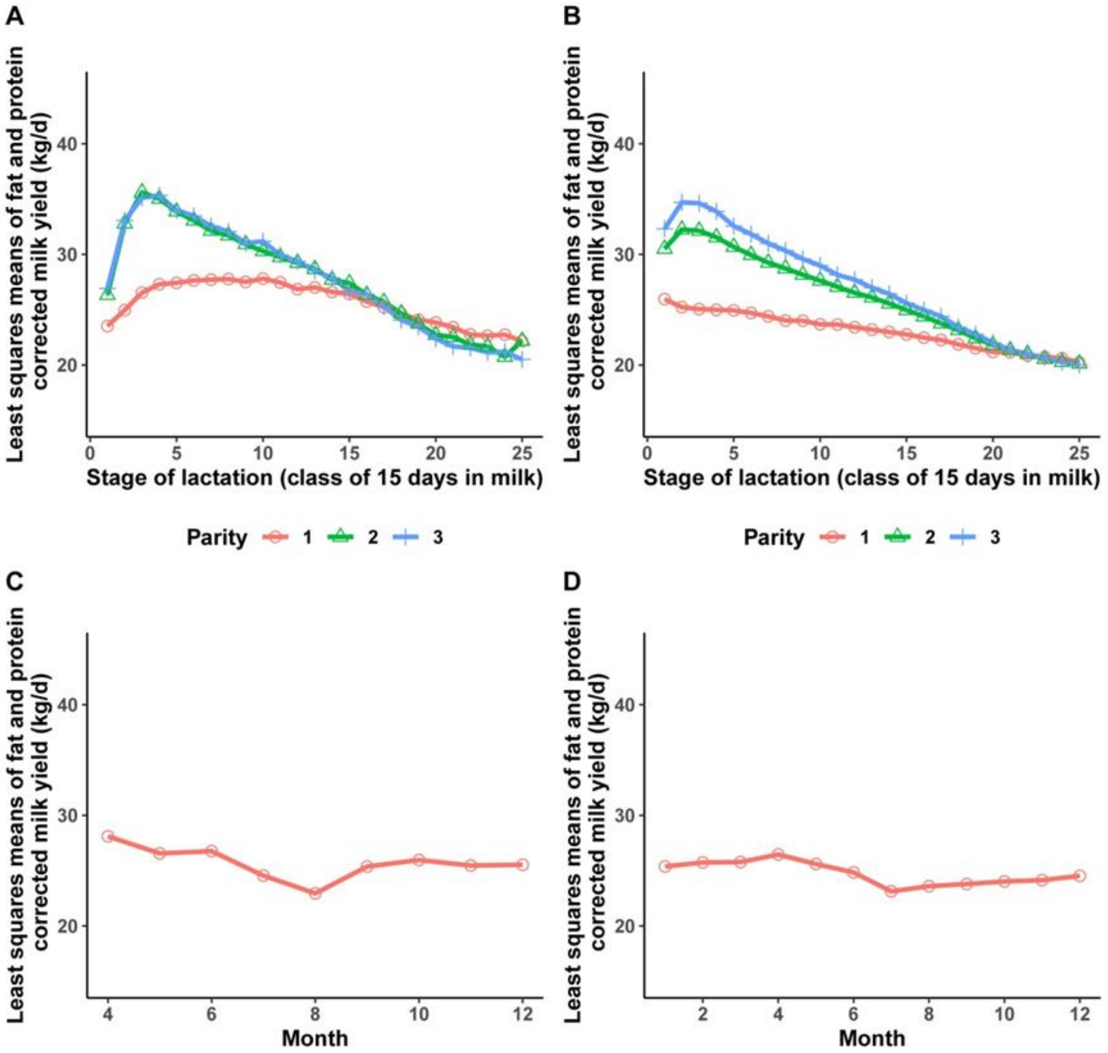


Figure 5-10. The least squares means evolution of the fat- and protein- corrected milk production (FPCM) within and between lactation across the lactation stage (A, B) as well as the least squares means evolution of FPCM by month of test (C, D). A and C represent the Hebei set; B and D respect the Walloon set.

The predicted DMI and FE had a positive relationship as the correlation coefficients were around 0.50 and around 0.20 from the Hebei and Walloon data set, respectively. From the literature, this correlation between DMI and FE differs a lot. For instance, there was a negative correlation ($r=-0.32$) between DMI and FE expressed as ratio of milk yield to DMI but this value was not significant ($p > 0.05$) (Britt et al., 2003). A negative correlation ($r=-0.51$) was also reported between DMI and FE expressed as ECM/DMI (Ben Meir et al., 2018). However, according to the result from Barros et al.(2017), there was positive ($r=0.53$) correlation between DMI and FE expressed as FPCM/DMI estimated for Holstein cows. This is in agreement with our findings and the results mentioned in Chapter 4. This could be due to the fact that the cows tend to partition the nutrients to the maintenance requirement first and then after a certain limit to the milk production. A linear relation was found between FE and DMI) (Figure 5-13).

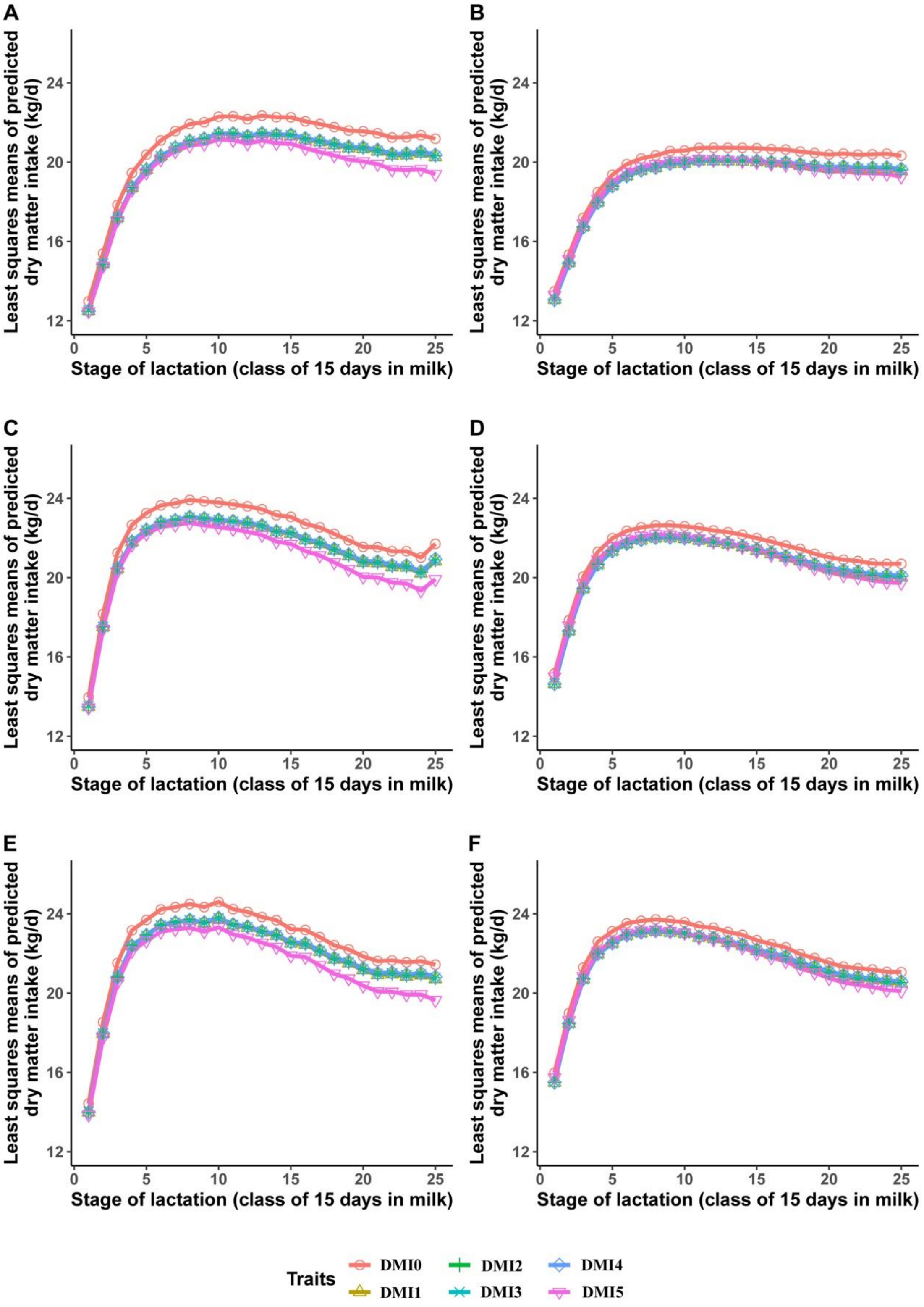


Figure 5-11. The least squares means evolution of the predicted dry matter intake (DMI) within lactation by the lactation stage. A, C, and E represent the first, second and third parity for the Hebei set, respectively; B, D, and F are the first, second and third parity for the Walloon set, respectively. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMI0 is the dry matter intake calculated using the predicted body weight of Mod0.

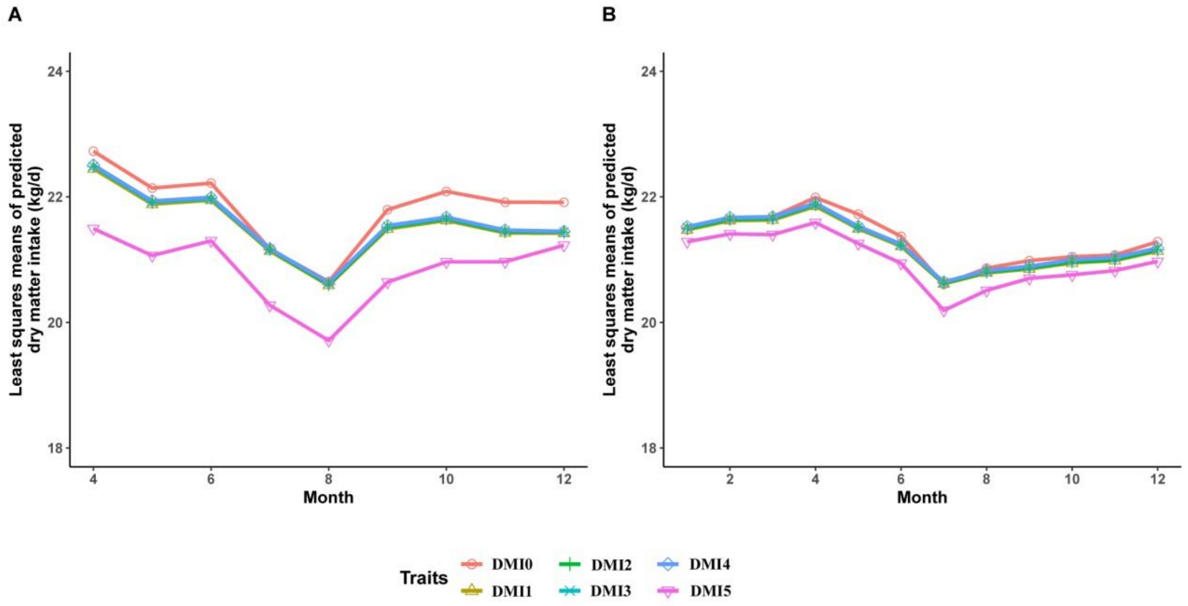


Figure 5-12. The least squares means evolution of predicted dry matter intake by month of test. A represents the Hebei set; B respects the Walloon set. DMI1 to DMI5 are the corresponding dry matter intake predicted using bodyweight records predicted from Mod1 to Mod5; DMI0 is the dry matter intake calculated using the predicted body weight of Mod0.

Table 5-8 The Pearson correlation coefficients of the feed efficiency related traits ($p < 0.001$).

Data set	**Traits**	**BW0**	**BW1**	**BW2**	**BW3**	**BW4**	**BW5**	**Milk yield (kg/d)**	**Methane yield (g/d)**	**Methane yield (g/kg MY)**	**Methane yield (g/kg DMI)**
Hebei	**FE0**	0.13	-0.16	-0.12	-0.2	-0.15	0.13	0.81	0.23	-0.64	-0.19
	FE1	0.18	-0.17	-0.13	-0.21	-0.16	0.15	0.82	0.23	-0.64	-0.20
	FE2	0.18	-0.17	-0.14	-0.21	-0.17	0.15	0.82	0.23	-0.64	-0.19
	FE3	0.18	-0.17	-0.13	-0.21	-0.17	0.15	0.82	0.23	-0.64	-0.20
	FE4	0.18	-0.18	-0.14	-0.21	-0.17	0.15	0.82	0.23	-0.64	-0.19
	FE5	0.18	-0.14	-0.1	-0.18	-0.13	0.08	0.83	0.22	-0.66	-0.21
Walloon	**FE0**	0.01	-0.19	-0.17	-0.21	-0.2	0.03	0.80	0.06	-0.69	-0.06
	FE1	0.02	-0.22	-0.2	-0.24	-0.23	0.01	0.80	0.04	-0.70	-0.08
	FE2	0.01	-0.22	-0.21	-0.24	-0.23	0.01	0.80	0.04	-0.69	-0.08
	FE3	0.02	-0.22	-0.2	-0.24	-0.23	0.01	0.80	0.04	-0.70	-0.08
	FE4	0.01	-0.22	-0.21	-0.24	-0.23	0.01	0.80	0.04	-0.69	-0.08
	FE5	0.02	-0.19	-0.18	-0.22	-0.2	0	0.81	0.03	-0.71	-0.10

The number 0 denotes the values predicted from the model proposed by Soyeurt et al.(2019) and 1 to 5 are the models (Mod1 to Mod5) proposed in the current study; Correlation values between FE and methane yield (g/kg DMI) were averaged among different models.

In conclusion, the observed lactation and month trends as well as the correlation values obtained with milk yield, BW, and DMI were in agreement with the literature confirming the interest of the developed FE prediction tool. As it requires only information based on the animal characteristics and milk MIR spectrum, this tool had promising usefulness to allow farmers to follow the FE routinely to order to optimize their milk production and to check the health status of their dairy cows.

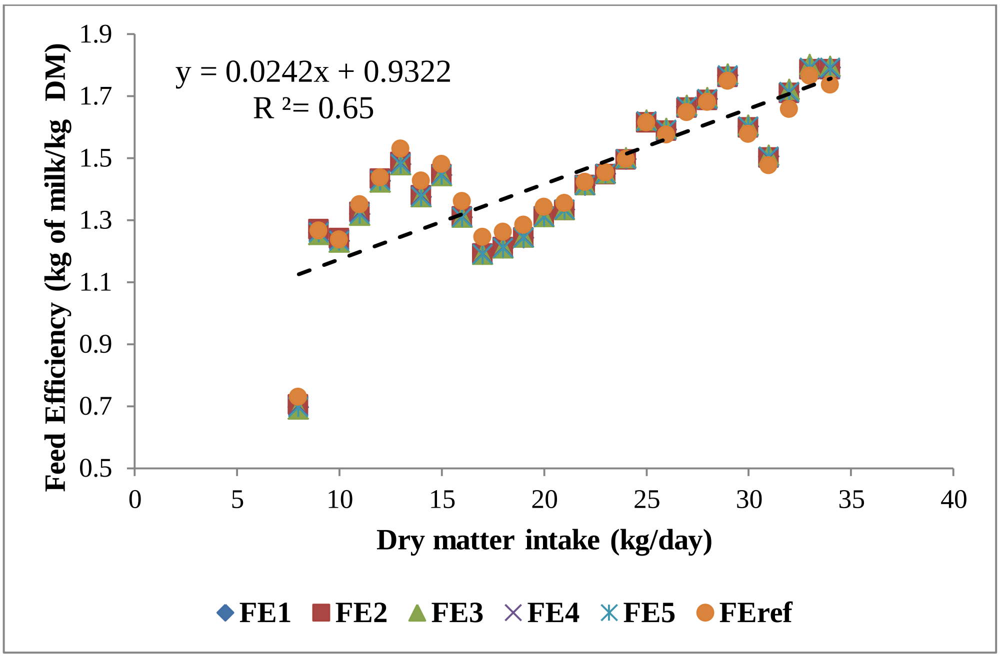


Figure 5-13. The FE evolution by dry matter intake from the calibration/test and validation sets.

Moreover, as we have mentioned in Chapter 1 that there is a relationship between the feed efficiency and the GHG emissions, the correlation values between the predicted FE and the predicted MIR methane yield (g/day) were calculated. A slightly positive relationship was observed when the methane yield was expressed as grams per day (Table 5-8). The correlation values varied between 0.03 and 0.20 for the Heibei and Walloon DHI datasets, respectively. To our knowledge, there is no available article already published mentioning this correlation. However, a moderate positive correlation (r=0.47) between daily methane production measured using respiration chambers and DMI was reported by Blore et al.(2020). Positive correlations of 0.18 and 0.33 were also found in this study using predicted MIR methane and predicted DMI from Hebei and Walloon datasets, respectively. Due to the positive correlations found in this study between DMI and FE, it is reasonable to observe a positive correlation between FE and methane yield. Moreover, the counterbalance between DMI and milk production might also result in a weaker correlation between FE and methane yield (Lassen et al., 2012). This explained the weak relationship between methane yield expressed as unit of DMI found in the current study (Table 5-8, expressed as average values among different models). The milk production was also reported to be positively related to the methane output; r^2 varied from 0.50 to 0.63 (Yan et al., 2010). Also, from the same study, the efficiency of utilization of metabolic energy for lactation was negatively linked to the methane energy output and demonstrated that the high productive cows would produce less methane. The variation of relationship between FE and methane yield expressed on different basis revealed the relevance to use the methane density (e.g., g/kg milk yield; g/kg DMI) instead of the methane yield (e.g., g/day) as the production reflects better the economic dimension or the cow FE (Vanlierde, 2019).

Conclusions and perspectives

To ensure the sustainability of dairy farming, knowing the feed efficiency of the cow on a routine basis is essential. So, the global aim of this thesis was to develop a FE prediction tool available at individual and large scales. To achieve this objective, it was of interest to use the data routinely collected by DHI organizations. So, this explains why the current thesis used the ratio of FPCM to DMI to assess the FE of a cow at a specific test-day.

Unfortunately, due to the cost and labor constraints, it is still difficult to collect the DMI records routinely in all commercial herds. So, in this thesis, we have tried to improve the robustness and the accuracy of a past model predicting the test-day BW of dairy cows from milk MIR spectrum and animal characteristics in order to estimate test-day DMI from NRC equation. Even if the prediction performances on calibration and validation sets did not seem to be improved, the extension of the calibration set from 721 to 1,658 records allowed to solve the issue of an observed over-estimation of BW within lactation. Moreover, we had formulated the hypothesis in this work that the use of a selection feature algorithm during the calibration process could improve the model robustness. However, no strong improvement was observed for the reduced models based on the calibration and validation sets. Indeed, the selection of features leads to keep only 2 spectral points in addition to the milk yield, parity and DIM (Mod1-Mod4) limiting the ability of the model to fit well the cow variability within lactation and per test month. There was a lack of BW loss at the early stage of lactation or during the grazing period for the models built using the feature selection. Therefore, the consideration of a higher number of spectral records as done in the past research and for Mod5 was more relevant. This could be related to the fact that the 277 spectral points used were, in a certain way, already selected as they were chosen based on the experience gained through the development of other MIR equations. However, the low number of records in the late lactation might result in an under-fitting of the model and lead to a problematic prediction of BW for the cows in late lactation. Moreover, as consecutive MIR points are known to be highly correlated, the bandwidth averaging around the two obtained final 2 MIR points could help to improve the model performance as well as the obtaining of expected BW evolution shape across DIM. This could be further tested. In conclusion, the results obtained in this thesis confirmed the possibility to predict a BW record with an average accuracy of around 50-60 kg using traits routinely recorded by DHI. This accuracy can be considered as good in regard with the within day variability of BW, as we know the other methods like 3D-carema also having an accuracy of around 43 kg (i.e., considering the cost and accuracy) (Chapter 2). However, even if some differences in the lactation or monthly patterns between BW models, the FE records estimated from those BW predictions were less sensible to a moderate change of BW values. Indeed, correlation coefficients greater than 0.99 were observed between FE records estimated from BW predicted using the different developed equations from both DHI datasets. The accuracy of FE estimated using the NRC equation and predicted BW was comprised between 0.05 and 0.06, which can be also considered as good as this represents less than 5% of the variation. Furthermore, the descriptive statistics as well as lactation and monthly tendencies observed for predicted FE were in agreement with the literature.

Predicting FE or BW indicators at individual and large scales is therefore currently feasible. However, independently of the inherent model accuracies, there is still a risk when applying those models to large datasets to obtain erroneous predictions. The reasons can be related to the use of poor quality of spectral data or the realization of a prediction from a spectral extrapolation. So, as it was not possible to implement the modalities of the ring test for BW prediction process as this trait cannot be measured using chemical milk analysis, we have tested 3 statistical methods easy to use in the DHI framework and their combinations to solve this issue. From the obtained findings, the combination of the method deleting the 1% of extreme low and high predicted values (M1) with the one deleting records if their GH distance with the BW calibration set is higher than 5 (M2) allowed to improve the RMSD without deleting a large amount of data. However, if the pursued aim is to have the lowest RMSD (i.e., the highest gain), the combination of M2 or another method deleting records if the absolute fat difference for the considered sample is higher than 0.30 g/dL of milk (M3) must be preferred. This cleaning procedure must be applied by the DHI organizations if they want to use BW and, therefore, FE predictions for practical applications.

In conclusion, the work done in this thesis reached the pursued objective as a FE prediction tool was developed and is now available for DHI organizations. Moreover, the work done in this thesis proves the feasibility to apply the developed models on real DHI databases and the relevancy of the obtained predictions. However, work is never finished and some further investigations can be proposed.

First, the estimation of FE used the equation developed by NRC to predict DMI. So, a reasonable questioning concerns the potentiality to develop directly a MIR based DMI equation. Indeed, some works are already published on this topic (Lahart et al., 2019). However, through the current research context, it is difficult to create a representative DMI calibration set. Indeed, the equipment is mainly available on research farms in which the animal genetics and/or the feeding or management practices are not representative of the dairy cow population. This lack of variability could have a detrimental effect on the model accuracy (Grelet et al., 2020). The development of an equation predicting BW did not present this inconvenience as the measurement method can be easily implemented in different farms. However, this does not mean that there is no interest to develop a DMI equation. This highlights only the need to have a large acquisition of DMI data. This could be reached by creating an international collaboration in order to share DMI records, similarly to the work done currently for the development of MIR methane equation. If a such calibration is available, a study could be conducted to compare the approach using NRC equation and predicted BW vs. a direct prediction of DMI.

Second, even if the current work improved the prediction performance of BW equation by adding new samples in the calibration set, the BW variability could be still better considered. Indeed, from the Australian and Hebei datasets, some samples to predict had a very high GH distance suggesting the interest to still increase the calibration set. However, a part of this variability could be also related to the fact that the spectral data of those samples were provided by a Bentley spectrometer. Indeed, a standardization procedure was applied on the spectral to convert them into a Foss basis, which is the spectral format used to build the equation. Unfortunately, this standardization was not operate using the most appropriate way. Indeed, it is better to have a short period between the estimation of the standardization coefficients and the spectral acquisition in order to reflect as close as possible the current existing spectral variability. So, if we want to add new samples in the calibration set, it must be important to standardize the spectral data in the best condition.

Third, it was shown in this thesis that Mod5 and Mod0 had a different behavior for the spectral data coming from the Bentley spectrometer. We have made the hypothesis that this was related to the standardization as mentioned previously. Grelet et al.(2020) recommended using the 212 spectra having the highest reproducibility between instruments. In this study, we have used 277 spectral points. Therefore, a potential improvement could consist of reducing a little bit the number of spectral points with the hope that this reduction will not be linked to a lack of model fit within the lactation or the year as we have observed for Mod1-Mod4. Additionally, the in time standardization could also benefit the accuracy of the predictive models.

Fourth, although the current study highlighted the feasibility to predict routinely the test-day FE at cow level using phenotypes such as milk yield, parity, days in milk, as well as the MIR spectral points, there exists also other variables that could be informative to improve the BW or FE predictive models. For instance, the diet information, the ruminal parameters, the reproductive status (Poncheki et al., 2015) are some of them. Age was reported as a key factor affected the BW of cows (Miller and Hooven, 1970) as well as the weather condition (Hooven et al., 1968). Indeed, the BW gain is lower during the summer time (Schultze and Davis, 1961). Temperature humidity index (THI) could affect the body temperature of cows and, thereby, the heat stress status and the maintenance of cows (Dikmen and Hansen, 2009). Accordingly, THI could affect the body maintenance status of cows, and thereby could be the latent predictor of energy balance (e.g., BW changes) of the dairy cows as well (Berman, 2005). It is interesting to note that, at the beginning of this study, the month of test was attempted to be adopted into the BW predictive models, however, the performance of the models was not improved (data was not shown) and it was eliminated. This lack of improvement might be due to the mix of different datasets obtained from several countries which lead to a potential diversity of weather conditions even if the records were collected at the same test month. Days open was also reported as another significant ($P < 0.01$) source of variation in BW change based on 318 Holstein cows (Miller and Hooven, 1970). The herd effect was also reported explaining around 30% of the variation in BW and also could be a latent predictor of BW or FE of dairy cows. The possible reason for that might be due to the difference in diet and management among herds (Hietanen and Ojala, 1995). Indeed, although results obtained by Veerkamp et al.(1995) showed no significant effect of diet on the BW or BCS during the first 26 weeks of lactation, results from other researches revealed the BW or

BCS differed significantly across different feeding systems (Berry et al., 2006). Similarly, the impact of the diet concentrate level on the dairy cow FE was highlighted. For example, using the definition of FE equal to FCM divided by the net energy consumed, the cows consumed all-forage ration were 1.5 times more efficient than the cows given forage plus concentrate diet even if the milk production only accounts for 83% of the later cows (Richardson et al., 1971). The fat percentage change at the early lactation was also an indicator of the negative energy balance (de Vries and Veerkamp, 2000) which was related to BW. So, fat content change could be a potential predictor of BW. However, this information is predicted accurately using the milk MIR spectral data. Therefore, this variable would be redundant if we consider the spectrum as features in the BW or FE equations. In addition, the age within parity, calving year/season, and breed are factors contributing to the changes of BCS (Bastin and Gengler, 2013). Additionally, the time of feed supply to animals and the excretion of feces and urine could influence the daily variation of BW (Cho et al., 2020). Consequently, all of those variables, mentioned above, could be considered in further studies to improve the BW or FE modelling. Based on the consideration of these additional factors, the profile of BW or FE could be more accurate. Unfortunately, many of those features are not recorded routinely. Finally, in the view of improving reference BW accuracy, defining a protocol could be of interest when measuring the BW of cows to determine whether the variation from feces and urine should be considered.

Another issue concerns the thresholds used to clean the dataset especially for method 3 which fixed the limitation of the absolute fat residual difference at 0.30 g/dL of milk. This threshold could be changed in function of the dataset used by using the methodology proposed in this thesis. Moreover, the typical trends of BW across the DIM and parity are quadratic. Due to the limited number of records for several DIM and the lack of data regarding the late parities, the parity and DIM were considered as linear variables in the current study. This might lead to a potential bias of the BW prediction. These typical trends across DIM and parity must be considered in a further study.

Lastly, a reflection must be started about the communication of the FE predictions to the farmers. (Jud Heinrichs and Ishler, 2016) mentioned that a FE value lower than 1.3 kg of milk/kg DM should be problematic. However, their calculation was based on the ECM/DMI ratio. The relevancy of this threshold must be discussed with nutritionists and veterinarians to confirm its potential. Moreover, due to the high throughput of the proposed FE prediction methodology, repeated measurements can be easily obtained for many cows allowing the possibility to compare herds between them and also to investigate the interest of considering this trait in a selection program and for the development of genomic equation.

References

Aguerre, M. J., M. C. Capozzolo, P. Lencioni, C. Cabral, and M. A. Wattiaux. 2016. Effect of quebracho-chestnut tannin extracts at 2 dietary crude protein levels on performance, rumen fermentation, and nitrogen partitioning in dairy cows. Journal of dairy science 99(6):4476-4486.

Aguerre, M. J., M. A. Wattiaux, J. M. Powell, G. A. Broderick, and C. Arndt. 2011. Effect of forage-to-concentrate ratio in dairy cow diets on emission of methane, carbon dioxide, and ammonia, lactation performance, and manure excretion. Journal of dairy science 94(6):3081-3093.

Badinga, L., R. J. Collier, C. J. Wilcox, and W. W. Thatcher. 1985. Interrelationships of Milk Yield, Body Weight, and Reproductive Performance1. Journal of dairy science 68(7):1828-1831.

Banos, G. and M. P. Coffey. 2012. Technical note: Prediction of liveweight from linear conformation traits in dairy cattle. Journal of dairy science 95(4):2170-2175.

Barros, T., M. A. Quaassdorff, M. J. Aguerre, J. J. O. Colmenero, S. J. Bertics, P. M. Crump, and M. A. Wattiaux. 2017. Effects of dietary crude protein concentration on late-lactation dairy cow performance and indicators of nitrogen utilization. Journal of dairy science 100(7):5434-5448.

Bastin, C. and N. Gengler. 2013. Genetics of body condition score as an indicator of dairy cattle fertility. A review. Biotechnologie, Agronomie, Société et Environnement 17(1):64-75.

Ben Meir, Y. A., M. Nikbachat, Y. Fortnik, S. Jacoby, H. Levit, G. Adin, M. Cohen Zinder, A. Shabtay, E. Gershon, M. Zachut, S. J. Mabjeesh, I. Halachmi, and J. Miron. 2018. Eating behavior, milk production, rumination, and digestibility characteristics of high- and low-efficiency lactating cows fed a low-roughage diet. Journal of dairy science 101(12):10973-10984.

Berman, A. 2005. Estimates of heat stress relief needs for Holstein dairy cows. Journal of animal science 83(6):1377-1384.

Berry, D. P., M. P. Coffey, J. E. Pryce, Y. de Haas, P. Løvendahl, N. Krattenmacher, J. J. Crowley, Z. Wang, D. Spurlock, K. Weigel, K. Macdonald, and R. F. Veerkamp. 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. Journal of dairy science 97(6):3894-3905.

Berry, D. P., R. F. Veerkamp, and P. Dillon. 2006. Phenotypic profiles for body weight, body condition score, energy intake, and energy balance across different parities and concentrate feeding levels. Livestock Science 104(1):1-12.

Blore, K., M. J. Bell, P. Garnsworthy, D. Mallis, R. Eckard, P. Moate, and T. Yan. 2020. Modified approach to estimating daily methane emissions of dairy cows by measuring filtered eructations during milking. Landbauforschung 69(1).

Britt, J. S., R. C. Thomas, N. C. Speer, and M. B. Hall. 2003. Efficiency of Converting Nutrient Dry Matter to Milk in Holstein Herds1. Journal of dairy science 86(11):3796-3801.

Camacho-Tamayo, J. H., Y. Rubiano S, and M. d. P. Hurtado S. 2014. Near-infrared (NIR) diffuse reflectance spectroscopy for the prediction of carbon and nitrogen in an Oxisol. Agronomía Colombiana 32:86-94.

Centner, V., D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, and C. Sterna. 1996. Elimination of uninformative variables for multivariate calibration. Analytical chemistry 68(21):3851-3858.

Charmley, E., S. R. O. Williams, P. J. Moate, R. S. Hegarty, R. M. Herd, V. H. Oddy, P. Reyenga, K. M. Staunton, A. Anderson, and M. C. Hannah. 2016. A universal equation to predict methane production of forage-fed cattle in Australia. Animal Production Science 56(3):169-180.

Cho, H., S. Jeon, M. Lee, K. Kang, H. Kang, E. Park, M. Kim, S. Hong, and S. Seo. 2020. Analysis of the Factors Influencing Body Weight Variation in Hanwoo Steers Using an Automated Weighing System. Animals 10(8):1270.

Dado, R. G. and M. S. Allen. 1994. Variation in and Relationships Among Feeding, Chewing, and Drinking Variables for Lactating Dairy Cows. Journal of dairy science 77(1):132-144.

Danielsson, R., A. Schnürer, V. Arthurson, and J. Bertilsson. 2012. Methanogenic Population and CH<sub>4</sub> Production in Swedish Dairy Cows Fed Different Levels of Forage. Applied and Environmental Microbiology 78(17):6172.

de Ondarza, M. B. and J. M. Tricarico. 2017. Review: Advantages and limitations of dairy efficiency measures and the effects of nutrition and feeding management interventions. The Professional Animal Scientist 33(4):393-400.

de Vries, M. J. and R. F. Veerkamp. 2000. Energy Balance of Dairy Cattle in Relation to Milk Production Variables and Fertility. Journal of dairy science 83(1):62-69.

Dikmen, S. and P. J. Hansen. 2009. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? Journal of dairy science 92(1):109-116.

Donoghue, K. A., T. Bird-Gardiner, P. F. Arthur, R. M. Herd, and R. F. Hegarty. 2016. Genetic and phenotypic variance and covariance components for methane emission and postweaning traits in Angus cattle12. Journal of Animal Science 94(4):1438-1445.

Fearn, T. 2002. Assessing Calibrations: SEP, RPD, RER and R2. NIR news 13(6):12-13.

Gengler, N., M. Hostens, and G. Consortium. 2018. How fast can we change resilience and efficiency through breeding and management ? Page 264 in Proc. The 69th Annual Meeting of the European Association for Animal Production, Dubrovnik, Croatia.

Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2020. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods.

Gulay, M. S., M. J. Hayen, K. C. Bachman, T. Belloso, M. Liboni, and H. H. Head. 2003. Milk Production and Feed Intake of Holstein Cows Given Short (30-d) or Normal (60-d) Dry Periods. Journal of dairy science 86(6):2030-2038.

Guyader, J., M. Doreau, D. P. Morgavi, C. Gérard, C. Loncke, and C. Martin. 2016. Long-term effect of linseed plus nitrate fed to dairy cows on enteric methane emission and nitrate and nitrite residuals in milk. Animal : an international journal of animal bioscience 10(7):1173-1181.

Haile-Mariam, M., O. Gonzalez-Recio, and J. E. Pryce. 2014. Prediction of liveweight of cows from type traits and its relationship with production and fitness traits. Journal of dairy science 97(5):3173-3189.

Hietanen, H. and M. Ojala. 1995. Factors Affecting Body Weight and Its Association with Milk Production Traits in Finnish Ayrshire and Friesian Cows. Acta Agriculturae Scandinavica, Section A — Animal Science 45(1):17-25.

Hooven, N. W., R. H. Miller, and R. D. Plowman. 1968. Genetic and Environmental Relationships Among Efficiency, Yield, Consumption and Weight of Holstein Cows. Journal of dairy science 51(9):1409-1419.

Jud Heinrichs and V. A. Ishler. 2016. Feed Efficiency in Lactating Cows and Relationship to Income Over Feed Costs. Vol. 2020.

Khaire, U. M. and R. Dhanalakshmi. 2019. Stability of feature selection algorithm: A review. Journal of King Saud University - Computer and Information Sciences.

Korver, S. 1988. Genetic aspects of feed intake and feed efficiency in dairy cattle: a review. Livestock Production Science 20(1):1-13.

Lahart, B., S. McParland, E. Kennedy, T. M. Boland, T. Condon, M. Williams, N. Galvin, B. McCarthy, and F. Buckley. 2019. Predicting the dry matter intake of grazing dairy cows using infrared reflectance spectroscopy analysis. Journal of dairy science 102(10):8907-8918.

Lassen, J., P. Løvendahl, and J. Madsen. 2012. Accuracy of noninvasive breath methane measurements using Fourier transform infrared methods on individual cows. Journal of dairy science 95(2):890-898.

Lovendahl, P., G. F. Difford, B. Li, M. G. G. Chagunda, P. Huhtanen, M. H. Lidauer, J. Lassen, and P. Lund. 2018. Review: Selecting for improved feed efficiency and reduced methane emissions in dairy cattle. Animal : an international journal of animal bioscience 12(s2):s336-s349.

Miller, R. H. and N. W. Hooven. 1970. Factors Affecting Body Weights in a Herd of Holstein Cattle. Journal of dairy science 53(5):554-560.

Miller, R. H., N. W. Hooven, and M. E. Creegan. 1969. Weight Changes in Lactating Holstein Cows. Journal of dairy science 52(1):90-94.

Nam, K., H. Lim, and B.-i. Ahn. 2020. Analysis of Consumer Preference for Milk Produced through Sustainable Farming: The Case of Mountainous Dairy Farming. Sustainability 12(7):3039.

NRC. 1988. Factors Affecting the Composition of Milk from Dairy Cows. Attributes of Animal Products. Designing Foods: Animal Product Options in the Marketplace. National Academies Press (US), Washington (DC).

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th ed. National Academies Press, Washington, DC.

Olijhoek, D. W., P. Løvendahl, J. Lassen, A. L. F. Hellwing, J. K. Höglund, M. R. Weisbjerg, S. J. Noel, F. McLean, O. Højberg, and P. Lund. 2018. Methane production, rumen fermentation, and diet digestibility of Holstein and Jersey dairy cows being divergent in residual feed intake and fed at 2 forage-to-concentrate ratios. Journal of dairy science 101(11):9926-9940.

Peeters, A. 2010. Country Pasture/Forage Resource Profiles. Food and agriculture organization: Rome (Italy).

Phuong, H. N., N. C. Friggens, I. J. M. de Boer, and P. Schmidely. 2013. Factors affecting energy and nitrogen efficiency of dairy cows: A meta-analysis. Journal of dairy science 96(11):7245-7259.

Poncheki, J. K., M. L. S. Canha, S. L. Viechnieski, and R. d. Almeida. 2015. Analysis of daily body weight of dairy cows in early lactation and associations withproductive and reproductive performance. Revista Brasileira de Zootecnia 44(5):187-192.

Ranjitkar, S., D. Bu, M. Van Wijk, Y. Ma, L. Ma, L. Zhao, J. Shi, C. Liu, and J. Xu. 2020. Will heat stress take its toll on milk production in China? Climatic Change:1-16.

Rees, H. V. 1964. Heat Stress in Dairy Cattle: Physiological Responses and Variations in Milk Composition andEquilibrium. University of Tasmania, Hobart, Australia.

Řehák, D., J. Volek, L. Bartoň, Z. Vodková, M. Kubešová, and R. Rajmon. 2012. Relationships among milk yield, body weight, and reproduction in Holstein and Czech Fleckvieh cows. Czech Journal of Animal Science 57(6):274-282.

Richardson, D. O., J. R. Owen, R. D. Plowman, and J. T. Miles. 1971. Importance of Sire X Ration Interactions in Production and Feed Intake Traits of Dairy Cattle. Journal of dairy science 54(10):1518-1525.

Saeys, W., A. M. Mouazen, and H. Ramon. 2005. Potential for Onsite and Online Analysis of Pig Manure using Visible and Near Infrared Reflectance Spectroscopy. Biosystems Engineering 91(4):393-402.

Schultze, A. B. and H. P. Davis. 1961. Changes in Body Weight During the First Pregnancy for Holstein Heifers Calving at Different Seasons. Journal of dairy science 44(9):1717-1720.

Seymour, D. J., A. Cánovas, T. C. S. Chud, J. P. Cant, V. R. Osborne, C. F. Baes, F. S. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. Journal of dairy science 103(2):1528-1540.

Soyeurt, H., E. Froidmont, I. Dufrasne, D. Hailemariam, Z. Wang, C. Bertozzi, F. G. Colinet, F. Dehareng, and N. Gengler. 2019. Contribution of milk mid-infrared spectrum to improve the accuracy of test-day body weight predicted from stage, lactation number, month of test and milk yield. Livestock Science 227:82-89.

St-Pierre, N. 2008. Managing measures of feed costs: Benchmarking physical and economic feed efficiency. Pages 22-23.

Tomas Naes, T. I., Tom Fearn and Tony Davies. 2002. A user-friendly guide to multivariate calibration and classification Journal of Chemometrics 17(10):571-572.

van Gastelen, S., E. C. Antunes-Fernandes, K. A. Hettinga, and J. Dijkstra. 2017. Relationships between methane emission of Holstein Friesian dairy cows and fatty acids, volatile metabolites and non-volatile metabolites in milk. Animal : an international journal of animal bioscience 11(9):1539-1548.

van Knegsel, A. T. M., H. van den Brand, J. Dijkstra, W. M. van Straalen, M. J. W. Heetkamp, S. Tamminga, and B. Kemp. 2007a. Dietary Energy Source in Dairy Cows in Early Lactation: Energy Partitioning and Milk Composition. Journal of dairy science 90(3):1467-1476.

van Knegsel, A. T. M., H. van den Brand, E. A. M. Graat, J. Dijkstra, R. Jorritsma, E. Decuypere, S. Tamminga, and B. Kemp. 2007b. Dietary Energy Source in Dairy Cows in Early Lactation: Metabolites and Metabolic Hormones. Journal of dairy science 90(3):1477-1485.

VandeHaar, M. J. and N. St-Pierre. 2006. Major Advances in Nutrition: Relevance to the Sustainability of the Dairy Industry. Journal of dairy science 89(4):1280-1291.

Vanlierde, A. 2019. Development of equations to predict methane eructed by lactating cows from milk mid-infrared spectra. University of liege, Gembloux, Belgium.

Vanlierde, A., F. Dehareng, N. Gengler, E. Froidmont, S. McParland, M. Kreuzer, M. Bell, P. Lund, C. Martin, B. Kuhla, and H. Soyeurt. 2020. Improving robustness and accuracy of predicted daily methane emissions of dairy cows using milk mid-infrared spectra. Journal of the Science of Food and Agriculture.

Vanlierde, A., M. L. Vanrobays, F. Dehareng, E. Froidmont, H. Soyeurt, S. McParland, E. Lewis, M. H. Deighton, F. Grandl, M. Kreuzer, B. Gredler, P. Dardenne, and N. Gengler. 2015. Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra. Journal of dairy science 98(8):5740-5747.

Vanrobays, M. L., J. Vandenplas, H. Hammami, E. Froidmont, and N. Gengler. 2015. Short communication: Novel method to predict body weight of primiparous dairy cows throughout the lactation. Journal of dairy science 98(1):692-697.

Veerkamp, R. F., G. Simm, and J. D. Oldham. 1995. Genotype by environment interactions: experience from Langhill. BSAP Occasional Publication 19:59-66.

Yan, T., C. S. Mayne, F. G. Gordon, M. G. Porter, R. E. Agnew, D. C. Patterson, C. P. Ferris, and D. J. Kilpatrick. 2010. Mitigation of enteric methane emissions through improving efficiency of energy utilization and productivity in lactating dairy cows. Journal of dairy science 93(6):2630-2638.

Zucali, M., G. Battelli, M. Battini, L. Bava, M. Decimo, S. Mattiello, M. Povolo, and M. Brasca. 2016. Multi-dimensional assessment and scoring system for dairy farms. Italian Journal of Animal Science 15(3):492-503.Zucali, M., G. Battelli, M. Battini, L. Bava, M. Decimo, S. Mattiello, M. Povolo, and M. Brasca. 2016. Multi-dimensional assessment and scoring system for dairy farms. Italian Journal of Animal Science 15(3):492-503.