

Expanding the potential of infrared spectroscopy as a tool of precision dairy farming: From on-site infrared milk analysis to spectral data mining for indicators of animal well-being

Mazen Bahadi

Department of Food Science and Agricultural Chemistry
McGill University, Montreal
January 2020

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree
of Doctor of Philosophy

©Mazen Bahadi, 2020

Short title

Expanding the potential of infrared spectroscopy as a tool of precision dairy farming

Abstract

In this thesis, two propositions were presented to expand the capabilities of infrared (IR) spectroscopy to monitor dairy production and milk quality. The first was to provide Canadian milk producers with suitable cost-effective and easy-to-use instruments for on-site milk analysis. Such instruments will help in realizing the proAction initiative of the Canadian Quality Milk Program (CQM) of the Dairy Farmers of Canada (DFC), which aims at enabling Canadian milk producers to self monitor fat and protein content, and other milk quality indicators through on-site milk sampling and inspection. Implementation of infrared milk analysis with the use of a portable Fourier transform infrared (FTIR) spectrometer equipped with a transmission cell and combined with an ultrasonic homogenizer proved to be effective in on-site prediction of milk components. An external validation study of the final prototype evaluated in this study yielded mean difference (MD) values that were ≤ 0.05 for fat, protein and lactose, which comply with the stipulations of the AOAC International official method 972.16, 33.2.31. Attenuated total reflectance (ATR) was evaluated as an alternative sample introduction method for raw milk analysis by FTIR spectroscopy without any homogenization. The ATR-FTIR calibration models developed by partial-least-squares regression (PLSR) yielded prediction error values of 0.06%, 0.07% and 0.06% for lactose, protein and non-fat solids, respectively, and 0.37% for milk fat, which makes ATR infeasible for the determination of milk fat. In addition, a linear variable filter (LVF) array spectrometer, which is a novel IR spectrometer, was evaluated for milk analysis. The prediction error values for PLSR models developed with raw milk IR spectra were 0.07%, 0.23% and 0.49% for lactose, protein and fat, respectively. Ultrasonic treatment of milk reduced the prediction errors for lactose and protein to 0.05% and 0.15%, respectively, but did not affect the prediction error for milk fat. Thus, use of this LVF IR spectrometer was deemed infeasible for milk fat determination. Nevertheless, the LVF IR spectrometer proved to be an effective tool in differentiating watered-down milk samples from genuine ones. In some countries, fraudulent dairy farmers might add water to increase the volume of milk. This practice can be detected by milk cryoscopy, which measures milk freezing point depression; however, the reading of a cryoscope can be restored to its legal values by the addition of true solutes, such as urea, ammonium sulfate and citrate. The LVF IR spectrometer differentiated between genuine milk samples and those that contained as low as 5% added water and different chemical adulterants. Quantitatively, PLSR detected added water in raw milk with a prediction error of 1.85% regardless of the identity of the chemical adulterant

that was added with water to mask its addition. These promising results convinced the manufacturer of the LVF IR spectrometer to produce a lab-in-box milk adulteration detector, which was commissioned for a major manufacturer of milk cryoscopes in Brazil. In addition, a transmission-based FTIR solution was developed to differentiate watered-down milk samples from genuine ones. For raw milk, the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) with 30 μm path length was used as a sample introduction method, and a multitiered prediction workflow was elaborated in the following manner. First, a classification model was developed to differentiate between the spectra of watered-down and genuine milk samples by applying principal component-based quadratic discriminant analysis. Second, soft independent modelling of class analogies (SIMCA) was employed to develop a classification model allowing for identification of the chemical adulterant present in the sample, if any. If the chemical adulterant is urea, citrate or sulfate, then added water and the chemical adulterant will be quantified by selection of the appropriate PLSR calibration models. If multiple chemical adulterants are present, then only added water will be quantified. The prediction error value of the PLSR calibration model for added water was 0.39 % for raw milk scanned with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) with 30 μm path length. The presence of different chemicals in milk did not undermine the capability of the water prediction model.

The second proposition that was discussed in this thesis to expand the capabilities of milk IR spectroscopy was the exploitation of milk FTIR spectra beyond the paradigm of predicting specific milk components by PLSR models that will be used in the decision making process on dairy farms. Combining principal component analysis (PCA) and mixed modeling proved to be a successful strategy to detect trends of subtle changes in milk FTIR spectra. This approach was applied to animal trials aimed at studying the effect of tie rail position, chain length, stall width, stall length and manger wall height on animal welfare level in the tie stall dairy farming system. For each trial, a spectral fingerprint was isolated, which represented changes to milk composition associated with the significant treatment effect and was interpreted in light of the behavioural data that was collected during the trial. Currently, animal welfare is assessed by trained technicians who visit dairy farms to evaluate animal injuries, quality of the cow's lifts and sets, body condition and lameness. The hybrid approach investigated in this thesis bridged the gap between two distinct scientific domains, namely, FTIR spectroscopy and animal behaviour science, and it will open the door to study animal welfare from a novel angle, which will eventually help dairy herd

improvement agencies provide new services for dairy farmers in the field of animal welfare based on milk FTIR spectra that are routinely recorded.

Résumé

Dans cette thèse, deux propositions ont été présentées pour augmenter la fonctionnalité de la spectroscopie infrarouge (IR) pour contrôler la production laitière et la qualité du lait. La première était de développer des instruments rentables et faciles à utiliser pour l'analyse d'échantillons de lait au site de collecte. De tels instruments aideront à réaliser le volet Qualité du Lait inclus au sein de l'initiative proAction des Producteurs laitiers du Canada, en permettant aux producteurs de lait canadiens de surveiller eux-mêmes et plus étroitement la teneur en matières grasses et en protéines ainsi que d'autres indicateurs de la qualité de lait. L'utilisation d'un spectromètre infrarouge portable à transformée de Fourier (IRTF) équipé d'une cellule de transmission et combiné à un homogénéisateur à ultrasons s'est avérée efficace et fiable pour l'analyse d'échantillons de lait directement à la ferme, dans le but de déterminer leur teneur en composantes. Une étude de validation externe du prototype final évalué dans cette étude a donné une différence moyenne ≤ 0.05 pour la matière grasse, la protéine et le lactose, ce qui correspond à la norme officielle de l'AOAC International 972.16, 33.2.31. La réflectance totale atténuée (RTA) a été évaluée comme méthode alternative d'introduction d'échantillons pour l'analyse du lait cru non homogénéisé par spectroscopie IRTF. Les modèles d'étalonnage RTA-IRTF développés par régression des moindres carrés partiels ont donné des valeurs d'erreur de prédiction de 0.06%, 0.07% et 0.06% pour le lactose, la protéine et les solides non gras, respectivement, et de 0,37% pour la matière grasse laitière, ce qui rend la RTA inutile pour évaluer la teneur en gras du lait. Un spectromètre à filtre variable linéaire (FVL), qui est un nouveau type de spectromètre IR, a été aussi évalué pour l'analyse du lait. Les valeurs d'erreur de prédiction des modèles de régression des moindres carrés partiels développés avec des spectres IR pour le lait cru étaient de 0.07%, 0.23% et 0.49% pour le lactose, la protéine et la matière grasse respectivement. Le traitement du lait par ultrasons a réduit les erreurs de prédiction du lactose et de la protéine à 0.05% et 0.15%, respectivement, mais n'a pas affecté l'erreur de prédiction pour la matière grasse. Ainsi, l'utilisation de ce spectromètre FVL IR a été jugée inutile pour évaluer la teneur en matière grasse du lait. Néanmoins, le spectromètre FVL IR s'est avéré un outil efficace pour différencier les échantillons de lait frelatés des échantillons authentiques. Dans certains pays, il arrive que des producteurs laitiers frauduleux ajoutent de l'eau pour augmenter le volume de lait expédié. Cette pratique peut être détectée par la cryoscopie du lait, qui mesure la dépression du point de congélation du lait ; cependant, la lecture d'un cryoscope peut être rétablie aux valeurs attendues via l'ajout de solutés tels que l'urée, le

sulfate d'ammonium et le citrate. Le spectromètre FVL IR a permis de différencier les échantillons authentiques de lait de ceux contenant aussi peu que 5% d'eau ajoutée et divers adultérants. Quantitativement, le modèle de régression des moindres carrés partiels a détecté l'ajout d'eau dans le lait cru avec une erreur de prédiction de 1.85% et ce, peu importe l'identité de l'adultérant mélangé à l'eau pour masquer son ajout. Ces résultats prometteurs ont convaincu le fabricant du spectromètre FVL IR de produire un détecteur d'adultération du lait, qui a été commandé pour un grand fabricant de cryoscopes à lait au Brésil. En outre, une solution basée sur la transmission IRTF a été développée pour différencier les échantillons de lait adultérés des échantillons authentiques. Pour le lait cru, l'accessoire DialPath (Agilent Technologies, Santa Clara, Californie, ÉU) avec une longueur de trajet de 30 μm a été utilisé comme méthode d'introduction d'échantillon, et un flux de travail de prédiction à plusieurs niveaux a été élaboré. Premièrement, un modèle de classification a été développé pour différencier les spectres d'échantillons de lait dilué et authentique en appliquant une analyse discriminante quadratique basée sur les composantes principales. Deuxièmement, une modélisation indépendante douce des analogies de classe (SIMCA) a été utilisée pour développer un modèle de classification permettant d'identifier l'adultérant présent dans l'échantillon, le cas échéant. Si l'adultérant est l'urée, le citrate ou le sulfate, l'eau ajoutée et l'adultérant chimique sont quantifiés en sélectionnant les modèles d'étalonnage de moindres carrés partiels appropriés. Si plusieurs adultérants sont présents, seule l'eau ajoutée est quantifiée. La valeur d'erreur de prédiction de l'eau ajoutée était de 0.39% pour le lait cru analysé avec l'accessoire DialPath (Agilent Technologies, Santa Clara, Californie, ÉU). La présence de différents adultérants dans le lait n'a pas affecté les résultats du modèle de prédiction de l'eau.

La deuxième proposition discutée dans cette thèse visait à augmenter la fonctionnalité de la spectroscopie IR du lait par l'exploitation des spectres IRTF du lait au-delà de leur utilisation traditionnelle, soit pour la prédiction des composantes du lait, en introduisant des modèles de moindres carrés partiels utilisables dans les processus décisionnels au sein des troupeaux laitiers. La combinaison de l'analyse en composantes principales (ACP) et de la modélisation mixte s'est avérée une bonne stratégie pour détecter même les changements subtils dans les spectres IRTF du lait. Cette approche a été appliquée aux essais sur les animaux visant à étudier l'effet de la position de la barre d'attache, de la longueur de la chaîne, de la largeur de stalle, de la longueur de stalle et de la hauteur du muret entre la stalle et la mangeoire sur le niveau de bien-être de vaches laitières

logées en stabulation entravée. Pour chaque essai, une empreinte spectrale représentant les changements dans la composition du lait associés à l'effet significatif d'un traitement a été isolée, puis interprétée à la lumière des données comportementales recueillies au cours de l'essai. Actuellement, le bien-être animal est évalué par des techniciens qualifiés qui visitent les fermes laitières pour évaluer les blessures, la qualité des levers et des couchers, la condition de chair et la boiterie chez les vaches. L'approche hybride étudiée dans cette thèse se veut combler l'écart entre deux domaines scientifiques distincts, soit la spectroscopie IRTF et la science du comportement animal, et ainsi ouvrir la porte à l'étude du bien-être animal sous un nouvel angle. Cette nouvelle approche pourrait éventuellement aider les agences d'amélioration des troupeaux laitiers à fournir aux producteurs de nouveaux services liés aux questions de bien-être animal, en se basant sur les spectres IRTF de ces mêmes échantillons de lait qui sont déjà régulièrement collectés à la ferme.

Acknowledgments

I would like to express my gratitude to Dr. Ashraf Ismail for his mentoring, guidance and immense support, academically and financially, during my time in McGill University. I appreciate all the opportunities that you offered me, be it in research or teaching. I would also like to thank Dr. Elsa Vasseur for her guidance and patience during the work on the animal trials' data. Thank you for making my experience in bridging two different domains an enjoyable one and for your help in expanding my horizon as a researcher. A special thank you is extended to Dr. Jacqueline Sedman whose advice and comments were essential to better understand FTIR spectroscopy, quantitative analysis and scientific writing. I also thank Véronique Boyer for reviewing the French translation of my abstract.

I would also like to thank my parents, Saïd and Sarah, who bore all kinds of circumstances that accompanied the start of our new life in Canada and my pursuit of an advanced degree; my sisters, Manal and Rasha, for their moral and emotional support and my nieces and nephews, George, Nathalie, Jean-Paul and Zoë. Your laughs always brightened my days.

Thank you to all the funding partners that made my research, degree, and travel possible: the Schulich graduate fellowship (McGill University), the Pilarczyk graduate fellowship (McGill University), the graduate excellence award (McGill University), the Globalink partnership award for graduate research in industry outside Canada (MITACS) and PZL Tecnologia.

Contributions of authors

The author was responsible for the experimental work and the development of PLS regression and classification models for chapters 3-5. In addition, the author developed the hybrid analysis approach of milk FTIR spectral data mining reported in chapters 6 and 7 and wrote all the programming codes required for it. Dr. Ashraf Ismail and Dr. Elsa Vasseur are thesis supervisor and co-supervisor, respectively, who provided direct advisory input into the work as it progressed. Dr. Jacqueline Sedman provided guidance related to FTIR spectroscopy and quantitative analysis.

Results reported in this thesis were or will be presented in the following conferences:

- 1- Oral presentation at Pittcon 2019, Philadelphia (PA, USA)
Infrared Spectroscopic Methods for Detection of Adulterants in Milk
Mazen Bahadi, Ashraf Ismail, Jacqueline Sedman
- 2- Poster presentation at the American Dairy Science Association (ADSA) 2019 annual meeting, Cincinnati (OH, USA)
Detecting welfare status in a milk sample: Effects of tie-rail placement on milk composition by Fourier transform infrared spectroscopy
Mazen Bahadi, Ashraf Ismail¹, Débora Santschi, Daniel Lefebvre, Raj Duggavathi, Elsa Vasseur
- 3- Oral presentation at Novalait Forum Techno 2020, Québec City (QC, Canada)
Une analyse de lait pourrait-elle prédire le niveau de bien-être et de santé des vaches ?

Contributions to knowledge

1. **Demonstrated that major milk components and some minor ones can be determined by portable FTIR spectrometer equipped with a transmission cell and combined with ultrasonic processing of milk and PLS prediction models.**

In this study, it has been proven that milk analysis by FTIR spectroscopy can be performed outside central dairy laboratories. The final prototype that was tested yielded MD values that were ≤ 0.05 for fat, protein and lactose, which comply with the stipulations of the AOACI official method 972.16, 33.2.31. This proof of concept opens the door for on-site milk analysis, which will expand the applications of FTIR spectroscopy in monitoring milk compositions on dairy farms.

2. **Evaluated the potential of ATR-FTIR spectroscopy in raw milk analysis accompanied by PLS regression models.**

In this study, it has been proven that ATR-FTIR spectroscopy can be a viable option for the determination of protein, lactose and solids non-fat in raw milk. The prediction errors for these milk components were better than those reported in the literature so far and comparable to what has been obtained by transmission based FTIR spectroscopy. As a sample introduction method, ATR can eliminate issues related to pumping raw milk through transmission cell with micrometric optical path length. Contrary to what has been reported in the literature, the study relied on big sample size of producer raw milk (N=360) that resulted in 1080 spectra. In addition, the study confirmed ATR-FTIR spectroscopy inability to capture sufficient chemical information related to milk fat.

3. **Evaluated the novel linear variable filter (LVF) array IR spectrometer accompanied by PLS regression for milk analysis.**

LVF array IR spectrometers do not rely on interferometry to resolve different wavelengths in the mid-IR range, which means they do not have moving parts. This fact makes them good candidate for on-site applications of milk analysis. In addition, they are significantly cheaper than their FTIR counterpart and they are equipped with an ATR sample introduction accessory. This spectrometer revealed acceptable performance in determining water content, lactose and protein in raw milk. However, it was incapable of capturing sufficient chemical information related to milk fat.

4. Demonstrated that different combinations of IR spectrometers and sample introduction methods can capture chemical information in milk containing extraneous water and chemical adulterants.

Several combinations of IR spectrometers and sample introduction methods were used to scan milk samples containing extraneous water and chemical adulterants. These chemicals are intended to restore the freezing point depression reading of milk cryoscopy to its legal value. This practice is a serious concern in some countries. All evaluated combinations revealed reliable capabilities in detecting this practice. These combinations were ATR-LVF spectrometer and FTIR spectrometer combined with ATR accessory, transmission cell and the DialPath accessory (Agilent Technologies, Santa Clara, California, USA), which is a transmission-based sample introduction method.

5. Contributed to the development of a novel lab-in-box milk adulteration detection instrument.

The manufacturer of the LVF spectrometer built a lab-in-box instrument dedicated to the detection of watered-down milk after proving that the LVF spectrometer can detect this practice. The new instrument can differentiate genuine milk samples from watered-down ones and can estimate the percentage of added water regardless of the chemical adulterant that might have been added to milk to restore its freezing point depression to the legal value.

6. Developed an FTIR based solution to detect watering down of milk.

A multi-tier solution was developed to detect milk that contains extraneous water and chemical adulterants using FTIR spectroscopy combined with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) as a sample introduction method for raw milk in a business-oriented context. The solution included a principle component based quadratic discriminant analysis (PC-QDA) classification model to differentiate between genuine milk samples and watered-down ones with chemical adulterants, a soft independent modelling by class analogy (SIMCA) classification model to determine the chemical adulterant that might be present in the milk sample and PLS regression models to predict the percentage of added water and chemical adulterant in case single chemical is present. The added water prediction model performed well regardless of the type or number of chemical adulterants present in the milk sample. This solution was the result of a capacity building project between the McGill IR

group and PZL Industria Eletrônica Ltda (Londrina, PR, Brazil), which is a lead manufacturer of milk cryoscopy instruments in South America.

7. Evaluated combinations of spectral preprocessing treatments and multivariate chemometric techniques to study the effect of housing treatments of dairy cows on milk composition.

Principle component analysis (PCA), hierarchical cluster analysis (HCA) and partial least squares discriminant analysis (PLS-DA) were applied to raw or preprocessed milk FTIR spectral data collected during an animal trial designed to study the effect of the tie rail position on welfare level of dairy cows. The objective was to study animal welfare from a novel angle, which is the effect of animal comfort on milk composition. The merits and drawbacks of each technique were evaluated to determine its usefulness for the purpose of such study.

8. Proved that PCA can isolate the spectral fingerprint that reflects changes in milk chemical composition resulting from a systemic factor affecting milk components.

PCA was applied to milk FTIR spectral data as a variable reduction method that separated meaningful information from noise. The isolated principal components were used as input variables for hypothesis testing with mixed models to study the effects of housing treatments of dairy cows on milk composition. This practice will extend the boundaries of milk FTIR spectroscopy beyond the paradigm of predicting specific milk components by PLS regression models to study the metabolic state of dairy cows.

9. Developed a hybrid approach to analyze milk FTIR spectral data in the context of designed experiment animal trials.

PCA-mixed modeling approach was successfully applied as a data mining tool to milk FTIR spectral data collected during animal trials aiming at studying the effect of tie rail position, chain length, stall width, stall length and manger wall height on milk composition as an indicator of animal comfort and welfare level. Changes to milk components were inferred from principal components that revealed significant treatment effect in situations where numerical data of individual milk components did not reveal such effect when tested by mixed modeling. These changes were explained in light of behavioural data that had been collected during the trial. This approach represents a novel angle to study animal welfare.

Table of content

Short title.....	ii
Abstract.....	iii
Résumé.....	vi
Acknowledgments.....	ix
Contributions of authors	x
Contributions to knowledge.....	xi
Table of content	xiv
List of tables.....	xxi
List of figures.....	xxvii
List of abbreviations	xxxii
Chapter 1: Introduction	1
1.1 General introduction.....	1
1.2 Rationale and research objectives	4
Chapter 2: Literature review	7
2.1 Introduction	7
2.2 Importance of monitoring milk composition	8
2.3 Milk analysis by infrared spectroscopy.....	10
2.4 Multivariate regression for milk analysis by FTIR spectroscopy	14
2.5 Spectral pre-treatments or preprocessing	17
2.6 Calibration and validation of IR milk analyzers	19
2.7 Novel applications in milk analysis by infrared spectroscopy	22
2.7.1 Prediction of milk fatty acids and milk fat composition.....	22
2.7.2 Prediction of fatty acid chain length and unsaturation.....	24
2.7.3 Prediction of fat globule particle size in homogenized milk	26

2.7.4	Prediction of milk protein composition and its genetic variants.....	27
2.7.5	Prediction of milk lactoferrin as an indicator of mastitis.....	28
2.7.6	Prediction of milk coagulation properties.....	30
2.7.7	Prediction of titratable acidity and pH.....	33
2.7.8	Estimation of major mineral contents in cow milk.....	34
2.7.9	Prediction of acetone, β -hydroxybutyrate and citrate	35
2.7.10	Prediction of cows' body energy status and feed efficiency.....	35
2.7.11	Prediction of methane emissions from dairy cattle.....	37
2.7.12	Detecting cows' pregnancy status.....	40
2.7.13	Verification of dairy farming production system.....	41
2.7.14	Suggested predictions of milk components by FTIR spectroscopy.....	42
2.8	Conclusion.....	44
Chapter 3: Transmission based FTIR spectroscopy for on-site milk analysis.....		45
Abstract		45
3.1	Introduction	46
3.2	Materials and Methods.....	49
3.2.1	Milk samples	49
3.2.2	Homogenization.....	49
3.2.3	Assembly of prototype analyzers.....	51
3.2.4	Evaluation of the spectral quality of the assembled prototypes.....	52
3.2.5	Spectral acquisition.....	53
3.2.6	Development of PLS calibration models for milk components.....	53
3.2.7	External validation of the prototype analyzers	54
3.3	Results and Discussion.....	55
3.3.1	Homogenization.....	55

3.3.2	Evaluation of the spectral quality of the assembled prototypes.....	59
3.3.3	PLS calibration models for milk components.....	63
3.3.4	External validation of the prototype analyzers	70
3.4	Conclusion.....	72
	Acknowledgement	74
	Funding	74
	Connecting statement.....	75
	Chapter 4: Evaluation of ATR-FTIR spectroscopy and ATR-LVF IR spectrometer for on-site milk analysis.....	76
	Abstract	76
4.1	Introduction	77
4.2	Materials and Methods	82
4.2.1	Milk samples	82
4.2.2	Spectral acquisition.....	83
4.2.3	Development of PLS calibration models for milk components.....	83
4.3	Results and Discussion.....	85
4.3.1	Mid ATR-FTIR milk analysis.....	85
4.3.2	Comparisons with previous ATR-FTIR studies for milk analysis.....	90
4.3.3	Mid ATR-IR milk analysis	94
4.4	Conclusion.....	100
	Acknowledgement	101
	Funding	101
	Appendix.....	102
	Connecting statement.....	107
	Chapter 5: Case study of on-site milk analysis infrared spectroscopic methods for detection of added water and chemical adulterants in Brazilian milk	108

Abstract	108
5.1 Introduction	110
5.2 Materials and Methods	113
5.2.1 Detection of extraneous water and added chemicals by ATR-FTIR	113
5.2.1.1 Mid-IR bands of chemical adulterants.....	113
5.2.1.2 Qualitative detection of chemically adulterated milk.....	113
5.2.2 Effect of extraneous water and chemical adulterants on milk freezing point depression	115
5.2.3 Detection of extraneous water and added chemicals by ATR-IR.....	115
5.2.3.1 Spectral acquisition.....	115
5.2.3.2 Mid-IR bands of chemical adulterants.....	115
5.2.3.3 Preparation of adulterated milk samples	116
5.2.4 Detection of extraneous water and added chemicals by transmission based FTIR	117
5.2.4.1 Mid-IR bands of chemical adulterants.....	117
5.2.4.2 Preparation of adulterated milk samples	117
5.2.4.3 Spectral acquisition.....	118
5.2.4.4 Classification models for differentiation of adulterated milk samples.....	118
5.2.4.5 Quantification models for extraneous water and added chemical adulterants	119
5.3 Results and Discussion.....	121
5.3.1 Detection of extraneous water and added chemicals by ATR-FTIR	121
5.3.2 Effect of extraneous water and chemical adulterants on milk freezing point depression	127
5.3.3 Detection of extraneous water and added chemicals by ATR-IR.....	130
5.3.3.1 Mid-IR bands of chemical adulterants.....	130

5.3.3.2	Spectral analysis and added water quantification model.....	132
5.3.4	Detection of extraneous water and added chemicals by transmission based FTIR	136
5.3.4.1	Mid-IR bands of chemical adulterants.....	136
5.3.4.2	Classification models for differentiation of adulterated milk samples.....	136
5.3.4.3	Quantification models for extraneous water and added chemical adulterants	142
5.4	Conclusion.....	147
	Acknowledgement	159
	Funding	159
	Connecting Statement	160
	Chapter 6: Evaluation of data analysis approaches of chemical milk composition and FTIR spectral data in the context of controlled-design trials testing dairy cattle housing welfare improvements	161
	Abstract	161
6.1	Introduction	163
6.2	Materials and Methods.....	166
6.2.1	The data.....	166
6.2.2	Analysis of milk composition numerical dataset.....	167
6.2.3	Analysis of milk spectral dataset	168
6.2.4	Hybrid approach for milk spectral dataset analysis	169
6.3	Results and Discussion.....	172
6.3.1	Analysis of milk composition numerical dataset.....	172
6.3.2	Analysis of milk Spectral dataset.....	178
6.3.3	Hybrid approach for milk spectral dataset analysis	188
6.4	Conclusion.....	201

Acknowledgment	204
Funding	204
Connecting statement.....	205
Chapter 7: Data mining of milk FTIR spectral data by combining mixed modeling and multivariate analysis to study the effects of housing treatments on cow welfare	206
Abstract	206
7.1 Introduction	207
7.2 Materials and Methods	209
7.2.1 The data.....	209
7.2.2 Experimental design and statistical model.....	210
7.2.2.1 Tie rail trial	210
7.2.2.2 Chain length trial	211
7.2.2.3 Stall width trial	212
7.2.2.4 Manger wall and stall length trial	213
7.2.3 Spectral data analysis.....	215
7.2.4 Interpretation of spectral features in loading spectra.....	216
7.3 Results and Discussion.....	218
7.3.1 Interpretation of spectral features in loading spectra.....	218
7.3.2 Tie rail trial	222
7.3.3 Tie chain length trial	233
7.3.4 Stall width trial.....	242
7.3.5 Manger wall and stall length trial	253
7.4 Conclusion.....	262
Acknowledgment	263
Funding	263

Appendix.....	264
Chapter 8: General discussion and conclusion	277
8.1 Discussion	277
8.2 Conclusion.....	282
References.....	283

List of tables

Table 2-1 Band assignments in the mid-IR range for FTIR milk analysis	13
Table 3-1 Values of ultrasonication parameters used for milk homogenization with the Fisher ultrasonic processor	50
Table 3-2 Mean values of D (0.90) from volume distribution of fat globules diameters in raw milk sonicated by the Fisher processor for 15, 30, 60 and 90s.	56
Table 3-3 Mean values of D (0.90) from volume distribution of fat globules diameters in raw milk sonicated by the Vibra Cell processor for 60 and 90s, raw milk and high-pressure homogenized milk	57
Table 3-4 Mean D (0.90) values for high fat milk samples >4% sonicated for 90s and 120s with the Vibra Cell processor.....	57
Table 3-5 SNR ratio comparison of spectra collected on a FOSS analyzer, prototype 1, prototype 2 and prototype 3	60
Table 3-6 Comparison of FOMs of prototype 1 calibration models for homogenized packed milk	61
Table 3-7 Comparison of prototype 1 cross validation FOMs for homogenized packed milk.....	61
Table 3-8 Comparison of calibration FOMs for milk components' PLS models developed using milk spectra collected on P1, P2, P3 and a FOSS milk analyzer.....	68
Table 3-9 Comparison of cross-validation FOMs for milk components' PLS models developed using milk spectra collected on P1, P2, P3 and a FOSS milk analyzer	69
Table 3-10 Mean differences and standard deviation of the difference for the external validation sets for the assembled prototypes	70
Table 4-1 Comparison of calibration models' FOMs for milk components. Calibration models were developed using milk spectra scanned by FTIR spectrometer equipped with ATR sample introduction accessory	89
Table 4-2 Comparison of cross-validation FOMs for milk components' models developed using milk spectra collected on FTIR spectrometer equipped with ATR sample introduction accessory	89
Table 4-3 Summary of reported literature on the application of ATR-FTIR spectroscopy in the determination of milk components.	93

Table 4-4 Comparison of calibration FOMs for milk components' models developed using milk spectra collected on portable ATR-IR spectrometer.....	99
Table 4-5 Comparison of cross-validation FOMs for milk components' models developed using milk spectra collected on portable ATR-IR spectrometer	99
Table 5-1 Volumes of stock solution used to prepare different levels of chemically adulterated milk samples	114
Table 5-2 Detected absorption bands in cm^{-1} for different chemical adulterants in milk and water solutions	123
Table 5-3 Calculated freezing point depression values for aqueous urea solutions with different concentrations	128
Table 5-4 Freezing point of raw milk adulterated with different adulteration solutions.	129
Table 5-5 Comparison of characteristic IR absorbance bands of chemical adulterants in milk and water observed in spectra collected on ATR-IR (LVF) and ATR-FTIR spectrometers. Spectra of genuine milk and pure water were subtracted from spectra of spike milk samples and aqueous solutions of these adulterants.	131
Table 5-6 Calibration FOMs of the added water PLS prediction models. In addition to added water, milk samples contained one, two or three chemical adulterants to compensate for the change in freezing point.	134
Table 5-7 Cross-validation FOMs of the added water PLS prediction models. In addition to added water, milk samples contained one, two or three chemical adulterants to compensate for the change in freezing point.	134
Table 5-8 Centers of characteristic IR absorption bands of four chemicals in aqueous solutions and in milk that are used to mask the addition of water or to neutralize milk acidity.....	136
Table 5-9 Confusion matrix of PCA-QDA classification model with no covariance shrinkage developed using VN raw spectral data of homogenized milk samples.	140
Table 5-10 Classification accuracy of SIMCA models developed using spectral data of raw and homogenized adulterated milk samples to classify adulterated samples according to the chemical adulterant that they contain.	142
Table 5-11 Comparison of performance FOM of classification models for differentiating genuine milk samples from those containing extraneous water and chemical adulterants using milk FTIR spectra collected by a transmission cell.	149

Table 5-12 Comparison of performance FOM of classification models for differentiating genuine milk samples from those containing extraneous water and chemical adulterants using milk FTIR spectra collected by the DialPath accessory 30µm.	150
Table 5-13 Example of SIMCA output for classification of adulterated milk samples according to the added chemical. SIMCA can assign unknown sample to one class, multiple classes or no class. Class membership at 5% significance level.....	152
Table 5-14 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using homogenized milk spectra scanned by Cary 630 (Agilent Technologies, Santa Clara, California, USA) spectrometer with transmission cell.....	153
Table 5-15 Comparison of cross-validation FOMs for milk adulterants models developed using milk spectra collected on Cary 630 (Agilent Technologies, Santa Clara, California, USA) with transmission cell.....	154
Table 5-16 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using homogenized milk spectra scanned by Cary 630 (Agilent Technologies, Santa Clara, California, USA) spectrometer with Dial Path accessory.	155
Table 5-17 Comparison of cross-validation FMOs for milk adulterants models developed using milk spectra collected on Cary 630 Dial Path accessory (Agilent Technologies, Santa Clara, California, USA).	156
Table 5-18 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using raw milk spectra scanned by Cary 630 spectrometer with Dial Path accessory (Agilent Technologies, Santa Clara, California, USA).	157
Table 5-19 Comparison of cross-validation FMOs for milk adulterants models developed using milk spectra collected on Cary 630 Dial Path accessory (Agilent Technologies, Santa Clara, California, USA).	158
Table 6-1 P values of different effects on milk components for the TR configuration trial. No significant effect was detected for treatment. Week has the strongest significant effect on almost all milk components followed by Block.	173
Table 6-2 Principal components extracted from the numerical dataset of milk analysis for the TR configuration trial.....	175
Table 6-3 VIP scores of predictors of the parsimonious PLS-DA classification model according to treatment.	177

Table 6-4 Model comparison summary of PLS-DA models developed for the four spectral dataset of milk samples. The vector normalized first derivative spectral dataset resulted in the most parsimonious model.	186
Table 6-5 P values obtained from the mixed procedure that was applied to individual spectral variables in the raw, FD, VN raw and VN FD spectral datasets of milk.	187
Table 6-6 Meaningful principal components isolated from the VN raw and VN FD spectral datasets of milk samples spiked with two levels of lactose.	189
Table 6-7 Estimates of the least squares means of the significant PC scores produced by the mixed model for lactose spiked milk samples and aqueous solutions for the VN raw and VN FD spectral datasets.	189
Table 6-8 Differences of least squares means of PC scores produced by the mixed model for lactose spiked milk samples and aqueous solutions for the VN raw and VN FD spectral datasets. Significant differences among treatment levels are revealed by P values.	190
Table 6-9 Peak fitting results for PC2 integrated loading spectrum that was extracted from VN FD spectral dataset of milk samples spiked with lactose. The most prominent peak is at 1086 cm^{-1} , which is very close to lactose peak at 1076 cm^{-1}	195
Table 6-10 The hybrid analysis approach applied to meaningful PCs extracted from raw, FD, VN raw and VN FD spectral datasets.	198
Table 6-11 The least squares means output of the mixed procedure that was applied to PC5 of the VN raw spectral dataset of the TR trial. It reveals significant interactions between some weeks and treatments.	200
Table 7-1 Concentrations of spiked milk samples and aqueous solutions of milk minor components	217
Table 7-2 IR peaks of minor milk components detected in FTIR spectra of spiked milk samples and aqueous solutions of the respective molecules.	222
Table 7-3 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral averages for the tie rail trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment, start and block effects that are tested in this trial.	226
Table 7-4 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral averages for the tie rail trial. The table also lists P values obtained	

from the SAS Mixed Procedure for the treatment, start and block effects that are tested in this trial.	227
Table 7-5 Least squares means produced by the Mixed procedure for the scores of PC7 extracted from long-term VN FD spectral average dataset and that revealed a significant treatment effect. This table shows that T3 scores were significantly different from the other treatments.	228
Table 7-6 Differences of least squares means for PC7's scores. The Scheffé adjusted P values show that T3 is significantly different from T1.	228
Table 7-7 Results of peak fitting of lumpy regions in the integral of the loading spectrum of PC7 and probable molecules that can be assigned to the resulting peaks.	229
Table 7-8 Variable importance for the projection scores of significant milk components obtained from parsimonious PLS-DA model for milk composition numerical data for weeks 8-10 to classify milk samples by tie rail treatment.	230
Table 7-9 Milk composition data $\pm$ SD for weeks 8-10 for the tie rail trial with T1 and T2 averages by treatment.	231
Table 7-10 Milk composition data $\pm$ SD for weeks 8-10 for the tie rail trial with T3 and T4 averages by treatment	232
Table 7-11 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral averages for the tie chain length trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.	237
Table 7-12 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral averages for the tie chain length trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.	238
Table 7-13 Least squares means produced by the Mixed procedure for the scores of PC6 extracted from long-term VN FD spectral average dataset and that revealed a significant treatment effect.	239
Table 7-14 Differences of least squares means for PC6's scores. The Scheffé adjusted P value shows that T1 is significantly different from T2.	239
Table 7-15 Results of peak fitting of region 1365-1160 cm^{-1} in the integral of the loading spectrum of PC6 and probable molecules that can be assigned to the resulting peaks.	240

Table 7-16 Milk composition data $\pm$ SD for weeks 8-10 for the tie chain trial with long-term averages by treatment.	241
Table 7-17 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.	245
Table 7-18 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.	246
Table 7-19 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of week 5 milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial	247
Table 7-20 Least squares means produced by the Mixed procedure for the scores of PC5 extracted from long-term VN Raw spectral average dataset and that revealed a significant treatment effect.	248
Table 7-21 Differences of least squares means for PC5's scores. The Scheffé adjusted P value shows that T1 is significantly different from T2.	248
Table 7-22 Results of peak fitting of region 1100-940 cm^{-1} in the loading spectrum of PC5 and probable molecules that can be assigned to the resulting peaks.	251
Table 7-23 Average $\pm$ SD milk composition data for week 6 for the stall width trial.....	252
Table 7-24 Block design on week 6 of the stall width trial.	253
Table 7-25 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of long-term milk spectral datasets for the stall length and manger wall trial. The table also lists P values obtained from the SAS Mixed Procedure for the effects that were tested in this trial. ...	257
Table 7-26 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral datasets for the stall length and manger wall trial.	258
Table 7-27 Least squares means produced by the Mixed procedure for the scores of PC6 extracted from long-term FD spectral average dataset and that revealed a significant length effect.	259
Table 7-28 Differences of least squares means for PC6's scores.	259
Table 7-29 Milk composition average $\pm$ SD data for week 6 for the stall length and manger wall trial.	261

List of figures

Figure 1-1 Business model of the milk recording system.....	3
Figure 3-1 Mid-FTIR milk spectrum acquired on a milk analyzer.....	48
Figure 3-2 Volume distribution of fat globules diameters in ultrasonicated raw milk by the Fisher processor.	58
Figure 3-3 Milk FTIR spectrum acquired on P3 with a transmission cell.....	60
Figure 3-4 Subtraction spectra of multiple measurements of homogenized packed skim milk before and after passing whole milk into the transmission cell.	61
Figure 3-5 Prototype 1 at the McGill IR group lab and Valacta.....	62
Figure 3-6 Prototype 2 at Valacta	62
Figure 3-7 Prototype 3 at the McGill IR group lab.....	62
Figure 3-8 Predicted values vs. Valacta analysis values of milk fat for spectra collected on P3 for the external validation set	71
Figure 3-9 Predicted values vs. Valacta analysis values of milk proteins for spectra collected on P3 for the external validation set	71
Figure 4-1 Design of an ATR measurement surface [22].....	80
Figure 4-2 Mid-IR ATR-FTIR milk spectrum.....	80
Figure 4-3 IRSphinx (Comline Elektronik Elektrotechnik GmbH, Germany) ATR-IR spectrometer.....	81
Figure 4-4 Schematic of an LVF ATR-IR spectrometer	81
Figure 4-5 Mid ATR-IR milk spectrum.....	98
Figure 5-1 ATR-FTIR spectra of homogenized milk spiked with chemical adulterants.....	124
Figure 5-2 Principal component scores showing the separation of genuine (red) and adulterated (blue) homogenized milk samples	124
Figure 5-3 HCA dendrogram showing a separate arm for genuine homogenized milk samples	125
Figure 5-4 Principle component scores showing a separate cluster for whey adulterated raw milk samples (green) and intercepting genuine raw milk cluster (red) with water adulterated raw milk samples (blue) at the 5% added water level.....	125
Figure 5-5 HCA dendrogram showing a separate arm for genuine raw milk samples (red-marked) which also contains two replicates of the 5% added water sample (blue-marked)	126

Figure 5-6 Principle component scores showing a cluster with raw milk samples and those that have less than 0.25% added urea	126
Figure 5-7 HCA dendrogram showing the arm of raw milk samples (red-marked) with chemically adulterated ones.	127
Figure 5-8 Subtraction spectra of homogenized milk samples spiked with sodium carbonate (blue), sodium citrate (purple), urea (green), ammonium sulfate (red). The spectrum of genuine milk was subtracted from the spectra of the spiked milk samples.	131
Figure 5-9 Principal component scores of homogenized genuine milk samples and adulterated milk samples with solutions of urea, sodium citrate, sodium carbonate, ammonium sulfate and different combinations of these solutions – genuine milk is in red	133
Figure 5-10 Principal component scores of raw genuine milk samples and adulterated milk samples with solutions of urea, sodium citrate, sodium carbonate, ammonium sulfate and different combinations of these solutions – genuine milk is in red	134
Figure 5-11 Lab-in-Box milk adulteration detector built using LVF ATR-IR spectrometer	135
Figure 5-12 PC3 loading spectrum isolated from the VN raw spectral data of homogenized milk samples collected by a transmission cell	142
Figure 5-13 Decision-making workflow for authenticating milk sample freezing point depression by transmission based FTIR spectroscopy.....	146
Figure 5-14 ROC curves for classification models to differentiate genuine milk samples from those containing extraneous water and chemical adulterants.....	151
Figure 6-1 Data analysis approaches that are considered to detect the tie rail position effect on milk composition numerical and spectral data.....	165
Figure 6-2 PCA scores plot projecting milk samples' scores for PC1 vs. PC2 for the TR configuration trial.....	175
Figure 6-3 PCA loading plot (PC1 vs. PC2) projecting the main groups of milk components that explains the majority of the variation in the milk numerical dataset of the TR configuration trial.	176
Figure 6-4 Coefficients of important milk components for classification of observations according to treatment by the parsimonious PLS-DA model.....	178
Figure 6-5 Spectral averages comparison.	179

Figure 6-6 HCA dendrogram of raw spectral data. Samples tend to cluster according to season red is season 1, blue is season 2.	184
Figure 6-7 HCA dendrogram of raw spectral data after forward feature selection.	185
Figure 6-8 PCA scores of the raw spectral dataset. Samples tend to cluster according to season.	186
Figure 6-9 Variable importance plot for the PLS-DA model for the vector normalized first derivative spectra dataset.	187
Figure 6-10 PCA scores of milk samples spiked with two levels of lactose.	191
Figure 6-11 Comparison between FTIR spectra of milk samples spiked with lactose and the principal component extracted from raw spectra that revealed significant effect for lactose concentration.	193
Figure 6-12 Comparison between FTIR spectra of milk samples spiked with lactose and the principal component extracted from first derivative spectra that revealed significant effect for lactose concentration.	194
Figure 6-13 Comparison of loading spectrum of PC2 obtained from VN FD spectral dataset of milk samples spiked with different levels of lactose (red) and the spectral integral of loading spectrum of PC2 (blue).	195
Figure 6-14 The loading spectrum of PC5 that was extracted from the VN raw spectral dataset of the TR trial.	199
Figure 6-15 Workflow of the hybrid data analysis approach of FTIR milk spectral data to detect a treatment effect.	203
Figure 7-1 The spectral integral of PC7 loading spectrum extracted from long-term VN FD spectral average dataset for the TR trial.	229
Figure 7-2 The spectral integral of PC6 loading spectrum extracted from long-term VN FD spectral average dataset for the TCL trial.	240
Figure 7-3 The loading spectrum of PC5 extracted from long-term VN Raw spectral dataset for the stall width trial.	249
Figure 7-4 FTIR spectra of BHB aqueous solutions.	250
Figure 7-5 FTIR spectra of Milk samples with different fat concentrations.	251
Figure 7-6 The loading spectrum of PC6 extracted from long-term FD spectral dataset for the stall length and manger wall trial.	260

Figure 7-7 Urea [141]	264
Figure 7-8 Raw milk spiked with urea. Blue: raw milk, purple: raw milk spiked with 0.5% urea, red: raw milk spiked with 1% urea.	264
Figure 7-9 β -Hydroxybutyric acid (BHBA or BHB) [142]	265
Figure 7-10 Raw milk spiked with BHB standard. Red: raw milk, blue: raw milk spiked with 0.5% BHB, purple: raw milk spiked with 1% BHB.....	265
Figure 7-11 Acetone [143].....	266
Figure 7-12 Raw milk spiked with acetone. Red: raw milk, blue: raw milk spiked with 0.5% acetone, purple: raw milk spiked with 1% acetone.....	266
Figure 7-13 Citrate [144]	267
Figure 7-14 Raw milk spiked with citrate. Red: raw milk, blue: raw milk spiked with 0.5% citrate, purple: raw milk spiked with 1% citrate.	267
Figure 7-15 Acetate [145].....	268
Figure 7-16 Raw milk spiked with acetate. Red: raw milk, blue: raw milk spiked with 0.5% acetate, purple: raw milk spiked with 1% acetate.....	268
Figure 7-17 Phosphate [146].....	269
Figure 7-18 Raw milk spiked with phosphate. Red: raw milk, blue: raw milk spiked with 0.5% phosphate, purple: raw milk spiked with 1% phosphate.....	269
Figure 7-19 From left to right: D-(+)-glucose, D-(+)-galactose, β -lactose [147-149]	270
Figure 7-20 Red: raw milk, blue: raw milk spiked with 1% galactose, purple: raw milk spiked with 1% glucose, green: raw milk spiked with 1% lactose.....	270
Figure 7-21 Ammonium ion [150].....	271
Figure 7-22 Raw milk spiked with ammonium chloride. Red: raw milk, blue: raw milk spiked with 0.5% ammonium chloride, purple: raw milk spiked with 1% ammonium chloride.	271
Figure 7-23 Linoleic acid C18:2 cis 9,12 [121].....	272
Figure 7-24 Raw milk spiked with linoleic acid. Red: raw milk, blue: raw milk spiked with linoleic acid 0.14%.....	272
Figure 7-25 Creatine [122].....	273
Figure 7-26 Raw milk spiked with creatine. Red: raw milk, blue: raw milk spiked with 0.5% creatine, purple: raw milk spiked with 1% creatine, green: raw milk spiked with 0.002% (milk mean) creatine.	273

Figure 7-27 Histamine [124].....	274
Figure 7-28 Raw milk spiked with histamine. Red: raw milk, blue: raw milk spiked with histamine 0.28%.	274
Figure 7-29 Orotic acid [126]	275
Figure 7-30 Raw milk spiked with orotic acid. Red: raw milk, blue: raw milk spiked with orotic acid 0.14%.....	275
Figure 7-31 Hippuric acid [128]	276
Figure 7-32 Raw milk spiked with hippuric acid. Red: raw milk, blue: raw milk spiked with hippuric acid 0.3%.	276

List of abbreviations

AD	Adulterated
ANNs	Artificial Neural Networks
ANOVA	Analysis of variance
AOAC	Association of Official Analytical Chemists
AOACI	Association of Official Analytical Chemists International
AS	Ammonium sulfate
ATR	Attenuated total reflectance
AUC	Area under the curve
BHB	β -hydroxybutyrate
BHBA	β -hydroxybutyric acid
CDHI	Canadian Dairy Herd Improvement
CDN	Canadian Dairy Network
CLS	Classical least squares
CQM	Canadian Quality Milk
DA	Discriminant analysis
DFC	Dairy Farmers of Canada
DIM	Days in milk
EMA	Economically motivated adulteration
ER	Error rate
FAO	Food and Agriculture Organization
FAs	Fatty acids
FD	First derivative

FFA	Free fatty acids
FOMs	Figures of merit
FP	Freezing point
FP ratio	Fat-to- protein ratio
FPD	Freezing point depression
FPR	False positive rate
FT	Fourier transform
FTIR	Fourier transform infrared
HCA	Hierarchical cluster analysis
IDT	International Journal of Dairy Technology
ILS	Inverse least squares
IR	Infrared
LCFA	Long-chain fatty acids
LDA	Linear discriminant analysis
LG	Lactoglobulin
LTF	Lactoferrin
LVF	Linear variable filter
MCFA	Mid-chain fatty acids
MCP	Milk coagulation properties
MD	Mean differences
MIR	Mid-infrared
MLR	Multiple linear regression
MSC	Multiplicative scatter correction

MUFA	Mono-unsaturated fatty acids
NA	Not Adulterated
NCFPD	National Center for Food Protection and Defense
NEFA	Non esterified fatty acids
NIR	Near infrared
NN	Neural networks
NPN	Nonprotein nitrogen
P1	Prototype 1
P2	Prototype 2
P3	Prototype 3
PCA	Principal component analysis
PC-DA	Principal component-based discriminant analysis
PC-QDA	Principle component quadratic discriminant analysis
PCR	Principal component regression
PCs	Principal components
PDF	Precision dairy farming
PLS	Partial least squares
PLS-DA	Partial least squares discriminant analysis
PLSR	Partial-least-squares regression
PRESS	Predicted residual sums of squares
PUFA	Poly-unsaturated fatty acids
QDA	Quadratic discriminant analysis
r	correlation coefficient

R ²	Coefficient of determination
RCT	Rennet coagulation time
REP	Repeatability file
RMS	Root mean square
RMSE	Root mean square error
RMSEC	Root mean square error of calibration
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
ROC	Receiver operating characteristic curve
RPD	Ratio performance to deviation
RSD	Relative standard deviation
SARA	Acute and subacute ruminal acidosis
SbC	Sodium bicarbonate
SC	Sodium citrate
SCC	Somatic cell count
SCFA	Short-chain fatty acids
SD	Standard deviation
SDD	Standard deviation of the difference
SEC	Standard error of calibration
SECV	Standard error of cross validation
SEP	Standard error of prediction
SFA	Saturated fatty acids
SIMCA	Soft independent modelling by class analogy

Sn	Sensitivity
SNF	Solids-non-fat
SNR	Signal-to-noise ratio
SNV	Standard normal variate
SOP	Standard operational procedure
Sp	Specificity
SVM	Support vector machines
SW	Stall width
TA	Titrateable acidity
TCL	Tie chain length
TFA	Trans fatty acids
TPR	True positive rate
TR	Tie rail
TUFA	Total unsaturated fatty acids
U	Urea
USP	United States Pharmacopeial Convention
VFAs	Volatile fatty acids
VN	Vector normalized

Chapter 1: Introduction

1.1 General introduction

Quantitative milk analysis by mid-infrared (IR) spectroscopy was developed in the 1960s as a rapid, cost-effective, green and non-destructive analytical method [1, 2], which does not involve the use of hazardous chemicals or sample preparation. Absorption peaks in the mid-IR range at 5.73 μm , 6.46 μm , 9.6 μm and 7.9 μm were used to determine fat, protein, lactose and solids-non-fat (SNF) content in milk, respectively [1]. In the 1990s, Fourier transform infrared (FTIR) spectroscopy was proposed for the quantitative analysis of milk in combination with partial least squares (PLS) regression [3] and later became an official method for milk analysis by IR spectroscopy [4]. Modern FTIR milk analyzers are renowned for their high sample throughput, which reaches 500 samples/h, fast analysis time (i.e., 6-30s), accuracy that is better than 1% relative on the main constituents and precision that is better than 0.5% relative on the main constituents [5]. Accuracy is the closeness of a measured value to the true or accepted value; on the other hand, precision is the closeness of results to others obtained in exactly the same way (i.e., replicates) [6]. These analyzers are currently the cornerstone of the milk recording system that is implemented in various countries to monitor milk quality and verify the commitment of dairy producers to the standards of dairy production and industry. The information that is obtained by the milk recording system contributes to achieve the goals of precision dairy farming (PDF), which is defined as “the use of information and communication technologies for improved control of fine-scale animal and physical resource variability to optimize economic, social, and environmental dairy farm performance” [7]. The main objectives of PDF are maximizing individual animal potential, early detection of disease, and minimizing the use of medication through preventive health measures. Perceived benefits of PDF technologies include increased efficiency, reduced costs, improved product quality, minimized adverse environmental impacts, and improved animal health and well-being [8].

On May 29th, 1995, the Canadian Dairy Network (CDN) was established with mandates that included genetic evaluation services for dairy breeds in Canada and the establishment of industry standards of milk recording services for publishable lactations and for genetic evaluation. These standards are monitored on a herd-by-herd basis and disciplinary sanctions are imposed if these standards are violated by dairy farmers [9]. In 2011, the latest version of the milk recording

standard was published under the title “Dairy Herd Recording Service Standards for Herds Qualifying for Publishable Lactations and/or Genetic Evaluations in Canada” [10]. These standards are applied by two central dairy laboratories of the Canadian Dairy Herd Improvement (CDHI) agency, which are Valacta and CanWest DHI.

Figure 1-1 summarizes the current business model for milk recording systems. Dairy farmers enroll all animals of a herd that have calved at least once, and they subscribe to services offered by the DHI agency. The herd must be tested at least 10 times within a 12-month period [10]. A technician from the agency visits several dairy farms where he/she supervises the collection of milk samples from all individual animals and/or bulk milk tanks. These milk samples are then transported to the central dairy laboratory where they are analyzed by approved milk analyzers. For example, more than 3 million milk samples are shipped to Valacta from Québec and the Maritimes for analysis every year.

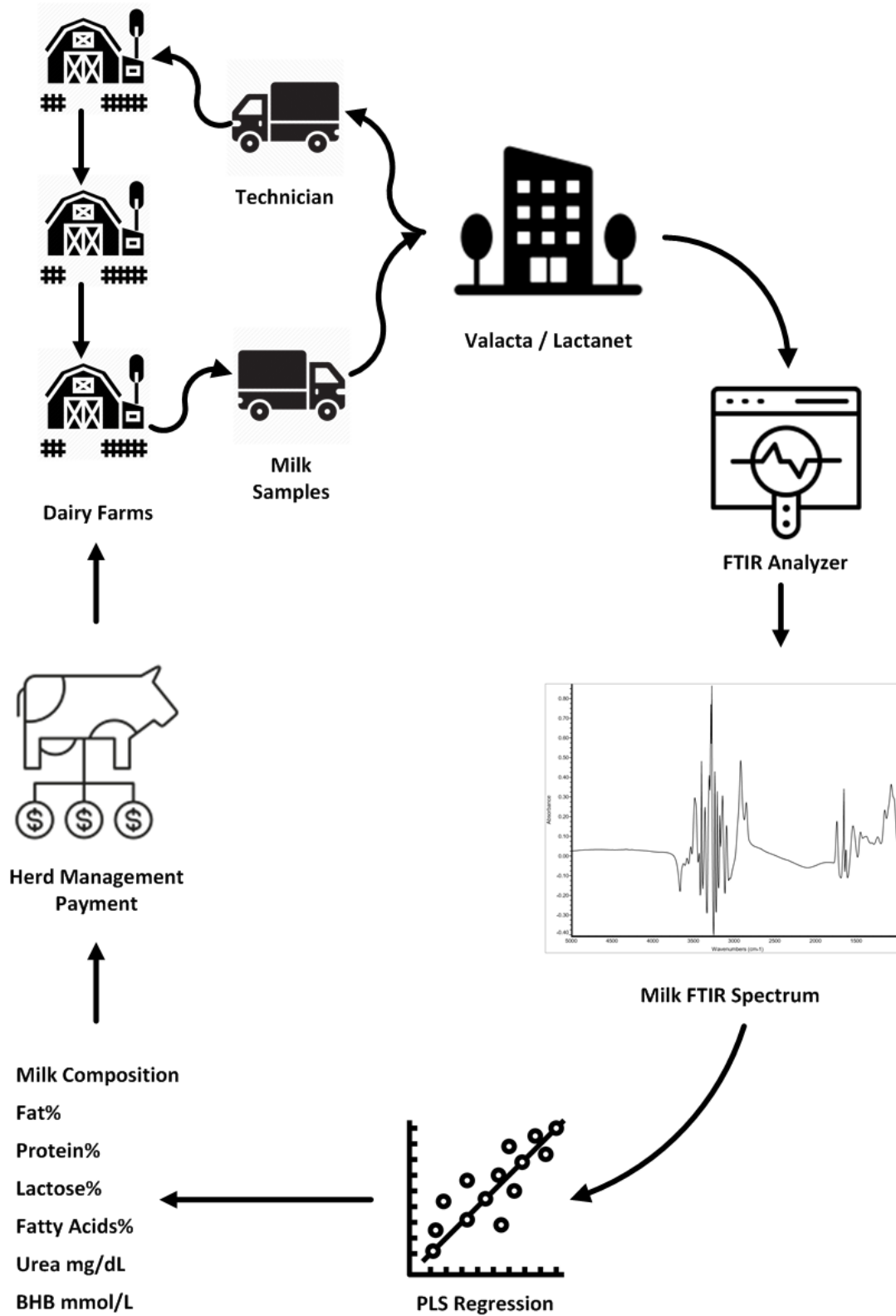


Figure 1-1 Business model of the milk recording system

In the central dairy laboratory, milk analyzers produce an FTIR spectrum for each milk sample. This spectrum is used as an input for multiple PLS regression models to predict the concentrations of specific milk components. The current list of milk components determined by Valacta includes: fat, protein, true protein, lactose, urea, β -hydroxybutyrate (BHB), myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), short-chain fatty acids (SCFA), mid-chain fatty acids (MCFA), long-chain fatty acids (LCFA), saturated fatty acids (SFA), total unsaturated fatty acids (TUFA), mono-unsaturated fatty acids (MUFA), poly-unsaturated fatty acids (PUFA), trans fatty acids (TFA), free fatty acids (FFA), *de novo* fatty acids, mixed fatty acids and preformed fatty acids. It must be noted that the determination of milk fatty acids is not considered standard analysis for monitoring milk composition for the time being. Depending on the concentrations of these milk components, different reports are generated to determine payments to dairy producers and to support dairy producers' decision-making process regarding the management of their herds. Milk fat and protein are the most important milk components and their levels will dictate daily decisions in the dairy production process. Urea is an indicator of the feed ration's protein use efficacy, where increased level of urea indicates a waste of the feed's protein and high level of nitrogen excretion in the urine. BHB is an indicator of ketosis, which is a metabolic illness that occurs at the beginning of the lactation period as a result of transition management issues in which cows rely on body fat reserves to fulfil increased energy needs at the beginning of the lactation for an extended period of time. Somatic cell count (SCC) is also determined in milk samples, which represents an indicator of udder health and milk quality. In addition to milk analysis, DHI agencies provide consultancy and support related to feeding, forage, heifer raising, reproduction, dry off and transition, animal comfort and welfare. Some of these services, such as the assessment of animal welfare, do not rely on milk recording data and they require visits by trained technicians to evaluate animal injuries, quality of the cow's lifts and sets, body condition and lameness.

1.2 Rationale and research objectives

The overall objective of this thesis is to present two propositions that will expand the capabilities of the milk recording system. The first is on-site milk analysis by IR spectroscopy and the second is the exploitation of milk FTIR spectra beyond the paradigm of predicting specific milk components by PLS regression models from those spectra.

Advances in manufacturing of FTIR spectrometers have led to the production of miniaturized portable FTIR spectrometers with reasonable cost. The objective of the research presented in Chapter 3 is to investigate the suitability of a portable FTIR spectrometer equipped with a transmission cell as a sample introduction method for on-site milk analysis, since the official method for milk analysis is currently based on transmission FTIR spectroscopy [4]. On-site analysis of milk will provide dairy farmers with a tool for self monitoring of milk composition without the need to increase the number of transported milk samples to the central dairy laboratory; hence, preventing an increase in the carbon footprint of this process. Milking robots provide data about milk composition; however, their predictions are based on near infrared (NIR) spectra which are characterized by broad and overlapped low intensity bands, between 10 and 100 times attenuated compared to the sharper mid-IR fundamental absorption bands produced by FTIR spectroscopy. The broad peaks in NIR spectra cannot be directly assigned to specific chemical compounds or interpreted in a straightforward manner as mid-IR spectra [11]. In addition, light scattering becomes a more serious issue for electromagnetic radiation in the NIR region due to shorter wavelengths (i.e., 2,500 to 750 nm) than the mid-IR region that are well below the diameter of fat globules in milk, which makes predictions of milk components that are based on NIR spectroscopy less accurate than their FTIR counterparts.

The objective of the research presented in Chapter 4 is to investigate the suitability of a linear variable filter (LVF) array IR spectrometer and attenuated total reflectance (ATR) as sample introduction method for on-site milk analysis. Linear variable filter arrays IR spectrometers are a novel type of IR spectrometers that do not rely on interferometry to resolve wavelengths in the mid-IR range. The absence of moving parts makes them good candidates for on-site milk analysis instruments. In addition, ATR provides a practical sample introduction method, especially for raw milk analysis. It eliminates the need to pump milk through a transmission cell with a micrometric path length; hence, avoiding issues related to cell clogging resulting from the size of milk fat globules in raw milk.

The objective of the research presented in Chapter 5 is to present a case study where on-site milk analysis by IR spectroscopy can provide a solution for a major issue that faces the dairy industry in Brazil, which is the addition of extraneous water and chemical adulterants to milk. In Brazil, milk cryoscopy is the official method for the detection of added water to milk; however, the

addition of some chemicals with the added water restores the readings of milk cryoscopy instruments to their legal values. IR spectroscopy combined with multivariate discriminate algorithms provides a viable solution for differentiation of genuine milk samples from watered-down ones.

The objective of the research in Chapters 6 and 7 is to evaluate multivariate algorithms to study the effects of different dairy cattle housing treatments that are intended to improve animal welfare on milk composition using milk FTIR spectral data directly without relying on predictions of concentrations of specific milk components. Data mining milk FTIR spectra to detect and analyze the spectral fingerprint of the welfare status of the animal on milk composition is a novel approach that will extend the capabilities of milk FTIR spectroscopy beyond the paradigm of predicting specific milk components by PLS regression models. Currently, milk FTIR spectra are not employed in the assessment of animal comfort and welfare.

Chapter 2: Literature review

2.1 Introduction

Precision dairy farming (PDF) is defined as “the use of information and communication technologies for improved control of fine-scale animal and physical resource variability to optimize economic, social, and environmental dairy farm performance” [7]. The main objectives of PDF are maximizing the economical return from dairy farming and minimizing cost of production and animal distress. This approach will lead to improved milk quality, minimal environmental impact and improved animal health and well-being [8].

Borchers and Bewley (2015) conducted a survey to assess the adoption, perception, effectiveness, and use of PDF technology by sending a questionnaire to 90 farms in the United States of America and 19 farms in other countries that included: Australia, Canada, India, Iran, Israel, Mexico, New Zealand, and the United Kingdom [12]. The authors’ list of parameters monitored by PDF technologies included: daily milk yield, milk components, step number, body temperature, milk conductivity, automatic estrus-detection monitors, and daily body weight measurements. In addition, the authors included proposed PDF parameters, such as jaw movements, ruminal pH, reticular contractions, heart rate, animal positioning and activity, vaginal mucus electrical resistance, feeding behavior, lying behavior, odor, glucose, acoustics, progesterone, individual milk components, milk color, infrared udder surface temperatures, and respiration rates [13].

According to the results of this survey, 68.8% of respondents indicated the use of technology on their dairy farms versus 31.2% not using technology at all. The percentages of producers measuring parameters by the implemented technologies were: 52.3%, 41.3%, 25.7%, 24.8% and 21.1% for daily milk yield, cow activity, mastitis, milk components (e.g., fat, protein and somatic cell count SCC) and standing estrus, respectively. In terms of parameter usefulness, the percentages of dairy producers who perceived the following parameters as useful were: 80.7%, 79.8%, 78.7%, 74.3%, 59.4% and 53.2% for daily milk yield, standing estrus, mastitis, cow activity, temperature and milk components (e.g., fat, protein and SCC), respectively. On the other hand, the surveyed dairy producers considered the benefit-to-cost ratio as the most important factor when making purchasing decisions regarding PDF technologies followed by total investment cost, simplicity and ease of use, proven performance through independent research, availability of local

support, compatibility with existing dairy practices and systems and time involved in using the technology [13].

2.2 Importance of monitoring milk composition

Monitoring the chemical composition of milk is currently implemented as a tool to manage dairy herds on dairy farms. Central dairy laboratories, such as Valacta (Sainte Anne de Bellevue, QC, Canada), are using FTIR milk analysers to determine milk fat, protein, lactose, urea, β -hydroxybutyric acid (BHBA or BHB), acetone, milk fatty acids and somatic cell count (SCC) in milk samples that are routinely collected from dairy farms [14]. Changes in milk composition of individual animals can be considered as an indicator of the animal's metabolic status or the efficacy of the feed management system. The early detection of these changes will allow dairy farmers to detect metabolic and management problems in early stages; hence, preventing any health disorders that might affect the well-being and productivity of the individual animal by taking preventive measures as early as possible.

Ketosis and mastitis are among the most economically relevant diseases in dairy cows [15]. At the beginning of lactation, the energy demand in dairy cows increases substantially. This energy demand is met by increasing feed intake and by fat mobilization from the cow's adipose tissue. When feed does not meet the increased energy demand, the cow can go through a period of negative energy balance. Increased dependency on body fat in providing the increased body energy needs results in increased production of major ketone bodies, such as acetone, acetoacetate and BHB; thus, their levels are increased in blood (i.e., ketonemia), urine (i.e., ketonuria), milk (i.e., ketolactia) and other body fluids. This case is referred to as Ketosis and it can be classified as clinical and subclinical, which causes economical losses to the dairy industry. Since cow blood sampling is not a convenient procedure for farmers, determining BHBA levels in milk is a viable alternative. In milk, BHBA concentrations that are <100 mol/L have been reported to be normal [16]. Other researchers have indicated that negative energy balance affects the fatty acids (FA) composition of milk fat in the early stages of lactation. Negative energy balance will increase the proportions of Omega-9 (C18:1 *cis*-9) and long chain fatty acids (LCFAs), notably C16:0 and C18:0, and will decrease the proportion of medium chain fatty acids (MCFAs) [15]. Other milk components will also be affected by negative energy balance or ketosis. In this case, protein, fat, milk fat-protein ratio, and lactose will be $<2.9\%$, $>4.8\%$, >1.4 and $<4.5\%$, respectively [15].

However, these thresholds should be cautiously used taking into consideration the characteristics of the herd in question.

Mastitis is another health problem that alters milk composition significantly. Milk secreted from inflamed mammary glands is characterized with high SCC, increased free fatty acids (FAs) concentrations, reduction in casein combined with an increase in whey protein, reduction in lactose concentration, changes in the concentration of minerals such as sodium, chloride, potassium and calcium, and increase in milk pH [15]. SCC can be elevated in healthy cows' milk, where macrophages and neutrophils constitute 66-88% and 1-11%, respectively. However, neutrophils will increase to almost 90% in case of inflammation of mammary glands. The proposed SCC threshold values for the Canadian Holstein population are: 500,000, 300,000 and 200,000 cells/mL for the following days in milk (DIM) classes: 5 to 10, 11 to 30, and 31 to 305 DIM, respectively. In addition, researchers reported an increase in bovine Lactoferrin (LTF) from 0.1-0.4 g/L up to 2.3 g/L in mature milk secreted from healthy and inflamed mammary glands, respectively [15]. Mastitis also decreases the lactose content of milk. Lactose is considered one of the most stable components of milk with a very low day-to-day variation of 0.9% compared to 7.7%, 7.0% and 2.0% for fat content, milk yield and SCC, respectively. Hence, a sudden decrease in lactose content that exceeds the day-to-day variation limit might be an early indication of mammary gland inflammation [15].

Acute and subacute ruminal acidosis (SARA) is another important health issue in dairy cows. It is associated with the accumulation of lactic acid and volatile fatty acids (VFAs), respectively, in the rumen and the subsequent decrease in the ruminal pH for several hours per day [17] that results from feeding high grain diets that are low in fiber to high yielding dairy cows under intensive livestock production systems that are adapted to digesting forage diets [18, 19]. In these cases, depressed milk fat [14, 18], reduced milk protein and increased milk non-protein nitrogen NPN [20] have been reported. In addition, decreased ruminal pH alters the biohydrogenation pathway of linoleic acid and increases the production of *trans-10* C18:1 fatty acid; thus, more trans fatty acids are absorbed, even if the intake of unsaturated fatty acids is not necessarily high [18, 21].

In addition to the animal's health and metabolic state, milk composition, specifically urea content, can provide indications about nutritional management, production, and economic variables in commercial dairy herds. Urea in milk originates from three sources, which are protein

decomposition, the digestion of NPN and the catabolism of amino acids in the mammary glands. True protein and NPN represent 95% and 5%, respectively, of milk protein and urea represents 30-35% of milk NPN [15]. Average herd milk urea concentrations are positively related to the dietary crude protein and rumen degradable protein; thus, milk urea concentration provides indication about the efficacy of nitrogen utilization in dairy cattle. Feeds high in protein lead to higher feed costs, environmental pollution and fertility problems. On the other hand, a very low milk urea content could indicate protein deficiencies in the diet of dairy cattle, potentially leading to a loss of production. The acceptable range for urea concentration in milk is between 200 and 400 mg/L or 20-40 mg/dL [15].

To summarize, milk composition provides rich information about the health, metabolic and nutritional state of cows and it can be employed as an on-site management tool for sustainable dairy production. However, dairy producers do not perceive milk composition as an important parameter as they do for milk yield [13], for example. One of the probable reasons for this perception is the lack of an on-site milk analyzer that is affordable and that can rapidly analyze milk with an acceptable accuracy and precision.

2.3 Milk analysis by infrared spectroscopy

Quantitative milk analysis by Fourier transform infrared (FTIR) spectroscopy is currently an official method of the Association of Official Analytical Chemists (AOAC) International [4], which is extensively used for producer payment, herd milk analysis and routine quality control in the dairy industry [22]. In 1964, Goulden described milk analysis by infrared (IR) spectroscopy [1]. He demonstrated that milk components absorb IR energy at specific wavelengths and that the intensities of the absorption peaks can be used for quantitative determination of these components. The reported wavelengths that were used to determine major milk components were 5.73 μm , 6.46 μm , 9.6 μm and 7.9 μm for fat, protein, lactose and solids-not-fat (SNF) content, respectively. In wavenumbers*, these values are equal to 1745.20 cm^{-1} , 1547.99 cm^{-1} , 1041.67 cm^{-1} and 1265.82 cm^{-1} , respectively. Goulden demonstrated that the absorbance intensities of wavelengths assigned to fat and lactose are a linear function of their concentrations, while the absorbance intensity of wavelength assigned to protein has a significant contribution from fat. For this reason, a correction equation was derived to calculate the concentration of protein at the corresponding wavelength. In

* The following equation was used to convert from wavelength to wavenumber: $y \text{ cm}^{-1} = 10,000,000 / x \text{ nm}$.

addition, it was demonstrated that scattering of light by fat globules is proportional to particle-size-to-wavelength ratio. Hence, attenuation of the IR beam due to scattering can be eliminated by decreasing fat globules size through milk homogenization. Temperature of milk samples was another important factor that was proved to affect the IR spectra of milk samples. The optimum reported temperature for recording milk IR spectra is $40\text{ }^{\circ}\text{C} \pm 1$. Change of $1\text{ }^{\circ}\text{C}$ led to a 1% change in the total signal. On the other hand, transmission cells with $40\text{ }\mu\text{m}$ path length and calcium fluoride windows were found to be the optimum measurement cells [1].

In 1967, Biggs described an IR milk analyser [2]. The reported precision of measurement of milk analysis by this analyzer is $\pm 0.03\%$ and the standard deviations between IR and chemical methods are $\pm 0.06\%$, $\pm 0.07\%$ and $\pm 0.06\%$ for fat, protein and lactose, respectively. IR milk analysis is rapid and less expensive than other laboratory chemical methods that are used for the determination of milk components. With an infrared milk analyzer (IRMA) two technicians can prepare samples and complete analyses for fat, protein, and lactose at the rate of one sample per minute. According to Biggs's calculations at that time, the labor cost for three component analyses is about 7 cents and the instrument cost per sample is less than 2 cents; thus, the cost of three component analyses is less than 10 cents per sample. This cost is less than one-tenth the cost of the equivalent analyses by accepted chemical methods. The IR milk analyzer, which was described by Biggs, consisted of a double beam IR spectrometer that produced two IR beams. One beam passes through a transmission cell containing milk and the other passes through a transmission cell containing distilled water. Both cells share equal optical path. After going through the cells, the IR beams enter a monochromator that includes a diffraction grating and a potassium bromide (KBr) prism, then radiation, within a narrowly selected range of wavelengths, is focussed on to a thermocouple detector. Biggs used wavelengths identical to those reported by Goulden. A micro-switch system was used to automatically switch between those wavelengths to obtain results for each of milk components. In addition, milk samples were homogenized and heated to a constant temperature before being pumped into the measurement cell. Milk fat homogenization is considered a crucial factor for the accurate determination of fat levels in raw milk samples. According to Biggs, raw milk fat should be homogenized once at 3000 psi or four times at 1500 psi. Biggs also suggested the use of sonic homogenization for this purpose. Another important factor that affects the accuracy of the obtained signal is temperature. According to Biggs, both milk sample and distilled

water should be heated to the same temperature, and the machine should be placed in a temperature-controlled environment to prevent any fluctuations in signal obtained from it. In addition, the milk cell should be flushed with at least 25 mL of the current sample in order to completely purge the remains of the previous one to avoid signal divergence [2].

In 1991, Van de Voort and Ismail proposed the use of Mid-FTIR spectroscopy for simultaneous multi-component proximate analysis of food to determine the levels of major food components, such as fat, protein and carbohydrates [22]. At that time, the application of IR spectroscopy in quantitative food analysis was limited due to the presence of water that has a tremendous IR absorbance, the necessity to use cells of short path length to minimize IR absorption by water and difficulties related to sample preparation and handling. FTIR spectroscopy differs from diffraction technology in the way different wavelengths are resolved. FTIR spectroscopy is based on interferometry, which records absorbance intensities at all wavelengths simultaneously. In an interferometer, an IR beam is divided into two beams by a beam-splitter. One of them is reflected to a fixed mirror and the other to a moving mirror. When the two beams recombine at the beam splitter, they undergo constructive and destructive interferences due to the differences in the path lengths from the two mirrors. After going through the sample, the IR energy reaches a detector, get digitized and an interferogram is produced. Fourier transform (FT) is then applied to the interferogram to convert the signal from time domain to frequency domain to produce a conventional IR spectrum [22].

In general, FTIR spectroscopy has several advantages. First, the IR beam does not lose its intensity because it does not go through a prism, grating or a slit; hence, the throughput or the Jacquinot's advantage is achieved [23]. Another advantage is the multiplexing or Fellgett's advantage, in which the signal-to-noise ratio (SNR) is improved by taking multiplexed measurements. In FTIR spectroscopy, SNR is proportional to the square root of the number of coadded scans of a specific spectrum [23]. In addition, FTIR spectrometers contain a laser that acts as an internal wavenumber standard, which means that the precision of wavenumbers in FTIR spectrum is $\pm 0.01 \text{ cm}^{-1}$ and that makes peak positions highly reproducible in FTIR spectrum [23]. This precision enables data manipulation operations such as spectral subtraction, spectral addition, the calculation of spectral ratios, spectral stripping, and the addition of scans [22]. Several factors might interfere with FTIR spectroscopy, such as Tyndall scattering, reflection losses, refractive index, temperature, sample

non-homogeneity and detector nonlinearity [22]. These effects can be minimized by maintaining the absorbance between 0.2 and 0.7, using environment-insensitive absorption bands, using peak areas rather than peak height and using more wavenumbers or calibration samples [22].

The IR bands that were proposed to determine the major food components were: the triglyceride $C = O$ ester linkage band and the $C - H$ stretching band to determine fat; the amide I and/or amide II bands to determine proteins; and the hydroxyl bending band to measure carbohydrates [22]. Water absorbs strongly across the IR spectrum; however, with the high spectral precision and spectral data manipulation routines available with FTIR, water can be subtracted out of the spectrum to reveal the spectrum associated with the other components [22]. For FTIR milk analysis, table 2-1 summarizes wavenumbers and band assignments that can be used to determine milk components.

Table 2-1 Band assignments in the mid-IR range for FTIR milk analysis

Wavenumbers cm^{-1}	Band assignment	Associated milk component	Reference
967	C=C–H bend and trans-C=C bonds	Fat	[24]
1400-800	Highly coupled stretching / bending modes	Carbohydrates mainly lactose	[24]
1100-1060	O=P–O stretch	Phosphate	[25]
1280-1200	Amide III	Protein	[25]
1565–1520	Amide II	Protein	[24]
1700–1600	Amide I	Protein	[24]
1650	H–O–H bend	Water	[24]
1745–1725	C=O stretch, triglyceride ester linkage	Fat	[24]
2300-2400		CO ₂ and water vapour	
2980–2800	C–H stretch	Fat	[24]
3007–3012	C–H stretch / cis-C=C bonds	Fat	[24]
3600–3200	O–H stretch	Water	[24]

2.4 Multivariate regression for milk analysis by FTIR spectroscopy

Several multivariate regression algorithms have been reported to develop calibration models for the determination of milk components' concentrations from milk FTIR spectra. These techniques include classical and inverse least squares regression (CLS and ILS), principal component regression (PCR) and partial-least-squares regression (PLSR) [26]. In CLS, the overall absorbance at a specific wavenumber is assumed to be a linear summation of absorptions of all components in the sample. Mathematically, it can be represented as follows [26]:

Equation 2-1

$$A_j = k_{0j} + \sum_{i=1}^l c_i k_{ij}$$

Where A_j is the overall absorbance at wavenumber j , c_i is the concentration of component i , l is the number of components present in the sample, k_{ij} is the product of the optical path length and the absorption coefficient of component i at wavenumber j . For calibration set that consists of multiple samples m , the formula can be represented in matrix notation as follows [26]:

Equation 2-2

$$A = CK$$

Where A is an $m \times n$ absorbance matrix, m is the number of samples or spectra, n is the number of wavenumbers in an individual spectrum and K is an $l \times n$ matrix of the product of the optical path length and the absorption coefficient for component l at wavenumber n . During the calibration process, matrix K is calculated, which requires an inversion operation for the product of $C^T C$ [26].

Equation 2-3

$$K = (C^T C)^{-1} C^T A$$

During the prediction operation, matrix C is calculated, which requires an inversion operation for the product of KK^T [26].

Equation 2-4

$$C = AK^T (KK^T)^{-1}$$

The dimensions of both matrices are $l \times l$. If the components' concentrations are collinear, the determinant of the matrices will become zero and they will not be invertible. Another drawback of

this algorithm is that CLS requires that concentrations of all components contributing to the absorbance at a specific wavenumber to be known, which cannot be achieved [26].

In ILS, which is also known as multiple linear regression (MLR), concentration of a specific milk component is assumed to be a function of absorbances at all wavenumbers in the spectrum and it is represented mathematically as follows [26]:

Equation 2-5

$$c_i = p_{i0} + \sum_{j=1}^n A_j p_{ij}$$

Where c_i is concentration of component i , A_j is absorbance of wavenumber j and p_{ij} is the proportionality constant of component i at wavenumber j . For the calibration set, the formula can be rewritten in matrix notation as follows [26]:

Equation 2-6

$$C = AP$$

During the calibration process, the product of $A^T A$ is inverted to calculate matrix P , which is then used in equation 2-6 to calculate the concentration C for unknown sample [26].

Equation 2-7

$$P = (A^T A)^{-1} A^T C$$

This inversion operation will only be possible if the number of samples equals or exceeds the number of wavenumbers. For this reason, ILS is not commonly used to develop prediction models based on full FTIR spectra due to the immense number of samples required to build such a model [26]. To overcome the obstacle of number of samples, a variable reduction method is applied to the spectral matrix. In PCR, principal component analysis (PCA) is applied to the spectral matrix and a new set of variables, called principal components (PCs), is calculated. These new variables are linear combinations of the spectral wavenumbers and they describe the same variation structure that is described by the original spectral matrix; however, only the first few PCs describe much of the variation and they can be used as predictors in a regression model. The spectral data matrix is decomposed into two matrices as follows [26]:

Equation 2-8

$$A = TB$$

Where B is an $h \times n$ loading matrix of h principal components and T is an $m \times h$ scores matrix. The loadings can be considered equivalent to regression coefficients and the scores are the new coordinates for each sample in the new principle component space. The scores matrix T' is then used in a regression process against milk component's concentration that is similar to the one of ILS to produce a prediction model for a milk component [26].

In PLS, the predictors matrix X, which is an $m \times n$ matrix, contains the spectral data where m is the number of the spectra or samples and n is the number of spectral variable or wavenumbers. The responses matrix Y, which is an $m \times k$ matrix, contains the concentrations of k milk components for m samples. A variable reduction process is applied to matrix X to extract new predictors that are referred to as latent variables, components or factors, which are linear combinations of the spectral variables (i.e., wavenumbers) [27]. These factors represent latent structures that maximize the covariance between X and Y; in other words, the PLS factors are calculated based on their relevance to Y, which means that the first factor extracted from matrix X explains the most variation in matrix Y and not necessarily the most variation in matrix X [27]. In PCR, PCs are calculated independently from the Y matrix and the first PC explains the most variation in matrix X, and that is the main difference between PLS and PCR. In PLS prediction model, matrix X is decomposed as follows:

Equation 2-9

$$X = TP'$$

Where T is the X scores matrix and P is the X loadings matrix. The Y matrix is then modeled on the X scores using linear regression [27]. The Y matrix can also be decomposed, if it contains multiple responses (i.e., the concentrations of multiple milk components):

Equation 2-10

$$Y = TQ'$$

Where T is the Y scores matrix and P is the Y loadings matrix. In this case, PLS is a powerful tool to study the relationship between the multiple responses. In FTIR spectroscopy, PLS is the preferred multivariate regression algorithm for developing prediction models because PLS can

deal effectively with collinear, noisy and wide data where the number of variables (i.e., wavenumbers in a spectrum) exceeds the number of observations (i.e., spectra of samples) [27]. However, it must be noted that PLS models are highly sensitive to systemic variations in FTIR spectra that are not related to the chemical composition of milk samples, such as differences in performance from one equipment manufacturer to another, homogenization efficiency, preservative type and concentration, and sample temperature [28]. For this reason, models are only considered realistic if their X loadings show high correlations with the spectral features related to the analyte of interest.

2.5 Spectral pre-treatments or preprocessing

In FTIR spectroscopy, a regression model is the inverse of Beer's law. It can be described mathematically as follows [29]:

Equation 2-11

$$\hat{y} = b_0 + \sum_{i=1}^n b_n x_n$$

Where $\hat{y}$ is the predicted concentration of the analyte of interest, x_n is the absorbance of wavenumber n and b_n is a regression coefficient that can be interpreted as a combination of the optical path length and the molar absorptivity coefficient ϵ that is modeled from the data. Considering this definition, spectral preprocessing can be defined as mathematical operations aiming at enhancing the chemical information represented by ϵ and reducing the physical effects of the prediction process in the spectral data, such as path length variability [29].

Smoothing

Smoothing aims at reducing noise in an FTIR spectrum that results from the instrumental measurement without reducing the number of variables (i.e., wavenumbers) in a spectrum; hence, it is a row-oriented transformation. Common smoothing transformations include moving average and Savitzky-Golay smoothing. In moving average smoothing, a central data point is replaced by the average absorbance values in a selected smoothing window in an FTIR spectrum. In Savitzky-Golay smoothing, data points in the averaging window are fitted to a polynomial and then the central point is replaced with the fitted value [29].

Normalisation

Normalisation is a row-oriented transformation that aims at putting all spectral variables (i.e., wavenumbers) on an even footing by rescaling a spectrum to a common sum, such as 1.00 or 100%. In area normalization, the sum of absolute absorbances of all spectral variables in a spectrum is calculated, which gives the scaling factor, and then each spectral variable is divided by it. In range scaling, the scaling factor is calculated for a specific spectral range. In peak normalization, the entire spectrum is scaled by the height of a specific peak of a known internal standard. In maximum normalization, the scaling factor is the maximum absorbance in the spectrum [29]. In vector normalization, a spectrum is divided by its norm, which is the square root of the sum of the squared values of all spectral variables. In this case, the spectrum is normalized to unit length vector [30]. Vector normalization is useful for pattern recognition applications [30]. Normalization after derivatization reduces path length variations [29].

Baseline correction

Baseline correction aims at subtracting a common offset from spectra to better overlay each other. The offset is an additive effect caused by path length differences of the measurement system or by differences in sample density. In cases of sloping linear baseline, a two point baseline correction method is applied, while the method of detrending can correct polynomial baseline shifts [29].

Derivatives

Derivatization or differentiation is calculated as the difference between the absorbance of a point in a spectrum and the absorbance of the previous point divided by a constant centring factor. The derivative centers the spectrum around the zero line and it measures the slope or the rate of change of the spectral variables, which means it can reveal more details in the spectral data, especially in regions where small peaks are present on shoulders of bigger ones. In a quadratic polynomial, the constant term represents the intercept, or the offset, of the equation and it disappears under first derivation, which means that the first derivative can eliminate additive effects in the spectral data. In addition, the first derivative removes the baseline and slope effects. However, in first derivative the maxima of an absorption band become zero, which makes the interpretation of the first derivative difficult. Second derivative defines the curvature in the spectral data and is used to correct for the quadratic baseline effects in near infrared spectroscopic data. Second derivative is

sometimes preferred over first derivative because the maxima of absorption bands becomes minima; hence, second derivative becomes easier to interpret than first derivative. In simple difference derivative, in which the constant centring factor is 1, noise level is increased in the spectral data. This noise originates from instrument electronics and mechanical vibrations, especially for second derivative. To reduce the noise that results from derivatization, segment-gap or Savitzky-Golay derivatives are used where a modified moving average or Savitzky-Golay smoothing algorithms, respectively, are applied to a spectrum before calculating the derivative [29].

Correcting multiplicative effects

Multiplicative effects result mainly from light scattering and they are wavelength specific, which makes them more of a concern in near infrared spectroscopy rather than mid-IR spectroscopy. Two algorithms are commonly used to correct for scattering effect, standard normal variate (SNV) and multiplicative scatter correction (MSC). SNV is a row-oriented algorithm where the mean and standard deviation are calculated for a specific spectral range, then spectral variables are centered by the calculated mean and divided by the calculated standard deviation. Mean centering removes the additive effects and division by the standard deviation reduce the residual scatter effect. MSC is a model-based scatter correction method that requires a training set to model the scatter effect and to calculate the mean spectrum of the training spectral set by least squares fitting process. The calculated parameters are used to modify a new spectrum to make it as similar as possible to the calculated mean spectrum of the training set. By doing so, multiplicative and additive effects are eliminated [29].

2.6 Calibration and validation of IR milk analyzers

Before calibrating IR milk analyzers, several procedures must be applied to verify that the instrument is in a good working condition, mechanically and electronically, and the readings are stable and optimized. These procedures are applied to filter based and FTIR milk analyzers and they include checks for: flow system integrity, homogenization efficiency, water repeatability, zero shift, linearity, primary slope, milk repeatability, purging efficiency and establishment of inter-correction factors for filter based milk analyzers [31]. Inter-correction factors are calculated to account for the water displacement effect on the absorption intensities in the IR spectrum and to account for the absorption contribution of other milk components to the intensity of a

wavelength, or a wavenumber, assigned to a specific milk component. For FTIR milk analyzers, this step is not needed if the analyzer uses PLS prediction models to determine the concentrations of milk components.

Among these procedures, verifying homogenization efficiency is the most crucial one. Milk fat globules cause light scattering, which leads to increased analytical error and poor repeatability. To reduce light scattering effect on milk fat predictions, raw milk is homogenized to reduce the size of fat globules to 1 μm or less [31]. For milk analyzers that contain high pressure homogenizer, raw milk is run through the milk analyzer homogenizer and the fat measurement is recorded. The milk that exists the analyzer, which is now homogenized, is collected and passed again through the homogenizer and the fat measurement is recorded a second time. Homogenization is considered adequate if the difference in readings between the first and second passes of the milk is $<0.05\%$ [31]. Another approach is the use of a laser light-scattering particle size analyzer capable of determining particle size distribution. Homogenization efficiency is determined by the value of $D(0.9)$, which is the mean globule diameter below which 90% of the fat volume is contained. If $D(0.9)$ is greater than 2 μm then homogenization is not acceptable, while $D(0.9) > 1.70 \mu\text{m}$ indicates wearing out of homogenizer [31].

As it was reported earlier, specific wavelengths or wavenumbers are used to determine the concentrations of specific milk components. These wavelengths or wavenumbers are 5.723 μm (1747.34 cm^{-1}), 3.48 μm (2873.56 cm^{-1}), 6.465 μm (1546.79 cm^{-1}) and 9.610 μm (1040.58 cm^{-1}) and they are assigned to Fat A, Fat B, protein and lactose, respectively. The Fat A and Fat B represents the IR bands of the carbonyl stretching in the ester linkage and the CH stretching in milk fat, respectively. In addition, 5.6 μm (1785.71 cm^{-1}), 3.6 μm (2777.78 cm^{-1}), 6.7 μm (1492.54 cm^{-1}), and 7.7 μm (1298.70 cm^{-1}) are used as reference wavelengths for Fat A, Fat B, protein and lactose, respectively. Absorption intensities of each milk component are ratioed against their respective reference wavelength to account for the effects of water absorption and light scattering [31]. Typically, 10 to 14 milk samples are used to calibrate milk analyzers. The chemical composition of these samples must always be determined by wet chemistry reference methods. These reference methods are: Kjeldahl protein N; methods 991.22, 991.23, modified Mojonnier ether extraction; method 989.05, oven dry; methods 990.20, 990.19 and by difference; method

990.21 for milk true protein, fat, total solids, and SNF, respectively. Lactose is determined either by difference or by spectrophotometric enzymatic analysis. Lactose by difference is calculated as

Equation 2-12 Calculating milk lactose by difference

$$\% \text{ anhydrous lactose} = \%TS - (\%fat + \%true \text{ protein} + 0.19 + \%ash)$$

where 0.19 is the NPN factor to convert true protein to crude protein (CP) and % ash is either directly measured (method 945.46) or indirectly calculated by the following equation:

Equation 2-13 Calculating ash in milk

$$\%ash = (\%true \text{ protein} \times 0.0596) + 0.5379$$

The spectrophotometric enzymatic analysis of lactose is based on AOAC method 984.15 modified by weighing all volume additions. All these methods are AOAC final action methods [32]. A 73:27 ratio of Fat B and Fat A is suggested to be used for fat predictions to give better agreement with reference chemistry than either Fat B or Fat A [33] for analyzers that do not use PLS calibration models. Distilled water containing 0.01% Triton X-100 is used as a zeroing solution and all zeroing solutions and milk samples must be maintained at $41 \pm 1^\circ\text{C}$ [31].

Calibration milk samples should have an orthogonal matrix of fat, protein, and lactose concentrations to eliminate high and moderate leverage samples. Leverage is an index of the relative contribution of each point to the regression line and it is considered large if it is $>2p/n$ and very large if it is $>3p/n$, where n is the number of data points and p is the number of predictors. For example, leverage can be considered very high, high, moderate and low if it is >1.000 , >0.500 , >0.333 and <0.333 , respectively. Samples with narrow range of concentration of components, nonuniform distribution of concentrations within the range, and with positive correlation between fat and protein content produce conditions where individual samples have significant influence on the calibration [34]. Producer raw milk samples can be used for calibration of milk analyzers in addition to modified milk samples. Milk components can be separated from each other and recombined together in specific ratios to produce modified milk samples whose matrices are orthogonal to each other to avoid having high leverage samples in the calibration set [34]. Calibration samples can be preserved by adding an aqueous 6.7% potassium dichromate solution at a level of 3 mL per 1,000 g of milk to achieve a final concentration in milk of 0.02% potassium dichromate [34]. It has been proved that Potassium Dichromate $\text{K}_2\text{Cr}_2\text{O}_7$ has little or no effect on

mid-IR test results and a bronopol-based preservative can also be used to preserve milk calibration samples [35]. Bronopol does not exhibit absorption bands in mid-IR regions related to milk composition. The pasteurized modified milk calibration samples, preserved with potassium dichromate, have 28 days shelf life and the raw producer milk calibration sets, preserved with potassium dichromate, have 15 days shelf life at 4 °C. Chemical preservatives only inhibit the microbial growth in milk samples, they do not inhibit the activity of milk native enzymes. However, the influence of lipolysis and proteolysis on milk components' predictions is either not significant, or significant but very small [36].

Calibration models are verified by testing an independent set of validation milk samples and comparing the predicted milk component to the chemical reference tests of these samples. The mean difference (MD) and standard deviation of the difference (SDD) for each milk component of the validation set is calculated. If MD is large, then the prediction model is biased. A positive MD value indicates overprediction and a negative MD value indicates underprediction. The official first action method for fat, lactose, protein, and solids in milk by IR spectroscopic method using fixed-filter wavelengths (AOACI, 2000; method 972.16, 33.2.31) indicates that the MD between instruments and reference method values should be $\leq 0.05\%$ for fat, protein, and lactose, and $\leq 0.09\%$ for total solids [37].

2.7 Novel applications in milk analysis by infrared spectroscopy

Milk analysis by infrared spectroscopy is currently an active field of research. New studies have been published on improving the predictions of milk components that are currently being determined by commercial milk analyzers. Other studies aimed at extracting new traits from the milk infrared data and at correlating the spectral data with indicators of interest for the dairy industry.

2.7.1 Prediction of milk fatty acids and milk fat composition

Gas chromatography (GC) is the standard method for quantifying FAs content in milk, where FAs are extracted, saponified and trans-methylated. Infrared analysis was developed as an alternative cost-effective and fast method to determine FAs during milk routine analysis [38]. Currently, the method of choice used for relating milk FAs to spectral data is PLS without any pre-treatment of the spectral data. The spectral regions that are used are located between 1736 and 1805 cm^{-1} and between 2823 and 3016 cm^{-1} [39].

Rutten *et al.* (2009) concluded that increased number of samples increases predictability of milk fat composition; however, it does not increase the predictability of FAs with low concentrations. This can be explained by the fact that increased number of samples will model better the relationship between milk fat and the major FAs present in milk triglycerides. In addition, it was concluded that the effect of season on validation coefficient of determination R^2 was limited but was occasionally large on prediction bias. It was also concluded that FTIR spectroscopy can be used to predict major FAs, combined groups of fatty acids, and the ratio of saturated to unsaturated fatty acids [40].

Afseth *et al.* (2010) demonstrated that the bands that are used in FTIR determination of FAs are: around 3010 cm^{-1} mainly relates to fatty acid unsaturation (cis =C–H stretch), region between 2800 cm^{-1} and 3000 cm^{-1} is related to the symmetric and asymmetric C–H stretch of methyl and methylene groups, strong band around 1745 cm^{-1} is related to C=O stretch, C–H deformations constitute dominating features between 1440 cm^{-1} and 1300 cm^{-1} , between 1200 cm^{-1} and 1000 cm^{-1} constitute the C–O stretch vibrations and a band around 966 cm^{-1} is related to the trans =C–H out-of-plane stretch, which has been proven to be an important feature in the quantification of trans contents of foods in general [41].

Soyeurt *et al.* (2011) investigated six mathematical approaches to improve the FA predictions in milk collected from multiple breeds (dual-purpose Belgian Blue, Holstein, Jersey, Normande, Montbeliarde, and Red and White), different European countries (Belgium, Ireland, UK) and different production systems. These approaches were: raw spectra and PLS, raw spectra in combination with PLS and repeatability file (REP), first derivative of spectral data and PLS, first derivative and REP and PLS, second derivative of spectral data and PLS and second derivative and REP and PLS. The repeatability file was generated by recording mid-infrared (MIR) spectra of several milk samples provided by different spectrometers and each spectrum was centered by subtracting the average of all spectra for the samples included in the repeatability file. The samples in the repeatability file were then used to extend the initial calibration set. This method decreased the repeatability error. Methods were compared using the cross-validation coefficient of determination (R^2_{cv}), the ratio of standard deviation of GC values to the standard error of cross-validation (RPD), and the validation coefficient of determination (R^2_v). Methods that used first derivative had, on average, the highest R^2_{cv} , RPD, and R^2_v . This can be explained by the fact that

first derivative reveals more details in the spectral data, especially in regions where small peaks are present on shoulders of larger ones. The spectral regions that were used were $926 - 1600\text{ cm}^{-1}$, $1712 - 1809\text{ cm}^{-1}$, and $2561 - 2989\text{ cm}^{-1}$ [39]. The best accuracy was observed for the infrared predictions of C4:0, C6:0, C8:0, C10:0, C12:0, C14:0, C16:0, C18:0, C18:1 trans, C18:1 cis-9, C18:1 cis, and for some groups of FAs studied in milk (saturated, monounsaturated, unsaturated, short-chain, medium-chain, and long-chain FAs). The study concluded that equations with R^2_{cv} greater than 95% will be useful in milk payment systems, while equations with R^2_{cv} greater than 75% will be useful for animal breeding purposes [39].

Eskildsen *et al.* (2014) demonstrated that variation associated with total fat content and breed was responsible for successful FTIR-based predictions of FAs in raw milk samples. The study could not assign signals to individual FAs in the FTIR measurement when several FAs were present in the same mixture. The study concluded that predicted concentrations of individual FAs in milk rely on indirect correlations, which are confined to covariance structures with total fat content rather than absorption bands directly associated with individual FAs. Since these covariance structures cannot be conserved in future samples or samples from different breeds, biased predictions will be obtained from such models. This conclusion was in agreement with the results of previously mentioned studies that showed only individual FA with high concentration or groups of FAs can be predicted from FTIR measurements [42].

Bonfatti *et al.* (2016) reported that the accuracy of MIR predictions decreased when FAs were expressed on a fat basis in Italian Simmental cows, which was consistent with the previous studies. Predictions of individual fatty acids PLS models were high (R^2_{cv}) for SFA, MUFA, short and medium-chain fatty acids, C12:0, C14:0, C16:0, Σ unsaturated C18, Σ C18:1, and C18:1n-7 cis-9. Other fatty acids, namely C14:1, C18:1 trans, C18:1n-7 trans-9, Σ CLA, C18:2 cis-9,trans-11, and C18:3n-3, were poorly predicted by MIR spectroscopy ($R^2_{cv} < 0.70$) [43]. This can be explained by the fact that milk samples from Holstein cows tend to dominate calibration sets used to develop prediction models for milk components, since Holstein is the most relevant breed for dairy farmers. Italian Simmental might have been underrepresented in these calibration sets.

2.7.2 Prediction of fatty acid chain length and unsaturation

Fat prediction of MIR analyzers are affected by chain length and unsaturation of fatty acids that make up milk triglycerides. Increasing the chain length increases the difference between MIR

prediction and reference chemistry by 0.0429% and by −0.0566% fat per unit of increase in carbon number per 1% change in fat, for fat B and fat A, respectively. Increasing the unsaturation decreases the difference between MIR prediction of fat and chemistry for fat B by −0.4021% and increased fat A by 0.0291% fat per unit of increase in double bonds per 1% change in fat concentration [44].

Wojciechowski and Barbano (2016) suggested an additional adjustment to improve the predictions of fat levels in milk samples. It was demonstrated that sample-to-sample variation in the differences in mean fatty acid chain length (CL) and mean unsaturation (UN) of the milk fat cause differences between MIR predictions of fat content of milk and the laboratory chemical reference analysis of the same sample. CL is defined as mean carbon number per fatty acid, while UN is defined as mean double bonds per fatty acid. Variation in fatty acid CL is reflected in the fat B, which is measured by using C–H stretch of fat, but not in fat A, which is measured by using C=O stretch, causing error in the fat A measure. Conversely, variation in UN of fatty acids only has a small effect on fat A measures, but causes larger errors in fat B measures of fat content of milk [28]. To overcome this weakness, they suggested to develop PLS models to determine the CL and UN of fatty acids for milk samples by using their MIR spectra. These values will be used to correct the fat A and fat B measurements according to the following equations [28]:

Equation 2-14 Suggested relations to adjust fat A and fat B readings according to chain length and saturation of milk fatty acids

$$fat\ A - ((-0.1756 \times CL) + 2.5591) + (0.16 \times UN) + 0.0452$$

$$fat\ B - ((-0.1422 \times UN) + 0.073)$$

The mean fatty acid CL (expressed as carbon number) was calculated by dividing the total fatty acid concentration weighted for carbon number by the total fatty acid concentration, as determined by gas liquid chromatography (GLC). The mean UN (expressed as double bonds per fatty acid) was calculated by dividing the total fatty acid concentration weighted for the number of double bonds by the total fatty acid concentration, as determined by gas liquid chromatography. The following equations show the calculation of fatty acid concentration weighted for carbon number and fatty acid concentration weighted for number of double bonds, respectively [28].

Equation 2-15 Calculation of fatty acid concentration weighted for carbon number and double bonds

$$fatty\ acid\ (mmol) \times number\ of\ carbons\ in\ fatty\ acid\ chain$$

$$\text{fatty acid (mmol)} \times \text{number of double bonds in fatty acid chain}$$

The PLS models for prediction of fatty acid CL and total UN were calculated using the spectral ranges 3000 – 2750 cm⁻¹, 1800 – 1700 cm⁻¹ and 1580 – 1000 cm⁻¹, applying mean centering of the data and 2-point baseline correction relative to average absorbance calculated over wavelength ranges of 2650 – 2550 cm⁻¹ and 1260 – 1200 cm⁻¹. The validation performance of the prediction model for CL produced a relative standard deviation (RSD) of 0.43% for CL and 3.3% for UN. [28].

2.7.3 Prediction of fat globule particle size in homogenized milk

Milk fat is present in the form of globules that range in diameters from <0.2 to >15 µm. The small fat globules represent 80% of the total number of fat globules but they contain only 3% of the mass of fat. On the other hand, large globules represent only 2% of the total number of fat globules but they contain 95% of the mass of fat [45]. The size of fat globules in milk needs to be reduced because large fat globules increase light scattering, leading to an inaccurate estimate of fat, protein and lactose content of milk. In addition, large fat globules can cause the Christiansen light scattering effect, which causes a change in the refraction of light at wavelengths near maximum absorption by the carbonyl and carbon-hydrogen groups. The Christiansen effect causes a shift in the apparent wavelength of maximum light absorption to a longer wavelength. This effect can be reduced by decreasing the fat globule diameter [46].

Di Marzo *et al.* (2016) developed a PLS model to detect absorbance shifts in MIR spectra of milk fat due to the Christiansen effect caused by systematic variations in fat globules size. This PLS model can predict particle size distribution and the reported parameters were: particle size distribution D (0.5), particle size distribution D (0.9), surface volume mean diameter D [3,2] (i.e. Sauter mean diameter) and volume moment mean diameter D [4,3] (i.e. De Broucker mean diameter). The spectral regions that were used were: 3000 – 2750 cm⁻¹, 1800 – 1700 and 1585 – 1000 cm⁻¹ and a systematic shift in the region of absorbance of fat B and fat A was observed. The mean MIR-predicted values for D (0.5), D[3,2], and D[4,3] were lower (P < 0.05) than laser light-scattering values (i.e. the official method for determining particle size), whereas no difference in the MIR-predicted D (0.9) versus laser light scattering was detected [46].

To obtain accurate analytical results for milk components, the fat globule diameter of milk should be less than one-third of the wavelength of fat B (3.48 μm), which is the shortest wavelength used for fat analysis. Particle size of the milk produced by a homogenizer within a milk analyzer should result in a mean globule diameter $D(0.9) < 1.7 \mu\text{m}$. If $D(0.9) \geq 1.7 \mu\text{m}$, then the homogenizer performance has deteriorated and should be replaced. In addition, different types of homogenizers have different efficiency at the same pressure. For example, one stage homogenizer will be less efficient than double stage homogenizer. Homogenization efficiency is also affected by milk temperature, pump speed, pump stroke length, fat content and time of usage. This prediction model can be used to warn milk analyzer operators that the homogenizer is near failure and needs to be replaced to ensure quality of results [46].

2.7.4 Prediction of milk protein composition and its genetic variants

Milk protein composition has been linked to cheese making properties. Associations between increased κ -casein (κ -CN) content and decreased rennet coagulation time, enhanced curd firmness and increased cheese yield have been reported in the literature. For this reason, developing a fast method for determining milk protein fractions by MIR spectroscopy is considered feasible by the cheese making industry [47].

Bonfatti *et al.* (2011) investigated the feasibility of MIR spectroscopy to predict milk protein fractions and their genetic variants. The reference method that was used to determine milk protein fractions was reversed phase (RP) HPLC and predictions models were built by using modified PLS (MPLS) and first derivatives of the samples' spectra. Whole spectrum was used and regions from 3470 to 3040 cm^{-1} and from 1700 to 1600 cm^{-1} were excluded [47].

Casein and whey protein fractions showed peaks around 1650 and 1550 cm^{-1} , which correspond to amide I and amide II bands, respectively. In addition, peaks were observed between 1300 and 1000 cm^{-1} due to ionic and covalent phosphate bonds. Sulfur amino acids were responsible for the S-H stretch vibration occurring between 2500 and 2600 cm^{-1} . Other vibrations, such as S-S and C-S, were observed below 1000 cm^{-1} and good correlation was also observed between 2800 and 3600 cm^{-1} due to C-H and O-H stretching [47].

The most accurate predictions were obtained for total protein, casein (CN), α_{S1} -CN, β -lactoglobulin (LG), glycosylated κ -CN, and whey protein content. The coefficients of determination between predicted and measured values in cross-validation ranged from 0.61 to

0.78. Less accurate results were obtained for β -CN, α_{S2} -CN and κ -CN. No accurate predictions were obtained for α -LA nor for γ -CN. No feasible predictions were produced by MIR spectroscopy for milk protein genetic variants κ -CN A and B; β -CN A1, A2, and B; and β -LG A and B. The coefficients of determination between predicted and measured values in cross-validation were <0.15 for the content of κ -CN genetic variants <0.01 for the content of β -CN variants. The best predictions were obtained for β -LG A and β -LG B contents. In addition, Unfavorable results were obtained when protein fraction contents were measured as percentage ratios of total protein or CN. The study concluded that MIR cannot be used to accurately predict individual milk protein composition; however, it can play a role as indicator traits in selective breeding to enhance milk protein composition. The inability of MIR spectroscopy to predict milk protein fractions accurately can be explained by the fact that protein fractions differ in the length of the amino acids chain and in the secondary and tertiary structure, rather than in the kind of chemical bonds. In addition, the heterogeneity within each protein fraction, such as the presence of glycosylated and phosphorylated forms, makes the prediction based on the use of spectra difficult. Previous study that used PLS and untreated spectra did not give better results [47].

Bonfatti *et al.* (2016) reported that PLS models had good predictive ability for overall protein and casein content, for which the R^2_{CV} of models were >0.80 . Values of R^2_{CV} for the content of the casein fractions ranged from 0.74 for α_{S1} -CN to 0.22 for unglycosylated κ -CN. Glycosylated κ -CN was also predicted with poor accuracy ($R^2_{CV} = 0.46$). For whey protein fractions, as well as for most caseins, the R^2_{CV} showed that models could only discriminate between high and low protein values. When protein fractions were expressed on a protein basis, results were even more unsatisfactory because prediction of protein fraction content relies indirectly on the relationship between the content of individual proteins and total milk protein [43].

2.7.5 Prediction of milk lactoferrin as an indicator of mastitis

Lactoferrin (LTF) is an iron-binding glycoprotein that is associated with the cow immune system and it is secreted by the mammary epithelial cells. Its normal level in milk ranges between 0.1 and 0.4 g/L of milk and it becomes higher during the later lactation stage (0.25-0.4 g/L). LTF can also be released by polymorphonuclear neutrophils during inflammation since it has antibacterial and antifungal properties. LTF can be involved in host defence mechanisms and it can modulate the inflammatory process. For this reason, high concentrations of LTF in milk can be considered as an

indicator for the presence of mastitis. In such cases LTF might reach 2.3 g/L of milk. Lindmark-Månsson *et al.* (2006) reported that the highest amounts of LTF were present in udder quarter milk samples that had the highest SCC, which agrees with previous reports in the literature. The correlation between LTF concentration and SCC was 0.918 ($P < 0.001$) and the increase in LTF in the quarter milk samples started at a SCC of 5000 cells/mL. In addition, she found that the correlation between LTF and polymorphonuclear neutrophils was weaker than the correlation between LTF and SCC, which indicates that LTF is produced by the udder tissue too. However, the author acknowledged that the correlation between SCC and LTF was lower (0.455, $P < 0.001$) in a previous study and that the cow that had high SCC quarter (i.e., 1,815,000 cells/mL) did not show signs of mastitis [48].

Soyeurt *et al.* (2012) conducted a study that aimed at developing calibration model to quantify LTF content in milk by using large number of samples collected in Belgium, Ireland and Scotland from cows of different breeds and from different production systems. The study also aimed at evaluating LTF predictions to identify the presence of clinical mastitis [49]. In this study, LTF was measured by a commercial ELISA kit and milk MIR spectra were collected by FOSS's MilkoScan FT6000. Six different PLS calibration models for LTF prediction were developed by using raw spectral data, raw spectral data with repeatability file, first derivative data, first derivative data with repeatability file, second derivative data, second derivative data with repeatability file. In addition, logistic regression models were developed to detect the presence of mastitis. These models included either predicted MIR LTF, somatic cell scores (SCS), whose distribution is more normal than SCC distribution, SCS and predicted MIR LTF, or SCS, predicted MIR LTF and their interaction. SCS was calculated by using the following formula:

Equation 2-16 Calculation of somatic cell scores

$$SCS = \left(\log_2 \left(\frac{SCC}{100000} \right) \right) + 3$$

These models were applied to spectral data from cows that had mastitis to assess their capability of detecting mastitis [49]. The PLS model that was built by using first derivative spectral data and the repeatability file gave the most accurate results. The cross-validation coefficient of determination R^2_{cv} , RPD, validation coefficient of determination R^2_v , standard error of cross validation (SECV) and standard error of prediction (SEP) were 0.72, 1.86, 0.60, 50.55 mg/L and

58.98 mg/L, respectively. Since SECV and SEP were relatively similar, the model was considered robust. Three spectral regions were considered relevant to the prediction of LFT. These regions were the region surrounding 1200 cm^{-1} , which is associated to C-O bonds, the region located around 1300 cm^{-1} , which is related to COOH, and the region between 1700 and 1800 cm^{-1} . On the other hand, the logistic model that only included predicted LTF could not accurately predict the presence of mastitis despite a moderate correlation between SCS and LTF ($R=50.54$). Model specificity was better when LTF was included in the regression along with SCS when compared with SCS alone [49].

Bonfatti *et al.* (2016) reported that PLS models could only discriminate between high and low values of lactoferrin, and they were not sufficiently accurate to lead to precise quantification. The previous study produced higher R^2_{cv} only when samples from 3 different countries were used. However, when the calibration was tested in external validation on Belgian samples only, the R^2_{cv} was consistent with the estimate obtained in the current study. In addition, the average lactoferrin content in the current study was $128 \pm 96\text{ mg/L}$, which was lower than the one reported in the previous study [43].

2.7.6 Prediction of milk coagulation properties

Milk coagulation properties (MCP) directly affect the cheese-making properties, such as cheese yield, moisture and quality [50]. MCP include good reactivity to rennet, high curd-firming capacity, good syneresis ability, and whey drainage [43]. These properties can be assessed by measuring milk rennet coagulation time (RCT, min), and curd firmness (a_{30} , mm). RCT is the time between the addition of the clotting enzyme and the beginning of the coagulation process, while a_{30} is curd firmness measured 31 minutes after addition of the clotting enzyme. MCP reference values are assessed by computerized renneting meter. This technic is time consuming and requires trained personnel; hence, it cannot be used on a large scale to determine MCP properties of milk samples [50].

De Marchi *et al.* (2009) investigated the potential of MIR milk analysis as an alternative method to predict MCP of milk samples. PLS models were developed by using untreated and pretreated spectra to predict these properties. Spectral treatment included: normalization (N), multiplicative scatter correction (MSC), first derivative (Savitzky-Golay, 3 data points each side), and N plus first derivative spectra. These models were enhanced by using the following spectral regions: 1600

to 900 cm^{-1} , $3040\text{ to }1700\text{ cm}^{-1}$ and $4000\text{ to }3470\text{ cm}^{-1}$. The best models were obtained by using raw spectral data. The root mean square errors of cross-validation for RCT and a_{30} were 2.36 min and 6.86 mm, respectively. R^2 for RCT was 0.59, which indicates the ability to discriminate between high and low values; however, it was not considered sufficiently accurate to be implemented in industry. On the other hand, R^2 for a_{30} was 0.37, which was not considered satisfactory [50].

The loading spectra of the prediction models showed peaks occurring at 968, 1115, 1146 and $1,180\text{ cm}^{-1}$, which are attributed to C-O and C-C stretching, respectively, whereas O-C-H, C-C-H, and C-O-H bending account for peaks at $1466\text{ and }1331\text{ cm}^{-1}$, respectively. A peak was also observed at 1240 cm^{-1} which may be due to amide III or phosphate bands. Dominant peaks are also observed at $1589\text{ and }1500\text{ cm}^{-1}$, which can be attributed to amide II. Peaks associated with lipids (2935 , 2839 , 1763 , and 1751 cm^{-1}) were also apparent in the loadings of the RCT model. These results can be explained by the effect of milk protein and fat content on milk coagulation. Increased protein level decreases significantly the curd firming time [50]. The study concluded that despite the models needed further improvement to their accuracy for any application in the industry, they can still be used in phenotypic-based selection programs to improve MCP at herd or individual level [50].

Dal Zotto *et al.* (2008) demonstrated that the assessment of MCP through computerized renneting meter provides repeatable and reproducible measures for RCT but not for a_{30} . In addition, adding bronopol preservative to milk has no detrimental effects on the reliability of RCT measures by computerized renneting meter [51].

De Marchi *et al.* (2012) reported an enhanced accuracy for the predictions of MCP by MIR spectroscopy [52]. In this study, a Formagraph was used as reference method for determining MCP values instead of computerized renneting meter and the testing-time of the analysis was set up at 60 minutes to investigate if milk not forming a curd within the conventional threshold of 30 minutes showed coagulation aptitude after this time. In addition to RCT and a_{30} , the study included the determination of curd-firming time k_{20} , which is the interval in minutes from the beginning of coagulation to the moment the width of the graph attains 20 mm, and a_{60} , which is curd firmness at 60 minutes after rennet addition. The most accurate models were those for RCT, k_{20} and a_{30} . Less favorable results were obtained for a_{60} . In addition, the study demonstrated that MCP were

strongly influenced by dairy cooperative and herd, suggesting the existence of different feeding and management conditions. It also showed that milk chemical composition and acidity had a large influence on MCP, where increased values of casein and titratable acidity yielded improved MCP. Another factor that affected MCP was season of sampling. Samples collected during the summer had better results in terms of RCT and k_{20} [52], which was attributed to a decrease in the content of protein in the summer that is usually associated with a decrease in the content of casein [53]. Finally, there were differences in MCP between different breeds. Brown Swiss produced milk with the shortest RCT and the highest a_{30} , whereas Holstein-Friesian showed the worst technological properties and Simmental cows were intermediate between the previously mentioned breeds [52].

Using the previously enhanced calibration equations, De Marchi *et al.* (2013) developed models to detect noncoagulating milk (NC), which is defined as milk not forming a curd within 30 min from rennet addition, in Holstein cows by using MIR spectroscopy. The reference method that was used to determine MCP was Formagraph and the traits that were studied were rennet coagulation time, curd-firming time, and curd firmness at 30 and 60 min after rennet addition [54]. Raw spectral data was used to build the models and the spectral regions 3040 to 3470 cm^{-1} and 1600 to 1700 cm^{-1} were excluded. The most accurate prediction model was developed for RCT, followed by k_{20} and a_{30} . Models for the prediction of a_{60} were not satisfactory. Results showed no specific spectral information distinguishing NC from coagulating samples. However, peaks associated with protein were found to be very dominant for RCT and k_{20} , whereas those associated with lipids seem to be more dominant for curd firmness traits [54].

In another study, Cecchinato *et al.* (2009) developed calibration equations to estimate heritability and genetic correlations for measured MCP and their predictions obtained from MIR spectra [55]. Studies have reported that exploitable additive genetic variation exists for RCT and a_{30} , hence enhancement of these traits through breeding is a viable option. Heritability is a statistic used in breeding and genetics studies that estimates how much variation in a phenotypic trait in a population is due to genetic variation among individuals in that population [56]. Point estimates of heritability ranged from 0.30 to 0.34 and from 0.22 to 0.24 for RCT and a_{30} , respectively. Heritability estimates for MCP predictions were larger than those obtained for measured MCP. Estimated genetic correlations between measures and predictions of RCT were very high and ranged from 0.91 to 0.96. Estimates of the genetic correlation between measures and predictions

of a30 were large and ranged from 0.71 to 0.87 [55]. Genetic correlation is the proportion of variance that two traits share due to genetic causes [57]. The study concluded that breeding strategies for the enhancement of MCP based on MIR predictions as indicator traits could be easily and immediately implemented for dairy cattle populations where routine acquisition of spectra from individual milk samples is already performed [55].

Bonfatti *et al.* (2016) reported that prediction accuracy was satisfactory for RCT ($R^2_{cv}=0.69$) but poor for other MCP traits ($R^2_{cv} < 0.42$), which was not consistent with the literature. Such inconsistency might be explained by the greater variability in MCP traits detected by other authors and by the different equipment used when measuring MCP since the accuracy of the prediction model depends on the accuracy of the reference analysis. Unsatisfactory predictions for curd firmness at 60 min were obtained, which were consistent with the literature. Among curd yield traits, the prediction of dry matter (DM) curd yield showed the greatest R^2_{cv} (0.85), followed by fat, raw, water, and protein curd yield, for which the R^2_{cv} was 0.62. Considering that curd yield partly depends on rheological property of milk, raw curd yield was predicted with relatively high accuracy ($R^2_{cv}=0.67$). Curd composition was predicted with poor accuracy, with R^2_{cv} values ranging from 0.35 for fat content to 0.61 for DM content [43].

2.7.7 Prediction of titratable acidity and pH

Titrateable acidity (TA) is defined as the volume of 0.25 molar sodium hydroxide solution required to achieve a color change of the pH indicator phenolphthalein to pink in a specific volume of milk sample. It is a dimensionless milk index and its unit is °SH (degree Soxhlet-Henkel). TA is different from pH, which is defined as the negative log of hydrogen ions concentration. Cheese-ready milk must have a Soxhlet-Henkel number between 6.0 and 7.4 °SH.

TA plays a fundamental role in the aggregation rate of para-casein micelles and the reactivity of rennet. It also influences the rate of syneresis and determines the suitability of milk for cheese making. In the production of premium cheeses, milk with low acidity (hypo-acid milk) is generally considered unsuitable for cheese making because of its negative effects on the rheology of the acid-rennet curd and on the textural properties of the cheese paste. On the other hand, the pH of milk affects both the enzymatic and aggregations reactions, which means lowering the pH decreases the colloidal stability of milk [50].

De Marchi *et al.* (2009) developed PLS models by using untreated and pretreated spectra to predict TA and pH. Spectral treatment included: normalization (N), multiplicative scatter correction (MSC), first derivative (Savitzky-Golay, 3 data points each side), and N plus first derivative spectra. These models were enhanced by using the following spectral regions: 1600 to 900 cm^{-1} , 3040 to 1700 cm^{-1} and 4000 to 3470 cm^{-1} . The best models were obtained by using the first derivative of the spectra. The root mean square errors of cross-validation for TA and pH were 0.25 SH $^{\circ}$ /50 mL and 0.07, respectively. The R^2 for the TA model was 0.66, which provides approximate prediction, whereas for pH R^2 was 0.62, which discriminate between high and low values. The study could not assign functional groups to individual peaks [50]. The study concluded that TA and pH had the potential to be predicted by MIR spectroscopy and multivariate data analysis [50]. Bonfatti *et al.* (2016) reported that prediction accuracy was satisfactory for pH ($R^2_{cv}=0.79$) [43].

2.7.8 Estimation of major mineral contents in cow milk

Soyeurt *et al.* (2009) investigated the possibility of predicting the content of several milk minerals by PLS models and milk MIR spectral data. The studied minerals were Ca, K, Mg, Na and P. The reference method that was used for determining the content of these minerals in milk was inductively coupled plasma atomic emission spectrometry (ICP-AES). The cross-validation coefficients of determination (R^2_{cv}) were 0.23 and 0.50 for K and Mg, respectively; hence, not showing any potential application. The external validation coefficients of determination were 0.97, 0.14, and 0.88 for Ca, Na, and P, respectively, suggesting a potential application for Ca and P but not for Na. The researchers concluded that MIR spectroscopy can only predict Ca and P content in milk [58]. The correlations between the Ca and P MIR predictions and the known milk components were calculated. All correlations were inferior to the correlation calculated based on the cross-validation. Therefore, the calibration equations established to predict the contents of Ca and P in milk were coming from real spectral absorbance. The spectral regions that were correlated to variability in Ca content were: 1454 and 1458 cm^{-1} and between 2831 and 2970 cm^{-1} , while variability in P content were correlated to the following spectral regions: 1200 and 1277 cm^{-1} , between 2841 and 2974 cm^{-1} with a maximum at 2974 cm^{-1} and between 1442 and 1469 cm^{-1} . High correlation was observed at 1242 cm^{-1} , which is related to the P=O bond present in phospholipids. On the other hand, the spectral regions associated with the bond between Ca and the carboxylate group of casein at 1410 cm^{-1} and 1575 cm^{-1} did not show high correlation. In addition, previously reported regions in the literature related to Ca and P in milk did not show any

correlations in this study. These regions are 956 to 946 cm^{-1} , which was related to the concentration of organic P, and 980 cm^{-1} , which was related to bound Ca [58]. Bonfatti *et al.* (2016) reported that low values of R^2_{cv} for minerals ranged between 0.41 and 0.48. This can be explained by reduced variability in mineral concentrations [43].

2.7.9 Prediction of acetone, β -hydroxybutyrate and citrate

It has been proved that acetone and BHB can be used as milk biomarkers to indicate ketosis. In addition, citrate is considered as an early indicator of negative energy balance. By using FTIR milk analysis and PLS, Grelet *et al.* (2016) developed equations to predict acetone, BHB and citrate levels in raw milk samples. The coefficient of determination (R^2) of cross-validation was 0.73, 0.71, and 0.90 for acetone, BHB and citrate, respectively, with root mean square error of 0.248, 0.109, and 0.70 mmol/L, respectively. The external validation R^2 were 0.67 for acetone, 0.63 for BHB, and 0.86 for citrate, with respective root mean square error of validation of 0.196, 0.083, and 0.76 mmol/L, respectively. Acetone content ranged from 0.020 to 3.355 mmol/L with an average of 0.103 mmol/L; BHB content ranged from 0.045 to 1.596 mmol/L with an average of 0.215 mmol/L; and citrate content ranged from 3.88 to 16.12 mmol/L with an average of 9.04 mmol/L. A first derivative was obtained and the spectral regions that were used were: 968.1 to 1,577.5 cm^{-1} , 1,731.8 to 1,762.6 cm^{-1} , 1,781.9 to 1,808.9 cm^{-1} , and 2,831.0 to 2,966.0 cm^{-1} . To evaluate the prediction model performance, ratio performance/deviation (RPD) was calculated. RPD is a criterion that shows simultaneously the accuracy of predictions and the global variability of the reference values. The RPD is defined as the ratio of the standard deviation (SD) to the root mean square error (RMSE), SD/RMSE , while RMSE can be the one calculated in cross-validation or the one of a validation set. When the RPD is between 1.5 and 2, the model can discriminate low from high values; an RPD between 2 and 2.5 indicates that rough quantitative predictions are possible, and RPD between 2.5 and 3 or above corresponds to good and excellent prediction accuracy. It was suggested to use these models for herd management, or at individual level by using thresholds or relative values to cope with low accuracy [59].

2.7.10 Prediction of cows' body energy status and feed efficiency

McParland *et al.* (2011) investigated the potential of regularly collected MIR milk spectra at central dairy laboratories to predict body energy status and related traits in lactating dairy cows, which included energy balance, body energy content, body condition score and energy intake. The

energy balance was defined as a function of milk yield, fat and protein content, dry matter intake, body weight and body condition score. On the other hand, the body energy content was defined as a function of body weight and body condition score, predicting body lipid and protein weight [60].

The aim of the study was to directly predict body energy status of cattle from MIR milk spectra without relying on milk composition components, such as fat-to-protein ratio or fatty acid composition of milk fat. The importance of determining body energy status stems from the fact that the duration of negative energy balance might be a precursor for impaired health and fertility in dairy cows. Data on 815,129 test days from 3,151 lactations of 1,145 Holstein cows were available to compute body energy status. These data included milk yield, milk composition, dry matter intake, body weight, body condition score, age of calving, season of calving and parities [60]. Random regression models were fitted to daily milk yield, fat percent, protein percent, dry matter intake, body condition score and body weight to provide daily solutions to calculate a lactation profile for energy balance. All random regression models were fitted within parity and included the fixed effects of genetic line, feeding group, year of calving by season of calving, age at calving, year of record by month of record, a fourth-order orthogonal polynomial on days post calving, and the random effect of the interaction of cow by a fourth-order orthogonal polynomial on days post calving. On the other hand, MIR milk spectra were collected for 6,665 milk samples from 18 test dates of 465 lactations from 277 cows. MIR milk spectral data were converted to absorbance and Boxcar smoothing (i.e., moving average) was applied to it by averaging the spectral data over spectral segments of 5 data points in length. In addition, the first derivative was calculated on both the smoothed and unsmoothed absorbance spectral data. Cows were stratified according to selection line, feeding treatment, and season of calving, and the data split randomly within stratum into 4 equally sized data sets. The calibration data set included 75% of the data and 25% were considered as validation set. PLS calibration models were developed to predict body energy balance, body energy content, body condition score and energy intake using unsmoothed and smoothed spectral data as well as the first derivative. Validation results showed that smoothing and first derivative did not enhance predictions. Among the body energy status indicators, energy balance was consistently the most accurately predicted one. The authors concluded that the average accuracy of predicting body energy status from MIR spectral data was as high as 75% when energy balance was measured across lactation. In addition, they considered predictions of body energy

status from MIR milk spectra more accurate than predictions obtained from milk composition indicators, such as fat-to-protein ratio in milk [60].

In a later study, McParland *et al.* (2014) investigated the potential of MIR milk spectroscopy to predict individual cow energy intake and efficiency. Feed efficiency was described by residual feed intake (RFI), which is the difference between actual energy intake and energy used, such as milk production, maintenance, and body tissue anabolism, or supplied from body tissue mobilization. RFI can be mathematically equivalent to energy balance. A total of 1,535 records for energy intake, RFI, and milk MIR spectral data were available across 36 different test days from 535 lactations on 378 cows. Data included dry matter intake and fecal analysis, which was used to calculate the effective energy intake (EEI), milk yield, milk chemical composition, cow body weight and cow body condition score. Using this data, RFI was calculated for each cow and energy balance (EB) was calculated as the difference between effective energy intake and effective energy expended through milk production and maintenance. In addition, MIR milk spectra were recorded periodically, and PLS models were developed to predict RFI, EB and EEI from the MIR milk spectra. The correlation coefficient r of models to predict RFI across lactation ranged from 0.48 to 0.60; however, the strongest one was obtained in early lactation $r=0.65$. They also found a very strong correlation between EB and RFI $r=0.85$ [61].

2.7.11 Prediction of methane emissions from dairy cattle

Methane emissions, which are eructed by cattle, represent a loss of use of energy intake that increases the feed cost and decreases profitability of dairy farms. Hence, mitigating methane emissions will improve feed efficiency, reduce feed cost and reduce feed related costs. Few studies investigated the potential of FTIR milk analysis as a cheap and fast method to predict dairy cattle methane emissions [62].

During digestion in the reticulorumen, carbohydrates are degraded by microbial fermentation into CO_2 , CH_4 , H_2 and volatile fatty acids (VFA), which are mostly acetic, propionic and butyric acid. Acetic and butyric acids are the primary precursor for milk fat, while propionic acid is used in lactose synthesis. Research shows that an increased butyrate/propionate ratio decreases lactose content and increases fat content in milk and the rumen CH_4 synthesis. On the other hand, the rumen VFA composition influences milk fat composition. Short and medium-chain fatty acids (4 to 14 carbons) are derived from *de novo* synthesis based on acetate and butyrate. Long-chain fatty

acids (>16 carbons) are collected from the circulating lipids and fatty acids of 16 carbons are obtained from the two sources. Thus, fatty acid composition of milk fat reflects the VFA produced during ruminal fermentation and CH₄ production. Since FTIR milk analysis can predict milk fat composition, Dehareng *et al.* (2012) evaluated the prediction of CH₄ emissions from MIR milk spectra [63]. Holstein cows of different parity (i.e. primiparous and multiparous) were offered different diets. Diet 1 consisted of freshly cut pasture grass, dried beet pulp, soybean meal and soybean hulls. Diet 2 consisted of: corn silage, meadow hay, cracked corn, rapeseed meal, palm meal, soybean meal and a 50:50 mix of coconut and flaxseed oil. Diet 3 consisted of grass silage, corn silage, cracked corn, soybean meal and dried beet pulp. Cows were given a 21 days adaptation period, then milk spectra and CH₄ measurements were collected. Sulfur hexafluoride (SF₆) gas tracer technique was used to measure enteric CH₄ production by the cows once a day. Milk spectra were recorded twice a day (i.e. morning and evening). Assuming that there is a delay between the production of fermentation products and their use to produce milk components, different ways were used to calculate the average of two milk spectra to be related to one-day CH₄ measurement. The averages were for two spectra collected on [63]:

- the same day of CH₄ measurement (day 0)
- evening of the same day and morning next day (day 0.5)
- next day (day 1)
- evening next day and morning 2 days later (day 1.5)
- 2 days later (day 2)

PLS model was developed for predicting CH₄ emission from milk MIR spectra, which did not undergo any pre-treatment. Different models were built for different milk spectral average sets. The spectral regions used for that were: 972 to 1589 cm⁻¹, 1720 to 1782 cm⁻¹ and 2746 to 2970 cm⁻¹ [63]. The best predictions were achieved by PLS model built by using CH₄ emission per Kg of milk, rather than CH₄ emission per day, and the milk average spectra with an interval between the measurement of CH₄ and the spectral data equal to 1.5 days. The cross-validation coefficient of determination R²_{cv} was 0.79 and the difference between R²_c and R²_{cv} was the lowest (0.08), which indicates a robust model. The study also showed that CH₄ predictions were not affected by the different diets and that R²_{cv} of CH₄ prediction model was higher than the correlation obtained between CH₄ emissions and milk production and MIR milk components. This fact indicates that

the prediction model is not based only on correlation between CH₄ production and milk components. However, the study concluded that further research is needed to produce more reliable prediction model [63].

Vanlierde *et al.* (2015) predicted CH₄ emissions from milk MIR spectra by taking lactation stage into consideration, which better reflects the metabolic status of the cow [62]. The author hypothesized that inclusion of the stage of lactation in the prediction model might improve the relationship between CH₄ emissions and MIR spectra because milk fatty acids profile is strongly influenced by the evolution of body tissue mobilization during lactation. The study included cows from different countries (i.e. Belgium, Ireland), different breeds (i.e. Holstein, Jersey, Holstein-Jersey crossbred), different parity (i.e. first, second, third and later), and different lactation stages. In addition, different diets were used to feed the cows, such as grass or high silage diets, with or without linseed supplementation and synchronized or not in terms of fermentable energy and nitrogen supplies in the rumen. CH₄ emissions were measured daily by the sulfur hexafluoride (SF₆) tracer gas technique and MIR milk spectra were collected twice daily for cows in two different countries. To obtain one milk spectrum per CH₄ measurement, the 2 spectra were averaged proportionally to the milk yield at each milking [62]. Two PLS calibrations models were developed to predict CH₄ emissions from MIR milk spectra after applying the first derivative and three spectral regions were used: 968-1577 cm⁻¹, 1720-1809 cm⁻¹, and 2561-2966 cm⁻¹. The first model was independent from lactation stage (ILS) and the second model was dependent on lactation stage (DLS), which included the days in milking (DIM) for each cow in addition to the MIR milk spectra. To obtain the DLS calibration equation, each first derivative value of the spectrum was multiplied by a constant (i.e., 1), a linear ($\sqrt{3} \times x$) and a quadratic $\left[\sqrt{5/4} \times (3x^2 - 1) \right]$ modified Legendre polynomial where

Equation 2-17 Equation used to embed lactation stage in constants that are multiplied by the first derivative value for each spectrum data point

$$x = -1 + 2 \left[(DIM - 5) / (365 - 5) \right]$$

Through this process, a modified spectral data set was generated containing 867 data points (289 data points for each constant, linear, and quadratic part) referred to as the DLS spectral data set. The linear and the quadratic parts of these modified spectra take into account the lactation stage of

cows. The range of application of the DLS equation was between 5 and 365 DIM given the range of lactation covered in the definition of the Legendre polynomials [62]. After applying the two prediction equations to an external validation spectral data set whose CH₄ emissions were not recorded, a statistical model was developed to evaluate the behaviour of the predicted CH₄ emissions and to verify that they comply with the biological processes in the cow's body. This model reflected several effects known to influence CH₄ production such as effects of herd, year and month of test-date, lactation number, lactation stage, animal, and milk yield [62].

Equation 2-18 Statistical model that reflects several effect which influence CH₄ immisions

$$y_{pqrstu} = \mu + h_p \times y_q + mo_r + l_s \times d_t + c_u + m\alpha + (m \times m)\beta + e_{pqrstu}$$

Where y_{pqrstu} was the predicted CH₄ trait, μ was the general mean, $h_p \times y_q$ was the crossed fixed effect of herd p and year of test-date q, mo_r was the fixed effect of month r, $l_s \times d_t$ was the crossed fixed effect of lactation numbers and lactation stage t, c_u was the random effect of cow u, m was the daily milk yield (kg/d), α was the linear regression coefficient on m, β was the quadratic regression coefficient on m and e_{pqrstu} was the associated random residual [62].

The R^2_c for the ILS and DLS calibration models were 0.77 and 0.75, respectively. The standard error of calibration (SEC) was 63 g/d for both ILS and DLS calibration models. These parameters show that both models had similar abilities to predict eructed CH₄. In addition, CH₄ emission predictions of both ILS and DLS models were significantly affected by all studied fixed effects. However, behavior of DLS predictions throughout the lactations was more in agreement with the literature than the predictions of the ILS model. Increasing values of CH₄ predictions were observed for the DLS model in the period 0-100 DIM, which accompanies an increase in feed intake in cows postpartum, and thereafter these values decreased. On the other hand, predictions of the ILS model showed a reversed pattern where emission predictions decreased at the beginning of the lactation and then increased until the end of the lactation. Therefore, the DLS model seems to better reflect biological processes that drive CH₄ emissions than the ILS model [62].

2.7.12 Detecting cows' pregnancy status

Lainé *et al.* (2014) investigated the potential of MIR milk spectra to detect cows' pregnancy status. They collected MIR milk spectra of open cows (i.e. non-pregnant cows) and they estimated all the relevant effects by statistical modeling. Later, they collected MIR milk spectra of cows with

unknown pregnancy status and they adjusted these spectra according to the estimated coefficients of the relevant effects to produce new spectra, which would be the expected spectra if these cows would have been open. The calculated spectra were then removed from the respective observed ones for the cows with the unknown pregnancy status, and the residual spectra were used to predict the pregnancy status. These residual spectra contained errors, pregnancy status, and unaccounted factors [64].

Before calculating the expected spectra, first derivative was calculated on the raw spectral data and noisy regions of the milk spectrum were removed. Outliers were flagged if samples had milk yield, fat content and protein content outside the acceptable ranges of the International Committee for Animal Recording (ICAR) or if they have a Mahalanobis distance greater than 3. The final edited dataset included a total of 411,406 spectra (114,338 spectra from pregnant cows and 297,018 spectra from open cows) from 68,998 cows in 1,045 herds. A predictive discriminant analysis was performed to produce a discriminant function that was validated using an external validation set of 14,883 residual spectra. Results were expressed in terms of specificity and sensitivity. Specificity is defined as the ability of the equation to predict correctly the non-event (open cows) among all observations which are not pregnant. Sensitivity is defined as the ability of the equation to predict correctly the event (pregnant cows) among all observations belonging to pregnant cows. [64]. The classification error was 0.7% and 55.5% when the predictive discriminant function was applied to the calculated residual and raw spectra, respectively, of samples with unknown pregnancy status after 50 days of insemination. Specificity was 86.2% and sensitivity was 99.7% for predictions based on the calculated residuals. The authors concluded that changes in the pregnancy status of the cow can be reflected in the MIR milk spectrum [64].

2.7.13 Verification of dairy farming production system

Capuano *et al.* (2014) investigated the potential of MIR milk spectra to distinguish between milk produced by fresh grass feeding, pasture grazing and organic farming by using partial least square discriminant analysis (PLS-DA). A total of 116 tank milk samples were collected from 30 different farms in the Netherlands. The grass intake with cut fresh grass fed indoor was estimated from the acreage of cut land and weight of fed fresh grass. The grass intake via grazing was estimated by the farmer through their experience from the animal energy needs minus what was fed next to the grassland. Samples were categorized in five groups: milk samples from cows at least 19 h outdoors

on pasture daily, milk samples from cows 6–9 h outdoors on pasture daily, milk samples from cows indoors with fresh grass in the diet, milk samples collected in spring from cows indoors with no fresh grass in the diet and milk samples collected in winter with no fresh grass in the daily ration and cows were indoor all the time. Fresh grass in the cows' daily ration was between 36% and 94%. In addition, milk samples were collected from two certified organic farms, three certified biodynamic farms and one farm converting to organic farming. MIR spectra were collected for all milk samples in the range of 925-5008 cm^{-1} and outliers were detected by principal component analysis (PCA). The following regions were eliminated due to the lack of chemical information or inadequate signal-to-noise ratio: 1800-2800 cm^{-1} , 3000-3600 cm^{-1} and 4000-5000 cm^{-1} . Various data pre-processing techniques were tested including mean-centering, auto scaling (scaling to unit variance), Pareto scaling (scaling to square root of the variance), first and second derivatives, smoothing and orthogonal signal correction (OSC) and a combination of them. PLS-DA was performed to verify the cows' feeding regime (fresh grass feeding vs. no fresh grass feeding), the housing management system (indoors vs. pasture) and the farming management system (organic vs. conventional). The training set included 75% of records of the original data set, while 25% of records were used as validation set [65]. The most satisfactory results in terms of number of misclassified samples in external validation was obtained by a PLS-DA model after auto scaling, smoothing and second derivative transformation of the raw data and the application of one OSC component [65].

The PLS-DA model discriminated between milk from cows that had fresh grass in the daily ration and milk from cows that had not fresh grass with sensitivity and specificity values of 88% and 83% in external validation and all the samples from cows that had no fresh grass collected in spring were correctly classified. For pasture grazing model, 75% of samples from cows in doors in spring were correctly classified. Discrimination of organic and conventional milk yielded 80% and 94% correct classification, respectively, in external validation. The authors concluded that milk FTIR spectra contain valuable information on cows' diet that can be used for authentication purposes [65].

2.7.14 Suggested predictions of milk components by FTIR spectroscopy

De Marchi *et al.* (2014) suggested that further research is needed to determine whether MIR milk spectra could be used for predicting the following constituents of milk [66]:

1. Potassium, magnesium, and zinc content, which are important for transmitting nerve impulses, for mineral structure of bones, for wound healing, and healthy immune systems
2. Phospholipids and acidic glycolipids, which are important for infant development
3. Vitamins A and B, which are important for healthy eyes and skin
4. Sensory features, which are important for the characterization of milk taste
5. Cheese yield
6. Whey components, such as glutathione, α -tocopherol and vitamin C

2.8 Conclusion

Monitoring milk composition is an important management tool for dairy producers. Changes in milk components can be an indicator of metabolic or management issues in individual cows or the dairy herd in general. Milk analysis by FTIR spectroscopy is a rapid and cost-effective method and it does not require the use of hazardous chemicals with no sample preparation, which qualifies it as a green analytical method. However, dairy producers do not perceive milk composition as an important parameter as they do for milk yield. One of the probable reasons for this perception is the lack of an on-site milk analyzer that is affordable and that can rapidly analyze milk with an acceptable accuracy and precision. Currently, central dairy laboratories send technicians to dairy farms to routinely collect milk samples from individual cows and milk tanks, which are shipped to the laboratory for analysis by FTIR commercial milk analyzers. The information about individual animals and the dairy herd in general that is available to dairy farmers is limited by the frequency of the analysis. Having an on-site milk analyzer will provide dairy farmer with a tool to self monitor milk composition more frequently, especially for animals with specific health issues. It must also be noted that the current payment system does not take into consideration elements of fine milk composition to determine payments for dairy farmers.

In this literature review, a summary was presented about the development of milk analysis by infrared spectroscopy and multivariate regression algorithms that are used to determine milk components from infrared spectral data. In addition, studies that intend to improve predictions of milk fatty acids and extract additional traits from milk spectral data were reviewed. The common drawback of the majority of these studies has been the fact that they were focused on developing PLS prediction models to predict fine milk composition components, such as milk fatty acids, different milk protein fractions and milk minerals, which do not necessarily produce a distinct signal in milk FTIR spectrum. Few studies relied on implementing classification algorithms to extract information from milk FTIR spectra that can be implemented in herd management or by the dairy industry. This observation suggests that the use of milk FTIR spectra can be extended beyond the paradigm of predicting specific milk components by PLS regression models, where classification models can be developed to predict the health and well-being status of dairy cows. Such field has been rarely investigated up to this date.

Chapter 3: Transmission based FTIR spectroscopy for on-site milk analysis

Abstract

Three Fourier transform infrared (FTIR) spectrometers were evaluated for transmission-based on-site milk analysis, two of which were portable FTIR spectrometers. The three spectrometers produced high quality milk spectra with excellent signal-to-noise ratio. Accordingly, three analyzer prototypes were assembled employing CaF_2 transmission cell as a sample introduction method with 40-50 μm path length. Pre-analyzed preserved producer raw milk samples were obtained from Valacta Inc. (Sainte Anne de Bellevue, QC, Canada) and their mid-infrared (MIR) spectra were collected by the assembled prototypes. Partial least squares (PLS) regression was employed to develop calibration models for the major milk component (i.e., fat, protein and lactose). To evaluate the goodness of fit and the prediction capability of the developed models, root mean square error of calibration (RMSEC), root mean square error of cross validation (RMSECV), root mean square error of prediction (RMSEP), number of factors, bias (if available), correlation coefficient r , the ratio performance/deviation (RPD) and spectral regions were used to compare the developed models.

In addition, several milk fat homogenization approaches before IR analysis were investigated to reduce the effect of light scattering on the accuracy of the IR predictions. Ultrasonication before the analysis proved to be an effective method for homogenizing milk fat and to be integrated with an on-site IR milk analysis process. The results show that applying 3000 joules of ultrasonication energy to 5 mL milk sample for 120s is sufficient to produce a homogenized raw milk sample in a consistent manner.

Prediction models developed using producer raw milk spectra scanned on prototype 3 gave the most consistent figures of merit (FOMs) (i.e., RMSEC, RMSECV and RMSEP) and RPD values for calibration models for fat, protein and lactose indicated excellent prediction capabilities. The inspection of the loading spectra of these PLS models revealed the conventional peaks at the expected positions for the major milk components. This observation confirms that the adequate FOMs obtained by prototype 3 are a result of spectral information that was captured by this prototype; hence, a portable FTIR spectrometer can be adapted to be used as an on-site milk analyzer.

3.1 Introduction

Quantitative milk analysis by Fourier transform infrared (FTIR) spectroscopy is currently an official method of the Association of Official Analytical Chemists (AOAC) [4], which is extensively used for producer payment, herd management and routine quality control in the dairy industry [22]. Goulden was the first to describe milk analysis by infrared (IR) spectroscopy [1]. He demonstrated that milk components absorb IR energy at specific wavelengths and that the intensities of the absorption peaks can be used for quantitative determination of these components. The reported wavelengths that were used to determine major milk components were 5.73 μm , 6.46 μm , 9.6 μm and 7.9 μm for fat, protein, lactose and solids-not-fat (SNF) content, respectively. In wavenumbers[†], these values are equal to 1745.20 cm^{-1} , 1547.99 cm^{-1} , 1041.67 cm^{-1} and 1265.82 cm^{-1} , respectively [1]. In addition, it was demonstrated that scattering of light by fat globules is proportional to particle-size-to-wavelength ratio. Hence, attenuation of the IR beam due to scattering can be eliminated by decreasing fat globules size through milk homogenization.

Modern milk analyzers in central dairy laboratories employ FTIR spectrometers and they use specific regions in milk FTIR spectrum to predict the concentrations of major and minor milk components (figure 3-1). These regions are 1) 1,200-900 cm^{-1} for lactose, 2) 1,280-1,200 cm^{-1} Amide III of proteins, 3) 1,500-1,200 cm^{-1} different absorbance bands originating from minor milk components 4) 1,565–1,520 cm^{-1} Amide II of proteins, 5) 1,745–1,725 cm^{-1} $C = O$ stretching in the triglyceride ester linkage of milk fat (i.e., Fat A), 6) 2,980–2,800 cm^{-1} $C - H$ stretching of the aliphatic chain in fatty acids in milk fat (i.e., Fat B). The Amide I band of proteins is dominated by noise from the $H - O - H$ bending band of water at 1650 cm^{-1} due to the immense water absorbance of IR energy. Similarly, the $O - H$ stretching band in water at 3600–3200 cm^{-1} is also dominated by noise due to the same reason.

Modern FTIR milk analyzers are renowned for their high sample throughput, which reaches 500 samples/h, fast analysis time (i.e., 6-30s), accuracy that is better than 1% relative on the main constituents and precision that is better than 0.5% relative on the main constituents [5]. To achieve these targets, a spectral resolution of 16 cm^{-1} is used to obtain high signal-to-noise ratio (SNR) since most absorption bands in aqueous solutions are quite broad. The sample introduction method

[†] The following equation was used to convert from wavelength to wavenumber: $y \text{ cm}^{-1} = 10,000,000 / x \text{ nm}$.

in these analyzers is a transmission cell with calcium fluoride (CaF_2) windows, whose refractive index is similar to that of milk [5], with an optimum path length of $37\text{ }\mu\text{m}$ [22] and samples are flushed at a very high velocity (i.e. 30m/s) [5] to keep the cell clean from residues buildup. Milk fat is homogenized by two stages high pressure homogenizer, with the pressure drop over each stage being approximately 10^7 Pa . The homogenizer consists of two chambers with ruby balls (Al_2O_3) blocking the inlet openings of each chamber. These balls are connected to a ceramic seat (ZnO_2) by springs. High-pressure pump pushes the milk through the inlet and against the ruby balls. This action generates force that breakdown fat globules and reduce their diameter [5]. Fat globules diameter should be reduced to $1\text{ }\mu\text{m}$ since the shortest wavelength that is used in milk analysis is $3\text{ }\mu\text{m}$ (3333 cm^{-1}). The high pressure within the analyzer prevent the release of dissolved CO_2 and air as small bubbles, which will disturb the analysis [5]. Water is used as a spectral background and temperature fluctuations of the entire system, the sample cell and the interferometer are kept within tight ranges to avoid the shifting of water spectrum and optical misalignment due to thermal expansion [5]. In addition, the interferometer is sealed against atmospheric humidity and it is equipped with rubber damper system to protect the mirror drive and other optical components from environmental vibrations. Due to the highly overlapping and co-linear structure of FTIR data, partial least squares (PLS) regression is used for milk calibration models [5]. PLS models both the X (i.e., the FTIR spectral data) and Y (i.e., the reference values) matrices simultaneously to find latent variables, or factors, in X that will best predict Y values [27].

Constant monitoring of milk composition is an effective tool to manage dairy herds. Changes in major milk components and some minor ones are indicators of health issues in dairy cows. In cases of negative energy balance in dairy cows, protein, fat and lactose are $<2.9\%$, $>4.8\%$ and $<4.5\%$, respectively [15]. In cases of ketosis, fat, lactose and BHB are $>4.8\%$, $<4.5\%$ and $>100\text{ mol/L}$, respectively [15]. In addition, several papers have reported the implementation of milk analysis by FTIR to predict: milk fatty acids and milk fat composition [39-43], fatty acid chain length and unsaturation [44], fat globule particle size in homogenized milk [46], milk protein composition and their genetic variants [47], milk lactoferrin as an indicator of mastitis [49], milk coagulation properties [50-52, 54, 55], milk titratable acidity and pH [50], milk major mineral content milk [58], acetone, β -hydroxybutyrate (BHB) and citrate [59], cow's body energy status and feed

efficiency [60, 61], methane emission from dairy cattle [62, 63], cow status pregnancy [64] and dairy farming production system [65].

The objective of this study was to investigate the suitability of different benchtop and portable FTIR spectrometers for on-site milk analysis. These spectrometers have several advantages including improved SNR for measurements acquired over a short period of time (i.e., the multiplex or the Fellgett's advantage) and high throughput with high resolution (i.e., the Jacquinot's advantage). We believe that providing dairy producers with an on-site tool to monitor milk composition will help them detect issues in dairy cows in its early stages without the need to constantly ship milk samples to centralized dairy laboratories, which will help dairy producers to reduce their environmental footprint. The guidelines that were mentioned in this introduction will be used to assemble prototypes for on-site milk analysis and to develop PLS calibration models that will predict major milk components and some minor ones. It must be noted that an on-site milk analyzer will not be used for payment purposes.

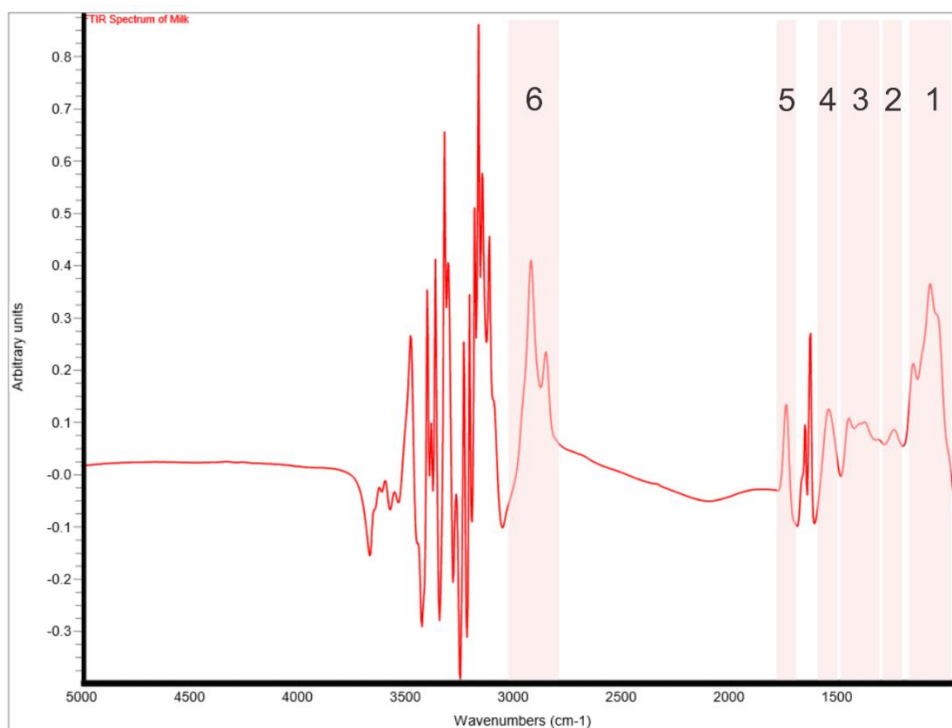


Figure 3-1 Mid-FTIR milk spectrum acquired on a milk analyzer. 1) 1,200-900 cm^{-1} lactose, 2) 1,280-1200 cm^{-1} Amide III of proteins, 3) 1,500-1,200 cm^{-1} different absorbance bands originating from minor milk components 4) 1,565-1,520 cm^{-1} Amide II of proteins, 5) 1,745-1,725 cm^{-1} C = O stretching in the triglyceride ester linkage of milk fat (i.e., Fat A), 6) 2,980-2,800 cm^{-1} C - H stretching of the aliphatic chain in fatty acids in milk fat (i.e., Fat B).

3.2 Materials and Methods

3.2.1 Milk samples

Milk samples that have been used in this chapter can be divided into 3 categories. The first category of milk samples includes industrially homogenized packed milk that was acquired from the local supermarkets in Montreal. Fat levels that were reported on the milk packs were used as the reference value for fat and they were 0%, 1%, 2%, 3.25% and 3.8%. The second category of milk samples includes pre-analyzed producer raw milk samples that were collected by Valacta Inc. (Sainte-Anne-de-Bellevue, Quebec, Canada) from different farms in Quebec. Valacta Inc. provided 370 raw milk samples that were preserved with a bronopol-based preservative along with their FTIR analysis results that included fat, protein, lactose, urea and β -hydroxybutyric acid (BHB) concentrations. The preservative inhibited microbial growth in milk samples; hence, it prevented changes to milk composition by milk microflora before analysis. These numbers were used as reference values for building the calibration models. The third category of milk samples includes two official calibration milk sets that were purchased from Valacta Inc. One set had industrially homogenized milk samples, while the other set had producer raw milk samples. The major milk components in both kits were determined by reference chemical methods.

3.2.2 Homogenization

Several homogenization approaches were evaluated for use in combination with the proposed on-site analyzer. These approaches included: hand-held rotator, CO₂ compression homogenizer, high-pressure air compression homogenizer, mechanical lever homogenizer and ultrasonication. Homogenization of raw milk was performed at room temperature for hand-held rotator, high-pressure air compression and mechanical lever homogenizers. On the other hand, raw milk was homogenized at room temperature and at 60 °C when the CO₂ compression homogenizer was used. For all homogenization approaches, the size of milk fat globules was visually inspected under optical microscope with 1000x magnification power before and after homogenization, and the size of these globules was compared with fat globules of industrially homogenized packed milk. In addition to microscope inspection, Beckman Coulter Delsa™ Nano C (Brea, California, USA) particle analyzer was used to measure the size of fat globules in ultra-sonicated raw milk to verify the homogenization efficacy of the ultrasonication approach. Milk was defined as a diluent whose refractive index (IR), viscosity (cP) and dielectric constant were set to 1.462, 0.8878 and 78.3,

respectively. Marquardt analysis method was used since the resulting data is expected to have multiple peaks, the lower limit for x axis was set to 195 nm to exclude the size of casein micelle from the measurement and the upper limit for x axis was set to 10000 nm, which is the maximum for this equipment. The measurement cell was set to disposable cell from Beckman Coulter size A54093 and its center was adjusted by the equipment. Samples particle distribution was measured at 25 °C in triplicates and D (0.90) (nm), the maximum diameter (nm) and its percentage from the volume distribution were reported. D (0.90) indicates that 90% of the total fat globules volume in the sample comes from particles with diameter that lies below the D (0.90) value. JMP 13 from SAS Inc (Cary, North Carolina, U.S.) was used to perform analysis of variance (ANOVA) to determine the significance of the sonication factors on the D (0.90) value. The significance level was $\alpha = 0.05$, which is the conventional level adopted in analytical chemistry and food science [6]. Two ultrasonic processors were evaluated in this study. The first was Fischer Model 500 homogenizer (115V) 750 Watts (Fisher Scientific, Hampton, New Hampshire, United States). Table 3-1 shows the different combinations of sonication factors that were evaluated for homogenizing raw milk.

Table 3-1 Values of ultrasonication parameters used for milk homogenization with the Fisher ultrasonic processor

Factors	Levels
Time	15s, 30s, 60s, 90s
Probe temperature	Not set, 50°C
Pulse	Note set, 5s on:1s off
Amplitude	100%
Volume	25 mL

The second ultrasonic processor was Vibra Cell Ultrasonic Liquid Processors Model VCX 130PB 130 Watt (Sonics & Materials Newtown, Connecticut, United States). The factors that were evaluated in this study were homogenization type (i.e., homogenization group) and storage, and the significance level was $\alpha = 0.05$. The tested homogenization types were raw milk that was homogenized by a high-pressure homogenizer at Valacta (HM), raw milk that was not homogenized (RM), ultra-sonicated milk for 60s (US1) and ultra-sonicated milk for 90s (US2). Samples were sonicated and their particle size distribution was measured on day 1 (i.e. upon

reception) and on day 2 after they were stored in the fridge for one day. In addition, the statistical differences in fat globules diameter were tested in milk samples that contain high fat level (i.e. >4) that were sonicated for 90s and 120s, and the significance level was $\alpha = 0.05$.

3.2.3 Assembly of prototype analyzers

Three prototypes were assembled that included: FTIR spectrometers from three different vendors, a transmission cell with CaF₂ windows as a sample introduction method and a pumping system to push/pull the milk and the cleaning solution into/from the cell under constant pressure to prevent changes in the cell path length. Prototype one (P1) had a Bomem MB150 (ABB, Montreal, Quebec, Canada) FTIR spectrometer and a temperature-controlled high-pressure transmission cell with 46 μm path length. The cell was also equipped with a bypass valve to divert the excess milk and cleaning solution from passing through the cell to prevent early erosion of the cell windows and to avoid unnecessary fat and protein deposits on the cell windows. The spectrometer was configured and operated by an inhouse software that was written by Thermal-Lube Inc (Pointe-Claire, QC, Canada). Prototype two (P2) had a portable Cary 630 (Agilent Technologies, Santa Clara, California, USA) FTIR spectrometer and a transmission cell with 46 μm path length. The cell was not temperature controlled or high pressure one. The spectrometer was configured and operated by MicroLab software (Agilent Technologies, Santa Clara, California, USA). Prototype three (P3) had a portable ALPHA II (Bruker, Billerica, Massachusetts, USA) FTIR spectrometer and a transmission cell with 47 μm path length. The transmission cell was connected to a pump and the inlet to that pump was covered by a 25 μm filter model X5002 (Qosina, Long Island, New York, USA). The spectrometer was configured and operated by an inhouse software that was written by Thermal-Lube Inc (Pointe-Claire, QC, Canada).

The CaF₂ windows of the transmission cell were separated by ~50 μm polytetrafluoroethylene (PTFE) spacer to create the space between the two windows (i.e. the optical path). To calculate the path length of the transmission cell, a spectrum of the empty dry cell was collected and ratioed against empty background. The number of fringes was counted within specific spectral range and the path length was calculated as follows:

Equation 3-1 Calculating the path length of a transmission cell using wavenumbers and the number of fringes in an FTIR spectrum

$$L = \frac{n \times 10}{2(W_1 - W_2)}$$

Where L = cell pathlength (in mm), W1 = starting wavenumber (cm⁻¹), W2 = ending wavenumber (cm⁻¹), n = number of fringes between W1 and W2.

Equation 3-2 Calculating the path length of a transmission cell using wavelength and the number of fringes in an FTIR spectrum

$$L = \frac{n \times W_1 \times W_2}{2(W_1 - W_2) \times 1000}$$

Where L = cell pathlength (in mm), W1= starting wavelength (in μm), W2= ending wavelength (in μm), n = number of fringes between W1 and W2.

The path length stability of the transmission cell was monitored by passing a solution that contained 10% ethanol and 1% sodium azide at the beginning and at the end of each data collection session. The heights of the following peaks were observed for changes in the path length: 1045 cm⁻¹ and 1085 cm⁻¹ for ethanol and 2048 cm⁻¹ for sodium azide. An aqueous solution of 0.01% triton was used to clean the transmission cell.

3.2.4 Evaluation of the spectral quality of the assembled prototypes

Several aspects were considered as indicators of spectral quality of FTIR spectra collected by the assembled prototypes. SNR was determined by the Analyze Noise functionality in Omnic 7.3 (Thermo Electron Corporation, Waltham, Massachusetts, USA) after selecting the spectral region 1920-1900 cm⁻¹. This spectral region did not contain any information related to milk components. SNR of spectra of homogenized packed milk collected on the assembled prototypes were compared to SNR of milk spectrum collected on a commercial milk analyzer (Foss, Hillerød, Denmark). To assess the potential of an FTIR spectrometer for milk fat analysis, a PLS calibration model was developed using spectra of homogenized packed milk with different fat levels (i.e. 0%, 1%, 2%, 3.25%, 3.8%), which were collected on P1. The sample set was scanned three times and the order of sample scanning was reversed each time. TQ Analyst Professional Edition 7.2.0.161 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to build the PLS models and no spectral pre-treatment was applied to the spectra. Spectra of the same skim milk sample were collected before and after each sample set run. These spectra were subtracted from each other to verify the absence of deposition buildup of milk components on the transmission cell windows, specifically milk fat and protein, after several milk samples were passed through the transmission cell.

3.2.5 Spectral acquisition

To acquire milk FTIR spectra, the spectrometers of P1 and P2 were set up as follows: resolution 16 cm^{-1} , the number of scans for background acquisition was 256, the number of scans for sample acquisition was 64 and milk spectra were ratioed against the spectrum of distilled water. Milk samples were heated in a water bath for 15 minutes at $50\text{ }^{\circ}\text{C}$ before acquiring the spectra. Warm solution of 0.01% triton in distilled water was passed through the cell before pumping the milk samples to condition the CaF_2 windows of the cell and to avoid the formation of air bubbles on these windows. The same solution was passed through the cell after the spectral acquisition to avoid the buildup of milk fat and protein depositions. The temperature of the transmission cell for P1 was set at $50\text{ }^{\circ}\text{C}$. P1 and P2 were moved to Valacta (Sainte-Anne-de-Bellevue, Quebec, Canada), one at a time, and spectra of producer raw milk samples were collected there. Two homogenization approaches were applied to raw milk samples prior to recording the spectra, ultrasonication and mechanical homogenization. Since the particle analyzer was not available at that time, determination of ultrasonication parameters was based on visual observation of fat globules under the optical microscope. Milk was ultrasonicated by a probe from Fisher Scientific (Hampton, New Hampshire, United States) Model 500 (115V) 750 Watts using the following settings: amplitude 100%, temperature $50\text{ }^{\circ}\text{C}$, sonication time 60s and pulse 5s on 1s off. The spectrometer of P3 was set up as follows: resolution 16 cm^{-1} , the number of scans for background acquisition was 32, the number of scans for sample acquisition was 32 and milk spectra were ratioed against the spectrum of distilled water. Warm solution of 0.01% triton in distilled water was passed through the cell before and after pumping the milk samples through the cell. Raw milk samples were homogenized by ultrasonic processor Vibra Cell Ultrasonic Liquid Processors Model VCX 130PB 130 Watt (Sonics & Materials Newtown, Connecticut, United States) for 90-120s. By the end of this treatment, the temperature of the milk was $\sim 50\text{ }^{\circ}\text{C}$, which eliminated the need for the water bath preheating treatment. All milk samples were scanned in triplicates and the total number of spectra was 1110.

3.2.6 Development of PLS calibration models for milk components

TQ Analyst Professional Edition 7.2.0.161 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to build PLS calibration models for major and some minor milk components using FTIR spectra of milk samples and their corresponding reference values for each milk component. The FTIR spectra were either kept raw, without applying any mathematical pre-

treatment, or they were subjected to the Savitzky–Golay first derivative (SG FD) algorithm prior to calibrating the model. The window size was 7 and the polynomial order was 3. After the raw spectra were loaded into the software, the Spectrum Outlier functionality in TQ Analyst was used to exclude all the spectra that were considered as spectral outliers. The refinement of each model went through several iterations. The first iteration was performed on the full FTIR spectrum. The loading spectra that resulted from this iteration were examined and the spectral regions that showed high loadings were kept for the subsequent iteration. The spectral regions that will be included in the model must be relevant to the milk component for which a calibration model is being developed. This process was repeated until a stable calibration model was obtained. For each iteration, cross-validation was performed in TQ Analyst using leave-one-out approach. In addition, Valacta provided spectra of a milk sample set along with the corresponding reference values for each sample. These spectra, which were collected on a commercial milk analyzer (Foss, Hillerød, Denmark), were used to build PLS calibration models for milk components. The figures of merit (FOMs) yielded by these models were considered as reference values to which FOMs produced by the PLS models that were developed for the prototypes were compared.

3.2.7 External validation of the prototype analyzers

The external validation of the assembled prototypes was performed on new milk sample sets, which were not part of the calibration sets that were used to develop the prediction models. The locations for the external validation for P1, P2 and P3 were the Macdonald Campus dairy farm, the McGill IR lab in Macdonald Campus and Léothé dairy farm (Saguenay, QC, Canada), respectively. The number of raw milk samples that were scanned by P1, P2 and P3 were 75, 40 and 40, respectively. For P1, milk samples were scanned over 2 days (D1 and D2). The raw milk samples were homogenized by ultrasonication, and the same procedures for ultrasonication and spectral acquisition were used for the respective prototypes. Major milk components and BHB were predicted by the PLS calibration models that were developed for each prototype. Urea was predicted for spectra collected on P2 and P3. For P3, principal component analysis was used to detect spectral outliers that were excluded from the external validation study. In addition, milk samples for the external validation sets were analyzed by Valacta and their analysis results were used as reference values. The prediction results of the prototypes were regressed against the reference values and the mean differences (MD) and the standard deviation of the differences (SDD) were calculated for each milk component within each external validation set.

3.3 Results and Discussion

3.3.1 Homogenization

Several milk homogenization approaches were evaluated with different conditions for each approach. The hand-held rotation homogenizer did not show any potential for effective homogenization. It was not powerful enough and when milk was tested at room temperature the homogenizer did not affect the size of fat globules. In addition, the CO₂ compression homogenizer did not show any potential for effective homogenization, it caused milk splashing and it was not practical for on-site analysis. Mechanical homogenization was the third approach that was evaluated. The homogenizer had a similar design to high pressure industrial homogenizers except that it had only one homogenization chamber instead of two. The high pressure was provided by either a mechanical lever that was operated manually or by compressed air. Fat globules size was reduced using this approach; however, it raised several concerns. First, the pressure that was generated by the compressed air was too high and it cracked the CaF₂ window of the transmission cell. Second, the high pressure generated by the mechanical lever might increase the path length of the transmission cell over time, which will significantly affect the accuracy of milk measurements. Third, this homogenizer had a weak design. The internal spring in the homogenization head broke under the generated pressure and the homogenizer started to leak from several points after using it for several weeks. Eventually, this approach was deemed impractical for on-site analysis.

Ultrasonication was the most effective homogenization approach. Microscopic inspection revealed that the size of milk fat globules was reduced when milk samples were sonicated for 60s by the Fisher processor. In addition, this approach has the potential to be properly integrated with an on-site milk analysis process. To determine the optimum parameters to adequately homogenize milk, fat globules diameter was measured by a particle analyzer. For the Fisher processor, full factorial ANOVA was applied to the log-transformed measurements of D (0.90) after excluding raw milk and homogenized packed milk measurements. Significant differences were revealed between the treatments that represented different combinations of sonication period, probe temperature and pulse cycle, $F(15, 32) = 3.8867$; $P = 0.0006$. Only the sonication period had a significant effect on the reduction of milk fat globules diameter ($P < 0.0001$), while the other factors and their interactions did not have any significant effect on the values of D (0.90) measurements. Sonication

for 60s and 90s achieved the least mean value for D (0.90), which means that 60s and 90s were the optimum sonication time for the Fisher processor without using a pulse cycle or a heated probe.

Table 3-2 Mean values of D (0.90) from volume distribution of fat globules diameters in raw milk sonicated by the Fisher processor for 15, 30, 60 and 90s.

Time s	Mean D (0.90)	Lower 95%	Upper 95%
15	4762	3036	7471
30	1250	797	1962
60	1118	713	1754
90	1068	681	1675

For the Vibra Cell processor, table 3-3 presents the mean D (0.90) for raw milk samples (RM), high-pressure homogenized milk samples at Valacta (HM), sonicated milk for 60s (US1) and sonicated milk for 90s (US2), whose particle size distribution was measured upon reception and after one-day storage in the fridge. ANOVA tested the differences in fat globules diameter among the four homogenization groups and the effect of refrigerated storage of samples for one day. Significant differences were revealed among the tested groups, $F(7,80) = 6.1157$; $P < 0.0001$. Homogenization type had a significant effect ($P = 0.0003$) on D (0.90) mean values, while the testing day and their interactions did not reveal any significant effect on D (0.90) mean value. This can be interpreted that cold storage of milk samples will not undermine the effectiveness of the sonication treatment in reducing the diameter of milk fat globules. The mean values for D (0.90) were ~1291 nm and ~906 nm for US2 and HM, respectively. This observation indicates that US2 treatment produced fat globules diameters close to that of HM group. In addition, one-tailed t test was performed on the log transformed values of D (0.90) to compare the means of HM group against the means of US1 and US2 groups. The alternative hypothesis of the first and second tests were $\mu_{US1} > \mu_{HM}$ and $\mu_{US2} > \mu_{HM}$, respectively. The first t test revealed that the mean D (0.90) value for US1 group was significantly greater than the mean D (0.90) value for HM group, $t(58) = 2.5832$; $P = 0.0062$; assuming equal variances according to Levene's test $F(1, 58) = 1.3553$; $P = 0.2491$. The second t test revealed that the mean D (0.90) value for US2 group was not significantly greater than the mean D (0.90) value for HM group, $t(58) = 0.7610$; $P = 0.2249$; assuming equal variances according to Levene's test $F(1, 58) = 2.2948$; $P = 0.1352$. These observations indicate that sonicating milk for 90s with the Vibra Cell processor produced particle

size distribution of fat globules similar to the high-pressure homogenization treatment applied to raw milk samples at Valacta.

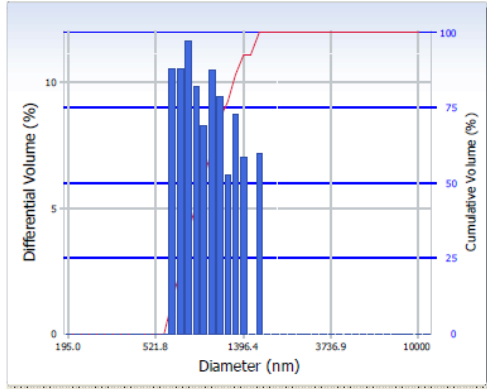
Table 3-3 Mena values of D (0.90) from volume distribution of fat globules diameters in raw milk sonicated by the Vibra Cell processor for 60 and 90s, raw milk and high-pressure homogenized milk

Treatment	Mean D (0.90)	Lower 95%	Upper 95%
HM	906	584	1407
RM	5590	2891	10808
US1	1762	946	3280
US2	1291	693	2403

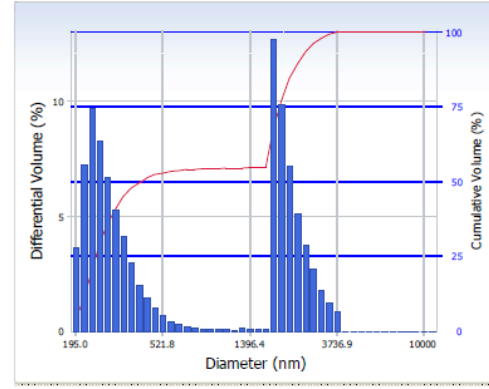
The correlation was not strong between the fat content of milk samples and their D (0.90) values ($R^2 = 0.15$). Nevertheless, milk samples with high fat levels (i.e., >4%) were sonicated by the Vibra Cell processor for 90s and 120s and the means of D (0.90) values were tested for significant differences. Levene's test revealed that there are significant differences between the variances of the two treatments, $F(1, 16) = 7.6991$; $P = 0.0135$. For this reason, Welch test was performed, and it did not reveal any significant differences between the mean values of D (0.90) for the two treatments, $F(1, 9) = 2.9151$; $P = 0.1216$. This observation indicates that sonicating milk samples with high fat content for 90s or 120s will yield statistically similar mean values for the fat globules' diameters. The mean D (0.90) values were ~2252 nm and ~1296 nm for the 90s and 120s, respectively. On the other hand, the standard deviations of the D (0.90) values of the two treatments were 1626 nm and 423 nm for the 90s and 120s, respectively. Since Levene's test revealed that the variances are significantly different, it can be concluded that sonicating milk samples for 120s (i.e., 3000 joules) will yield a particle size distribution of fat globules that is more uniform than sonicating for 90s.

Table 3-4 Mean D (0.90) values for high fat milk samples >4% sonicated for 90s and 120s with the Vibra Cell processor

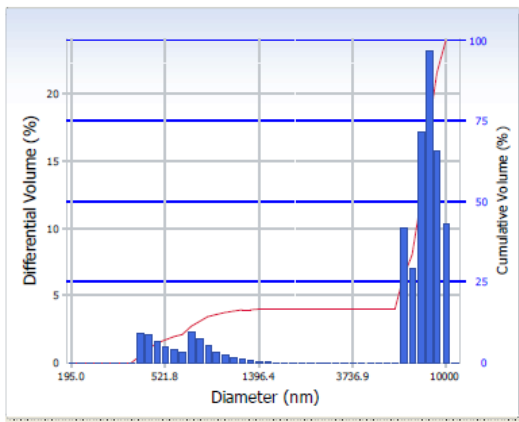
Time s	Mean D (0.90)	Std. Dev.	Lower 95%	Upper 95%
90	2252	1626	1002	3501
120	1296	423	971	1621



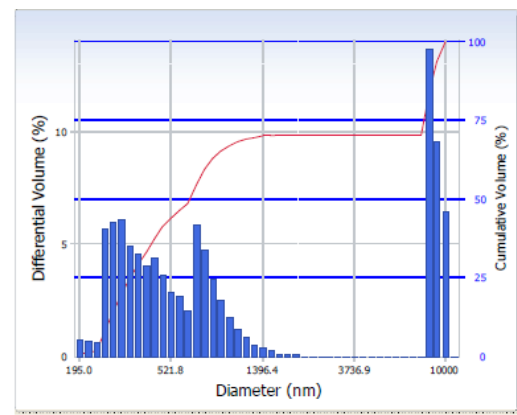
A



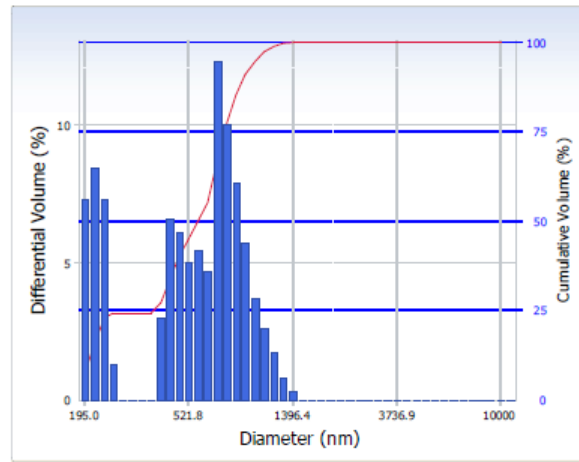
B



C



D



E

Figure 3-2 Volume distribution of fat globules diameters in ultrasonicated raw milk by the Fisher processor. A: raw milk, B: homogenized milk, C: Raw milk ultra sonicated for 15s at 50 °C with pulsing cycle of 5s on and 1s off, D: Raw milk ultra sonicated for 90s at 50 °C with pulsing cycle of 5s on and 1s off, E: Raw milk ultra sonicated for 90s with pulsing cycle of 5s on and 1s off with no heating

3.3.2 Evaluation of the spectral quality of the assembled prototypes

Noise was expressed as Peak-to-Peak noise and root mean square (RMS) noise [67]. Peak-to-peak noise is the difference between the corrected intensities of the highest and lowest noise peaks in the selected spectral region, which does not contain information related to the chemical composition of milk. On the other hand, RMS noise is a statistical analysis of the noise variation, which is equal to the square root of the average of the squares of the data points in the selected spectral region. It is mathematically expressed as follows:

$$RMS = \sqrt{\frac{\sum (i_v - \bar{i})^2}{n}}$$

Where i_v is the intensity at the wavenumber v , $\bar{i}$ is the mean intensity in the selected spectral region and n is the number of analyzed wavenumbers. Smaller values for both indicators mean lower level of noise in the FTIR spectrum. The RMS noise values were: 0.00004481, 0.000111, 0.000126 and 0.00001603 for spectra collected on the commercial milk analyzer, P1, P2 and P3, respectively (Table 3-5). These numbers indicate that the ALPHA II (Bruker, Billerica, Massachusetts, USA) spectrometer produced spectra with the highest spectral quality among the three spectrometers that were evaluated in this study and that its SNR ratio was at the same level as that of the commercial milk analyzer. Milk spectra collected on all prototypes revealed the typical spectral regions that will be used in PLS models to predict the concentrations of major and some minor milk components (Figure 3-3).

Figure 3-4 presents the subtraction results of skim milk spectra that were collected at the beginning and at the end of each cycle of milk spectra acquisition that was performed to assess the potential of FTIR spectrometer for milk fat analysis. No peaks are observed in Fat A, Fat B, Amide II and Amide III regions, which confirms the absence of milk fat and protein depositions. This observation proves that the above-mentioned sample scanning procedure is not causing fat and protein build-up on the CaF₂ windows of the transmission cell in the short-term.

Table 3-6 and Table 3-7 present a comparison of the calibration and cross-validation FOMs for fat PLS models that were developed using spectra of homogenized packed milk collected on P1. Milk fat calibration models gave excellent FOMs for fat predictions by these models. The cross-validation correlation coefficients (r) were 0.99, 0.99 and 0.99 when Fat A, Fat B and Fat A&B

spectral regions were used in these models, respectively. The root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and root mean square error of cross validation (RMSECV) were 0.02%, 0.02%, 0.04% and 0.03%, 0.04%, 0.05% and 0.06%, 0.06%, 0.6% for Fat A, Fat B and Fat A&B models, respectively. These numbers indicate that an FTIR spectrometer can capture information related to the chemical composition of milk. In addition, it can be concluded that the model that used Fat A spectral region produced the most robust prediction model since its RMSEP was the lowest and almost equal to the RMSECV of that model.

Table 3-5 SNR ratio comparison of spectra collected on a FOSS analyzer, prototype 1, prototype 2 and prototype 3

Spectrometer	Spectral Range	Peak-to-Peak Au	RMS
FOSS	1920-1900	0.00009151	0.00004481
P1	1920-1900	0.0002092	0.000111
P2	1920-1900	0.0003006	0.000126
P3	1920-1900	0.00003911	0.00001603

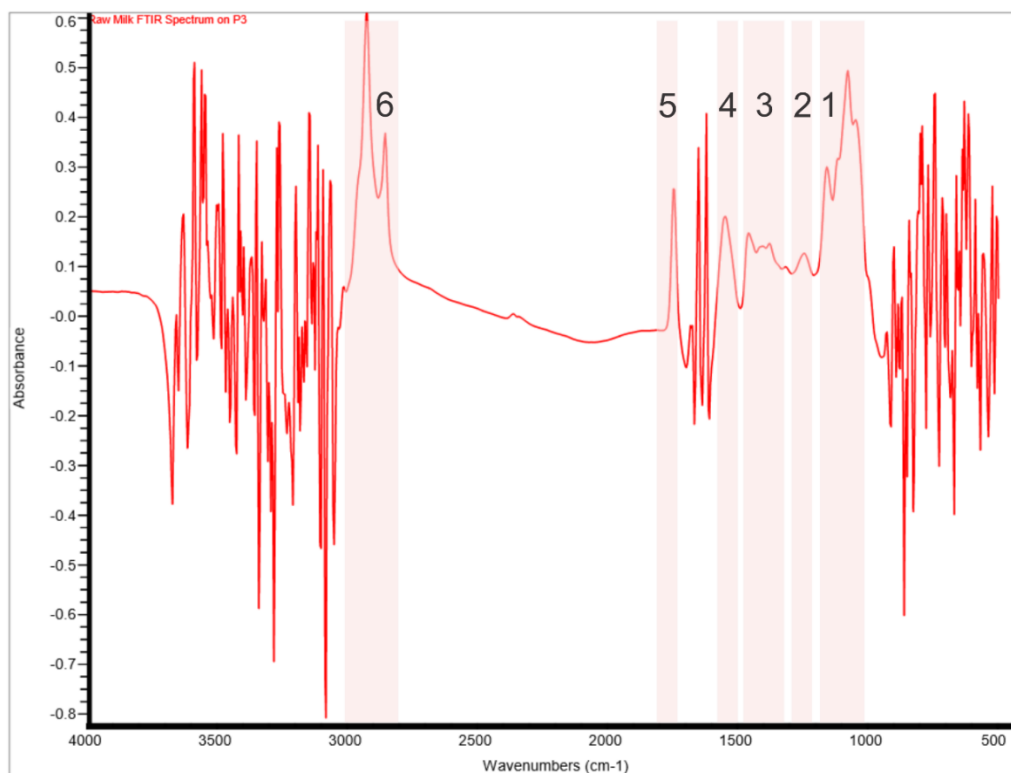


Figure 3-3 Milk FTIR spectrum acquired on P3 with a transmission cell. It reveals the same regions as in milk spectrum collected on commercial milk analyzer with matching spectral quality. 1) 1,200-900 cm^{-1} lactose, 2) 1,280-1,200 cm^{-1} Amide III of proteins, 3) 1,500-1,200 cm^{-1} different absorbance bands originating from minor milk components 4) 1,565-1,520 cm^{-1} Amide II of proteins, 5) 1,745-1,725 cm^{-1} C = O stretching in the triglyceride ester linkage of milk fat (i.e., Fat A), 6) 2,980-2,800 cm^{-1} C – H stretching of the aliphatic chain in fatty acids in milk fat (i.e., Fat B),

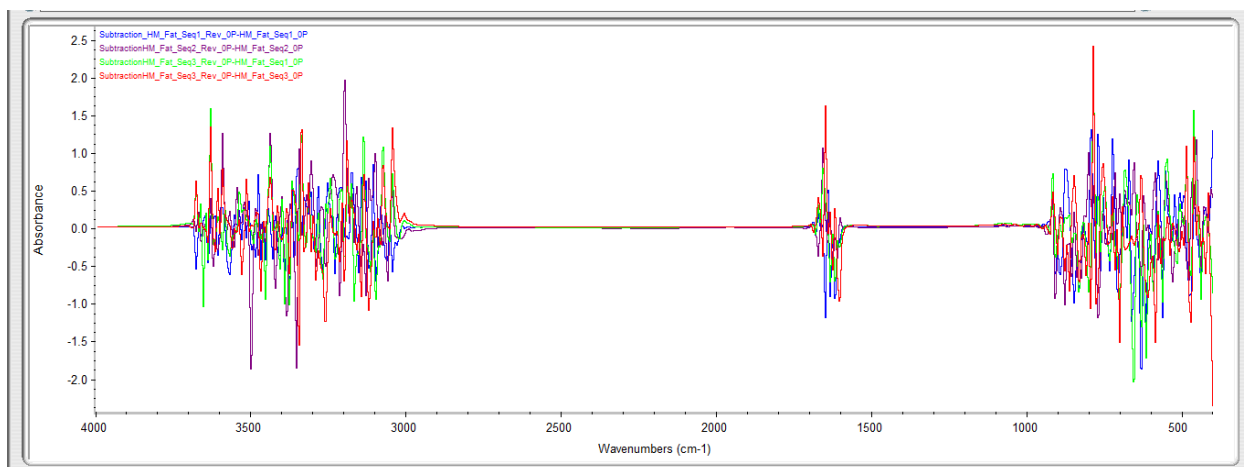


Figure 3-4 Subtraction spectra of multiple measurements of homogenized packed skim milk before and after passing whole milk into the transmission cell. The subtraction spectra do not show any milk fat or protein build up on the cell windows

Table 3-6 Comparison of FOMs of prototype 1 calibration models for homogenized packed milk

Component	Region	r	RMSEC%	RMSEP%	Factors	Bias
Fat (Fat A)	1,778.64-1,735.15	0.99	0.02	0.02	3	-
Fat (Fat B)	2,955.00-2,892.70	0.99	0.03	0.04	4	-
	2,884.62-2,846.52					
Fat (Fat A & B)	1,735.15- 1,735.15	0.99	0.06	0.06	2	-
	2,955.00- 2,892.70					
	2,884.62- 2,846.52					

Table 3-7 Comparison of prototype 1 cross validation FOMs for homogenized packed milk

Component	r	RMSECV %	Factors	Bias
Fat (Fat A)	0.99969	0.04	3	-
Fat (Fat B)	0.99939	0.05	4	-
Fat (Fat A & B)	0.99910	0.06	2	-

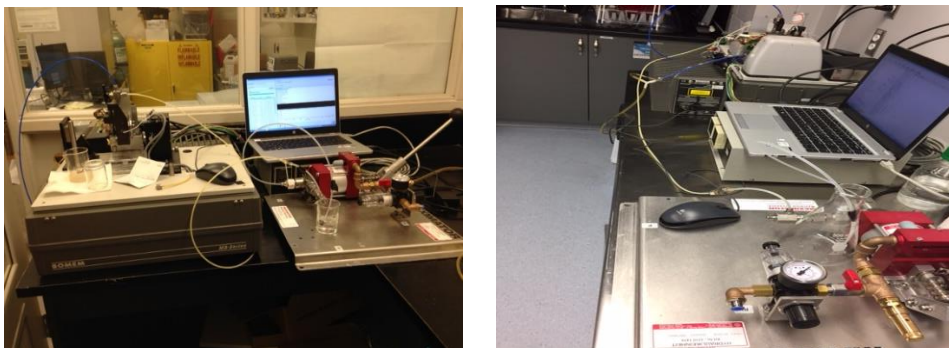


Figure 3-5 Prototype 1 at the McGill IR group lab and Valacta

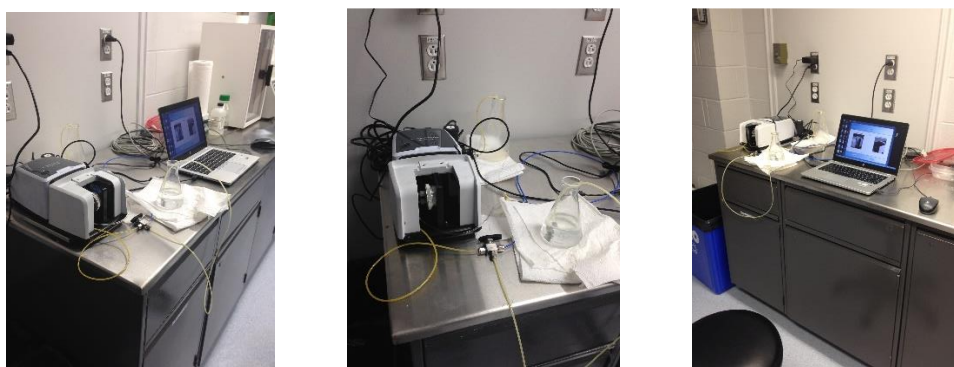


Figure 3-6 Prototype 2 at Valacta

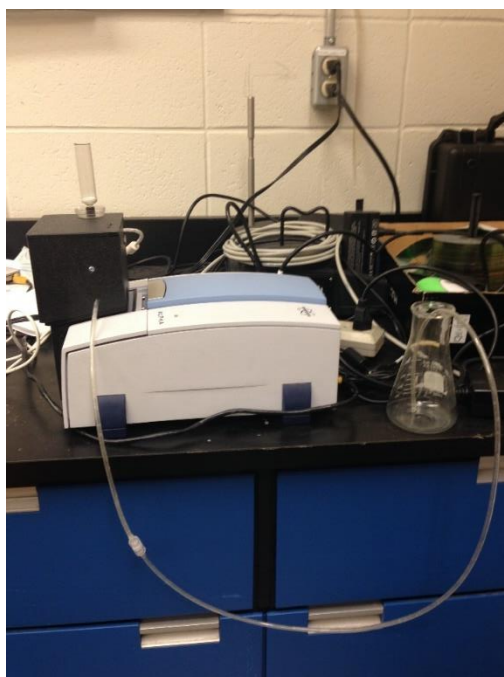


Figure 3-7 Prototype 3 at the McGill IR group lab

3.3.3 PLS calibration models for milk components

Inspection of loading spectra obtained from the initial calibration models revealed high loadings for spectral regions whose absorbance intensities strongly correlate with the corresponding covalent bonds in the analyte of interest. All PLS calibration models for milk fat revealed high loadings for regions 3,000-2,800 cm^{-1} and 1,765-1,725 cm^{-1} . The first region is known as Fat B and it is assigned to the asymmetrical stretching ($\nu_{as}CH_2$) and symmetrical stretching (ν_sCH_2) of the methylene group in milk fat [24]. The second region is known as Fat A and it is assigned to the $C = O$ stretching vibration of the ester linkage in milk fat [24]. All PLS calibration models for milk protein revealed high loadings for regions 1,580-1,500 cm^{-1} and 1,400-1,200 cm^{-1} . The Amide II band and the Amide III band of milk proteins are located at 1,565-1,520 cm^{-1} [24] and at 1,280-1,200 cm^{-1} [25]. All PLS calibration models for lactose revealed high loadings for region 1,200-1,000 cm^{-1} that is assigned to carbohydrates coupled stretching and bending [24].

Several FOMs were used to compare the performance of the calibration models that were developed for milk components. These FOMs included: correlation coefficient (r) for calibration and cross-validation, root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), root mean square error of cross validation (RMSECV), predicted residual sums of squares (PRESS) and number of factors used, bias (if available) and the spectral regions that were used for each model. The root mean square errors (RMSE) of calibration, prediction and cross validation are measures of error for the prediction model. In other words, they represent the average uncertainty that can be expected when predicting the output values for new samples expressed in the same units as the response variable [29]. Mathematically, they are expressed as follows:

Equation 3-3

$$RMSEC = \sqrt{\frac{\sum_{i=1}^N (y_{i,cal} - \hat{y}_{i,cal})^2}{N}}$$
$$RMSEP = \sqrt{\frac{\sum_{i=1}^N (y_{i,val} - \hat{y}_{i,val})^2}{N}}$$

$$RMSECV = \sqrt{\frac{\sum_{j=1}^Z \sum_{i=1}^N (y_{i,j} - \hat{y}_{i,j})^2}{N}}$$

Where $y_{i,cal}$ is the reference value for sample i in the calibration set, $\hat{y}_{i,cal}$ is the predicted value for sample i in the calibration set, $y_{i,val}$ is the reference value for sample i in the validation set, $\hat{y}_{i,val}$ is the predicted value for sample i in the validation set, N is the number of samples [29], $y_{i,j}$ the reference value for sample i in the cross validation group j, $\hat{y}_{i,j}$ the predicted value for sample i in the cross validation group j and Z is the number of cross validation groups [39]. The closer the values of RMSEC, RMSEP and RMSECV to each other, the more robust the model will be. Larger differences between these values indicate that the calibration model is overfitting. PRESS is an additional measure of error [29] and it is mathematically expressed as follows:

Equation 3-4

$$PRESS = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Where y_i is the reference value for sample i and $\hat{y}_i$ is the predicted value for sample i. PRESS is calculated for the PLS calibration model for each iteration when an additional PLS factor or latent variable is added to the model. A satisfactory model should yield a decreasing PRESS for the first few PLS factors. To avoid overfitting, only the minimum number of PLS factors that achieve a significant reduction in PRESS will be eventually included in the final model. Bias is the average value of the differences between the reference and predicted values for a set of replicated sample measurements [29]. Mathematically, it is expressed as follows:

Equation 3-5

$$BIAS = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{N}$$

Once the calibration model has been finalized for a specific milk component, the ratio performance/deviation (RPD) was calculated for that model as follows and models with RPD >3 were considered having excellent prediction capability [59].

$$RPD = \frac{SD}{RMSECV}$$

Where SD is the standard deviation of the reference values for the samples that are used to develop the calibration model. A value close to 1 indicates that the model is not capturing the variation related to the chemical information of the analyte of interest and the model is only describing the standard deviation of the calibration set, which means it will have a poor prediction power. Table 3-8 and table 3-9 present the FOMs of calibration models for milk fat, protein and lactose that were developed using milk spectra collected on the prototypes using Valacta's analysis results as reference values. Raw milk samples that were scanned on P1 were homogenized either mechanically or by ultrasonication, while those that were scanned on P2 and P3 were only ultrasonicated.

Regarding milk fat, all 3 prototypes produced good calibration correlation coefficients $r > 0.96$, which indicates a strong statistical relationship between samples' fat level and the spectral signal. However, the model that was developed using ultra-sonicated Valacta's raw milk calibration kit spectra collected on P3 (Alpha II spectrometer, Bruker, Billerica, Massachusetts, USA) produced the most consistent FOMs. RMSEC, SMSEP and RMSECV were 0.01%, 0.01% and 0.02%, respectively. The small difference between RMSEC and RMSECV indicates a low leverage for the samples that were used in developing the model and the small difference between RMSEC and RMSEP indicates the robustness of this model. The low RMSEP and high RPD values ($RPD = 53.18$) for this model confirm its excellent prediction capability. By comparing the FOMs of fat calibration models that were developed for P1, it can be noticed that ultrasonication produced more consistent FOMs (RMSEC, RMSEP and RMSECV) than mechanical homogenisation, even though the transmission cell temperature was controlled when mechanical homogenisation was used. These observations emphasise the importance of proper milk fat homogenization for milk FTIR analysis.

Regarding milk protein, P3 with ultrasonicated producer raw milk samples gave the most consistent FOMs in comparison to the other models. The values for calibration correlation coefficient, RMSEC, RMSECV and RMSEP were >0.96 , 0.08%, 0.07% and 0.09%, respectively. The small difference between RMSEC and RMSECV indicates a low leverage for the samples that

were used in developing the model and the small difference between RMSEC and RMSEP indicates the robustness of this model. In addition, the RPD of this model is >3 , which indicates an excellent prediction capability. The calibration model that were developed using the spectra of milk samples of Valacta official calibration kits gave inconsistent FOMs. While the RMSEC, RMSEP and RMSECV were all acceptable, the values of the correlation coefficient of the cross-validation models were not acceptable. This can be explained by the interference of atmospheric water vapor on the Amide II band in milk FTIR spectrum, which is close to the OH bending band that is located at $\sim 1650\text{ cm}^{-1}$. To verify the effect of water vapor suppression on milk protein PLS calibration model, a new set of milk samples were scanned on the Nicolet iS5 FTIR spectrometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) that is equipped with a function to suppress the interference of water vapor. The resulting PLS model for milk protein yielded improved FOMs. The values for calibration correlation coefficient, RMSEC, RMSEP and the number of factors were 0.96, 0.03%, 0.01% and 7. The values for cross validation correlation coefficient, RMSECV and the number of factors were 0.93, 0.03% and 7. This observation suggests that the IR measurement chamber in an on-site milk analyzer must be hermetically sealed against the surrounding environment.

Regarding milk lactose, the calibration model that was developed using spectra of milk samples of the industrially homogenized Valacta calibration kit gave the most consistent FOMs. Its correlation coefficient was >0.97 and the values of RMSEC, RMSEP and RMSECV were 0.02%, 0.02% and 0.02%, respectively. In addition, its RPD value confirms its excellent prediction capability. By taking all the calibration and cross-validation FOMs for the three main components of milk, P3 proved to be a good candidate for a portable FTIR milk analyzer.

Regarding urea and BHB, the PLS calibration models did not yield excellent FOMs. The correlation coefficients were 0.63 and 0.76 and the RMSEP values were 1.14 mg/dL and 0.03 mmol/L for urea and BHB, respectively.

To summarize, three FTIR spectrometers were evaluated in this study for on-site milk analysis. The ALPHA II (Bruker, Billerica, Massachusetts, USA) FTIR spectrometer had the best SNR, which was at the same magnitude as the SNR of a commercial FTIR milk analyzer. This observation explains its superior performance to the other spectrometers in this study. Industrial two stages homogenization and ultrasonication were the most effective methods to reduce the size

of fat globules and to eliminate light scattering. Ultrasonication of raw milk samples for 90-120s gave the same particle size profile as that of industrial two stages homogenization. The results of this study show that a portable FTIR spectrometer can be used for on-site milk analysis to determine major milk components. The PLS calibration models for major milk components that were developed using milk FTIR spectra collected on the ALPHA II (Bruker, Billerica, Massachusetts, USA) and iS5 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) spectrometers gave calibration FOMs that were similar to those obtained from models that were developed using milk spectra collected on commercial milk FTIR analyzers. Eliminating the water vapor interference will enhance the performance of these spectrometers for on-site milk analysis.

Table 3-8 Comparison of calibration FOMs for milk components' PLS models developed using milk spectra collected on P1, P2, P3 and a FOSS milk analyzer

Milk Samples	System	Component	Region	r	RMSEC	RMSEP	Factors
Producer raw milk Mechanical homogenization	P1	Fat %	1,766- 1,735; 2,877- 2,846; 2,962- 2,908	0.97	0.17	0.31	2
		Protein %	1,581- 1,511	0.91	0.21	0.17	2
		Lactose %	1,200-1,009	0.67	0.10	0.13	1
Producer raw milk Ultrasonic homogenization	P1	Fat %	1,766-1,735; 2,876-2,846; 2,954-2,915	0.96	0.17	0.15	3
		Protein %	1,589-1,400	0.94	0.15	0.17	3
		Lactose %	1,172-1,018	0.39	0.11	0.23	1
Producer raw milk Ultrasonic homogenization	P2	Fat %	1,759-1,734; 2,860-2,846; 2,929-2,921	0.96	0.16	0.17	4
		Protein %	1,577-1,200 Baseline: linear removed	0.98	0.09	0.11	5
		Lactose %	1,176-1,018 Baseline: fixed two points	0.70	0.17	0.18	3
Valacta milk calibration set Industrial homogenization	P3	Fat %	1,766- 1,727; 2,974- 2,834	0.99	0.02	0.03	3
		Protein %	1,450- 1,199 Baseline: linear removed	0.93	0.03	0.04	2
		Lactose %	1,200- 1,000	0.97	0.02	0.02	1
Valacta raw milk calibration set Ultrasonic homogenization	P3	Fat %	1,761- 1,732; 2,948- 2,838	0.99	0.01	0.01	3
		Protein %	1,457- 1,199 Baseline: linear removed	0.99	0.01	0.05	4
		Lactose %	1,200- 1,000	0.81	0.04	0.04	2
Producer raw milk Ultrasonic homogenization	P3	Fat %	1,766-1,727; 2,950-2,834 Baseline: Fixed two points	0.99	0.06	0.06	2
		Protein %	1,585- 1,504 Baseline: Fixed two points	0.97	0.08	0.07	2
		Lactose %	1,200- 1,006 Baseline: Fixed two points	0.95	0.04	0.07	5
		Urea mg/dL	1488-1454 SG FD	0.63	2.35	1.14	2
		BHB mmol/L	1500-1000 SG FD	0.76	0.01	0.03	2
Producer raw milk	FOSS	Fat %	2,971-2,823	0.99	0.01	0.02	7
		Protein %	1,577-1,498	0.99	0.01	0.01	8
		Lactose %	1,200-960	0.99	0.02	0.03	7

Table 3-9 Comparison of cross-validation FOMs for milk components' PLS models developed using milk spectra collected on P1, P2, P3 and a FOSS milk analyzer

Milk Samples	System	Component	r	RMSECV %	Factors	RPD
Producer raw milk Mechanical homogenization	P1	Fat %	0.95	0.22	2	3.45
		Protein %	0.78	0.33	2	1.64
		Lactose %	0.18	0.15	1	0.96
Producer raw milk Ultrasonic homogenization	P1	Fat %	0.92	0.25	3	2.61
		Protein %	0.84	0.24	3	2.09
		Lactose %	0.01	0.14	1	2.14
Producer raw milk Ultrasonic homogenization	P2	Fat %	0.95	0.18	4	3.28
		Protein %	0.97	0.10	5	4.02
		Lactose %	0.63	0.18	3	1.28
Valacta milk calibration set Industrial homogenization	P3	Fat %	0.99	0.06	3	24.73
		Protein %	0.70	0.07	2	1.39
		Lactose %	0.94	0.02	1	3.09
Valacta raw milk calibration set Ultrasonic homogenization	P3	Fat %	0.99	0.02	3	53.18
		Protein %	0.48	0.05	4	1.07
		Lactose %	0.56	0.05	2	1.20
Producer raw milk Ultrasonic homogenization	P3	Fat %	0.99	0.07	2	8.63
		Protein %	0.95	0.09	2	3.33
		Lactose %	0.88	0.06	5	2.15
		Urea mg/dL	0.32	3.09	4	-
		BHB mmol/L	0.52	0.02	2	-
Producer raw milk	FOSS	Fat %	0.99	0.02	7	39.41
		Protein %	0.99	0.01	8	38.60
		Lactose %	0.99	0.02	7	6.67

3.3.4 External validation of the prototype analyzers

External validation of the assembled prototypes revealed that P3 had the best analytical performance (Table 3-10). The MD values for fat, protein and lactose were ≤ 0.05 , which comply with the stipulations of the AOACI official method 972.16, 33.2.31 [37]. Fat and protein predictions agreed very well with the reference values provided for the samples (Figure 3-8 and Figure 3-9). This level of performance for P3 can be attributed to the high SNR ratio of the Alpha II (Bruker, Billerica, Massachusetts, USA) spectrometer that was comparable to the SNR ratio of the commercial milk analyzer. In addition, the homogenization procedure by the ultrasonic probe was fine tuned when it was applied to milk samples scanned by P3. On the other hand, the PLS prediction model for urea had the least accuracy, which was not surprising considering that the calibration and cross validation indicators of the urea prediction model were the poorest.

Table 3-10 Mean differences and standard deviation of the difference for the external validation sets for the assembled prototypes

FOM	Milk Component	P1		P2	P3
		D1	D2		
MD	Fat	0.139	0.239	0.140	-0.001
	Protein	-0.287	-0.523	0.048	-0.008
	Lactose	0.062	0.036	0.047	0.047
	Urea	-	-	0.935	-2.992
	BHB	0.007	0.010	-0.006	-0.008
SDD	Fat	0.345	0.345	0.174	0.105
	Protein	0.214	0.237	0.111	0.073
	Lactose	0.226	0.292	0.170	0.091
	Urea	-	-	0.857	3.407
	BHB	0.042	0.048	0.033	0.041

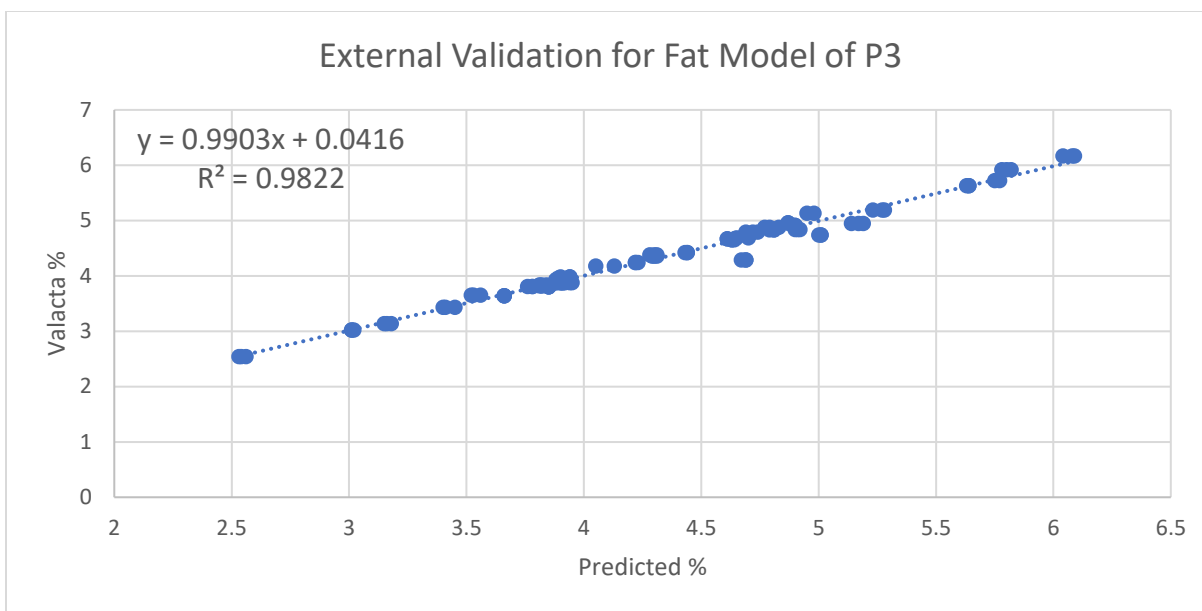


Figure 3-8 Predicted values vs. Valacta analysis values of milk fat for spectra collected on P3 for the external validation set

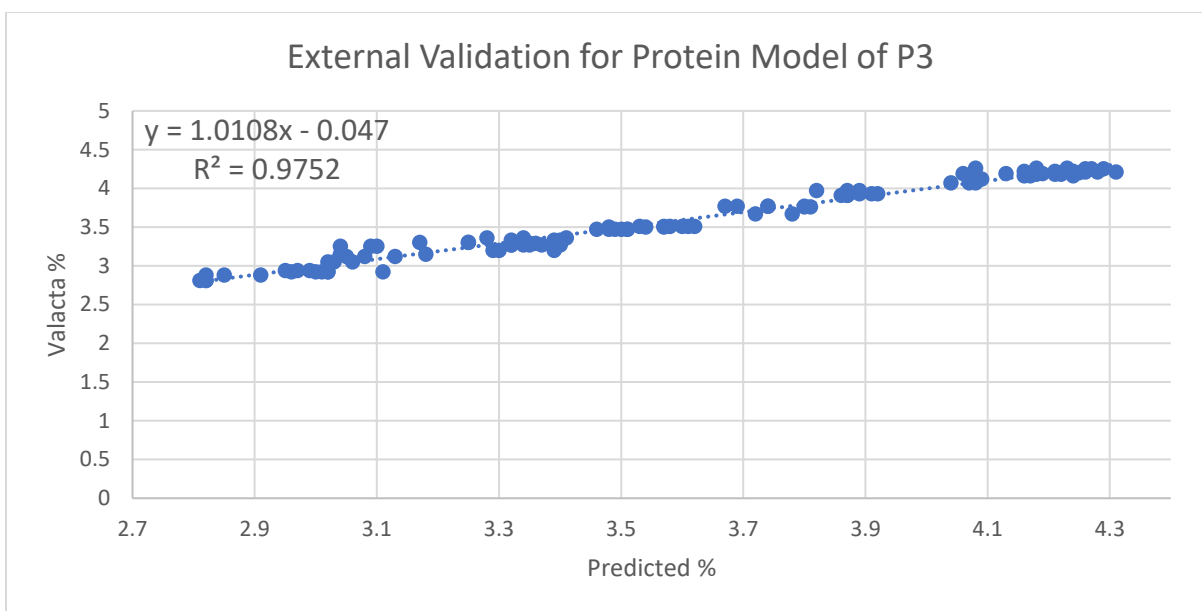


Figure 3-9 Predicted values vs. Valacta analysis values of milk proteins for spectra collected on P3 for the external validation set

3.4 Conclusion

Three FTIR spectrometers were evaluated for potential use for on-site milk analysis. All spectrometers produced high quality milk spectra with adequate SNR; hence, three prototypes were assembled using these spectrometers, which were Bomem MB150 (ABB, Montreal, Quebec, Canada), portable Cary 630 (Agilent Technologies, Santa Clara, California, USA) and portable Alpha II (Bruker, Billerica, Massachusetts, USA). PLS calibration models were developed using spectra of producer raw milk samples collected on P3, which had the Alpha II (Bruker, Billerica, Massachusetts, USA) as a spectrometer. These models revealed excellent prediction capabilities for major milk components and acceptable prediction capabilities for some minor components. The external validation study of P3 revealed that MD values for fat, protein and lactose were -0.001%, -0.008% and 0.047%, respectively. These numbers comply with the stipulations of the AOACI official method 972.16, 33.2.31.

On the other hand, evaluation of different homogenization approaches revealed that the two stages high pressure homogenization and ultrasonication were the most effective methods to reduce the diameter of fat globules. Ultrasonication had the potential to homogenize milk and be integrated within on-site IR milk analysis process. The results showed that applying 3000 joules of ultrasonication energy to 5 mL milk sample for 120s would be sufficient to produce a consistent particle size profile for raw milk samples that was similar to the one obtained by industrial homogenizers.

This study represents a proof of concept that miniaturized FTIR spectrometers can be implemented for on-site milk analysis. The expected benefits of such approach are as follows:

- 1- It will increase the frequency of the determination of milk components using the same technology that is implemented in central dairy laboratories without the need to increase the number of shipped samples to those laboratories.
- 2- The increased frequency of milk analysis on dairy farms will provide more details about the nutrition, health and metabolic state of the cow, which will make herd management decision-making closer to a real time process. This approach will help dairy farmers adopt proactive strategies in managing their herds rather than reacting to emerging issues.

- 3- Increasing the amount of information about milk components using FTIR spectroscopy, which is the same technology implemented in milk analyzers in central dairy laboratories, will better reflect the compliance of dairy farmers with industry standards.
- 4- Finally, having an on-site milk analyzer will guarantee the continuation of the milk analysis process during times when sample shipments might stop for prolonged period, such as the lockdown period of the province of Québec due to the Covid-19 pandemic.

Acknowledgement

The author thanks the following parties for their contribution to this project:

- Dr. Ashraf Ismail (McGill University) as the PI of this project.
- Dr. Jacqueline Sedman (McGill University) for her guidance on spectroscopy and quantitative analysis.
- Valacta Inc. (Sainte-Anne-de-Bellevue, Quebec, Canada) for providing milk samples along with analysis results.
- Thermal-Lube Inc (Pointe-Claire, QC, Canada) for providing the Cary 630 (Agilent Technologies, Santa Clara, California, USA) and the Alpha II (Bruker, Billerica, Massachusetts, USA) portable FTIR spectrometers and their operating software. In addition, Thermal-Lube designed and produced the pumping head used with P3.
- Dr. Emmanuel Akochi-Koblé from Thermal-Lube Inc for his help during the trip to Saguenay.
- Dr. Salwa Karboune for authorizing access to the Beckman Coulter Delsa™ Nano C particle analyzer.
- Dr. Varoujan Yaylayan for authorizing access to the Fischer ultrasonic processor.
- The Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada) and Léothé dairy farm (Saguenay, QC, Canada) for providing the location and milk samples for the external validation of prototype analyzers.

Funding

The project was funded by Innov'Action agroalimentaire program of the ministère de l'agriculture, des pêcheries et de l'alimentation du Québec (MAPAQ) project no. IA216684.

Connecting statement

In the previous chapter, it was proven that milk analysis by FTIR spectroscopy can be performed by a portable FTIR spectrometer equipped with a transmission cell in combination with ultrasonic homogenization of raw milk. It was also proven that this setup can be utilized for on-site milk analysis, which reduces the need to constantly ship milk samples to central dairy laboratories. However, the cost of portable FTIR spectrometers and ultrasonic probes is significant. In this chapter, attenuated total reflectance (ATR) will be evaluated as an alternative to transmission cells as a sample introduction method for raw milk analysis by FTIR spectroscopy. ATR will eliminate the issue of cell clogging that is encountered when passing raw milk through a transmission cell with micrometric optical path length. In addition, a new type of IR spectrometers, linear variable filter (LVF) array spectrometers, will be evaluated for milk analysis. The cost of this IR spectrometer is significantly less than its FTIR counterpart and it does not contain moving parts, which makes it more suitable for a portable on-site analyzer.

Chapter 4: Evaluation of ATR-FTIR spectroscopy and ATR-LVF IR spectrometer for on-site milk analysis

Abstract

Attenuated total reflectance (ATR) is a sample introduction method that is currently used with IR spectrometers in research due to its practicality. The objective of this study was to evaluate the performance of ATR as a sample introduction method with an FTIR spectrometer for on-site milk analysis. In addition, the study aimed at evaluating the potential of a novel miniaturized infrared (IR) spectrometer that does not use interferometry to resolve the different wavelengths for on-site milk analysis. The major advantage of this type of spectrometer is that it has no moving parts due to the absence of an interferometer, which makes it a good candidate for on-site applications of milk analysis. Additional advantages include small size, reduced energy consumption and low cost, which makes it a strong candidate for the mass production of on-site milk analyzers. Drawbacks of this IR spectrometer include limited spectral range, decreased signal-to-noise ratio and reduced resolution, which is 36 cm^{-1} or 18 cm^{-1} at 1800 cm^{-1} , which might decrease the number of milk components that can be determined. ATR-FTIR spectroscopy yielded accurate prediction models for water soluble and colloidal components of milk. RMSECV values for PLS prediction models of lactose and protein were 0.06% and 0.07% respectively, in raw milk samples. Concerning water, ATR-FTIR spectroscopy showed acceptable results with RMSECV of 0.5%. Due to the high percentage of water in raw milk, the prediction error was deemed satisfactory. However, ATR-FTIR spectroscopy gave poor results for the determination of milk fat in raw milk. RMSECV value was 0.39%, which is not acceptable for milk fat determination. The miniaturized IR spectrometer gave an acceptable performance in analyzing milk and it could capture chemical information related to milk lactose, water and protein. Among the four major components of milk, lactose gave the most accurate results with both raw and homogenized milk. The prediction error of lactose was 0.06%. For water, the prediction error was 0.5%, which will not be problematic for applications, such as the determination of extraneous added water to milk. For milk protein, processing raw milk samples with an ultra-sonic probe prior to spectral acquisition produced predication models with reasonable accuracy. The prediction error was 0.16%. On the other hand, the miniaturized IR spectrometer proved to be completely inefficient in capturing chemical information related to milk fat. The prediction error of milk fat varied between 0.2% and 0.5%.

4.1 Introduction

Attenuated total reflectance (ATR) is a sample introduction method that is currently used with IR spectrometers in research due to its practicality. In this method, the sample, liquid or solid, is simply deposited on the measurement surface for spectral acquisition, which simplifies sample preparation procedure prior to the IR measurement. ATR is a phenomenon that results from the effect of the angle of incidence of a light beam on its reflection from a surface. When the angle of incidence is small, partial reflection and partial refraction occurs. However, if the angle of incidence exceeds a critical value, total internal reflection occurs at the surface [22]. In this sample introduction method, the IR radiation does not pass through the sample itself; instead, it is directed through a crystal with a high refractive index that is in contact with the sample (Figure 4-1). Inside the crystal, the beam is reflected multiple times before reaching the detector [25]. The number of reflections that the light will undergo is controlled by the thickness and the length of the crystal and the actual angle of incidence, as shown in Equation 4-1 [68].

Equation 4-1 Calculation of the number of reflections that an IR beam will undergo in an ATR crystal

$$N = \left(\frac{\text{length of ATR crystal}}{\text{thickness of ATR crystal}} \right) \cotangent(\theta)$$

An evanescent wave is generated when the IR beam hits the reflecting surface. This wave penetrates the sample up to a depth of approximately 0.1λ , where λ is the wavelength of the IR radiation. For mid-IR, the penetration depth is less than 10 μm , which is similar to transmission cells with thin optical path length, and by comparison to transmission cells, repeatability is enhanced because the optical path of the evanescent wave is a function of the wavelength of the IR radiation and it is not affected by the sample dimension or the sampling system settings [25]. Having said that, ATR can be considered as an alternative sample introduction method to transmission cell for milk analysis by mid-IR. The advantages of using ATR in milk analysis can be summarized in the following points: 1) the stability of the optical path will enhance the accuracy of milk analysis due to elimination of path length fluctuations, 2) reduced noise that results from the immense water absorbance of IR energy due to the thin optical path length of the evanescent wave, specially for IR bands that are close to the $O - H$ stretching and $O - H$ bending bands of water, 3) ATR yields a spectrum that closely resembles the transmission spectrum of a milk sample [22]. Major milk components are assigned to the same IR bands as in milk FTIR

spectrum obtained by a transmission cell. These assignments are as follows: 1200-900 cm^{-1} for carbohydrates (i.e., mainly lactose) [24], 1100-1060 cm^{-1} for phosphate $O = P - O$ stretching [25], 1280-1200 cm^{-1} for Amide III of proteins [25], 1565–1520 cm^{-1} for Amide II of proteins [24], 1700–1600 cm^{-1} for Amide I of proteins [24], 1650 cm^{-1} for $H - O - H$ bending of water [24], 1745–1725 cm^{-1} for $C = O$ stretching in the triglyceride ester linkage of milk fat (i.e., Fat A) [24], 2400-2300 cm^{-1} for CO_2 , 2980–2800 cm^{-1} for $C - H$ stretching of the aliphatic chain in fatty acids in milk fat (i.e., Fat B) [24] and 3600–3200 cm^{-1} for $O - H$ stretching in water [24] (Figure 4-2). However, there might be one concern when using ATR in milk analysis and that is milk fat. Fat in milk is present in the form of globules that range in diameters from <0.2 to >15 μm . The small fat globules represent 80% of the total number of fat globules but they contain only 3% of the mass of fat. On the other hand, large globules represent only 2% of the total number of fat globules but they contain 95% of the mass of fat [45]. Taking into consideration that the path length of the evanescent wave at the ATR crystal surface can be as small as 1 μm , less than 20% of a large fat globule will be probed by the evanescent wave, which might make the accurate determination of milk fat challenging using ATR as a sample introduction method. Reduction of fat globules diameter through homogenization might be a solution to overcome this obstacle.

Few studies have been reported in the literature on the use of ATR for milk analysis. They are mainly focused on determination of fat and protein in milk. One objective of this study is to investigate the potential of ATR as a sample introduction method for on-site milk analysis using FTIR spectroscopy. In such an application, ATR will eliminate the need for pumping accessories, avoid clogging of transmission cells with thin path length and simplify sample handling since one drop of milk on the ATR surface will suffice for the analysis. On the other hand, the major disadvantage of ATR is surface contamination, which will introduce interferences to the sample spectrum. This issue might be addressed by collecting a background before each sample measurement; however, this procedure will double the time needed for analysis.

Another objective of this study was to evaluate the potential of a novel miniaturized infrared (IR) spectrometer that does not use interferometry to resolve the different wavelengths for on-site milk analysis. Recently, the McGill IR group received a miniaturized filter-based IR spectrometer, the IRSphinx (Comline Elektronik Elektrotechnik GmbH, Germany) (Figure 4-3). This miniaturized spectrometer is based on a linear variable filter (LVF) as a dispersive element mounted on top of

a pyroelectric line sensor. LVFs are small wedged Fabry-Pérot etalons that filter mid-IR radiation at specific wavelengths and that lay on standard detector of 128-pixel pyroelectric line array [69]. This spectrometer is equipped with zinc selenide (ZnSe) ATR crystal as a sample introduction method (Figure 4-4). This spectrometer will be referred to as ATR-IR spectrometer in this study.

The major advantage of the ATR-IR spectrometer over ATR-FTIR ones is that it has no moving parts due to the absence of an interferometer, which makes it a good candidate for on-site applications of milk analysis. Additional advantages of this type of spectrometer include small size, reduced energy consumption and low cost, which makes it a strong candidate for the mass production of on-site milk analyzers. However, the ATR-IR spectrometer operates in the spectral range of 1800-900 cm^{-1} , which might reduce its efficacy in milk fat determination due to the absence of Fat B region from its spectra. Compared to ATR-FTIR spectrometers, the main disadvantages of this type of spectrometers are limited spectral range, decreased signal-to-noise ratio and reduced resolution, which is 36 cm^{-1} or 18 cm^{-1} at 1800 cm^{-1} .

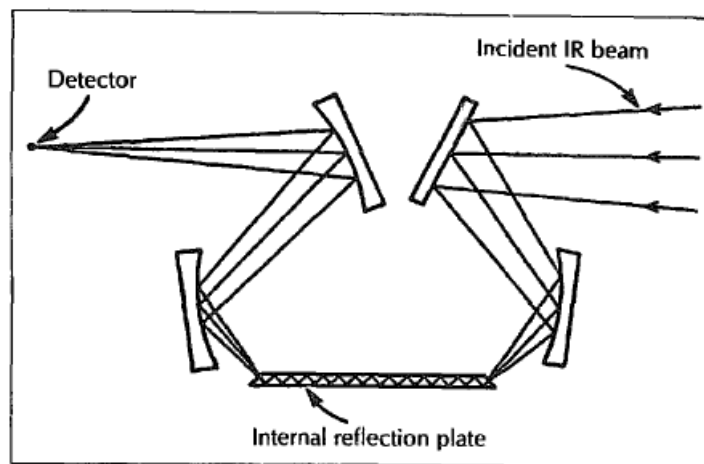


Figure 4-1 Design of an ATR measurement surface [22]

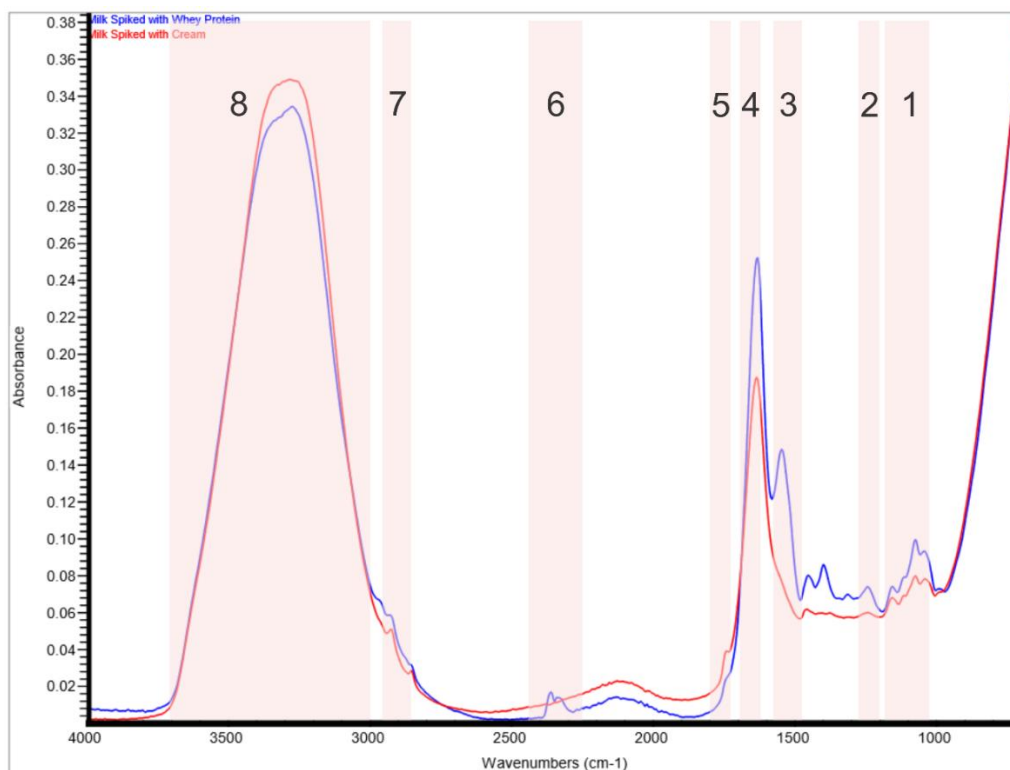


Figure 4-2 Mid-IR ATR-FTIR milk spectrum. 1) 1200-900 cm^{-1} lactose, 2) 1280-1200 cm^{-1} Amide III of proteins, 3) 1565-1520 cm^{-1} Amide II of proteins, 4) 1700-1600 cm^{-1} for Amide I of proteins, 1650 cm^{-1} for $\text{H} - \text{O} - \text{H}$ bending of water, 5) 1745-1725 cm^{-1} $\text{C} = \text{O}$ stretching in the triglyceride ester linkage of milk fat (i.e., Fat A), 6) 2400-2300 cm^{-1} CO_2 , 7) 2980-2800 cm^{-1} $\text{C} - \text{H}$ stretching of the aliphatic chain in fatty acids in milk fat (i.e., Fat B), 8) 3600-3200 cm^{-1} $\text{O} - \text{H}$ stretching in water. Blue: milk spiked with whey protein leads to increased intensities at Amide I and Amide II bands, red: milk spiked with cream leads to increased intensities at Fat A and Fat B bands



Figure 4-3 IRSphinx (Comline Elektronik Elektrotechnik GmbH, Germany) ATR-IR spectrometer. It does not rely on interferometry to generate different wavelengths. Instead it relies on linear variable filter and pyroelectric line sensor to detect the different IR wavelengths.

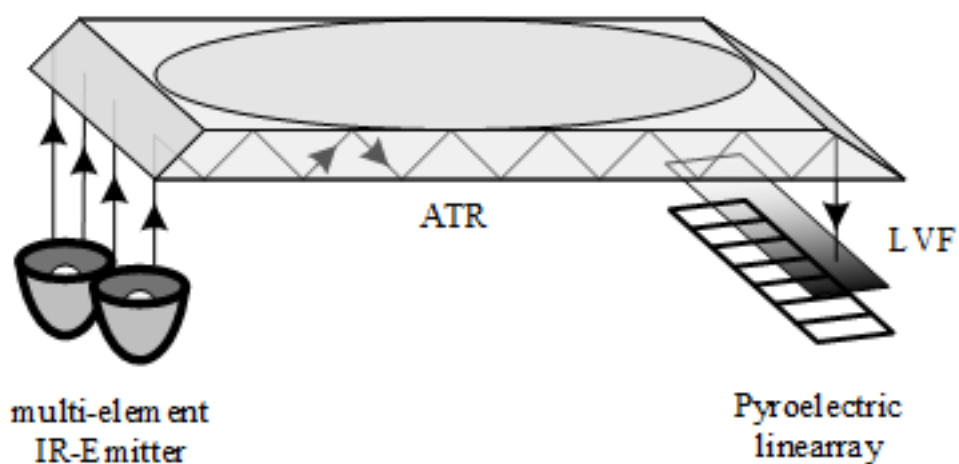


Figure 4-4 Schematic of an LVF ATR-IR spectrometer

4.2 Materials and Methods

4.2.1 Milk samples

Pre-analyzed bovine raw milk samples were received from Valacta Inc (Sainte Anne de Bellevue, Quebec, Canada) along with their milk composition data. We received 11 batches of samples and each batch contained 40 milk samples. The total number of milk samples was 440. These samples were preserved with a bronopol-based preservative and each batch was kept in the fridge till the IR measurement was acquired. For ATR-FTIR measurements, samples of the first four batches were heated to 35 °C before the IR measurement. Samples of the fifth batch were sonicated for 10 minutes at 35 °C in a sonication bath Branson 5200 (Branson Ultrasonics, Danbury, Connecticut, USA). Samples of the sixth batch were homogenized using a laboratory high pressure homogenizer. Each sample was heated to 40 °C prior to IR measurement. Samples of the seventh batch were also homogenized the same way as the previous batch but they were heated to 60 °C prior to IR measurement. Samples of the eighth batch were not homogenized and they were heated to 60 °C prior to IR measurement. Samples of the ninth batch were heated to 60 °C prior to high pressure homogenization, then they were heated to 35 °C before recording the spectra.

For ATR-IR measurements, samples were kept at 40-50 °C in a water bath before the measurement process. After each measurement, the milk was wiped off the ATR surface and it was cleaned with soap solution and water and then wiped dry with a paper towel. Samples of batch ten were kept raw, while samples of batch eleven were homogenized by an ultra-sonic probe from Fisher Scientific Model 500. Fat globules were examined under conventional laboratory microscope with 1000x magnification power.

In addition to the Valacta samples, homogenized packed milk (i.e., 1%, 2%, 3.25% and 3.8%), skimmed milk and 10% cream were purchased from the local market in Montreal. Different mixtures of skimmed milk and 10% cream were prepared to produce homogenized milk mixtures that contained the following fat levels: 0%, 1%, 2%, 3%, 4%, 5%, 6% and 7%. In addition, skim milk was mixed with whole milk 3.8% to produce the following fat levels: 0.76%, 1.14%, 1.52%, 1.9%, 2.28%, 2.66%, 3.04%, 3.23%, 3.42% and 3.8%. Whole milk was added to skim milk by the following percentages (v/v): 20, 30, 40, 50, 60, 70, 80, 85 and 90%.

4.2.2 Spectral acquisition

For the first eight batches of samples, ATR-FTIR spectra were recorded using FTIR Excalibur 3000 spectrometer (Agilent Technologies, California, USA) equipped with single bounce diamond ATR (Specac, UK). The spectra of the ninth batch of Valacta's samples were recorded using Alpha ATR-FTIR equipped with a single bounce diamond ATR (Bruker, Germany) and on a transmission cell CaF_2 windows with 50 μm spacer COAT system (Thermal Lube, Montreal, Quebec, Canada). For ATR-FTIR measurements, mid-IR range of $4000\text{--}700\text{ cm}^{-1}$ was used and a total of 64 scans at 8 cm^{-1} resolution were acquired and ratioed against a background of the clean ATR crystal. One background scan was performed each day before starting the measuring process. A milk droplet was deposited on the ATR measurement surface and the droplet size was fixed at 50 μL using a micropipette. After each measure, the drop of milk was wiped off and the crystal surface was cleaned with soap solution and water and then wiped dry with a paper towel. For the COAT system, the resolution was 16 cm^{-1} for the transmission cell measurements at Thermal Lube.

For batches ten and eleven, the IR spectra were recorded by IRSphinx (Comline Elektronik Elektrotechnik GmbH, Wackersdorf, Germany) equipped with ZnSe ATR crystal. Mid-IR range of $1794\text{--}919\text{ cm}^{-1}$ was used and a total of 300 scans were collected and ratioed against a background of the clean ATR crystal. One background with 500 scans was collected each day before starting the measuring process. The recorded spectra were converted to absorbance and they were exported to comma-separated-values (CSV) file format. The spectra of the first four batches were collected in triplicates, while the rest of the spectra were collected in duplicates. A total of 1040 spectra were collected for the quantitative calibration models of milk components.

4.2.3 Development of PLS calibration models for milk components

TQ Analyst Professional Edition 7.2.0.161 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to build PLS calibration models for major milk components using FTIR spectra of milk samples and their corresponding reference values for each milk component. The FTIR spectra were either kept raw, without applying any mathematical pre-treatment, or they were subjected to the Savitzky–Golay first derivative (SG FD) algorithm prior to calibrating the model. The window size was 7 and the polynomial order was 3. After the raw spectra were loaded into the software, the Spectrum Outlier functionality in TQ Analyst was used to exclude all the spectra that were considered as spectral outliers. The refinement of each model went through

several iterations. The first iteration was performed on the full FTIR spectrum. The loading spectra that resulted from this iteration were examined and the spectral regions that showed high loadings were kept for the subsequent iteration. The spectral regions that will be included in the model must be relevant to the milk component for which a calibration model is being developed. This process was repeated until a stable calibration model was obtained. For each iteration, cross-validation was performed in TQ Analyst using leave-one-out approach. Several figures of merit (FOMs) were used to compare the performance of the calibration models that were developed for milk components. These FOMs included: correlation coefficient (r) for calibration and cross-validation, root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), root mean square error of cross validation (RMSECV), predicted residual sums of squares (PRESS) and number of factors used, bias (if available) and the spectral regions that were used for each model.

For the first four batches of ATR-FTIR spectra, calibration models were developed for lactose, protein, fat, water, non-fatty solids, total solids, all-but-fat. For batches number five, six, seven, eight and nine, only fat calibration models were developed. For ATR-IR spectra, models for fat, protein, lactose and water were developed.

4.3 Results and Discussion

4.3.1 Mid ATR-FTIR milk analysis

Using the ATR-FTIR spectra of the first four batches of the Valacta's samples, PLS models were developed for the following milk components: lactose, protein, fat, water, non-fatty solids, total solids and all-but-fat (Table 4-1 and Table 4-2). Milk fat was not homogenized in these samples. The rationale behind this approach is to evaluate whether an ATR sample interface will produce approximate numbers for milk components. An ATR sample interface will be convenient for on-site analysis applications because it will not require a pumping system to introduce the milk sample to the spectrometer. In addition, eliminating the homogenization process of milk fat will be more convenient for on-site milk analysis applications. It must be mentioned that this type of on-site analyzer will not be used for payment purposes. All spectral regions that are reported in this study showed high loadings in the initial PLS models that were developed for the majority of the respective milk components.

For milk fat, the best model was obtained using raw spectra (Table 4-1 and Table 4-2). Additional spectral pre-treatments (i.e., SG FD) did not enhance the prediction power of the model. The correlation coefficient, RMSEC and RMSEP for the calibration model were 0.85, 0.29% and 0.39%, respectively. The correlation coefficient and RMSECV for cross validation were 0.75 and 0.37%, respectively. Despite the close values that were obtained for the measurements of error (i.e. RMSEC, RMSEP and RMSECV), which is usually a good indicator of model's stability, the model prediction power was undermined by the spectral regions that showed high loadings. These regions were 2,707-3,833 cm^{-1} and 1,820-795 cm^{-1} , which do not represent the regions where milk fat shows the strongest IR absorbance (i.e., Fat A and Fat B). The baselines of these two regions were corrected using two points and one point for the first and the second regions, respectively. The use of such wide spectral regions to produce close values for the measurements of error might be an indicator of overfitting.

Water soluble and colloidal components gave high correlation coefficients for models developed using producer raw milk spectra (Table 4-1 and Table 4-2). The correlation coefficients values of cross validation were 0.95, 0.98 and 0.98, and RMSECV values were 0.06%, 0.07% and 0.06% for lactose, protein and non-fatty solids, respectively. Among the three measurements of error, RMSECV is the most realistic because it represents the average RMSE calculated over several

iterations of model validation. In each iteration, one sample is held out of the calibration set, the model is calculated, the held-out sample is predicted by the model and then RMSE is calculated for that iteration. RMSECV values for lactose and protein are comparable to those reported in the previous chapter and that were obtained by PLS models developed using spectra collected on FTIR spectrometer equipped with transmission cell. The spectral regions that showed the highest loadings for lactose and protein models were 1,141-1,010 cm^{-1} and 1,701-1,484 cm^{-1} , respectively. For non-fatty solids, two spectral regions showed high loadings, which were 1,202-974 cm^{-1} and 1,706-1,476 cm^{-1} . All these regions contain the IR bands that show the maximum IR absorbance by the respective milk component, which means that an ATR-FTIR spectrometer can capture the chemical information related to the water soluble and colloidal components of milk. On the other hand, the water prediction model (Table 4-1 and Table 4-2) gave less accurate results where the correlation coefficient of cross validation and RMSECV values were 0.76 and 0.60%, respectively. Due to the high percentage of water in milk, this error can be acceptable. The spectral region in the water model that gave the highest loadings was 3,370-2,800 cm^{-1} , which spans the OH stretching band of water and the milk fat CH stretching region located at 3000-2800 cm^{-1} . This observation suggests that an ATR-FTIR spectrometer will not be capable of capturing chemical information from the CH stretching region that is located at the shoulder of the broad OH stretching band of water, which might explain the unsatisfactory performance of the fat PLS model. The high variability in water content prediction and the incapability of an ATR-FTIR spectrometer to capture information related to CH stretching might explain the high RMSECV values for all-but-fat and total solids prediction models, which were 0.45% and 0.45%, respectively.

These results show that FTIR spectrometers equipped with ATR accessory can be implemented to determine the non-fatty solids in raw milk samples without any prior treatment. In addition, these spectrometers can approximately determine the water content in raw milk samples. The least satisfactory FOMs were obtained from the raw milk fat calibration model. The correlation coefficient and RMSECV for cross validation values were 0.75 and 0.37%, respectively. The low accuracy of the fat model has also affected the accuracy of the total solids model. Such an error would mean that a sample with 3% fat might be predicted to contain 2.63% - 3.37% fat, which is not an acceptable range of error. This variability of the fat model can be explained by the fact that fat in milk is present in the form of globules that range in diameters from <0.2 to >15 μm . The

small fat globules represent 80% of the total number of fat globules but they contain only 3% of the mass of fat. On the other hand, large globules represent only 2% of the total number of fat globules but they contain 95% of the mass of fat [45]. Taking into consideration that the path length of the evanescent wave at the ATR crystal is only 2-3 μm [70], almost 20% of a large fat globule is probed by that wave. For this reason, reducing the size of fat globules might improve the fat ATR measurement and the accuracy of the PLS fat model. Because of the unsatisfactory calibration model that was obtained for raw milk fat, Valacta's raw milk samples were subjected to different treatments to investigate the effects of these treatments on milk fat predictions. Samples of the fifth batch were sonicated, while samples of the sixth batch were homogenized and scanned by an ATR-FTIR spectrometer and by FTIR COAT system (Thermal Lube, Quebec, Canada) with a 50 μm CaF_2 transmission cell. It must be mentioned that the COAT system was not designed for milk analysis. In addition, different temperatures were used to pre-heat milk samples to convert milk fat into liquid state to minimize light scattering.

Table 4-2 shows that none of these treatments could achieve a RMSECV value comparable to that obtained by FTIR spectrometer and a transmission cell that was reported in the previous chapter. Bath sonication demonstrated a weak linear relationship between spectral intensities and fat levels in milk samples where the cross-validation correlation coefficient was 0.52. Similarly, homogenization alone was not enough to improve the accuracy of fat predictions as demonstrated by the homogenized fat treatment. This treatment produced a weak linear relationship where the cross-validation correlation coefficient and RMSECV values were 0.59 and 0.45%, respectively. The results of both treatments prove that the size of fat globules in raw milk samples is not compatible with the ATR optical path length. The only treatment that showed a significant improvement of the linear relationship between fat levels in milk and spectral data were the transmission cell measurement of homogenized milk fat and the ATR measurements of cream-skimmed milk mixture, in which the cross-validation correlation coefficients were 0.91 and 0.99, respectively. This improvement can be explained by the fact that homogenized fat globules are completely probed by the IR beam in the transmission cell setting and that two stages industrial homogenization of milk fat reduces the size of fat globules to levels compatible with the ATR effective pathlength.

To summarize, an ATR-FTIR spectrometer can capture the chemical information related to water soluble and colloidal components and water but not milk fat in raw milk. PLS calibration models developed for lactose and protein revealed FOMs that were close to those FOMs obtained from calibration models for the same milk components of prototype 3 that was evaluated in the previous chapter, which had a transmission cell as a sample introduction method. For predictions of milk fat, FTIR spectrometer equipped with a transmission cell and ultra-sonic probe proved to be superior to an ATR-FTIR spectrometer. RMSEP for prototype 3 from the previous chapter was 0.01%, while it is 0.39% for the ATR-FTIR spectrometer that was evaluated in this study. Ultra-sonication can reduce fat globules size to a size that is as small as 0.725 μm [71].

Table 4-1 Comparison of calibration models' FOMs for milk components. Calibration models were developed using milk spectra scanned by FTIR spectrometer equipped with ATR sample introduction accessory

Component	Component	r	RMSEC %	RMSEP %	Factors
Producer raw milk - ATR	Lactose	0.96	0.06	0.10	6
	Protein	0.98	0.06	0.11	8
	Solids Non-Fat	0.99	0.05	0.11	8
	Water	0.86	0.47	0.85	8
	Fat	0.85	0.29	0.39	7
	Total Solids	0.89	0.40	0.58	7
	All-but-Fat	0.77	0.39	0.48	6
Sonicated producer raw milk - ATR	Fat	0.64	0.42	0.43	2
Homogenized producer milk – ATR	Fat	0.77	0.34	0.46	3
Homogenized producer milk - Transmission Cell	Fat	0.96	0.16	0.14	2
Cream-Skim milk mix - ATR	Fat	0.99	0.24	0.17	2

Table 4-2 Comparison of cross-validation FOMs for milk components' models developed using milk spectra collected on FTIR spectrometer equipped with ATR sample introduction accessory

Milk – FTIR Measurement	Component	r	RMSECV%	Factors
Producer raw milk - ATR	Lactose	0.95	0.06	6
	Protein	0.98	0.07	8
	Solids Non-Fat	0.98	0.06	8
	Water	0.76	0.60	8
	Fat	0.75	0.37	7
	Total Solids	0.85	0.45	7
	All-but-Fat	0.68	0.45	6
Sonicated producer raw milk - ATR	Fat	0.52	0.47	-
Homogenized producer milk – ATR	Fat	0.59	0.45	-
Homogenized producer milk - Transmission Cell	Fat	0.91	0.23	-
Cream-Skim milk mix - ATR	Fat	0.99	0.33	-

4.3.2 Comparisons with previous ATR-FTIR studies for milk analysis

There have been several reports in the literature concerning the use of ATR-FTIR for the determination of milk components. The following section will discuss the main differences between these reports and the current study in terms of methodology and results.

Etzion *et al.* (2004) investigated the determination of protein concentration in raw milk by mid FTIR combined with ATR as a sample introduction method [25]. The two main differences between our study and theirs are: 1) our study investigated the determination of all major milk components, while theirs was focusing on determining milk protein 2) the type of milk samples that were used to develop the prediction model. Our study relied on 360 producer raw milk samples from different farms across Quebec that resulted in 1080 spectra, while theirs relied on 26 milk standards that were mixed in the laboratory that resulted in 235 spectra. Due to the number of samples that we covered; we can say that our models were more capable of capturing the realistic variation in milk components. In addition, our study focused on PLS, which is the algorithm that is employed in commercial milk analyzers for determining milk components, while in Etzion study they used PLS and Artificial Neural Networks (ANNs) that had PCA scores of the decomposed milk spectra as an input in addition to fat and lactose levels in their samples. ANNs are used to model non-linear relationships between predictors (i.e., spectral data) and responses (i.e., milk component concentrations) and they have two drawbacks. First, it is a black box algorithm, which means if it works it does not provide any tools to explain how it does so. By using information related to milk protein, fat and lactose, the author might have been modeling the relationship between these different components. The author did not mention the statistical approach that was used to randomize those mixes, which means any non-linear relations between these three components might have been modeled by the ANNs algorithm. Second, ANNs are highly susceptible to overfitting, which requires using a large data set to produce a realistic model and that was not the case with the Etzion study. The root mean square prediction error values for the PLS model and the ANNs model that was combined with fat and lactose concentrations were 0.22% and 0.08%, respectively. In our study, RMSEC, RMSEP and RMSECV values were 0.06%, 0.10% and 0.06%, respectively, for raw milk protein PLS model. All these values are less than the reported PLS prediction error for the Etzion study. Regarding their PLS model, the author did not analyze the loading spectra produced by the PLS algorithm, which are extremely important in determining whether the model is capturing the information from the correct spectral regions.

Nevertheless, both studies agreed that ATR-FTIR has a potential for determining protein in raw milk samples.

Iñón et al. (2004) investigated the predictive capability of milk ATR-FTIR spectra in PLS models for predicting total fat, total protein, total carbohydrates, calories and calcium [72]. Their study was focused on determining milk components in commercially packed milk, which is homogenized by a two-stage homogenizer, while our study was focused on determining these components in raw milk. Theoretically, milk fat homogenization will reduce light scattering and that should improve the predictive performance of the PLS models. Their reported values for RMSECV were 0.47%, 0.18%, and 0.5% for fat, protein and lactose, respectively, for PLS models that employed variable selection algorithm to determine the spectral regions to include as an input for the model. Our results reported in table 4-2 show that we achieved lower RMSECV without homogenizing our samples. In addition, Iñón did not report the correlation coefficients of their models, which makes it difficult to assess how well these models were fitting the data and undermines the reliability of their reported RMSECVs. In addition, the number of their samples was limited to 83 that included all types of milk that were available in the local market, while our study included 360 raw milk samples. Iñón also compared the predictive performance of PLS1 vs. PLS2. In the first algorithm, the Y matrix include the reference values for one component; on the other hand, the Y matrix for PLS2 include the reference values for more than one milk component. In this case, the PLS algorithm will decompose the Y matrix and produce a second set of latent variables and the prediction model will be fitting the latent variables of the Y matrix against those of the X matrix. The problem with this approach is that the latent variables of the Y matrix will be based on the correlations between milk components. These correlations are affected by several factors, such as the breed of the cow, the lactation stage and the age of the animal, to mention few, and none of these factors were controlled in their study.

Linker and Etzion (2009) investigated the potential of ATR-FTIR for real-time analysis of raw milk in milking lines by collecting 189 milk samples from 70 cows over a duration of 18 months [73]. The focus of their study was online determination of milk fat and protein during the milking process, while our focus was on determining those two components in an offline setting. In addition, they did not use PLS to build the prediction models. Instead, they used PCA and discrete wavelet to decompose the spectral data in regions where the milk component of interest shows the

most intense absorbance in order to reduce the dimensionality of the data and then they used the PCA scores and the wavelet coefficients as input to build feedforward neural networks (NN) with sigmoid activation functions prediction models for fat and protein. For some protein models, they also included fat concentration as input. The author did not give a compelling explanation for this unusual approach to develop these prediction models. NN are useful for non-linear relationships between predictors (i.e., spectral data) and responses (i.e., milk component concentrations); however, the relationship between absorbance and the concentration of the analyte of interest is linear according to Beer's law, which does not justify their approach. The protein NN models that included fat concentrations as input might have been modeling the correlations between these two components and that might explain the variable prediction errors that they obtained for models developed for different seasons. In general, their prediction error for milk protein was between 0.27% and 0.32%, while our RMSECV for raw milk protein was 0.07%. On the other hand, the prediction error for milk fat varied between 0.1% and 0.8%. Both our studies concluded that ATR-FTIR is not applicable for the determination of milk fat in raw milk samples; however, our study showed promising results for the offline determination of milk protein in raw milk samples.

Bassbasi *et al.* (2014) investigated the potential of ATR-FTIR for determining solid non-fat (SNF) in raw milk using partial least squares (PLS) and support vector machine (SVM), which might be helpful for the current payment system. Their study included 56 milk samples. For PLS, the reported RMSEC and RMSEP values were 0.20%-0.46% and 0.24%-0.51%, respectively, while in our study, RMSEC and RMSEP values for SNF were 0.05% and 0.11%, respectively. This observation suggests that increased number of milk samples is needed to capture the natural variability in SNF in raw milk samples by the PLS prediction model. On the other hand, the reported RMSEC and RMSEP values were 0.18%-0.26% and 0.25%-0.33%, respectively, for the SVM model. However, SVM is appropriate for cases where the relationship between the predictors (i.e., spectral data) and responses (i.e., milk component concentrations) is nonlinear, which is not the case with ATR-FTIR spectroscopy. In addition, SVM did not dramatically reduce the prediction error; hence, SVM is not justified for this application. Nevertheless, both of our studies agree that ATR-FTIR has the potential for the determination of SNF in raw milk samples.

To summarize, the prediction models that were developed during this study using milk ATR-FTIR spectra to predict lactose, protein and SNF gave better prediction errors with raw milk samples in

comparison to what have been reported in the literature so far. As for milk fat, the results of our study agreed with the literature that ATR-FTIR is not appropriate for milk fat analysis due to the limited optical path length of the evanescent wave at the ATR surface, which is shorter than the diameter of raw milk fat globules.

Table 4-3 Summary of reported literature on the application of ATR-FTIR spectroscopy in the determination of milk components. Our study yielded better prediction errors for lactose, protein and SNF in raw milk samples. Our study agreed with these reports that ATR-FTIR spectroscopy is not suitable for milk fat determination in raw milk samples.

Ref.	Algorithm	Milk Samples	Samples No.	Component	Prediction Error
[25]	PLS	Remixed	26	Protein	0.22-0.29%
	ANNs	milk standards			0.08-0.29%
[72]	PLS with variable selection	Commercial packed milk	83	Fat	0.35%
				Protein	0.27%
				Carbohydrate	0.34%
[73]	Feedforward neural networks (NN) with sigmoid activation functions with PCA scores or wavelet coefficients as input	Raw milk	189	Fat	0.1-0.8%
				Protein	0.27-0.32%
[74]	PLS	Raw milk	56	SNF	0.24-0.51%
	SVM				0.25-0.33%

4.3.3 Mid ATR-IR milk analysis

The milk spectrum obtained from the ATR-IR spectrometer shows two main regions of IR absorbance (Figure 4-5). The first region is at 1,200-950 cm^{-1} that can be attributed to lactose and the second one is at 1,700-1,600 cm^{-1} , which contains the Amide I band of proteins and the H – O – H bending band of water at 1,650 cm^{-1} . We can notice that the second absorbance region overlaps with the Amide II band of proteins at 1,565-1,520 cm^{-1} , which is not clearly observed. The ATR-IR spectrum of milk completely lacks absorbance band at the Fat A region located at 1,745-1,725 cm^{-1} that originates from the C = O stretching in the triglyceride ester linkage of milk fat. This observation suggest that the ATR-IR spectrometer might not be capable of capturing chemical information related to milk fat. In addition, the spectrum data spacing is not consistent within the spectral range of the detectors used in this spectrometer. The difference between two consecutive wavenumbers was 13-9, 9-6 and 6-3 for spectral ranges 1,800-1,500 cm^{-1} , 1,500-1,200 cm^{-1} and 1,200-920 cm^{-1} , respectively. This observation suggests that this spectrometer might not produce enough data points in key spectral ranges, such as Fat A, required for the determination of milk fat.

Using Valacta's raw milk samples, PLS models were developed for the following milk components: lactose, protein, water and fat. Raw spectra were used to develop these models and no baseline correction was applied. Table 4-4 and Table 4-5 summarizes the calibration and the cross-validation FOMs of the developed models, respectively. It can be noticed that lactose, which is a water-soluble component, gave the most consistent calibration and cross-validation FOMs. For raw milk samples, the calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.97, 0.05%, 0.07%, 0.95, and 0.07%, respectively. For ultra-sonicated milk samples, the calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.98, 0.04%, 0.07%, 0.96, and 0.05%, respectively (Table 4-4 and Table 4-5). The loading spectra of both lactose calibration models revealed high loadings for spectral region 1,299-987 cm^{-1} , which is assigned to carbohydrates in milk IR spectrum. This observation confirms the fact that ATR-IR spectrometer could capture the chemical information related to lactose in milk that was demonstrated in absorbance bands between 1,200 cm^{-1} and 950 cm^{-1} in the ATR-IR spectrum of milk. The measurements of error (i.e., RMSEC, RMSEP and RMSECV) are comparable to those obtained from lactose prediction model that was developed using ATR-FTIR milk spectra, which

means that this type of spectrometer might be a good candidate for on-site applications that require the determination of water soluble components.

Milk protein, which is a colloidal component, gave greater values for the measurements of error than those of the lactose models; however, the calibration and cross-validation FOMs were consistent, which can be considered as a sign of model stability (Table 4-4 and Table 4-5). The calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.86, 0.21%, 0.25%, 0.81 and 0.23%, respectively. RMSECV that is equal to 0.2% is considered relatively high when compared to RMSECV obtained from calibration models developed with milk FTIR spectra obtained by a transmission cell. In addition, the loading spectra revealed high loadings for spectral region 1,600-1,500 cm^{-1} , which includes the Amide II band that is used for determining milk proteins. Ultra-sonication of raw milk samples enhanced the linearity and improved the prediction capability of the milk protein model and led to a reduction in the measurements of error (Table 4-4 and Table 4-5). The calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.96, 0.10%, 0.16%, 0.91 and 0.15%, respectively. It can be noticed that these FOMs are consistent; however, the prediction error in this case is still higher than that obtained from spectra collected with an ATR-FTIR spectrometer or with FTIR spectrometer equipped with a transmission cell. The loading spectra revealed high loadings for the spectral region 1,658-1,241 cm^{-1} , which includes the Amide II and Amide III bands of protein. This observation suggest that ultra-sonication helped the ATR-IR spectrometer capture more chemical information related to milk proteins. As a result, ATR-IR spectrometer can be used for applications that require the determination of milk protein with reasonable accuracy. The accuracy of the predictions can be enhanced by ultra-sonicating the milk samples.

Water calibration models revealed acceptable performance (Table 4-4 and Table 4-5). For raw milk samples, the calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.98, 0.14%, 0.55%, 0.75 and 0.51%, respectively. It can be noticed that RMSEP and RMSECV are similar, which suggest that the prediction error for water is going to be around 0.5% for the ATR-IR spectrometer. Ultra-sonication of milk samples made the linearity and the FOMs of calibration and cross-validation more consistent; however, it did not reduce the prediction error. The calibration correlation

coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.93, 0.36%, 0.45%, 0.90 and 0.42%, respectively. It must be noted that the entire spectrum that spanned the spectral region $1,780\text{-}1000\text{ cm}^{-1}$ was used to develop these models, which means that these models were indirectly modeling water content information by using information related to other milk components. Water is not a component that determine milk value except in cases of adulteration where extraneous water is added to increase the volume of milk. In such scenario, water must be added with amounts that exceed 10% to generate substantial economic gain, which means a prediction error of 0.5% might be acceptable for detecting this fraudulent practice. Having said that, ATR-IR spectrometer might be a viable option for on-site determination of added water to milk.

Among the four major milk components, fat gave the least satisfactory prediction models (Table 4-4 Table 4-5). For raw milk samples, the calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.39, 0.45%, 0.46%, 0.20 and 0.49%, respectively. This model shows lack of linearity and high values for measurements of error for both calibration and cross-validation data sets. In addition, the loading spectrum did not detect the correct spectral region whose absorbance intensity is correlated with fat content (i.e., Fat A). The spectral region that showed high loadings was $1,639\text{-}1,376\text{ cm}^{-1}$. This observation suggests that the ATR-IR spectrometer under investigation was not capable of capturing any chemical information related to milk fat. This can be explained by the limited spectral range of this spectrometer, which only contains the Fat A region, or the ester linkage stretching band centered at 1745 cm^{-1} . This region did not show any absorbance in milk spectrum due to the limited number of data points that were recorded by the spectrometer (Figure 4-5). In addition, the spectral range of this spectrometer lacks the Fat B region, or the CH stretching bands located at $3000\text{-}2800\text{ cm}^{-1}$, which is a significant contributor to the correct prediction of milk fat in PLS models dedicated for this purpose. For milk samples that were homogenized with ultrasonic probe, the calibration correlation coefficient, RMSEC, RMSEP, the cross-validation correlation coefficient and RMSECV values were: 0.80, 0.42%, 0.59%, 0.69, 0.51%, respectively. Ultrasonic processing of milk samples prior to acquisition of the ATR-IR spectra improved the linearity of the model and the loading spectrum showed high loadings for spectral region $1,754\text{-}1,241\text{ cm}^{-1}$. This spectral region in the loading spectrum is still considered very wide and overlaps with IR bands assigned to other milk components, such as protein. Most importantly, the values of the measurements of

error remain high, which indicates that ultrasonic homogenization did not improve the predictive power of the fat PLS model. The stability and the prediction power of the model were not improved when spectra of industrially homogenized milks samples were used to develop the model. The RMSECV was 0.75% and the model exhibited linearity only when the full spectrum (i.e., 1,770-950 cm^{-1}) was used as an input for the calibration model. We can conclude that the ATR-IR spectrometer under investigation is not a viable option for milk fat determination.

To summarize, the ATR-IR spectrometer that was evaluated in this study could capture chemical information related to milk lactose, water and protein. For lactose, the measurements of error (i.e., RMSEC, RMSEP and RMSECV) that were obtained for the PLS models were either identical or close to those obtained from models developed using ATR-FTIR spectra or FTIR spectra collected with a transmission cell. For water, a prediction error of 0.5% will not be problematic for applications, such as the determination of extraneous added water to milk, that will require the addition of 10% or more of water to generate substantial economic gain. For milk protein, processing raw milk samples with an ultra-sonic probe prior to spectral acquisition will produce predication models with reasonable accuracy. For this component, the ATR-FTIR spectrometer produced better prediction models for milk protein using raw milk samples. The RMSECV values were 0.07% and 0.23% for protein PLS models that were developed using spectra of raw milk samples collected on ATR-FTIR and ATR-IR spectrometers, respectively. On the other hand, this type of spectrometer proved to be completely inefficient in capturing chemical information related to milk fat; hence, it cannot be considered for applications that require milk fat predictions. The main reason for this conclusion regarding milk fat was the anemic number of data points that the spectrometer recorded in the Fat A region.

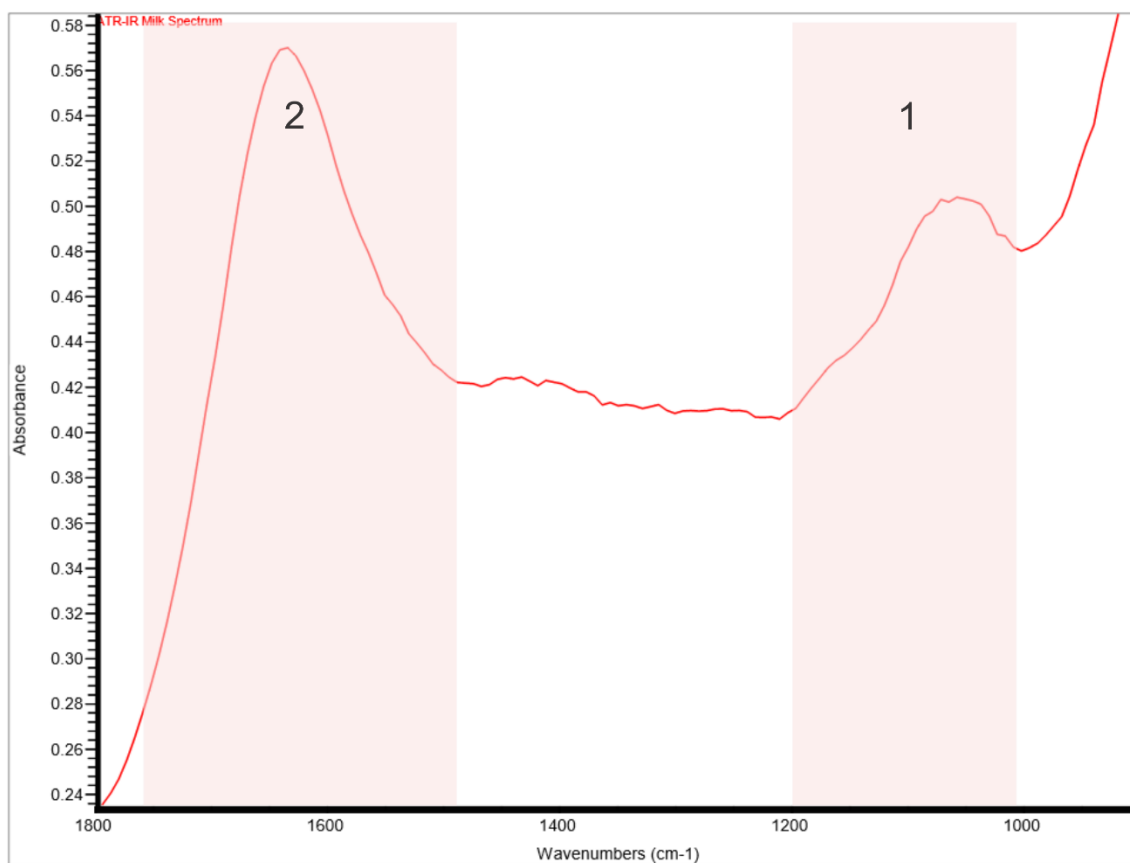


Figure 4-5 Mid ATR-IR milk spectrum. 1) 1200-900 cm^{-1} lactose, 2) 1700-1600 cm^{-1} for Amide I of proteins, 1650 cm^{-1} for H – O – H bending of water. We can notice the absence of absorbance band at 1745-1725 cm^{-1} that originates from the C = O stretching in the triglyceride ester linkage of milk fat (i.e., Fat A).

Table 4-4 Comparison of calibration FOMs for milk components' models developed using milk spectra collected on portable ATR-IR spectrometer

Milk	Component	Corr. Coeff.	RMSEC%	RMSEP%	Factors
Producer raw milk	Lactose	0.97	0.05	0.07	4
	Protein	0.86	0.21	0.25	2
	Water	0.98	0.14	0.55	8
	Fat	0.39	0.45	0.46	1
Ultrasonicated raw milk	Lactose	0.98	0.04	0.07	3
	Protein	0.96	0.10	0.16	3
	Water	0.93	0.36	0.45	3
	Fat	0.80	0.42	0.59	2
Packed milk	Fat	0.98	0.25	0.36	2
Whole milk & skim milk mixture	Fat	0.99	0.12	0.19	3

Table 4-5 Comparison of cross-validation FOMs for milk components' models developed using milk spectra collected on portable ATR-IR spectrometer

Milk	Component	Corr. Coeff.	RMSECV%	Factors
Producer raw milk	Lactose	0.95	0.07	4
	Protein	0.81	0.23	2
	Water	0.75	0.51	8
	Fat	0.20	0.49	1
Ultrasonicated raw milk	Lactose	0.96	0.05	3
	Protein	0.91	0.15	3
	Water	0.90	0.42	3
	Fat	0.69	0.51	2
Packed milk	Fat	0.92	0.75	2
Whole milk & skim milk mixture	Fat	0.98	0.19	3

4.4 Conclusion

ATR-FTIR spectroscopy yielded accurate prediction models for water soluble and colloidal components of milk. RMSECV values for PLS prediction models of lactose and protein were 0.06% and 0.07% respectively, in raw milk samples. These measurements of error were similar to those that were obtained from PLS models developed using milk spectra collected on FTIR spectrometer equipped with a transmission cell as a sample introduction method that was evaluated in the previous chapter. Concerning water, ATR-FTIR spectroscopy showed acceptable results with an acceptable RMSECV of 0.5%. On the other hand, ATR-FTIR spectroscopy gave poor results for the determination of milk fat in raw milk. RMSECV value was 0.39%, which is not acceptable for milk fat determination. This error is due to fat globules size that is considerably larger than the path length of the evanescent wave at the ATR crystal surface. Homogenization did not improve the fat prediction model. For predictions of milk fat, FTIR spectrometer equipped with a transmission cell and ultra-sonic probe proved to be superior to an ATR-FTIR spectrometer. RMSEP for prototype 3 from the previous chapter was 0.01%, while it is 0.39% for the ATR-FTIR spectrometer that was evaluated in this chapter.

The prediction models that were developed during this study using milk ATR-FTIR spectra to predict lactose, protein and SNF gave better prediction errors with raw milk samples in comparison to what have been reported in the literature so far. As for milk fat, the results of our study agreed with the literature that ATR-FTIR is not appropriate for milk fat analysis due to the limited optical path length of the evanescent wave at the ATR surface, which is shorter than the diameter of raw milk fat globules. The main advantage of our study was the large number of raw milk samples whose ATR-FTIR spectra were acquired during this study.

The ATR-IRSphinx spectrometer (Comline Elektronik Elektrotechnik GmbH, Germany) gave an acceptable performance in analyzing milk and it could capture chemical information related to milk lactose, water and protein. Among the four major components of milk, lactose gave the most accurate results with both raw and homogenized milk. The prediction error of lactose was 0.06%. For water, the prediction error was 0.5%, which will not be problematic for applications, such as the determination of extraneous added water to milk, that will require the addition of significant amounts of water to milk to generate substantial economic gain. For milk protein, processing raw milk samples with an ultra-sonic probe prior to spectral acquisition produced predication models

with reasonable accuracy. For this component, the ATR-FTIR spectrometer produced better prediction models for milk protein using raw milk samples. The RMSECV values were 0.07% and 0.23% for protein PLS models that were developed using spectra of raw milk samples collected on ATR-FTIR and ATR-IR spectrometers, respectively. On the other hand, this type of spectrometer proved to be completely inefficient in capturing chemical information related to milk fat; hence, it cannot be considered for applications that require milk fat predictions. The prediction error of milk fat varied between 0.2% and 0.5%.

Acknowledgement

The author thanks the following parties for their contribution to this project:

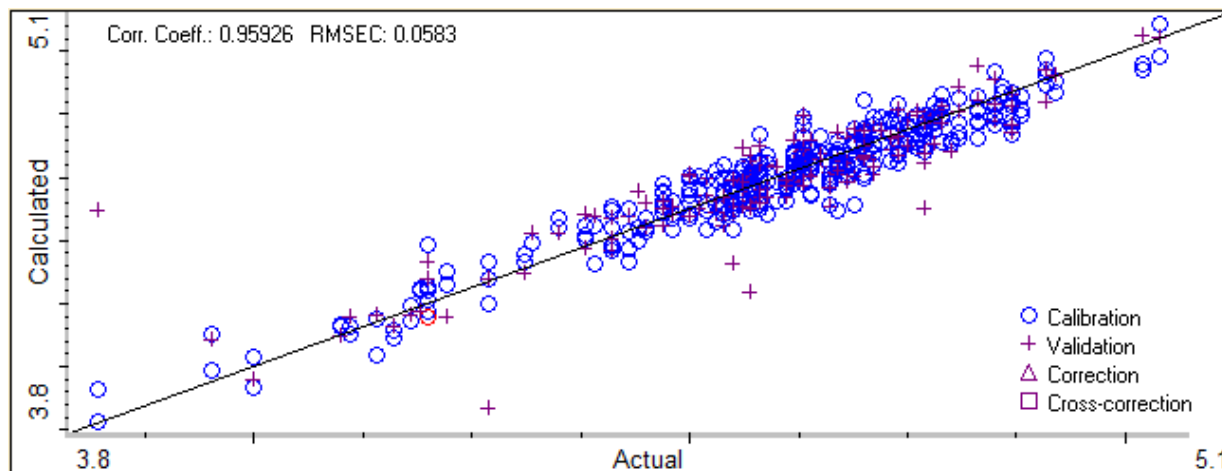
- Dr. Ashraf Ismail (McGill University) as the PI of this project.
- Dr. Jacqueline Sedman (McGill University) for her guidance on spectroscopy and quantitative analysis.
- Comline Elektronik Elektrotechnik GmbH (Wackersdorf, Germany) for providing the IRSphinx spectrometer and its operating software.
- Valacta Inc. (Sainte Anne de Bellevue, Quebec, Canada) for providing the raw milk samples along with their analysis results.

Funding

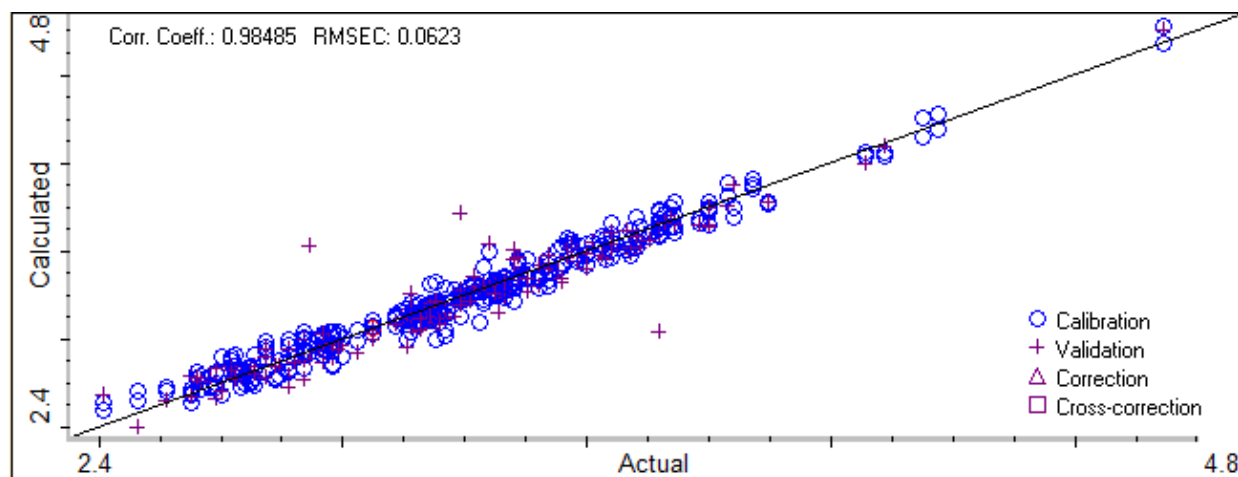
The project was funded by Innov'Action agroalimentaire program of the ministère de l'agriculture, des pêcheries et de l'alimentation du Québec (MAPAQ) project no. IA216684.

Appendix

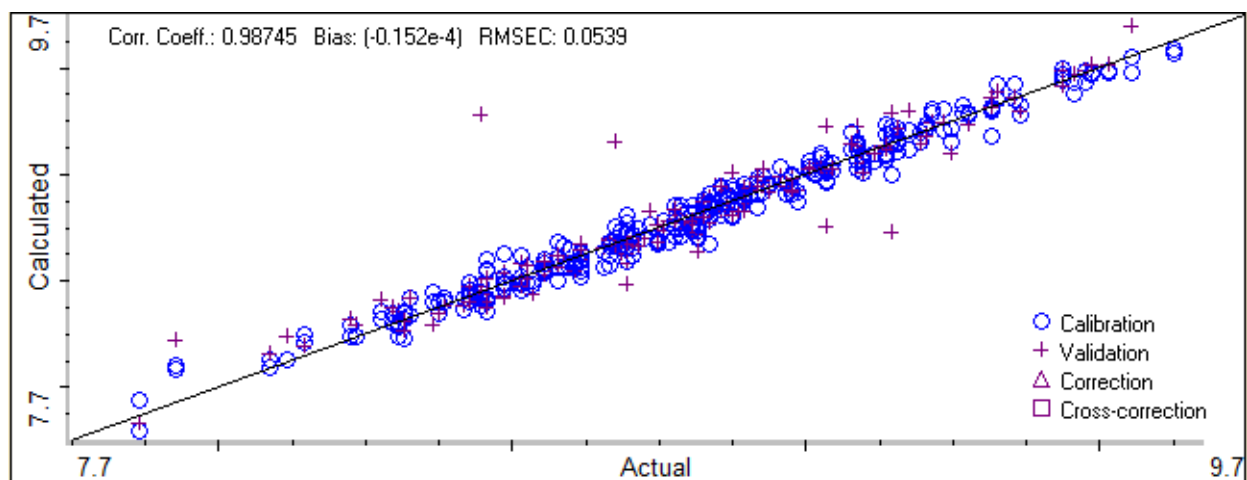
Calculated vs. Reference values for PLS calibration models for major milk components developed using spectra collected on FTIR spectrometer equipped with ATR sampling accessory.



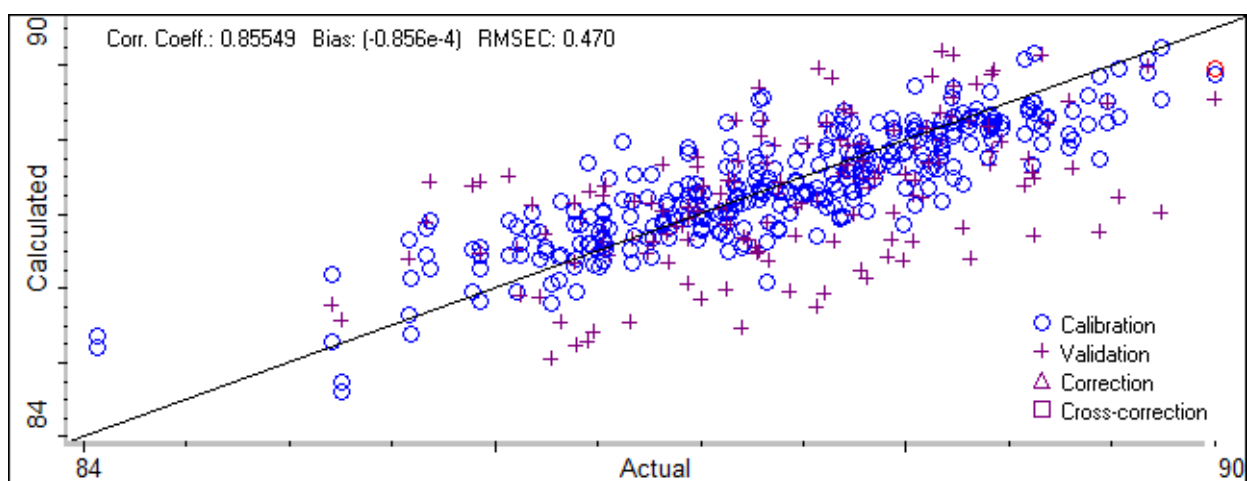
Lactose



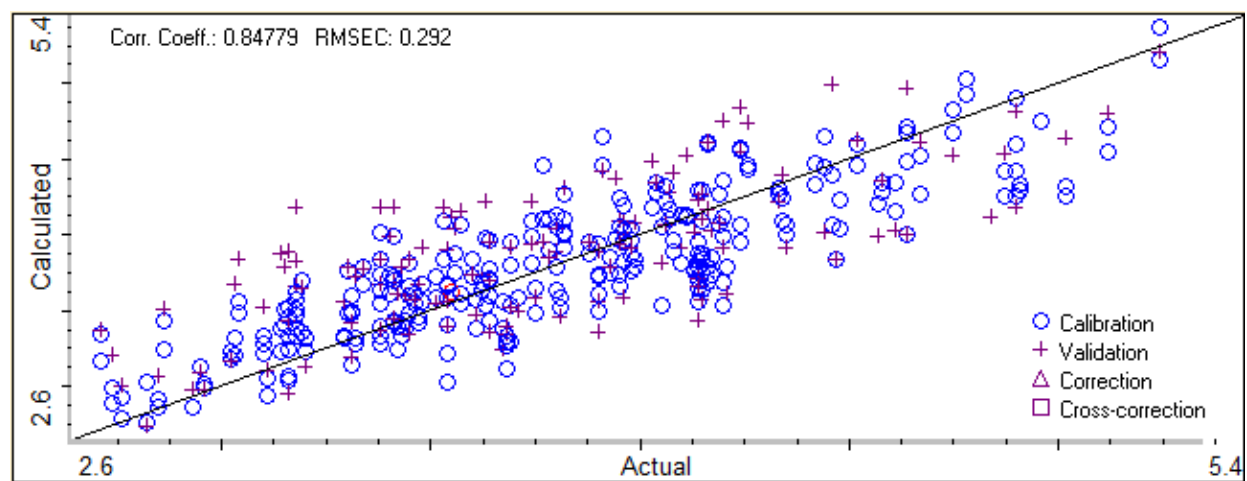
Protein



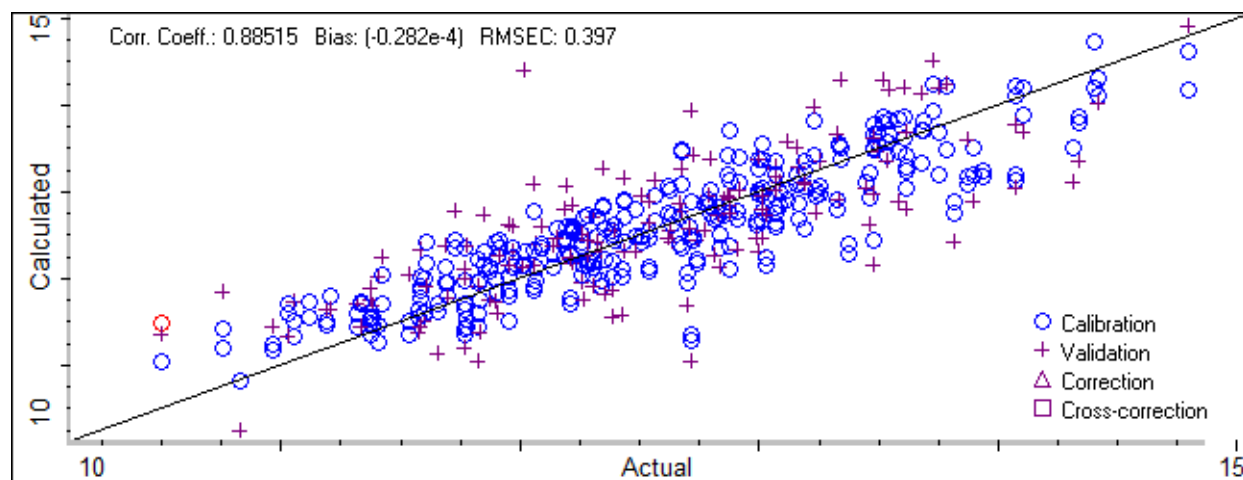
Solids Non-fat



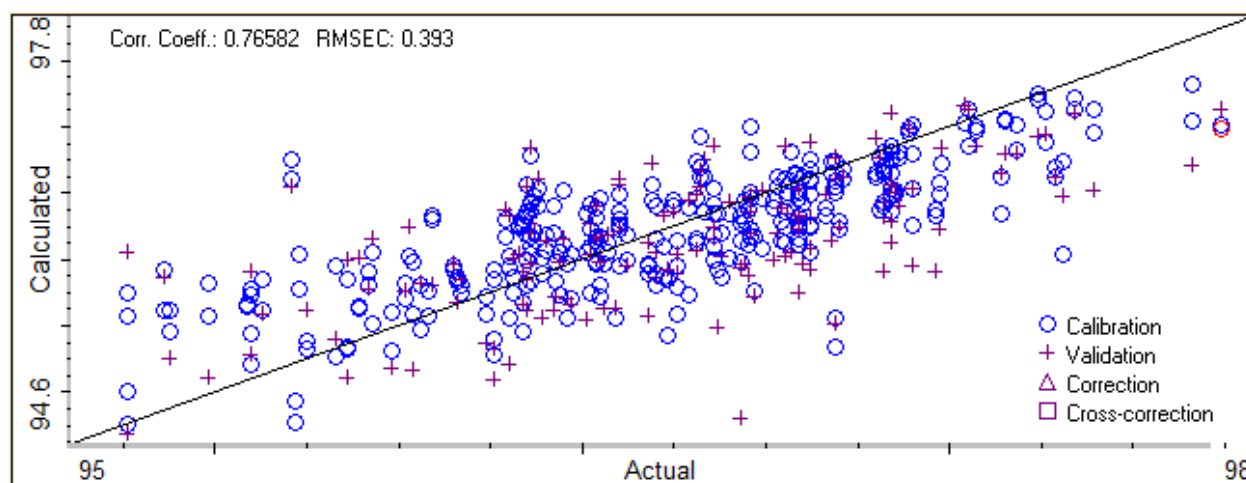
Water



Fat



Total Solids



All-but-fat

Corr. Coeff.: 0.97403 RMSEC: 0.0490

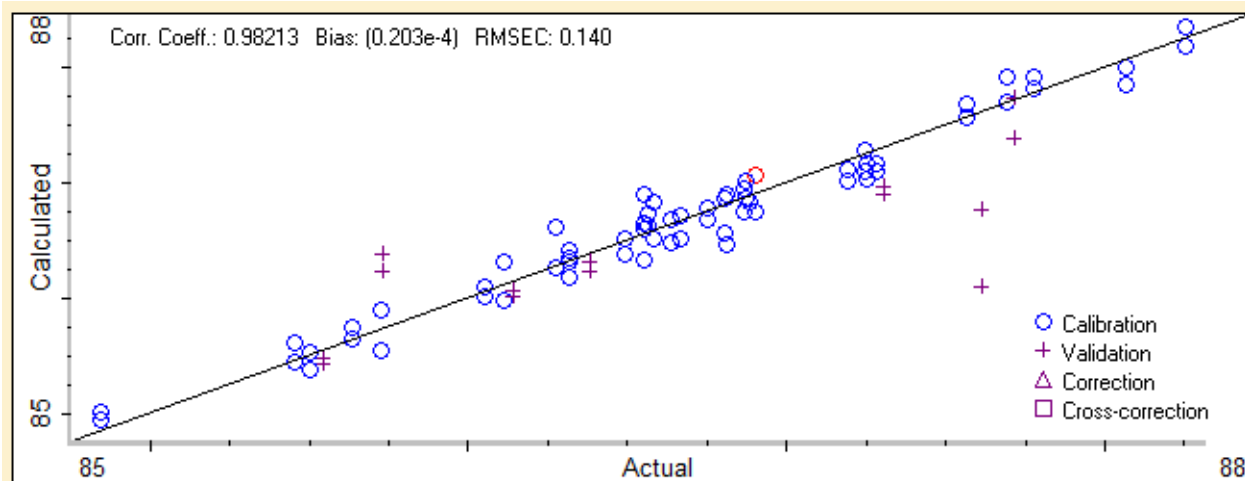
Calculated

Actual

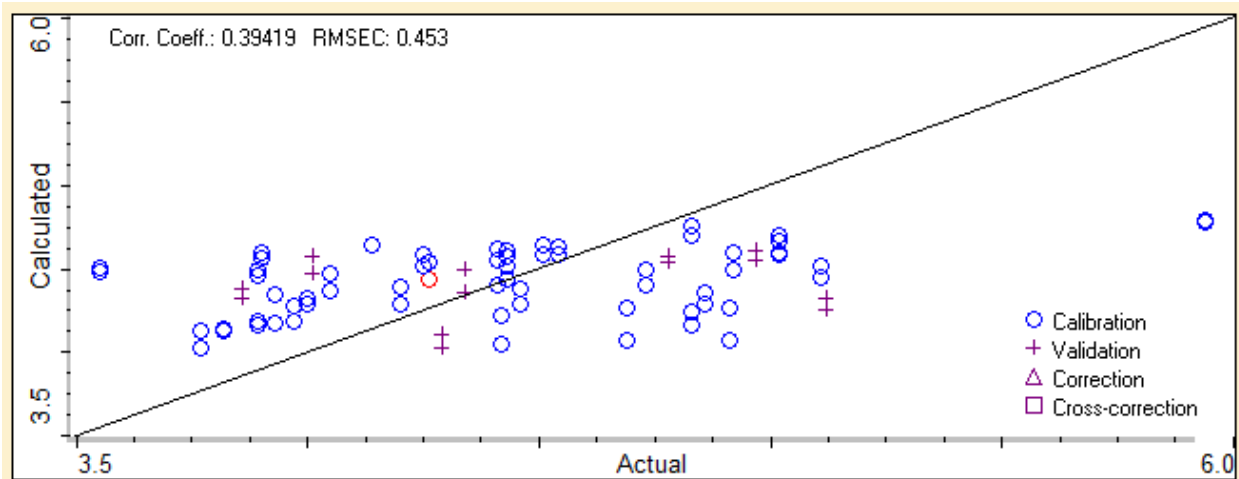
Legend:

- Calibration
- + Validation
- △ Correction
- Cross-correction

105



Water



Fat

Connecting statement

In chapter 3, portable FTIR spectrometers equipped with transmission cell revealed excellent analytical performance for on-site determination of major milk components. In the previous chapter, the novel LFV ATR-IR spectrometer proved to be an effective instrument in capturing chemical information related to water content, true solutes and colloidal components in milk. In this chapter, FTIR and LVF IR spectrometers along with ATR and two transmission-based sample introduction methods (i.e., transmission cell and DialPaht) will be evaluated for the detection of extraneous water and chemical adulterants that are added to milk to mask the addition of water. In some countries, such as Brazil where milk cryoscopy is the official method to detect added water, fraudulent dairy farmers add chemicals to watered-down milk to restore the milk freezing point depression reading of a cryoscope to its legal value. Having a portable IR based instrument that can confirm the authenticity of a milk sample is a crucial solution for the dairy industry in these countries. This solution will employ classification models to differentiate genuine milk samples from watered-down ones and quantitative models to determine added water and chemical adulterants.

Chapter 5: Case study of on-site milk analysis infrared spectroscopic methods for detection of added water and chemical adulterants in Brazilian milk

Abstract

Milk has always been a valuable commodity and the target of fraudulent practices aiming to achieve greater economic gains. Adding water is the most common form of milk adulteration, which can be detected by measuring the freezing point depression using milk cryoscopy. However, the readings of cryoscopy instruments can easily be manipulated by adding water and then dissolving a chemical compound that will compensate for the change in milk freezing point, which is directly related to the concentration of dissolved solids. In this study, it has been proven that the addition of aqueous solution of urea 1.75%, sodium citrate 2.75%, ammonium sulfate 1.5% or sodium carbonate 1.25% to raw milk will maintain its freezing point within the accepted legal limits, even when the addition of the adulteration solution is as high as 50%. To authenticate the milk freezing point readings, multiple combinations of IR spectrometers, sample introduction methods and chemometric algorithms were investigated to detect the addition of water and chemical adulterants to milk. ATR-FTIR spectroscopy combined with PCA and HCA could detect as low as 5% added water, 25 mg/dL added ammonium sulfate, and sodium citrate and 100 mg/dL added urea in raw milk.

In addition, a novel filter-based IR spectrometer equipped with ZnSe ATR crystal was proven to capture chemical information related to the authenticity of milk samples. PCA revealed that the cut off limit for detecting raw milk samples with adulteration solutions was 5%. Its spectra were used to develop PLS model to predict the percentage of extraneous water in milk regardless of the identity of the chemical adulterant that was present in milk with a prediction error of 1.85%.

Homogenizing milk and the use of a transmission cell as a sample introduction method gave the best performance for the classification and quantitative models employed to detect milk adulteration by the addition of water and chemical adulterants. Nevertheless, the use of the DialPath sample introduction accessory (Agilent Technologies, Santa Clara, California, USA) with raw milk gave satisfactory results especially when the path length was set at 30 μm . For raw milk, PCA-QDA classification algorithm with no covariance shrinkage developed using first

derivative spectral data collected with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) gave the best accuracy and error rate. The accuracy values were 98.84% and 100% and the error rate values were 1.16% and 0% for the training and validation sets, respectively. This model yielded perfect specificity for differentiation of adulterated milk samples from genuine ones for the training and the validation set. Sensitivity was 0.99 for the training set. SIMCA algorithm successfully classified adulterated samples into groups according to the chemical adulterant present in these samples. The accuracy for the raw milk SIMCA classification model was 89.90% when first derivative was used. Quantitatively, PLS prediction errors values were 0.39 %, 6.73 mg/dL, 10.3 mg/dL, 4.50 mg/dL and 0.014% for water, urea, citrate, ammonium sulfate and carbonate, respectively, for raw milk scanned with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) with 30 μm path length.

5.1 Introduction

Milk has always been a valuable commodity. According to the Food and Agriculture Organization (FAO) dairy market review [75], global milk output has increased 2.2% in 2018 and it was estimated at 843 million tonnes of milk. Milk output increased in all regions of the world with Asia registering the highest output expansion followed by Europe, North America and South America. In addition, world trade in dairy products expanded 2.9% in 2018 with North America being the largest contributor to dairy products export expansion followed by South America, Central America and the Caribbean [75]. Taking into consideration its economical importance, milk has always been the target of fraudulent practices aiming to achieve greater economic gains. According to the scholarly records of the Food Fraud Database of the United States Pharmacopeial Convention (USP), milk is ranked third among food commodities targeted by fraud [76]. Milk adulteration represents 11.5% of scholarly articles reporting fraud while adulteration of dairy products represents 2.6% [76]. In the economically motivated adulteration (EMA) incident database of the National Center for Food Protection and Defense (NCFPD), cases of dairy products adulteration represent 5.6% of the recorded cases and they are ranked fifth among reported food fraud incidents [76]. According to the same database, the most two common practices of food adulteration are “substitution and dilution” and “unapproved additives”, which represent 65% and 13.4% of recorded cases, respectively [76].

Adding water is the most common form of milk adulteration. It is usually added to increase the volume of milk to achieve greater profits. Addition of water will alter milk composition and reduces its specific gravity, foamy appearance and its nutritional value and it will increase the freezing point of milk [77]. In 1921, Julius Hortvet investigated the effect of adding different amounts of water on the freezing point of milk. He concluded that the addition of 1% of water will increase the freezing point of milk by ~ 0.005 °C and that the measurement of milk freezing point depression (FPD), or cryoscopy of milk, will accurately detect as low as 3% added water [78]. In 1990, the International Journal of Dairy Technology (IDT) published a procedure for detecting extraneous water by measuring the FPD of milk [79]. Currently, FPD of milk is measured by using the thermistor cryoscopic method, which relies on the following principle. When an aqueous solution is cooled without stirring to below the freezing point (i.e., super-cooling), the temperature of the solution initially is reduced to induce nucleation and then reaches a point when enough nuclei become available to trigger auto-crystallization. As a result, the temperature rises rapidly

from the super-cooling temperature to a plateau of relatively constant temperature level that corresponds to the freezing point of the sample. The highest temperature which occurs at the plateau of this freezing curve is known as the solution's freezing point [80]. Since milk contains salts and lactose, the freezing point of its water is depressed by a little over 0.5°C [79].

Milk cryoscopy has its limitations. For example, the freezing point of milk can be affected by several factors, such as breed, lactation and interaction of year and period. The differences in the freezing point of milk due to these factors are 0.0022 °C, 0.0032 °C and 0.0053 °C, respectively [81]. The most significant drawback of this method is that the readings of a cryoscope can easily be manipulated by adding water and then dissolving a chemical compound that will compensate for the change in milk's FPD because the freezing point of milk is directly related to the concentration of dissolved solids [78, 80, 82]. For example, the increase of urea content in milk by 20 mg/L will decrease the freezing point of milk by 0.0003 – 0.0004 °C [81]. For this reason, determination of freezing point of milk is not enough to judge the authenticity of milk.

In 2015, a major manufacturer of milk cryoscopy instruments in South America, PZL Industria Eletrônica LTDA (Londrina, PR, Brazil), contacted the McGill IR group requesting an infrared (IR) based solution to authenticate milk FPD readings and the corresponding added water reported by their instruments. According to senior officials at PZL, the main chemicals used to tamper with their instruments' readings of added water are urea, ammonium sulfate and citrate. Urea is used as an adulterant in milk because it is relatively cheap, easily available, rich in nitrogen and it is naturally present in milk. It represents 55% of the non-protein nitrogenous (NPN) content of milk. Its typical concentration in milk is 18-40 mg/dL and the upper limit is 70 mg/dL. Urea is added to milk to provide whiteness, increase the consistency and shelf life of milk, and for standardizing the content of non-fatty solids present in natural milk. In addition, such type of adulterated milk remains intact for 2 or more days [83]. Ammonium sulfate is generally recognized as safe (GRAS) compound by the US Food and Drug Administration and it is a certified food additive in Japan and the European Union. However, its addition to milk is considered as adulteration. It is a water-soluble nitrogenous compound and is used as a milk adulterant to mask the effects of dilution of added water and to increase the apparent protein content of milk. When added to milk, it increases the lactometer reading and thus the density of milk [83]. Citrate is naturally present in milk [84], it is not routinely determined by dairy control laboratories and its presence in milk only affects its

freezing point [85]. In addition to added water, PZL officials wanted to detect another practice, which is the addition of carbonate or bicarbonate to neutralize milk acidity. Sodium carbonate/bicarbonates are used to neutralize the natural acidity of milk and developed acidity by bacteria responsible for milk spoilage. Sodium bicarbonate is a GRAS food compound and can be used in food at levels up to 2%. According to the Codex Alimentarius, sodium carbonate can be used in all dairy products and it is the only legally permitted preservative, and it is used (<0.3%) as a stabilizer in condensed, evaporated, or powdered milk [83].

Several studies have investigated the potential of FTIR spectroscopy in detecting milk adulteration. Santos *et al.* (2013) investigated the potential of different mid and near IR spectrometers to identify and quantify milk adulteration by the addition of tap water, whey, hydrogen peroxide, synthetic urine, urea, and synthetic milk in different concentrations. The concentration of adulterant in milk samples ranged from 1.87 to 30 g/L for whey, from 0.78 to 12.5 g/L (i.e., 78 – 1250 mg/dL) urea for synthetic urine and urea, from 0.05 to 0.8 g/L urea for synthetic milk, and from 0.009 to 0.15 g/L for hydrogen peroxide. They developed classification models using soft independent modeling of class analogy (SIMCA) for detecting adulterated milk samples, and PLS models to quantitatively predict the adulterant of interest [86]. Jha *et al.* (2015) investigated the potential of ATR-FTIR for detection and quantification of added urea in milk with the following concentrations: 100 ppm, 500 ppm, 700 ppm, 900 ppm, 1300 ppm, and 2000 ppm. Their SIMCA classification model yielded well-separated clusters, which were pure milk, urea<900 ppm and urea>900 ppm. The RMSEP value for the urea quantification model was 254.23 ppm (i.e., 25.423 mg/dL) [87]. Botelho *et al.* (2015) investigated the potential of ATR-FTIR spectroscopy in the simultaneous detection of five adulterants in raw milk using partial least squares discriminant analysis (PLS-DA). These adulterants were water, starch, sodium citrate, formaldehyde and sucrose in the range of 0.5–10% w/v (i.e., 500 – 10000 mg/dL) [88].

The objective of this chapter is to develop an IR-based solution to authenticate milk FPD readings by differentiating adulterated milk from genuine one and quantifying the added chemical adulterant, if possible, in a business-oriented context. The focus of this work will be on milk containing added water and urea, ammonium sulfate and citrate as masking agent for the addition of water. In addition, the scope of this work will include milk containing carbonate as an acidity

masking agent. This IR-based solution will be highly appreciated by the Brazilian dairy section since Brazil is the world's fifth largest producer of milk [75].

5.2 Materials and Methods

5.2.1 Detection of extraneous water and added chemicals by ATR-FTIR

5.2.1.1 Mid-IR bands of chemical adulterants

A 10% solution in water (w/w) was prepared for the following adulterants: ammonium sulfate (AS), sodium bicarbonate (SbC), sodium citrate (SC) and urea (U). In addition, raw milk samples were spiked with the dry form of the previously mentioned adulterants at 10% (w/w). In addition, a 6% (w/w) whey in water solution and whey in raw milk was prepared.

IR spectra were recorded in the mid-IR range of 4000-700 cm^{-1} and a total of 64 scans at 8 cm^{-1} resolution were acquired and ratioed against a background of the clean ATR crystal. A spectrum of water was also recorded, and it was subtracted from the spectra of the water solutions of the above-mentioned adulterants. Omnic 7.3 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to determine the bands of each adulterant in the spectra of the water solution, spiked milk and the subtraction result spectrum.

5.2.1.2 Qualitative detection of chemically adulterated milk

5.2.1.2.1 Samples preparation

A stock solution of 10% (w/w) was prepared for the following adulterants: sodium bicarbonate, sodium citrate, urea and ammonium sulfate. Table 5-1 shows volumes of stock solution that were added to milk samples in order to prepare each level of adulteration for each compound. For each level, the sample's volume was complemented to 10 ml of milk.

In addition, whey and water were also used to adulterate milk samples at the following levels: whey 5%, 10% and 20% (w/w), and water 5%, 10%, 15%, 20%, 25%, 30% and 40% (v/v). For each adulterant, two sets of adulterated samples were prepared using raw milk that was obtained from the Macdonald campus dairy farm and homogenized whole milk 3.25% acquired from local supermarkets. In addition, pure milk samples (raw and homogenized) were used as controls. Samples were refrigerated till the IR measurement time. All milk samples did not contain preservatives.

Table 5-1 Volumes of stock solution used to prepare different levels of chemically adulterated milk samples

Adulterant level in milk sample %	Volume of stock solution μl
0.025	25
0.05	50
0.10	100
0.25	250
1	1000
2	2000

5.2.1.2.2 Spectral acquisition

Samples were heated to 35°C before recording the spectra. ATR-FTIR spectra of raw and homogenized milk samples were recorded using FTIR Excalibur 3000 spectrometer (Agilent Technologies, California, USA) equipped with single bounce diamond ATR (Specac, UK). Raw milk samples ATR-FTIR spectra were also recorded on Alpha ATR-FTIR equipped with a single bounce diamond ATR (Bruker Germany). The mid-IR range of 4000-700 cm^{-1} was used and a total of 64 scans at 8 cm^{-1} resolution were acquired and ratioed against a background of the clean ATR crystal. One background scan was performed each day before starting the measuring process. The milk droplet volume was fixed at 20 μL using a micropipette. After each measure, the drop of milk was wiped off and the crystal surface was cleaned with soap and water and then wiped dry with a paper towel. Homogenized adulterated milk spectra were collected in triplicates, while raw adulterated milk spectra were collected in quadruplicates on the Agilent spectrometer. Raw milk adulterated samples were collected in triplicates on the Alpha spectrometer. A total of 643 spectra were collected for the qualitative detection of chemical milk adulterants by ATR-FTIR.

5.2.1.2.3 Spectral analysis

An inhouse written software at the McGill IR group, DataAnalysis [89], was used to analyze the collected spectra. Hierarchical cluster analysis (HCA) was employed to create dendrograms to differentiate between pure milk, and adulterated samples. In addition, principal component analysis (PCA) was used to detect clustering trends of samples into groups according to their authenticity and to their adulteration level. Forward search feature selection algorithm was applied

to the spectral data to determine the spectral regions that will enhance the clustering trends. Outliers were visually detected on the PCA scores plot and excluded from the analysis.

5.2.2 Effect of extraneous water and chemical adulterants on milk freezing point depression

Aqueous solutions of urea, sodium citrate (i.e., trisodium citrate), ammonium sulfate and sodium carbonate were prepared with concentrations from 1% (w/w) to 3% (w/w) with 0.25% increments. The freezing points of these aqueous solutions were measured by the thermistor cryoscope PZL-7000 (PZL, Paraná, Brazil) and the solutions that gave freezing points between -0.500 °C to -0.540 °C were tested on milk. The chosen solutions were added to milk as follows (v/v): 10%, 20% and 30%. The solutions that gave consistent freezing point close to that of the genuine milk sample for the three tested levels were further investigated. The final concentration that was chosen for each chemical adulterant was used to prepare the adulteration solution that was added to milk samples as follows: 0-50% (v/v) with an increment of 5% and 60-90% with an increment of 10%. The freezing points of these preparations were measured by the PZL cryoscope.

5.2.3 Detection of extraneous water and added chemicals by ATR-IR

5.2.3.1 Spectral acquisition

ATR-IR spectra were collected by the IRSphinx spectrometer (Comline Elektronik Elektrotechnik GmbH, Germany) equipped with a zinc selenide (ZnSe) ATR crystal. The mid-IR range 1794-919 cm^{-1} was used and 300 scans were collected and ratioed against a background of the clean ATR crystal. One background with 500 scans was collected each day before starting the measurement process. The recorded spectra were converted to absorbance and they were exported to comma-separated-values (CSV) file format. Milk samples and aqueous solutions were kept at room temperature 20-23 °C and after each measurement, the milk was wiped off the crystal surface and it was cleaned with soap and water and then wiped dry with a paper towel. A total of 542 spectra were collected for this part of the study.

5.2.3.2 Mid-IR bands of chemical adulterants

Aqueous solutions of urea, sodium citrate (i.e., trisodium citrate), sodium carbonate and ammonium sulfate were prepared. The concentrations of these solutions were 10% (w/v). In addition, homogenized milk samples were spiked with 10% (w/v) of the dry form of these chemicals. The IR spectra of these preparations were recorded by IRSphinx spectrometer (Comline

Elektronik Elektrotechnik GmbH, Germany) equipped with a zinc selenide (ZnSe) 9 bounces ATR crystal. The spectra of genuine milk and pure water were also recorded. The spectra of these preparations were collected in duplicates. Omnic 7.3 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to average the duplicate spectra and to subtract the genuine milk and water spectra from the spectra of the spiked milk samples and the aqueous solution of the chemical adulterants, respectively. The Find Peak functionality in Omnic 7.3 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to locate the centers of the characteristic IR absorption bands for each chemical adulterant.

5.2.3.3 Preparation of adulterated milk samples

Bags of homogenized 3% milk were obtained from Cativa - Cooperativa Agroindustrial de Londrina (Paraná, Brazil) and batches of raw milk were obtained from farms in the vicinity of the city of Londrina (Paraná, Brazil). Adulteration solutions of 1.75% (w/w) urea, 2.75% (w/w) trisodium citrate, 1.5% (w/w) ammonium sulfate and 1.25% (w/w) sodium carbonate were prepared and mixed with raw and homogenized milk as follows (v/v): 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90%. In addition, different combinations of two adulteration solutions (1:1) and three adulteration solutions (1:1:1) were mixed and used to prepare raw and homogenized adulterated milk samples. All milk samples did not contain preservatives.

5.2.3.3.1 Spectral analysis and added water quantification model

An inhouse written software at the McGill IR group, DataAnalysis [89], was used to analyze the collected raw spectra. Principal component analysis (PCA) was used to detect clustering trends of samples into groups according to their authenticity and to their adulteration level. Outliers were visually detected and excluded from the analysis.

TQ Analyst Professional Edition 7.2.0.161 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to build PLS calibration models to predict added water in adulterated raw and homogenized milk samples. The percentage of the added solution was used as the reference value of added water in these models. The spectra were raw, raw with Savitzky–Golay (SG) smoothing or they were subjected to the Savitzky–Golay first derivative (SG FD) algorithm prior to calibrating the model. The window size was 11 and the polynomial order was 3. After the raw spectra were loaded into the software, the Spectrum Outlier functionality in TQ Analyst was used to exclude all the spectra that were considered as spectral outliers. The entire

spectral region was used since the addition of water will have a diluting effect on all milk components, which is a source of variation that can be captured by the PLS algorithm. The models were cross-validated using one-leave-out approach.

5.2.4 Detection of extraneous water and added chemicals by transmission based FTIR

5.2.4.1 Mid-IR bands of chemical adulterants

Homogenized milk samples were spiked with urea, trisodium citrate, ammonium sulfate and sodium carbonate with levels ranged from 1% to 10% and aqueous solutions of these chemicals with similar concentrations were prepared. The FTIR spectra of these preparations were collected by a portable FTIR spectrometer Cary 630 (Agilent Technologies, Santa Clara, California, USA) equipped with a transmission cell with 46-50 μm path length at room temperature. The spectra were ratioed against water background, the resolution was 16 cm^{-1} and the number of coadded scans was 32 scans. The Find Peak functionality in Omnic 7.3 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to determine the centers of the characteristic IR absorption bands of the chemical adulterants. This functionality was applied to raw spectra and to the subtraction spectra of pure milk and water from spiked samples and aqueous solutions, respectively.

5.2.4.2 Preparation of adulterated milk samples

Bags of homogenized 3% milk were obtained from Cativa - Cooperativa Agroindustrial de Londrina (Paraná, Brazil) and batches of raw milk were obtained from farms in the vicinity of the city of Londrina (Paraná, Brazil). Adulteration solutions of 1.75% (w/w) urea, 2.75% (w/w) trisodium citrate and 1.25% (w/w) ammonium sulfate were prepared and mixed with raw and homogenized milk as follows (v/v): 2, 4, 5, 6, 8, 9, 10, 12, 13, 14, 16, 18, 19, 20, 22, 23, 24, 26, 28, 29 and 30%. For ammonium sulfate, a solution of 1.25% was used instead of 1.50% because the addition of this solution at low percentage maintained the freezing point of a milk sample within accepted legal limits. The percentage of added water was 2-30% and the concentration of added chemicals were 35-528 mg/dL, 25-373 mg/dL and 56-791 mg/dL for urea, ammonium sulfate and trisodium citrate, respectively. Sodium carbonate was added in dry form as follows (w/w): 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8%. The sodium carbonate was added in dry form because it is mainly used to neutralize milk acidity and not to

mask the addition of water to milk. Each set of samples for each adulterant was divided into two subsets, a training or calibration set and a validation set. The training or calibration set included samples with the following levels of the adulteration solution: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30% , while the validation set included samples with 5, 9, 13, 16, 19, 23, 26 and 29% adulteration solution. In addition, a mixture was prepared from the three adulteration solutions (1:1:1) and it was added to milk samples at 10%, 20% and 30% (v/v). These samples were included in the validation set. All milk samples did not contain preservative.

5.2.4.3 Spectral acquisition

A portable FTIR spectrometer Cary 630 (Agilent Technologies, Santa Clara, California, USA) was used to collect the spectra of adulterated milk samples and genuine ones using two transmission-based sample introduction methods: a transmission cell with 46-50 μm path length, which was used with homogenized milk samples, and the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) with a path length set at 50 or 30 μm , which was used with raw milk samples. Samples were scanned at room temperature, the spectra were ratioed against water background, the resolution was 16 cm^{-1} and the number of coadded scans was 32 scans. An aqueous solution of 0.01% triton was used to clean the transmission cell and the measurement surface of the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) between samples. Samples were scanned in triplicates and the total number of collected spectra was 1319 for this study.

5.2.4.4 Classification models for differentiation of adulterated milk samples

Two-tier approach was used to differentiate adulterated milk samples from genuine ones. In the first step, a classification model was developed to determine whether a milk sample was genuine or adulterated. In the second step, another classification model was developed to determine the type of chemical adulterant that was added to the sample, if any. MATLAB codes were written to calculate differential first derivative (FD) of the spectra with a derivative window of 1, to vector normalize (VN) the spectra and to load individual spectra into a matrix. The spectral region that was retained for the classification models was $1600\text{-}950\text{ cm}^{-1}$. For each step, classification models were developed using raw, VN raw, FD and VN FD spectral datasets.

For the first step, principal component-based discriminant analysis (PC-DA) algorithms were chosen to differentiate genuine milk samples from adulterated ones. Principal component analysis

(PCA) was applied to each spectral dataset and principal components (PCs) with eigenvalue ≥ 1 and that explained $\geq 1\%$ of the variation were considered meaningful and were used as predictors for the classification model. The response was either Adulterated (AD) or Not Adulterated (NA). Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) with or without covariance shrinkage were used to develop models for the first step differentiation. Entropy R^2 , sensitivity (S_n) or true positive rate (TPR), specificity (S_p), false positive rate (FPR), accuracy, error rate (ER) and receiver operating characteristic (ROC) curve were used to compare the performance of the different combinations of spectral pre-treatments and classification algorithms. AD was considered the positive outcome while NA was the negative one. JMP Pro 13.2.1 was used to develop the classification models for the first step.

For the second step, soft independent modelling of class analogies (SIMCA) algorithm was used to determine the type of chemical adulterant that was present in the adulterated milk sample, if any. To implement this algorithm, a separate PCA model was created for milk samples that contained specific chemical adulterant after eliminating spectral outliers that were detected by Hotelling's T^2 distribution at 5% significance level. The Unscrambler X (Camo Software, Oslo, Norway) was used to develop the classification model for the second step. All models for both steps were evaluated by the validation dataset.

5.2.4.5 Quantification models for extraneous water and added chemical adulterants

TQ Analyst Professional Edition 7.2.0.161 (Thermo Electron Corporation, Waltham, Massachusetts, USA) was used to build PLS calibration models for extraneous water and added chemical adulterants using FTIR spectra of adulterated milk samples. The reference values were calculated based on the weights of the prepared samples. The FTIR spectra were either kept raw, without applying any mathematical pre-treatment, or they were subjected to the Savitzky–Golay first derivative (SG FD) algorithm prior to calibrating the model. The window size was 7 and the polynomial order was 3. After the raw spectra were loaded into the software, the Spectrum Outlier functionality in TQ Analyst was used to exclude all the spectra that were considered as spectral outliers. For quantification of added water, all spectral regions containing information related to milk composition were used, which included $3000\text{--}2800\text{ cm}^{-1}$, $1800\text{--}1700\text{ cm}^{-1}$ and $1600\text{--}950\text{ cm}^{-1}$. For quantification of chemical adulterants, each model went through several iterations of refinement. The first iteration was performed on the full FTIR spectrum. The loading spectra that

resulted from this iteration were examined and the spectral regions that showed high loadings were kept for the subsequent iteration. The spectral regions that will be included in the final model must be relevant to the chemical adulterant for which a calibration model is being developed. This process was repeated until a stable calibration model was obtained. For each iteration, cross-validation was performed in TQ Analyst using leave-one-out approach. Several figures of merit (FOMs) were used to compare the performance of the calibration models that were developed for milk components. These FOMs included: correlation coefficient (r) for calibration and cross-validation, root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP), root mean square error of cross validation (RMSECV), predicted residual sums of squares (PRESS) and number of factors used, bias (if available) and the spectral regions that were used for each model. In addition, the models were externally validated using the validation set's samples. The predictions of the analyte of interest were regressed against the reference values and the slope and the coefficient of determination (R^2) were reported.

5.3 Results and Discussion

5.3.1 Detection of extraneous water and added chemicals by ATR-FTIR

Table 5-2 summarizes the characteristic IR absorption bands for ammonium sulfate, sodium bicarbonate, sodium citrate, urea and milk whey protein in ATR-FTIR spectra of spiked milk samples. All these chemicals, except whey protein, show absorption bands in regions not related to major milk components (Figure 5-1), which means that ATR-FTIR should theoretically be able to detect milk samples that are adulterated by these chemicals. Milk samples spiked with whey protein show increased absorption intensity at $\sim 1548\text{ cm}^{-1}$, which is the Amide II band that is used for the quantification of milk protein by IR.

PCA and HCA were applied to ATR-FTIR spectra of raw and homogenized milk samples that were adulterated by the addition of milk whey, extraneous water and aqueous solutions of the above-mentioned chemicals. The adulterant levels in milk samples were 0.025%, 0.05%, 0.1%, 0.25%, 1% and 2%. The spectral region $2800\text{--}1800\text{ cm}^{-1}$ was excluded from the analysis because it does not contain information related to the chemical composition of milk and it contains information related to environmental interferences, such as CO_2 . For homogenized milk, the PCA scores plot and the HCA dendrogram show a complete separation between genuine and adulterated milk samples in general (Figure 5-2 and Figure 5-3). This observation confirms that ATR-FTIR can detect the differences in chemical composition resulting from adulteration in homogenized milk. However, the PCA scores plot and the HCA dendrogram of raw milk spectra did not reveal a similar clear separation between spectra of adulterated and genuine milk samples (Figure 5-4 and Figure 5-5). The reason behind the different clustering trends of raw and homogenized milk might be light scattering that is caused by large milk fat globules in raw milk. On the PCA scores plot, milk samples with 5% extraneous water created an intersection between the clusters of adulterated and genuine milk samples (Figure 5-4). On the HCA dendrogram, two replicates of milk sample that contained 5% extraneous water fell in the lower boundary of the arm of genuine raw milk, while the rest of the replicates of the 5% added water milk samples fell in the adulterated samples arm (Figure 5-5). This observation might suggest that 5% added water may represent a cut off limit for detection of added water in raw milk by ATR-FTIR. Such an addition will not be a tempting margin of profit for committing fraud and a fraudulent producer might add more than 10% of extraneous water for this practice to be economically lucrative. Having said that, ATR-

FTIR can be a viable option for differentiating raw milk samples with added water in a business context. On the other hand, milk samples adulterated with milk whey revealed a distinct cluster and a distinct arm from the remaining samples on the PCA scores plot and the HCA dendrogram, respectively, which can be explained by the increased intensity at the Amide II band (Figure 5-4 Figure 5-5).

Raw milk samples with added water and chemical adulterants showed different separation trends on the PCA scores plot and the HCA dendrogram depending on whether the adulterant is an endogenous component of milk or not. Raw milk samples adulterated with ammonium sulfate, sodium bicarbonate and sodium citrate solutions revealed a clear cluster on the PCA scores plot (Figure 5-6). On the HCA dendrogram, two replicates of raw milk samples adulterated with 0.025% sodium citrate fell in the arm of genuine raw milk (Figure 5-7). These two replicates might be considered as outliers or 0.025% might be considered as a cut off limit for differentiating milk samples adulterated with added water and sodium citrate. On the other hand, milk samples adulterated with added water and urea, which is a minor milk component, were more challenging to differentiate. Raw milk samples that contained less than 0.25% added urea were intercepting with genuine milk samples on the PCA scores plot (Figure 5-6). On the HCA dendrogram, raw milk samples that contained less than 0.1% added urea fell in the genuine raw milk arm. This observation suggests that ATR-FTIR might be capable of detecting raw milk samples with added water and >0.1% added urea or 100 mg/dL added urea. In a business-oriented context, a fraudulent user would need to add at least 10% water to achieve lucrative profit margin. The addition of 10% added water to one liter of milk will increase its freezing point by 0.054 °C [79]. Since the addition of 20 mg/L urea will decrease the freezing point of milk by 0.0003 – 0.0004 °C [81], 2.7-3.6 g of added urea are needed to compensate for the addition of 10% water to one litre of milk. This amount represents 0.27-0.36% or 270-360 mg/dL of added urea and it is already above the cut off limit for PCA which was 0.25%. Having said that, ATR-FTIR spectroscopy can be a viable option for differentiating chemically adulterated samples in a business-oriented context. It must be mentioned that these results were obtained after applying the forward search variable selection algorithm to the spectral data, which restricted the spectral regions that were used in PCA and HCA to 1022-1061, 1065-1084, 1119-1138, 2801-2839, 3657-3676, 3707-3726 and 3977-3996 cm^{-1} . These spectral regions are related more to the major milk components than the characteristic IR bands of the chemical adulterants. This observation suggests that the dilution of major milk

components by added water might be the main factor for the previously mentioned clustering trends.

To summarize, ATR-FTIR spectroscopy can be used to detect added water and chemicals to raw milk. The minimum detection limit of added water is 5%. Raw milk that contains chemical adulterants, which are not chemical components of milk, can be differentiated at levels as low as 0.025% or 25 mg/dL. For urea, which is a minor milk component, the cut off limit for differentiating milk samples with added urea is 0.1% or 100 mg/dL. However, taking into consideration the cost of acquiring an ATR-FTIR spectrometer and that it contains moving parts, it might not be an appropriate candidate for on-site milk adulteration detection instrument. For this reason, a cheaper alternative will be investigated for on-site detection of milk adulteration and transmission-based FTIR spectroscopy will be investigated for in-lab detection of milk adulteration.

Table 5-2 Detected absorption bands in cm^{-1} for different chemical adulterants in milk and water solutions

Adulterants	Water subtraction	In water	In milk
Ammonium Sulfate	1093, 1450	1093, 1450	1090, 1450
Sodium Bicarbonate	1359	1359	1359
Sodium Citrate	1279, 1389, 1569	1279, 1389, 1569	1279, 1389, 1569
Urea	1157, 1464	1158, 1465	1156, 1463
Whey	1548	1548	1548

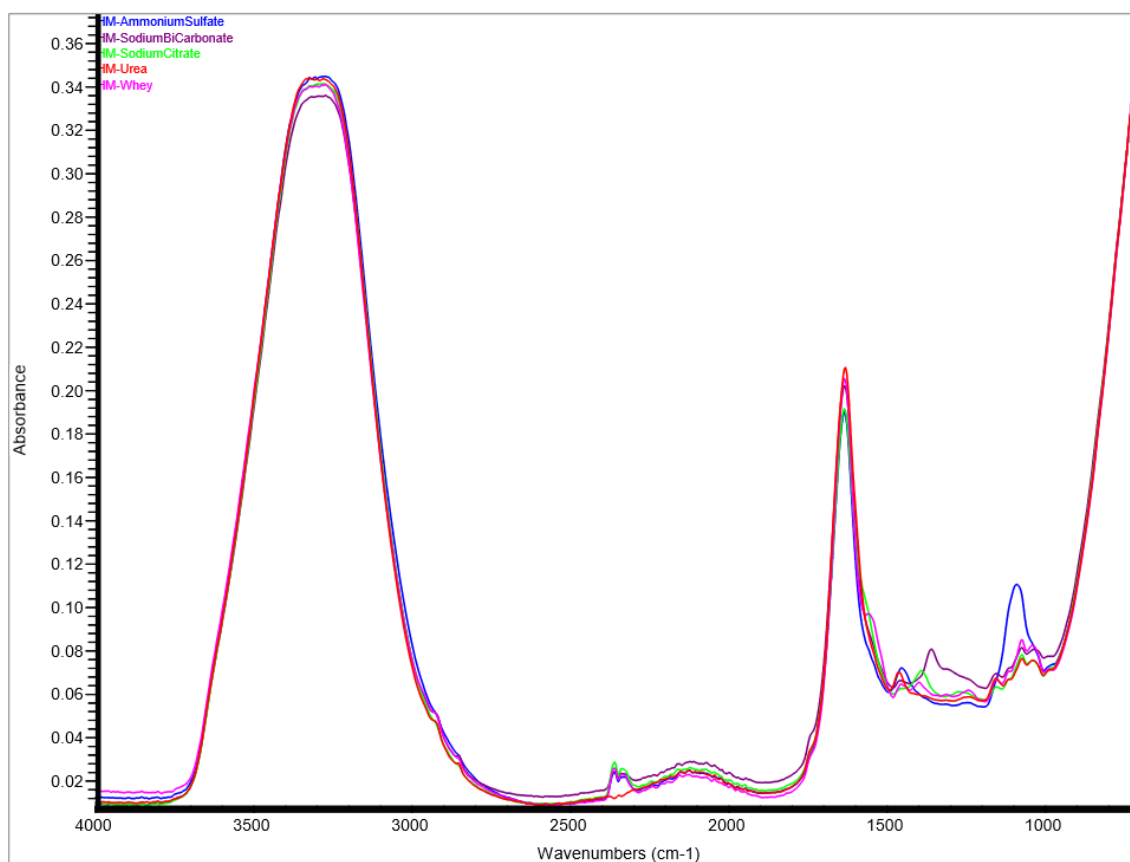


Figure 5-1 ATR-FTIR spectra of homogenized milk spiked with chemical adulterants. Blue ammonium sulfate, purple sodium bicarbonate, green sodium citrate, red urea, pink whey

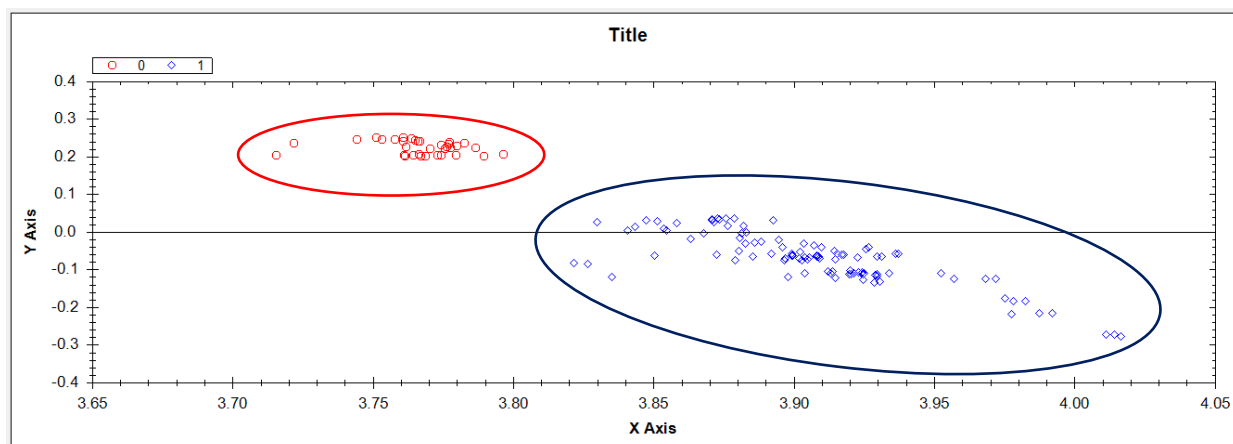


Figure 5-2 Principal component scores showing the separation of genuine (red) and adulterated (blue) homogenized milk samples

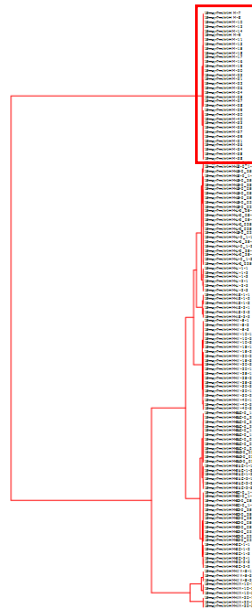


Figure 5-3 HCA dendrogram showing a separate arm for genuine homogenized milk samples

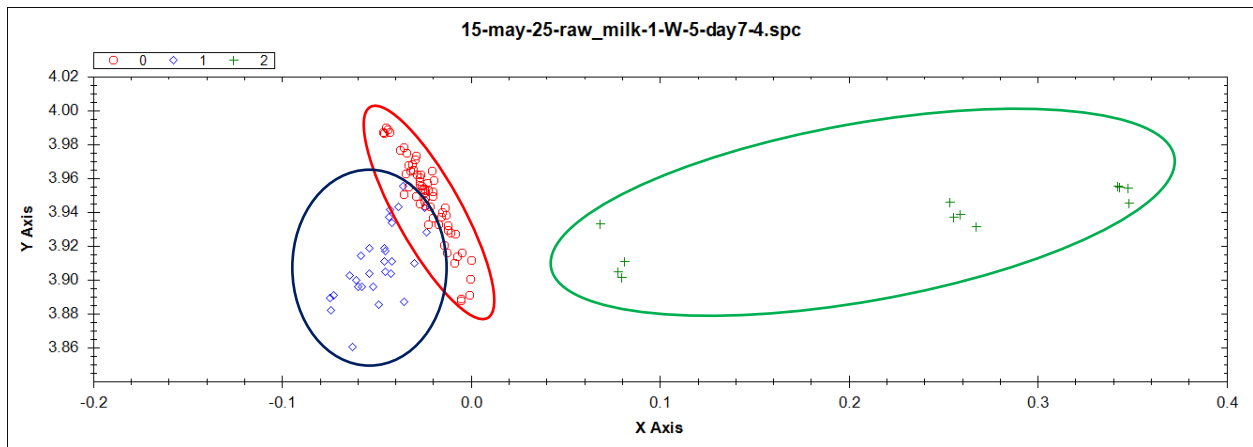


Figure 5-4 Principle component scores showing a separate cluster for whey adulterated raw milk samples (green) and intercepting genuine raw milk cluster (red) with water adulterated raw milk samples (blue) at the 5% added water level

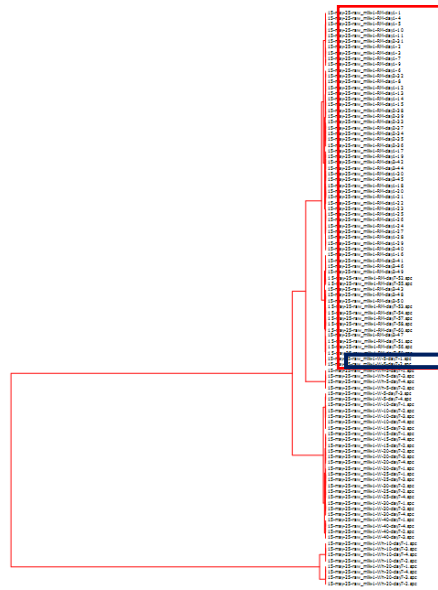


Figure 5-5 HCA dendrogram showing a separate arm for genuine raw milk samples (red-marked) which also contains two replicates of the 5% added water sample (blue-marked)

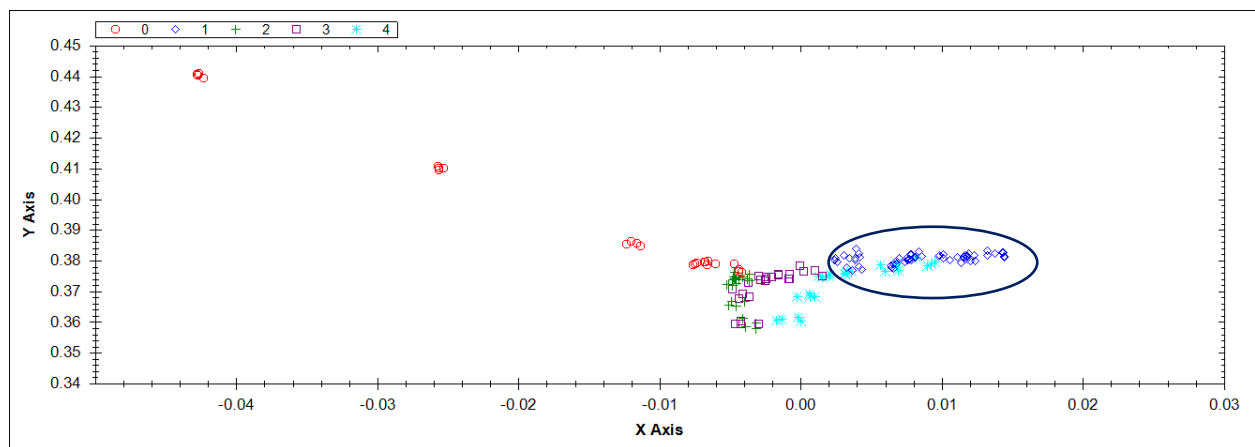


Figure 5-6 Principle component scores showing a cluster with raw milk samples and those that have less than 0.25% added urea

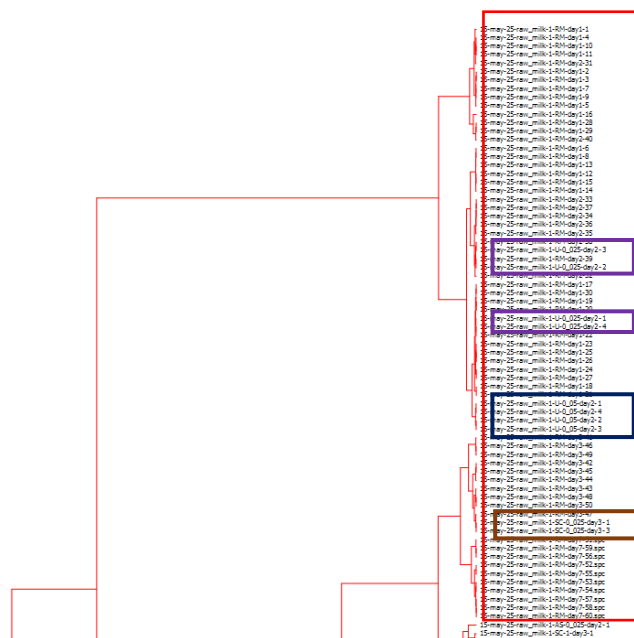


Figure 5-7 HCA dendrogram showing the arm of raw milk samples (red-marked) with chemically adulterated ones. 0.025% added urea (purple-marked), 0.05% added urea (blue-marked) and 0.025% added sodium citrate (brown-marked)

5.3.2 Effect of extraneous water and chemical adulterants on milk freezing point depression

According to the director of PZL Tecnologia, the chemicals that are frequently used to tamper with their cryoscope readings of milk freezing point are urea, sodium citrate and ammonium sulfate. These chemicals are added to compensate for the increase in milk freezing point as a result of adding extraneous water. They also suspected that sodium carbonate may have been used for the same purpose in addition to neutralizing milk acidity. For this reason, this study will be focusing on detecting raw milk adulterated with added water and these chemicals. It must be noted that milk adulteration is done in a way to avoid detection; hence, the exact practice will not be known. We had to assume probable scenarios that fraudulent milk producers might be using to adulterate milk. The premise of these scenarios is that a fraudulent producer would prepare an adulteration solution whose freezing point is greater than -0.500°C using one of the above-mentioned chemicals with a concentration that will keep the freezing point of raw milk within the legally accepted levels regardless of the amount of added solution to milk. It might be difficult to control the final freezing point of raw milk by adding water and the chemical adulterant separately.

The theoretical freezing point depression values were calculated for the proposed adulteration solutions, which are aqueous solutions of urea, sodium citrate, ammonium sulfate and sodium carbonate with concentrations ranging from 1% to 3%. The following equations were used to calculate these values

$$a) \Delta T = -\frac{X}{M}K, b) \Delta T = -\frac{X}{M}iK$$

Where ΔT is the freezing point depression value of the aqueous solution, X is the concentration of ionic or non-ionic substance g/Kg, M molecular weight of the substance, K is the cryoscopic constant for water, which equals to 1.853 and i is the number of ions produced when the substance is dissolved [90]. For example, the theoretical freezing point depression of 1.75% urea solution is -0.539 °C, which is close to the experimentally measured value of -0.534 °C (Table 5-3).

Table 5-3 Calculated freezing point depression values for aqueous urea solutions with different concentrations

Urea Solution Concentration %	Theoretical Freezing Point Depression
1.5	-0.463
1.75	-0.539
2	-0.617
2.25	-0.694
2.5	-0.771
3	-0.926

The initial screening of adulteration solutions with different concentrations revealed that solutions of urea, sodium citrate, ammonium sulfate and sodium carbonate with concentrations of 1.75%, 2.75%, 1.5% and 1.25%, respectively, had freezing points between -0.500 °C to -0.540 °C and that mixing of raw milk with 10-30% of these solutions maintained the freezing point of raw milk within legal limits. In fact, Table 5-4 shows that these solutions can be added to raw milk with a percentage up to 40% without noticing any visible changes on milk while maintaining the freezing point within legal limits. However, high concentrations of sodium carbonate caused changes in milk consistency, which undermined the claim that sodium carbonate is used to compensate the change in freezing point of milk resulting from extraneous water. In this case, neutralizing milk acidity would be the plausible reason for the presence of sodium carbonate in adulterated milk samples that are described by PZL director.

Table 5-4 Freezing point of raw milk adulterated with different adulteration solutions. Freezing point of genuine milk is between -0.520 °C and -0.550 °C (depending on the jurisdiction). In Brazil, added water is tolerated up to 2% = F.P. -0.510 °C
Concentration of adulteration solutions: a 1.75 g urea/dL; b 2.75 g sodium citrate/dL; c 1.5 g ammonium sulfate/dL; d 1.25 g sodium bicarbonate/dL

Percentage of adulteration solution added to raw milk	Freezing points (F.P.) and adulterant concentrations (mg/dL milk) of raw milk adulterated with								
	Water	Urea solution ^a		Citrate solution ^b		Sulfate solution ^c		Carbonate solution ^d	
	F.P. (°C)	F.P. (°C)	mg/dL	F.P. (°C)	mg/dL	F.P. (°C)	mg/dL	F.P. (°C)	mg/dL
0	-0.518	-0.553	-	-0.599	-	-0.519	-	-0.651	-
5	-0.485	-0.545	87.5	-0.590	137.5	-0.520	75	-0.626	62.5
10	-0.437	-0.545	175	-0.588	275	-0.520	150	-0.606	125
15	-0.431	-0.545	262.5	-0.585	412.5	-0.519	225	-0.594	187.5
20	-0.405	-0.543	350	-0.584	550	-0.519	300	-0.589	250
25	-0.379	-0.540	437.5	-0.579	687.5	-0.516	375	-0.587	312.5
30	-0.353	-0.539	525	-0.576	825	-0.517	450	-0.593	375
35	-0.325	-0.539	612.5	-0.570	962.5	-0.515	525	-0.590	437.5
40	-0.312	-0.540	700	-0.561	1100	-0.512	600	-0.584	500
45	-0.277	-0.540	787.5	-0.556	1237.5	-0.511	675	-0.577	562.5
50	-0.247	-0.536	875	-0.546	1375	-0.510	750	-0.570	625
60	-0.199	-0.534	1050	-0.534	1650	-0.502	900	-0.557	750
70	-0.147	-0.536	1225	-0.526	1925	-0.501	1050	-0.550	875
80	-0.100	-0.536	1400	-0.515	2200	-0.502	1200	-0.540	1000
90	-0.054	-0.536	1575	-0.507	2475	-0.494	1350	-0.531	1125

5.3.3 Detection of extraneous water and added chemicals by ATR-IR

The discussion in section 5.3.1 concluded that ATR-FTIR spectrometers might not be appropriate candidate for the on-site detection of added water and chemical adulterants in milk by IR due to the presence of an interferometer that contains moving parts and due to its substantial cost. In addition, the previous chapter concluded that linear variable filter (LVF) IR spectrometer can predict water and soluble components in milk with acceptable accuracy and it would be an appropriate candidate for applications that require on-site analysis of milk, such as the current scenario in this study. This type of spectrometers provides enough data points in the mid-IR range $1500\text{-}950\text{ cm}^{-1}$, which is considered the fingerprint region in mid-IR. For these reasons, this part of the study will investigate the capabilities of the LVF spectrometer in the qualitative and quantitative detection of watered-down milk samples. In addition to the acronym “LVF”, ATR-IR will be used to refer to this type of spectrometer throughout this study.

5.3.3.1 Mid-IR bands of chemical adulterants

The subtraction spectra of spiked milk samples and the aqueous solutions of the chemical adulterants revealed their characteristic IR absorption bands (Figure 5-8). These bands were centered at ~ 1586 , 1474 and 1167 cm^{-1} for urea, at ~ 1567 , 1408 and 1289 cm^{-1} for sodium citrate, at ~ 1463 and 1095 cm^{-1} for ammonium sulfate and at $\sim 1377\text{ cm}^{-1}$ for sodium carbonate (Table 5-5). This observation suggests that the ATR-IR spectrometer can detect changes in milk composition resulting from the addition of these chemicals, which makes it a good candidate for an on-site milk adulteration detection instrument. It must be noted that the position of some IR absorption bands of the chemical adulterants are shifted when compared with their FTIR counterparts (Table 5-5). Nevertheless, we do not expect this issue to undermine the functionality of the ATR-IR spectrometer for this application since its spectra will not be used for research purposes.

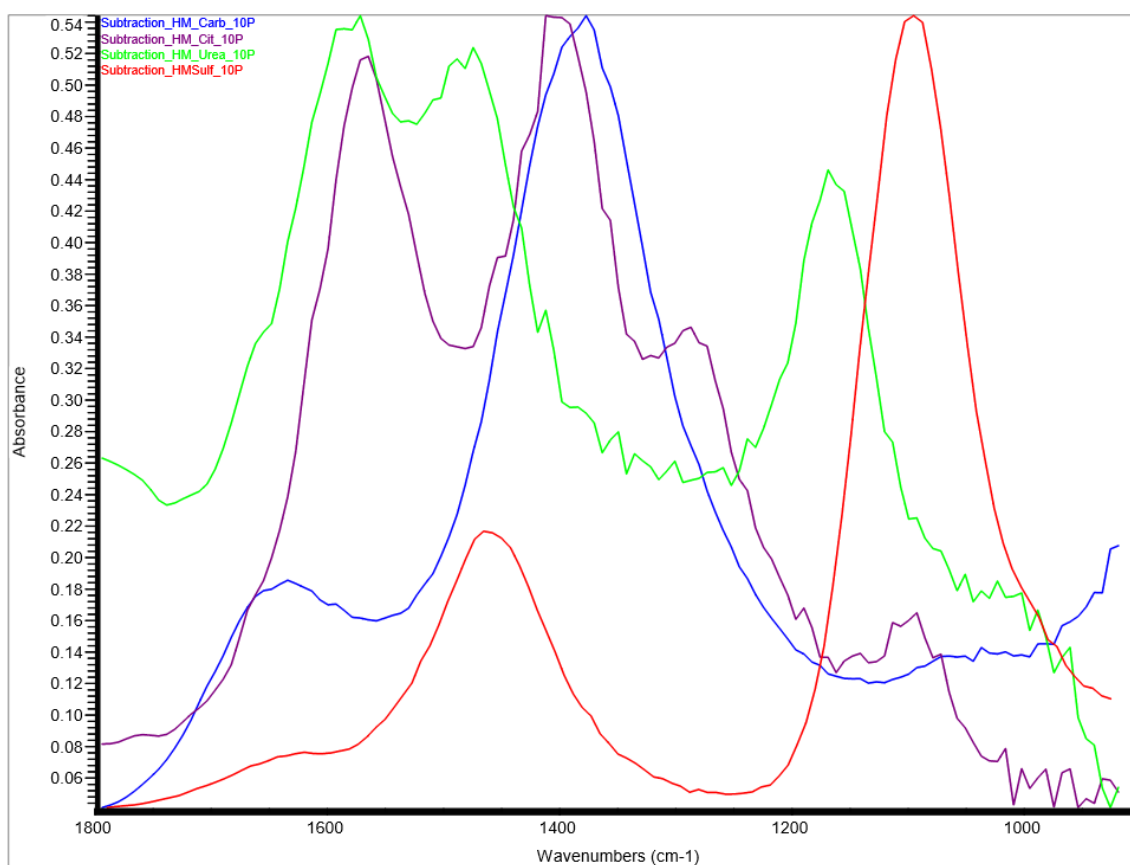


Figure 5-8 Subtraction spectra of homogenized milk samples spiked with sodium carbonate (blue), sodium citrate (purple), urea (green), ammonium sulfate (red). The spectrum of genuine milk was subtracted from the spectra of the spiked milk samples.

Table 5-5 Comparison of characteristic IR absorbance bands of chemical adulterants in milk and water observed in spectra collected on ATR-IR (LVF) and ATR-FTIR spectrometers. Spectra of genuine milk and pure water were subtracted from spectra of spike milk samples and aqueous solutions of these adulterants.

Adulterant	In Water - LVF	In Water - FTIR	In Milk - LVF	In Milk - FTIR
Urea	1166, 1477, 1581	1157, 1464	1167, 1474, 1586	1156, 1463
Sodium citrate	1288, 1398, 1563	1279, 1389	1289, 1408, 1567	1278, 1390, 1576
Sodium carbonate	1388	1359	1377	1359
Ammonium sulfate	1098, 1457	1093, 1450	1095, 1463	1090, 1450

5.3.3.2 Spectral analysis and added water quantification model

Principal component score plots show two clear clusters, one for genuine milk samples and another for the adulterated ones when PCA was applied to the raw full spectral range of the collected spectra. However, a minor intersection is observed between genuine milk samples and those that contained 5% adulteration solution. This intersecting region was more evident with raw milk samples. No specific trend was observed in the intersection region that is related to the identity of the chemical adulterant. The separation between the two clusters was enhanced when PCA was restricted to the spectral region $1100\text{-}1000\text{ cm}^{-1}$ (Figure 5-9 and Figure 5-10), which can be explained by the dilution effect that added water had on milk lactose. For homogenized milk, the separation between the two clusters was complete; on the other hand, the two clusters were adjacent to each other for raw milk. It must be noted that the number of spectral mathematical treatments was kept at a minimum to avoid creating mathematical artifacts, which will overestimate the capabilities of the LVF spectrometer. In addition, the freezing point values of some authentic milk samples were already low. Comparing the Brazilian legislation [91] with the standard procedure for determining added water by measuring milk freezing point [79], we notice that the accepted Brazilian freezing point for milk is below its counterpart in the standard procedure by $0.02\text{ }^{\circ}\text{C}$. This difference indicates a 4-5% added water in milk. Some raw milk samples had a freezing point below $-0.512\text{ }^{\circ}\text{C}$, which indicates 2% added water according to Brazilian legislations.

The spectra of the adulterated milk samples that were collected on the ATR-IR spectrometer were used to develop PLS models to predict the percentage of added water. In addition to added water, these samples contained one, two or three chemical adulterants to compensate for the change in the freezing point of milk. The entire spectral region was used since the addition of water will dilute all milk components, which is a source of variation that can be captured by the PLS algorithm. All models gave excellent correlation coefficients (i.e., $r = 0.99$), which indicates that the presence of the chemical adulterants did not significantly interfere with the modeling capability of the PLS algorithm for added water (Table 5-6 and Table 5-7). SG smoothing and derivatization did not improve the predictive capability of the PLS model. In general, homogenized milk gave lower values for the measurements of error; nevertheless, models developed for raw milk samples revealed excellent predictive performance. The model that gave the most consistent FOMs was the one that used raw spectra. The values of RMSEC, RMSEP and RMSECV were 1.41%, 1.85% and

1.85%, respectively. All these values of the measurements of error are less than 2%, which is the tolerated level of added water in milk according to the Brazilian legislation. It must be noted that cryoscopy can detect a minimum of 3% added water [78]. In addition, these models were developed using the percentage of added water as the reference values because the actual water content in these samples was not known. The accuracy of the added water model can be enhanced by accurately determining the water content in the calibration sample set. In light of these promising results, the German manufacturer of the LVF ATR-IR spectrometer produced a prototype of a lab-in-box milk adulteration detector (Figure 5-11). The spectrometer was controlled by a tablet and it was connected to a portable printer that produced printouts of the analysis results.

To summarize, the LVF spectrometer has the potential to capture chemical information related to the authenticity of milk samples; hence, the LVF spectrometer can be used for on-site detection of extraneous water and chemical adulterants in milk using a proper classification algorithm. In addition, its spectra can be used to develop PLS model to predict the percentage of extraneous water in milk regardless of the identity of the chemical adulterant that might have been added to milk to compensate for the change in the freezing point of milk.

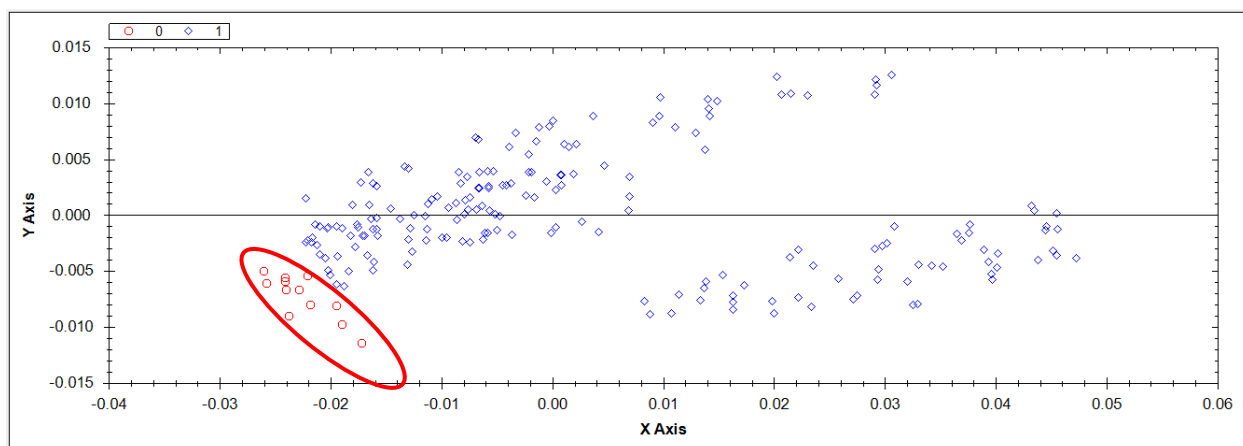


Figure 5-9 Principal component scores of homogenized genuine milk samples and adulterated milk samples with solutions of urea, sodium citrate, sodium carbonate, ammonium sulfate and different combinations of these solutions – genuine milk is in red

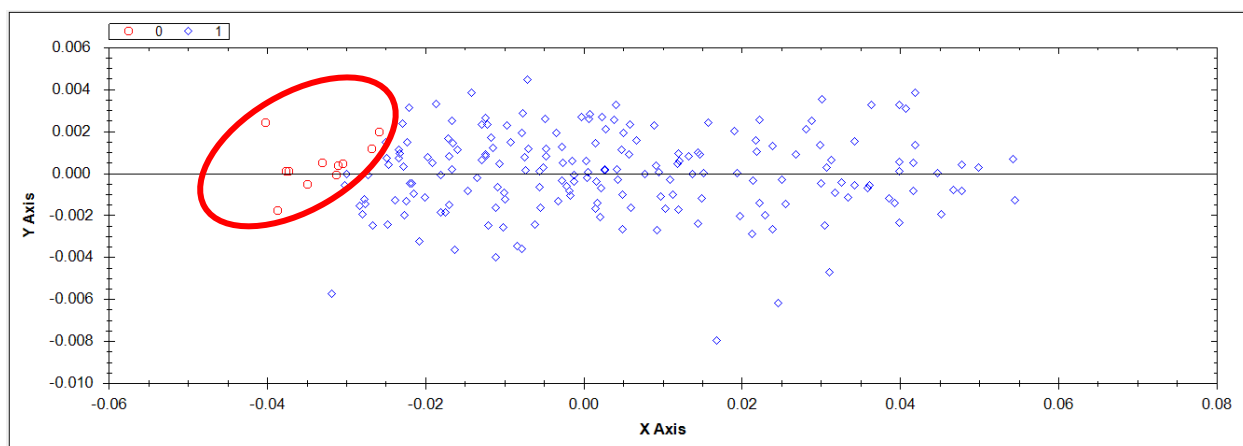


Figure 5-10 Principal component scores of raw genuine milk samples and adulterated milk samples with solutions of urea, sodium citrate, sodium carbonate, ammonium sulfate and different combinations of these solutions – genuine milk is in red

Table 5-6 Calibration FOMs of the added water PLS prediction models. In addition to added water, milk samples contained one, two or three chemical adulterants to compensate for the change in freezing point.

Milk	Spectrum	Region	Corr. Coeff.	RMSEC %	RMSEP %	Factors
Homogenized	Raw	1770-980	0.99	1.18	1.52	3
Raw	Raw	1770-980	0.99	1.41	1.85	7
Homogenized	Raw with SG smoothing	1770-980	0.99	1.18	1.52	3
Raw	Raw with SG smoothing	1770-980	0.99	1.56	1.93	7
Homogenized	SG FD	1770-980	0.99	1.04	1.91	4
Raw	SG FD	1770-980	0.99	1.84	2.02	3

Table 5-7 Cross-validation FOMs of the added water PLS prediction models. In addition to added water, milk samples contained one, two or three chemical adulterants to compensate for the change in freezing point.

Milk	Spectrum	Corr. Coeff.	RMSECV %	Factors
Homogenized	Raw	0.99	1.23	3
Raw	Raw	0.99	1.85	7
Homogenized	Raw with SG smoothing	0.99	1.21	3
Raw	Raw with SG smoothing	0.99	2.05	7
Homogenized	SG FD	0.99	1.54	4
Raw	SG FD	0.99	2.32	3



Figure 5-11 Lab-in-Box milk adulteration detector built using LVF ATR-IR spectrometer

5.3.4 Detection of extraneous water and added chemicals by transmission based FTIR

5.3.4.1 Mid-IR bands of chemical adulterants

Table 5-8 summarizes the centers of the characteristic IR absorption bands that are observed in the aqueous solutions and milk samples spiked with urea, trisodium citrate, ammonium sulfate and sodium carbonate. These bands are not overlapping with absorption bands of major milk components except for 1557 cm^{-1} and 1092 cm^{-1} that are observed in milk samples spiked with citrate and sulfate and that overlaps with the Amide II band of proteins and lactose absorption band, respectively. This observation suggests that FTIR spectra of milk samples adulterated with these chemicals should contain additional features in comparison to genuine milk samples that will facilitate differentiation between genuine and adulterated milk samples. In addition, these absorption bands will be used to verify the performance of qualitative and quantitative models that will be used to detect added water and chemical adulterants in milk. For example, if the loading spectra of a PCA model show increased loadings in spectral regions that contain these centers then the model is capturing the correct chemical information and the model is not overfitting the data.

Table 5-8 Centers of characteristic IR absorption bands of four chemicals in aqueous solutions and in milk that are used to mask the addition of water or to neutralize milk acidity.

Adulterant	Aqueous solution cm^{-1}	Milk cm^{-1}
Urea	1466, 1158	1465, 1158
Citrate	1554, 1391, 1280	1557, 1392, 1279
Sulfate	1455, 1093	1455, 1092
Carbonate	1394	1377
Mix of 4 adulterants	-	1462, 1394, 1279, 1097

5.3.4.2 Classification models for differentiation of adulterated milk samples

In this part of the study, we used transmission-based sample introduction methods because the official FTIR milk analysis method relies on transmission cell. However, transmission cell is only compatible with homogenized milk. For on-site detection of milk adulteration, homogenization might not be an available option. For this reason, we tested another transmission accessory that is offered by Agilent, which is the DialPath (Agilent Technologies, Santa Clara, California, USA).

It consists of two parts, the lower one is a fixed flat surface that contains a ZnSe window. The upper one is a rotating part that contains 3 ZnSe windows. After placing a droplet of milk on the window of the flat surface, the upper part is rotated and one of its windows faces the window of the lower part at a fixed distance that constitutes the optical path length. The IR beam will pass through the sample from the upper window to the lower one and then hits the detector in the spectrometer. The upper part can be rotated manually to create a 30, 50 or 100 μm path length. In addition of being a transmission-based accessory, the advantages of this sample introduction method are similar to ATR, mainly, it does not require pumping accessory, raw milk does not cause any clogging issues and it is easy to clean. These features make this accessory a good candidate for on-site raw milk analysis. In this study, the transmission cell was used for the homogenized milk and the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) was used for raw milk.

In addition, a two-tier approach was used to differentiate adulterated milk samples from genuine ones. In the first step, a classification model was developed to determine whether a milk sample was genuine or adulterated. In the second step, another classification model was developed to determine the type of chemical adulterant that was added to the sample, if any. Unlike what has been reported in the literature [86-88], we kept the crucial step of determining the authenticity of the milk sample within the scope of a binary classification model. Increasing the number of classes or groups in one classification model usually reduces its stability. The spectral range that was used to develop the classification models was limited to 1600-950 cm^{-1} because this region contains the characteristic IR absorption bands of milk proteins, lactose and the chemical adulterants. Spectral regions associated with milk fat were excluded due to milk fat variability, which might be an unnecessary confounding factor in this application.

For the first step, a binary classification algorithm that gives an answer to a closed question was needed. In this case, the algorithm must classify the milk sample as either Adulterated (AD) or Not Adulterated (NA). There are several algorithms that have been reported in the literature for developing multivariate classification models, such as discriminant analysis (DA), partial least squares discriminant analysis (PLS-DA), support vector machines (SVM) and artificial neural networks (ANN) [92]. SVM and ANN are applicable to situations where the relationship between class membership and the predictors (i.e., the spectra) is not linear. In the case of milk adulteration,

class membership will be the result of changes in concentrations of milk components that result from the dilution effect of added water or increased concentration of added chemical adulterants. According to Beer's law, the relationship between concentration and absorbance is a linear one, which eliminates the need for SVM and ANN for this application. PLS-DA is a powerful multivariate classification algorithm that maximizes the separation of observations according to their class membership to achieve maximum class separation. As a result, any sources of variation in milk composition that are not related to milk authenticity might contribute to class separation, which makes this algorithm prone to overfitting. DA represents a more realistic option in terms of modeling power; however, DA requires orthogonal predictors whose number is less than the number of observations used to create the model, which is not necessarily the case with spectral data. To overcome these issues, PCA is applied to the spectral data and PC scores can be used as predictors for DA. PCA decomposes the variance in the dataset into a number of new variables that is equal to the number of observed variables (i.e., wavenumbers) called principle components (PCs) [93]. These PCs are orthogonal to each other, which means that the correlation is 0 between any two PCs, and they describe the same variance structure as the original variables in the dataset (i.e., the spectra). Each PC is a linear combination of optimally weighted observed variables (i.e., wavenumbers). However, only the few first PCs account for a significant portion of the variability in the original dataset. In this case, PCs with eigenvalue ≥ 1 and that explained 1% of the variation or more were included in the model; hence, PCA reduced the number of predictors and eliminated collinearity among them.

DA employs different discrimination methods. In this study, LDA and QDA were considered for developing classification models for the first step. LDA uses the same covariance matrix for all classes, while QDA uses a different covariance matrix for each class [94]. In addition, DA can employ covariance shrinkage that improves the stability of underrepresented classes and reduces their prediction variance. This option is recommended when some classes have a small number of observations, such as the NA class in comparison to the AD class [95].

Table 5-11 and table 5-12 summarize the performance FOMs of milk authenticity classification models that were developed using different combinations of sample introduction methods, milk type (i.e., raw and homogenized), spectral pre-treatments and discrimination methods of DA. Table 5-11 and Table 5-12 summarizes FOMs of classification models developed using FTIR spectra

collected by a transmission cell and the DialPath accessory (Agilent Technologies, Santa Clara, California, USA), respectively. Entropy R^2 is a measure of fit that takes values between 0 and 1 where larger values indicate better fit and 1 indicates perfect prediction power. These values can be negative. Table 5-11 and table 5-12 show that PCA-QDA classification models with no covariance shrinkage that were developed using VN raw spectral data of homogenized and raw milk and FD spectral data of raw milk yielded the best fit. Entropy R^2 values were 0.99 and 0.77 for homogenized and raw milk, respectively, for the training set. Accuracy is the simplest measure of the quality of a classification model and it represents the percentage of correctly assigned observations. Error rate (ER) is the complementary index for accuracy and it is the percentage of wrongly assigned observations [92]. For homogenized milk, the PCA-QDA model with no covariance shrinkage developed using VN raw spectral data gave the best accuracy and error rate, which were 100% and 0%, respectively, for the training and the validation sets. For raw milk, the PCA-QDA model with no covariance shrinkage developed using FD spectral data gave the best accuracy and error rate. The accuracy values were 98.84% and 100% and the error rate values were 1.16% and 0% for the training and validation sets, respectively.

In addition, there are indicators that are related to the classification quality of a single class and that can be calculated from the confusion matrix that is produced by the DA algorithm (Table 5-9). In this scenario, the adulterated samples (AD) and genuine milk samples (NA) were considered the positive and the negative outcomes, respectively. Sensitivity (S_n) or true positive rate (TPR) describes the model ability to correctly recognize objects belonging to a specific class (Equation 5-1), which is AD in this case. If all observations are correctly assigned to a specific class, then $S_n=1$ [92]. On the other hand, specificity (S_p) characterizes the ability of a specific class to reject observations belonging to all other classes (Equation 5-2). If observations not belonging to a specific class are never assigned to that class then $S_p=1$ [92]. False positive rate (FPR) is a complementary indicator for specificity (Equation 5-3). These indicators can be presented in the receiver operating characteristic (i.e., ROC) curve, which plots TPR against FPR on y and x axes, respectively, for a specific class in a classification model (Figure 5-14). The calculated area under the curve (AUC) summarizes the predictive capability of the model for a specific class [92]. If the ROC curve is higher than the diagonal line, then it will perform better than a random classifier and the best performance is obtained when the upper left corner is reached by the ROC curve.

Table 5-9 Confusion matrix of PCA-QDA classification model with no covariance shrinkage developed using VN raw spectral data of homogenized milk samples. AD and NA are the positive and negative outcomes, respectively.

Actual	Predicted	
Milk Type	AD (+)	NA (-)
AD (+)	230 (TP)	0 (FN)
NA (-)	0 (FP)	12 (TN)

Equation 5-1

$$Sn = \frac{TP}{TP + FN}$$

Equation 5-2

$$Sp = \frac{TN}{FP + TN}$$

Equation 5-3

$$FPR = \frac{FP}{FP + TN} = 1 - Sp$$

So far, three models have yielded good fit and high accuracy. These models are: PCA-QDA with no covariance shrinkage developed using VN raw spectral data of homogenized milk (M1), PCA-QDA with no covariance shrinkage developed using VN raw spectral data of raw milk (M2) and PCA-QDA with no covariance shrinkage developed using FD spectral data of raw milk (M3). Among these, M1 and M3 yielded perfect specificity for differentiation of adulterated milk samples for the training and the validation set. For M2, specificity values of the AD class were 1 and 0.80 for the training and the validation sets, respectively. On the other hand, sensitivity values of the AD class were 1, 0.98 and 0.99 for M1, M2 and M3, respectively, for the training sets and the AUC values for the AD class were 1, 0.9971 and 1 for M1, M2 and M3, respectively. We can conclude from these observations that PCA-QDA with no covariance shrinkage is an effective algorithm to differentiate adulterated milk samples from genuine ones with high sensitivity and specificity. First derivative can be applied to the spectral data of raw milk to improve the prediction power of this model. Running forward stepwise feature selection in DA revealed that PC3 ($P < 0.0001$), PC1 ($P < 0.0001$) and PC2 ($P < 0.0001$) were the most significant predictors for M1, M2 and M3, respectively. Inspection of the loading spectra of these principal components revealed the

influential spectral features that are mainly responsible for the discrimination power of these models. For example, PC3 loading spectrum revealed increased loadings in the regions 1527-1246 cm^{-1} and 1155-1080 cm^{-1} that include the characteristic IR absorption bands of the chemical adulterants and decreased loadings in the regions 1600-1527 cm^{-1} and 1080-1000 cm^{-1} , which include the Amide II and lactose absorption regions, respectively (Figure 5-12). This loading spectrum reflects the fact that chemical adulterants' concentrations are increasing, while milk protein and lactose concentrations are decreasing as a result of water addition. We can conclude that PCA was successful in decomposing the variation in the spectral dataset and isolating the spectral fingerprint that reflects the changes in milk chemical composition resulting from the adulteration process. This successful decomposition helped the DA algorithm capture the correct chemical information related to the status of the milk sample; hence, avoid overfitting.

The second step of this classification process will be applied to adulterated samples, in which the sample will be assigned to a group that corresponds to the chemical adulterant that might be present in it. For this purpose, soft independent modelling of class analogies (SIMCA) algorithm was considered the optimum choice because the predictions produced by this algorithm are not restricted to one class. SIMCA has already been reported in the literature as an effective algorithm to classify adulterated samples into groups according to the chemical adulterant [86, 87, 96]. In SIMCA, a PCA model is developed independently for each class, then an unknown sample is projected in each subspace and a distance from the class model is calculated. The sample assignment is determined by comparing its distances from the classes' PCA models. As a result, an unknown sample might be assigned to one class, multiple classes or none (Table 5-13). For example, spectrum 26 from the validation set was not assigned to any class; on the other hand, spectra 17, 18, 29, 30, 41, 42 and 43 were assigned to two classes, the correct chemical adulterant and the mix class. For homogenized milk, five classes were included in the SIMCA model, which are carbonate, citrate, sulfate, urea and mix. The accuracy for unique classification considered only samples that were assigned to one group that corresponded to the correct adulterant. The accuracy for correct classification considered samples that were correctly assigned to the chemical adulterant group regardless of whether the sample was also assigned to the mix group or not. Since the mix contained all chemical adulterants, this situation was considered as a correct classification. For raw milk samples, the mix group was eliminated; hence, only samples assigned to their corresponding chemical adulterant group were considered correctly classified. Table 5-10 reveals

that raw spectral data yielded good performance with homogenized milk, while first derivative enhanced the accuracy of SIMCA predictions for raw milk samples.

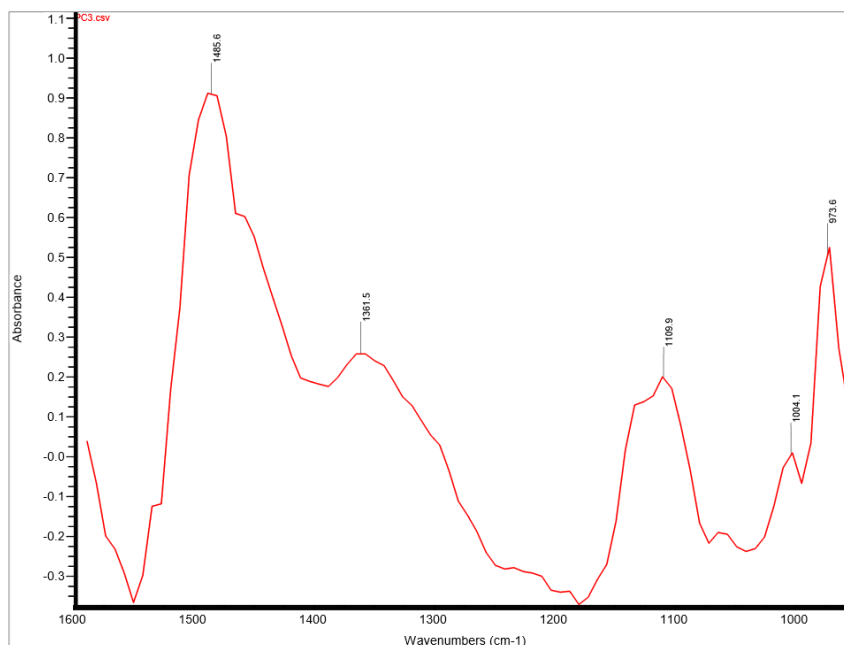


Figure 5-12 PC3 loading spectrum isolated from the VN raw spectral data of homogenized milk samples collected by a transmission cell

Table 5-10 Classification accuracy of SIMCA models developed using spectral data of raw and homogenized adulterated milk samples to classify adulterated samples according to the chemical adulterant that they contain. TC: transmission cell, DP: dial path, FD: first derivative, VN: vector normalized.

Accessory	Spectra	Accuracy % for unique classification	Accuracy % for correct classification
TC	Raw	81.40	97.67
	VN Raw	81.40	97.67
	FD	25.58	97.67
	VN FD	25.58	97.67
DP	Raw	-	81.63
	VN Raw	-	79.59
	FD	-	89.80
	VN FD	-	89.80

5.3.4.3 Quantification models for extraneous water and added chemical adulterants

PLS models that were developed to quantify chemical adulterants used spectral regions that showed high loadings in the PLS loading spectra during the first iteration of model development. These regions corresponded to the characteristic IR absorption bands of the analyte of interest (Table 5-8), which indicates that the models were capturing information related to the analyte of

interest and they were not overfitting the data. For added water prediction models, region 1,600-950 cm^{-1} was mostly used. In general, water and chemical adulterants prediction models relied on raw and FD spectral data, respectively. FD derivative helps in resolving overlapping bands and enhance small absorption bands on shoulders of bigger ones; hence, exposing more details in the spectral data [29].

All PLS models that are developed to predict added water in milk revealed excellent prediction capabilities. These models were developed using FTIR spectra of adulterated milk samples that contained added water and different added chemical adulterants, which means that the presence of the chemical adulterant did not undermine the predictive capabilities of the added water PLS models. The prediction error (i.e., RMSEP) values were 0.28%, 0.55%, 0.33% and 0.39 % for homogenized milk scanned with a transmission cell (Table 5-14 and Table 5-15), for homogenized milk scanned with the DialPaht set at 50 μm path length (Table 5-16 and Table 5-17), for homogenized milk scanned with the DialPaht set at 30 μm path length (Table 5-16 and Table 5-17) and for raw milk scanned with the DialPaht set at 30 μm path length (Table 5-18 and Table 5-19), respectively. The transmission cell yielded the lowest prediction error, which might be explained by the fact that the windows of the transmission cell are CaF_2 , while the transmission window of the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) are ZnSe , which has a refractive index higher than that of milk; hence, light refraction might be contributing to the prediction error. However, the difference between the prediction error of the models of the transmission cell and the DialPaht accessory becomes minimal when the path length of the DialPaht is set at 30 μm . This reduced path length yielded similar prediction errors for models developed using spectra of homogenized and raw adulterated milk samples.

For the added chemical adulterants, a similar trend has been observed for the prediction errors of the PLS models of these adulterants. Models developed with spectra of homogenized milk scanned with the transmission cell yielded the least prediction error. This error increased when the DialPath (Agilent Technologies, Santa Clara, California, USA) was set at 50 μm path length with homogenized milk and it was similar or close to that of the transmission cell when the path length was set at 30 μm with homogenized milk. A minor increase in the prediction error was observed when the DialPath (Agilent Technologies, Santa Clara, California, USA) was set at 30 μm with raw milk.

For urea, the prediction error values were 7.57 mg/dL, 8.62 mg/dL, 2.95 mg/dL and 6.73 mg/dL for homogenized milk scanned with a transmission cell (Table 5-14 and Table 5-15), for homogenized milk scanned with the DialPaht set at 50 μ m path length (Table 5-16 and Table 5-17), for homogenized milk scanned with the DialPaht set at 30 μ m path length (Table 5-16 and Table 5-17) and for raw milk scanned with the DialPaht set at 30 μ m path length (Table 5-18 and Table 5-19), respectively. All these prediction errors are lower than the prediction error reported by Santos *et al.* (2013), which was 0.232 g/L (i.e., 23.2 mg/dL) for urea in raw milk obtained with Cary 630 (Agilent Technologies, Santa Clara, California, USA) FTIR spectrometer equipped with DialPaht accessory set at 30 μ m path length.

For citrate, the prediction error values were 6.32 mg/dL, 9.63 mg/dL, 5.43 mg/dL and 10.3 mg/dL for homogenized milk scanned with a transmission cell (Table 5-14 and Table 5-15), for homogenized milk scanned with the DialPaht set at 50 μ m path length (Table 5-16 and Table 5-17), for homogenized milk scanned with the DialPaht set at 30 μ m path length (Table 5-16 and Table 5-17) and for raw milk scanned with the DialPaht set at 30 μ m path length (Table 5-18 and Table 5-19), respectively.

For sulfate, the prediction error values were 2.14 mg/dL, 2.79 mg/dL, 1.72 mg/dL and 4.50 mg/dL for homogenized milk scanned with a transmission cell (Table 5-14 and Table 5-15), for homogenized milk scanned with the DialPaht set at 50 μ m path length (Table 5-16 and Table 5-17), for homogenized milk scanned with the DialPaht set at 30 μ m path length (Table 5-16 and Table 5-17) and for raw milk scanned with the DialPaht set at 30 μ m path length (Table 5-18 and Table 5-19), respectively.

For carbonate, the prediction error values were 0.009%, 0.004%, 0.005% and 0.014% for homogenized milk scanned with a transmission cell (Table 5-14 and Table 5-15), for homogenized milk scanned with the DialPaht set at 50 μ m path length (Table 5-16 and Table 5-17), for homogenized milk scanned with the DialPaht set at 30 μ m path length (Table 5-16 and Table 5-17) and for raw milk scanned with the DialPaht set at 30 μ m path length (Table 5-18 and Table 5-19), respectively. It must be mentioned that the predictions of carbonate were not affected by the presence of added water or other chemical adulterants; on the other hand, the predictions of urea, citrate and sulfate were not accurate when more than one chemical adulterant was present in the milk sample.

To summarize, homogenizing milk and using a transmission cell as a sample introduction method gave the best performance for the classification and quantitative models employed to detect milk adulteration by the addition of water and chemical adulterants. Nevertheless, the use of the DialPath sample introduction accessory (Agilent Technologies, Santa Clara, California, USA) with raw milk gave satisfactory results especially when the path length was set at 30 μm . As a result, a decision-making workflow has been proposed to differentiate milk adulterated with added water and chemical adulterants from genuine one (Figure 5-13). After collecting the FTIR spectrum of an unknown sample, a DA classification model will determine the authenticity of the sample. If the sample is adulterated, then a SIMCA model will determine the chemical adulterant that is present in the sample, if any. If the chemical adulterant is carbonate, then most probably the sample does not contain added water and carbonate has been added to neutralize milk acidity, which means only carbonate will be quantified. If the milk sample contains urea, citrate or sulfate then added water and the chemical adulterant will be quantified. If multiple chemical adulterants are present, then only added water will be quantified.

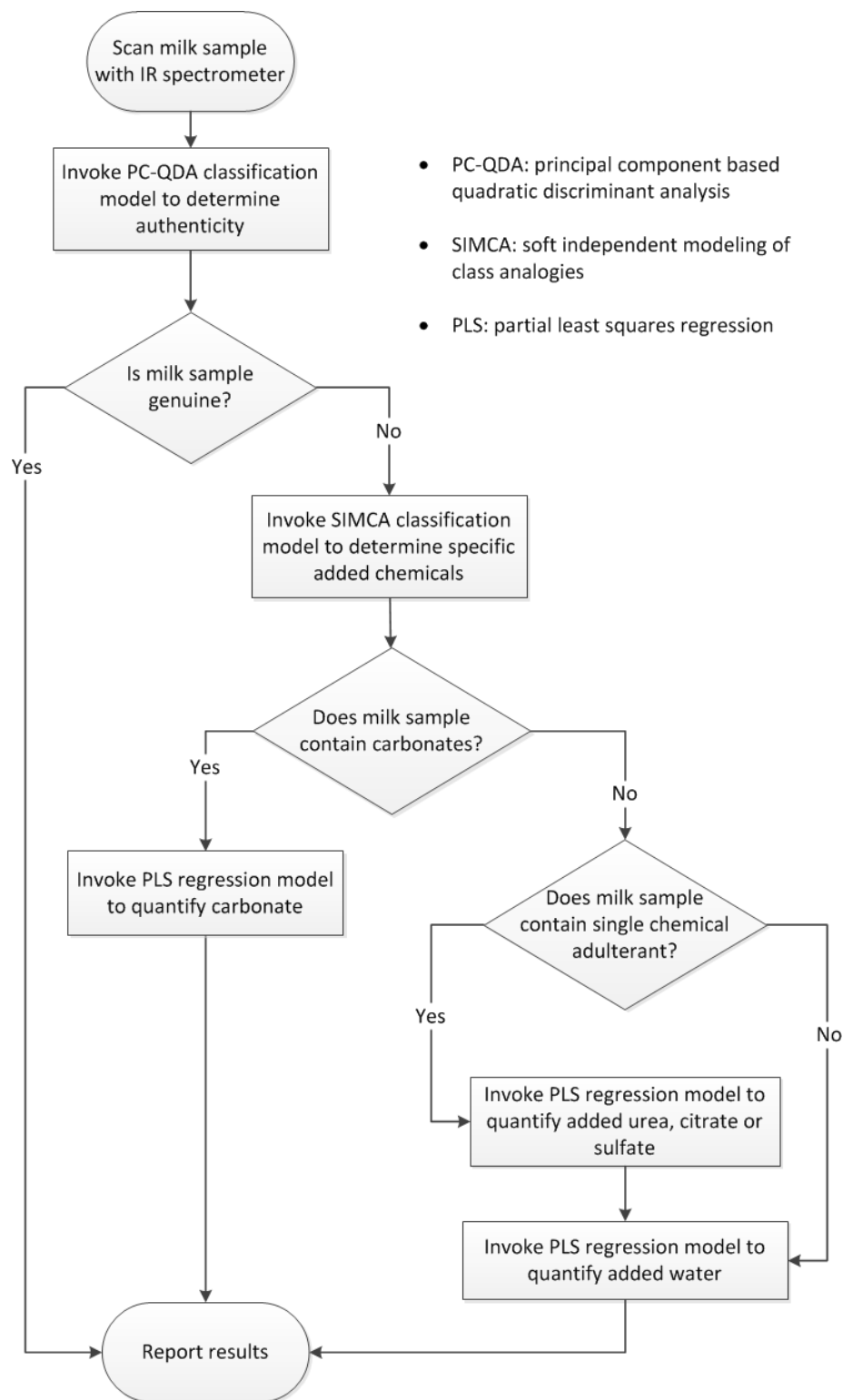


Figure 5-13 Decision-making workflow for authenticating milk sample freezing point depression by transmission based FTIR spectroscopy

5.4 Conclusion

In this study, multiple combinations of IR spectrometers, sample introduction methods and chemometric algorithms were investigated to detect the addition of water and chemical adulterants to milk. ATR-FTIR spectroscopy combined with PCA and HCA can be used to detect added water and chemicals to raw milk. The minimum detection limit of added water is 5%. Raw milk that contains chemical adulterants, such as ammonium sulfate, sodium bicarbonate and sodium citrate, can be differentiated at levels as low as 0.025% or 25 mg/dL. For urea, which is a minor milk component, the cut off limit for differentiating milk samples with added urea is 0.1% or 100 mg/dL.

A cheaper alternative was investigated for on-site detection of milk adulteration, which was LVF spectrometer equipped with ZnSe ATR crystal. This spectrometer has the potential to capture chemical information related to the authenticity of milk samples. PCA revealed that the cut off limit for detecting raw milk samples with adulteration solutions was 5%. In addition, its spectra can be used to develop PLS model to predict the percentage of extraneous water in milk regardless of the identity of the chemical adulterant that might have been added to milk to compensate for the change in the freezing point. Using the raw spectra of this spectrometer gave a model with the most consistent FOMs. The values of RMSEC, RMSEP and RMSECV were 1.41%, 1.85% and 1.85%, respectively.

Since the official FTIR milk analysis method is based on transmission, a portable FTIR spectrometer with a transmission cell and the DialPath (Agilent Technologies, Santa Clara, California, USA) accessory were evaluated for this application. Homogenizing milk and using a transmission cell as a sample introduction method gave the best performance for the classification and quantitative models employed to detect milk adulteration by the addition of water and chemical adulterants. Nevertheless, the use of the DialPath sample introduction accessory (Agilent Technologies, Santa Clara, California, USA) with raw milk gave satisfactory results especially when the path length was set at 30 μm . For raw milk, PCA-QDA classification algorithm with no covariance shrinkage developed using FD spectral data collected with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) gave the best accuracy and error rate. The accuracy values were 98.84% and 100% and the error rate values were 1.16% and 0% for the training and validation sets, respectively. This model yielded perfect specificity for differentiation

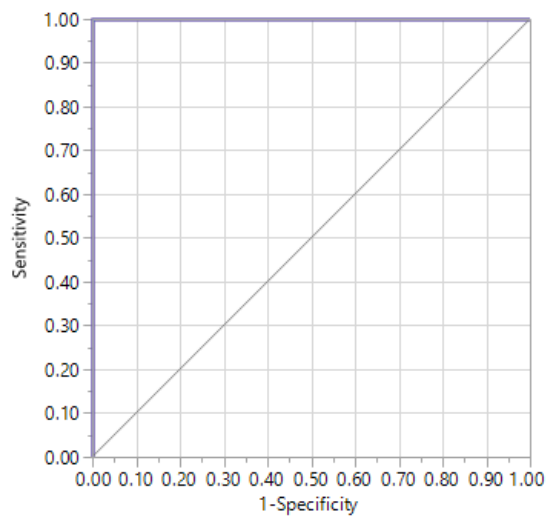
of adulterated milk samples for the training and the validation set. Sensitivity was 0.99 for the training set. PCA was successful in decomposing the variation in the spectral dataset and isolating the spectral fingerprint that reflected the changes in milk chemical composition resulting from the adulteration process. This successful decomposition helped the DA algorithm capture the correct chemical information related to the status of the milk sample; hence, avoiding overfitting. SIMCA algorithm successfully classified adulterated samples into groups according to the chemical adulterant present in these samples. The accuracy for the raw milk SIMCA classification model was 89.90% when FD was used. Quantitatively, PLS prediction errors values were 0.39 %, 6.73 mg/dL, 10.3 mg/dL, 4.50 mg/dL and 0.014% for water, urea, citrate, ammonium sulfate and carbonate, respectively, for raw milk scanned with the DialPath accessory (Agilent Technologies, Santa Clara, California, USA) set at 30 μm .

Table 5-11 Comparison of performance FOM of classification models for differentiating genuine milk samples from those containing extraneous water and chemical adulterants using milk FTIR spectra collected by a transmission cell. Sn: sensitivity, TPR: true positive rate, Sp: specificity, FPR: false positive rate, ER: error rate, AUC: area under the curve, AD: adulterated, NA: not adulterated. FD: first derivative, VN: vector normalized, T: training dataset, V: validation dataset.

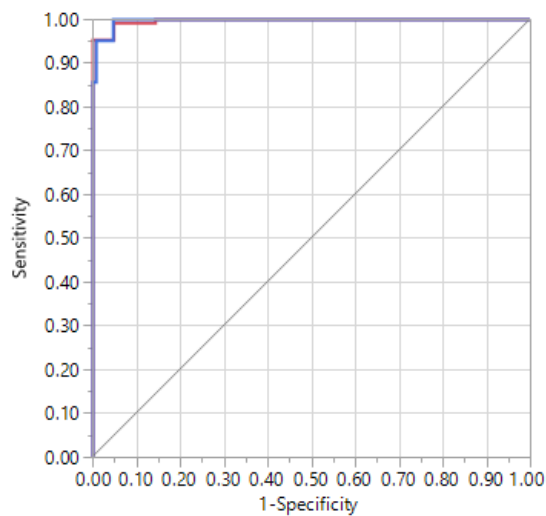
Transmission cell											
Spectra	Algorithm	Shrink Covariance	Dataset	Entropy R ²	Sn (TPR)	Sp	FPR	Accuracy	ER	AUC AD	AUC NA
Raw	PCA-LDA	No	T	-0.73	0.80	1.00	0.00	81.41	18.60	0.9761	0.9761
	PCA-LDA	No	V	-0.95	0.79	1.00	0.00	80.43	19.57		
	PCA-LDA	Yes	T	-0.71	0.79	1.00	0.00	80.17	19.83	0.9822	0.9822
	PCA-LDA	Yes	V	-0.83	0.79	1.00	0.00	80.43	19.57		
	PCA-QDA	No	T	0.99	1.00	1.00	0.00	100.00	0.00	1.0000	1.0000
	PCA-QDA	No	V	1.00	1.00	1.00	0.00	100.00	0.00		
	PCA-QDA	Yes	T	0.57	0.98	1.00	0.00	98.35	1.65	1.0000	1.0000
	PCA-QDA	Yes	V	1.00	1.00	1.00	0.00	100.00	0.00		
Raw VN	PCA-LDA	No	T	-0.73	0.80	1.00	0.00	81.41	18.60	0.9761	0.9761
	PCA-LDA	No	V	-0.95	0.79	1.00	0.00	80.43	19.57		
	PCA-LDA	Yes	T	-0.71	0.79	1.00	0.00	80.17	19.83	0.9822	0.9822
	PCA-LDA	Yes	V	-0.83	0.79	1.00	0.00	80.43	19.57		
	PCA-QDA	No	T	0.99	1.00	1.00	0.00	100.00	0.00	1.0000	1.0000
	PCA-QDA	No	V	1.00	1.00	1.00	0.00	100.00	0.00		
	PCA-QDA	Yes	T	0.57	0.98	1.00	0.00	98.35	1.65	1.0000	1.0000
	PCA-QDA	Yes	V	1.00	1.00	1.00	0.00	100.00	0.00		
FD	PCA-LDA	No	T	-0.81	0.80	1.00	0.00	80.58	19.42	0.9773	0.9773
	PCA-LDA	No	V	-0.71	0.81	1.00	0.00	82.61	17.39		
	PCA-LDA	Yes	T	-0.81	0.80	1.00	0.00	80.58	19.42	0.9773	0.9773
	PCA-LDA	Yes	V	-0.71	0.81	1.00	0.00	82.61	17.39		
	PCA-QDA	No	T	1.00	1.00	1.00	0.00	100.00	0.00	1.0000	1.0000
	PCA-QDA	No	V	0.60	1.00	0.67	0.33	97.83	2.17		
	PCA-QDA	Yes	T	0.88	1.00	1.00	0.00	99.59	0.41	1.0000	1.0000
	PCA-QDA	Yes	V	1.00	1.00	1.00	0.00	100.00	0.00		
FD VN	PCA-LDA	No	T	-0.82	0.80	1.00	0.00	80.58	19.42	0.9773	0.9773
	PCA-LDA	No	V	-0.73	0.81	1.00	0.00	82.61	17.39		
	PCA-LDA	Yes	T	-0.81	0.80	1.00	0.00	80.58	19.42	0.9773	0.9773
	PCA-LDA	Yes	V	-0.71	0.81	1.00	0.00	82.61	17.39		
	PCA-QDA	No	T	1.00	1.00	1.00	0.00	100.00	0.00	1.0000	1.0000
	PCA-QDA	No	V	0.60	1.00	0.67	0.33	97.83	2.17		
	PCA-QDA	Yes	T	0.88	1.00	1.00	0.00	99.59	0.41	1.0000	1.0000
	PCA-QDA	Yes	V	1.00	1.00	1.00	0.00	100.00	0.00		

Table 5-12 Comparison of performance FOM of classification models for differentiating genuine milk samples from those containing extraneous water and chemical adulterants using milk FTIR spectra collected by the DialPath accessory 30 μ m. Sn: sensitivity, TPR: true positive rate, Sp: specificity, FPR: false positive rate, ER: error rate, AUC: area under the curve, AD: adulterated, NA: not adulterated. FD: first derivative, VN: vector normalized, T: training dataset, V: validation dataset.

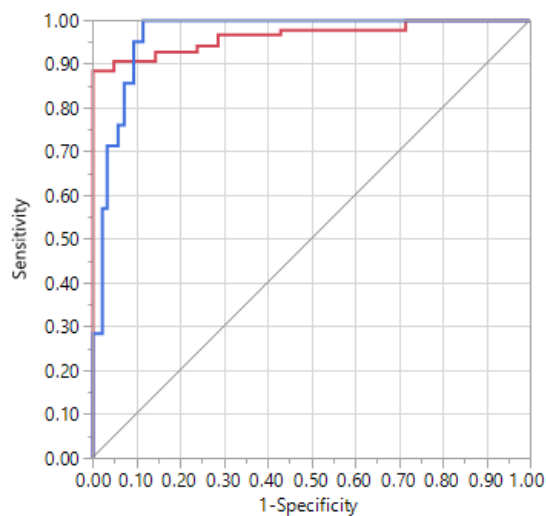
DialPath accessory											
Spectra	Algorithm	Shrink Covariance	Dataset	Entropy R ²	Sn (TPR)	Sp	FPR	Accuracy	ER	AUC - AD	AUC - NA
Raw	PCA-LDA	No	T	0.48	0.99	0.69	0.31	96.88	3.13	0.9809	0.9809
	PCA-LDA	No	V	0.42	0.92	0.40	0.60	85.71	14.29		
	PCA-LDA	Yes	T	0.57	0.99	0.69	0.31	96.88	3.13	0.9802	0.9802
	PCA-LDA	Yes	V	0.49	0.92	0.40	0.60	85.71	14.29		
	PCA-QDA	No	T	0.44	0.96	1.00	0.00	96.09	3.91	0.9998	0.9998
	PCA-QDA	No	V	1.00	1.00	1.00	0.00	100.00	0.00		
	PCA-QDA	Yes	T	-0.42	0.89	1.00	0.00	89.84	10.16	0.9971	0.9971
	PCA-QDA	Yes	V	-0.01	0.86	1.00	0.00	88.10	11.90		
Raw VN	PCA-LDA	No	T	0.07	0.98	0.63	0.38	95.37	4.63	0.9655	0.9655
	PCA-LDA	No	V	0.32	0.92	0.40	0.60	85.71	14.29		
	PCA-LDA	Yes	T	0.20	0.98	0.63	0.38	95.37	4.63	0.9636	0.9636
	PCA-LDA	Yes	V	0.37	0.92	0.40	0.60	85.71	14.29		
	PCA-QDA	No	T	0.77	0.98	1.00	0.00	98.46	1.54	0.9971	0.9971
	PCA-QDA	No	V	0.64	1.00	0.80	0.20	97.62	2.38		
	PCA-QDA	Yes	T	-0.51	0.87	1.00	0.00	88.03	11.97	1.0000	1.0000
	PCA-QDA	Yes	V	0.52	0.95	1.00	0.00	95.24	4.76		
FD	PCA-LDA	No	T	0.34	0.99	0.69	0.31	96.91	3.09	0.9869	0.9869
	PCA-LDA	No	V	0.41	0.92	0.80	0.20	90.48	9.52		
	PCA-LDA	Yes	T	0.41	0.99	0.69	0.31	96.91	3.09	0.9832	0.9832
	PCA-LDA	Yes	V	0.46	0.92	0.80	0.20	90.48	9.52		
	PCA-QDA	No	T	0.77	0.99	1.00	0.00	98.84	1.16	1.0000	1.0000
	PCA-QDA	No	V	1.00	1.00	1.00	0.00	100.00	0.00		
	PCA-QDA	Yes	T	-0.29	0.92	1.00	0.00	92.66	7.34	1.0000	1.0000
	PCA-QDA	Yes	V	0.99	1.00	1.00	0.00	100.00	0.00		
FD VN	PCA-LDA	No	T	0.20	0.96	0.69	0.31	94.21	5.79	0.9626	0.9626
	PCA-LDA	No	V	-0.19	0.92	0.40	0.60	85.71	14.29		
	PCA-LDA	Yes	T	0.26	0.96	0.69	0.31	94.21	5.79	0.9617	0.9617
	PCA-LDA	Yes	V	0.05	0.92	0.40	0.60	85.71	14.29		
	PCA-QDA	No	T	0.65	0.98	1.00	0.00	97.68	2.32	1.0000	1.0000
	PCA-QDA	No	V	1.00	1.00	1.00	0.00	100.00	0.00		
	PCA-QDA	Yes	T	-0.77	0.90	1.00	0.00	90.73	9.27	0.9985	0.9985
	PCA-QDA	Yes	V	-0.10	0.78	1.00	0.00	80.95	19.05		



A



B



C

Figure 5-14 ROC curves for classification models to differentiate genuine milk samples from those containing extraneous water and chemical adulterants. A) PCA-QDA developed using vector normalized raw spectra of homogenized milk samples collected with a transmission cell AUC=1.0000, B) PCA-QDA developed using vector normalized raw spectra of raw milk samples collected with Dial Path accessory AUC=0.9971, C) PCA-LDA developed using vector normalized raw spectra of raw milk samples collected with Dial Path accessory AUC=0.9655.

Table 5-13 Example of SIMCA output for classification of adulterated milk samples according to the added chemical. SIMCA can assign unknown sample to one class, multiple classes or no class. Class membership at 5% significance level.

Spectrum	Actual Adulterant ‡	Carbonate	Citrate	Ammonium sulfate	Urea	Mix
1.	CARBONATE0P7	*				
2.	CARBONATE0P7	*				
3.	CARBONATE0P7	*				
4.	CARBONATE0P5	*				
5.	CARBONATE0P5	*				
6.	CARBONATE0P5	*				
7.	CARBONATE0P2	*				
8.	CARBONATE0P2	*				
9.	CARBONATE0P2	*				
10.	CITRATE22P		*			
11.	CITRATE22P		*			
12.	CITRATE22P		*			
13.	CITRATE22P		*			
14.	CITRATE14P		*			
15.	CITRATE14P		*			
16.	CITRATE14P		*			
17.	CITRATE6P		*			*
18.	CITRATE6P		*			*
19.	CITRATE6P		*			
20.	MIX20P					*
21.	MIX20P					*
22.	MIX20P					*
23.	SULFATE18P			*		
24.	SULFATE18P			*		
25.	SULFATE18P			*		
26.	SULFATE12P					
27.	SULFATE12P			*		
28.	SULFATE12P			*		
29.	SULFATE4P			*		*
30.	SULFATE4P			*		*
31.	SULFATE4P			*		
32.	UREA28P				*	
33.	UREA28P				*	
34.	UREA28P				*	
35.	UREA18P				*	
36.	UREA18P				*	
37.	UREA18P				*	
38.	UREA12P				*	
39.	UREA12P				*	
40.	UREA12P				*	
41.	UREA4P				*	*
42.	UREA4P				*	*
43.	UREA4P				*	*

‡ Carbonate is added in dry form while urea, sulfate, citrate and mix are added as an adulteration solution.

Table 5-14 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using homogenized milk spectra scanned by Cary 630 (Agilent Technologies, Santa Clara, California, USA) spectrometer with transmission cell.

Milk	System	Adulterant	Region cm ⁻¹ Pre-treatment	r	RMSEC	RMSEP	Factors
H. Milk	Cary 630 – transmission cell	Sulfate mg/dL	1,461-1,446	0.99	1.61	2.14	8
			SG (window 7, polynomial 3) FD 1,114-1,069				
			SG (window 7, polynomial 3) FD				
H. Milk	Cary 630 – transmission cell	Citrate mg/dL	1,384-1,354	0.99	3.78	6.32	8
			SG (window 7, polynomial 3) FD 1,191-1,159				
			SG (window 7, polynomial 3) FD				
H. Milk	Cary 630 – transmission cell	Urea mg/dL	1,195-1,153	0.99	4.14	7.57	9
			SG (window 7, polynomial 3) FD 1,484-1,458				
			SG (window 7, polynomial 3) FD				
H. Milk	Cary 630 – transmission cell	Carbonate %	1,372-1,332	0.99	0.008	0.009	4
			SG (window 7, polynomial 3) FD				
H. Milk	Cary 630 – transmission cell	Water %	1,544-1,010	0.99	0.16	0.28	8
			1,796- 1,707				
			2,943-2,835				
			Raw				

Table 5-15 Comparison of cross-validation FOMs for milk adulterants models developed using milk spectra collected on Cary 630 (Agilent Technologies, Santa Clara, California, USA) with transmission cell.

Milk Samples	System	Adulterant	Cross-validation			External validation	
			r	RMSECV	Factors	R ²	Slope
H. Milk	Cary 630 – transmission cell	Sulfate mg/dL	0.99	2.27	8	0.99	0.99
H. Milk	Cary 630 – transmission cell	Citrate mg/dL	0.99	5.84	8	0.99	0.99
H. Milk	Cary 630 – transmission cell	Urea mg/dL	0.99	6.93	9	0.99	1.03
H. Milk	Cary 630 – transmission cell	Carbonate %	0.99	0.009	4	0.99	0.98
H. Milk	Cary 630 – transmission cell	Water %	0.99	0.21	8	0.99	1.00

Table 5-16 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using homogenized milk spectra scanned by Cary 630 (Agilent Technologies, Santa Clara, California, USA) spectrometer with Dial Path accessory. The same sample sets were scanned at 30 and 50 μm pathlength. a) milk samples contain added water and chemical adulterant, b) milk samples contain added water and urea only.

Milk Samples	System	Adulterant	Region cm^{-1} / Pre-treatment	r	RMSEC	RMSEP	Factors
H. Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Water % ^a	1,581-960 Raw	0.99	0.66	0.55	6
H. Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Water % ^a	1,581-960 Raw	0.99	0.26	0.33	9
Raw Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Water % ^b	1,500-960 FD	0.99	0.59	0.55	1
Raw Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Water % ^b	1,500-960 Raw	0.99	0.64	0.72	2
H. Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Urea mg/dL	1,484-1,457, 1,191-1,153 FD	0.99	6.4	8.62	3
H. Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Urea mg/dL	1,481-1,454, 1,191-1,137 Raw	0.99	2.76	2.95	7
Raw Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Urea mg/dL	1,481-1,427, 1,191-1,138 Raw	0.99	3.06	7.63	6
Raw Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Urea mg/dL	1,485-1,453, 1,195-1,141 FD	0.99	4.28	4.88	8
H. Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Sulfate mg/dL	1,454-1,442, 1,192-1,149 FD	0.99	2.65	2.79	9
H. Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Sulfate mg/dL	1,461-1,446, 1,114-1,069 Raw	0.99	1.99	1.72	5
H. Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Citrate mg/dL	1,380-1,350, 1,288-1,261 FD	0.99	6.77	9.63	6
H. Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Citrate mg/dL	1,399-1,369, 1,284-1,261 FD	0.99	5.90	5.43	2
H. Milk	Cary 630 - Dial Path Accessory – 50 μm pathlength	Carbonate %	1,381-1,326 FD	0.99	0.009	0.004	2
H. Milk	Cary 630 - Dial Path Accessory – 30 μm pathlength	Carbonate %	1,380-1,326 FD	0.99	0.009	0.005	2

Table 5-17 Comparison of cross-validation FMOs for milk adulterants models developed using milk spectra collected on Cary 630 Dial Path accessory (Agilent Technologies, Santa Clara, California, USA). a) milk samples contain added water and chemical adulterant, b) milk samples contain added water and urea.

Milk Samples	System	Adulterant	Cross-validation			External validation	
			r	RMSECV	Factors	R ²	Slope
H. Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Water % ^a	0.99	0.749	6	0.99	0.97
H. Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Water % ^a	0.99	0.347	9	0.99	0.99
Raw Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Water % ^b	0.99	0.757	1	-	-
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Water % ^b	0.99	0.968	2	-	-
H. Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Urea mg/dL	0.99	9.90	3	0.99	1.05
H. Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Urea mg/dL	0.99	5.43	7	0.99	0.98
Raw Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Urea mg/dL	0.99	5.06	6	0.99	1.00
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Urea mg/dL	0.99	7.23	8	0.99	1.00
H. Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Sulfate mg/dL	0.99	3.76	9	0.99	0.99
H. Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Sulfate mg/dL	0.99	3.35	5	0.99	1.00
H. Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Citrate mg/dL	0.99	12.3	6	0.98	0.98
H. Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Citrate mg/dL	0.99	8.90	2	0.98	0.97
H. Milk	Cary 630 - Dial Path Accessory – 50 µm pathlength	Carbonate %	0.99	0.013	2	-	-
H. Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Carbonate %	0.99	0.013	2	-	-

Table 5-18 Comparison of calibration models' FOMs for milk chemical adulterants and extraneous water. Calibration models were built using raw milk spectra scanned by Cary 630 spectrometer with Dial Path accessory (Agilent Technologies, Santa Clara, California, USA).

Milk	System	Adulterant	Region cm ⁻¹ Pre-treatment	r	RMSEC	RMSEP	Factors
Raw Milk	Cary 630 - Dial Path Accessory 30 µm pathlength	Sulfate mg/dL	1,484-1,457	0.99	2.48	4.50	5
			1,190-1,126 SG (window 7, polynomial 3) FD				
Raw Milk	Cary 630 - Dial Path Accessory 30 µm pathlength	Citrate mg/dL	1,445-1,347 SG (window 7, polynomial 3) FD	0.99	5.97	10.3	6
Raw Milk	Cary 630 - Dial Path Accessory 30 µm pathlength	Urea mg/dL	1,485-1,457 SG (window 7, polynomial 3) FD	0.99	4.77	6.73	8
Raw Milk	Cary 630 - Dial Path Accessory 30 µm pathlength	Carbonate %	1,380-1,322 SG (window 7, polynomial 3) FD	0.99	0.012	0.014	2
Raw Milk	Cary 630 - Dial Path Accessory 30 µm pathlength	Water %	1,575-1,010	0.99	0.25	0.39	9
			1,762-1,728				
			2,949-2,834 Raw				

Table 5-19 Comparison of cross-validation FMOs for milk adulterants models developed using milk spectra collected on Cary 630 Dial Path accessory (Agilent Technologies, Santa Clara, California, USA).

Milk Samples	System	Adulterant	Cross-validation			External validation	
			r	RMSECV	Factors	R ²	Slope
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Sulfate mg/dL	0.99	8.65	5	0.99	0.99
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Citrate mg/dL	0.99	11.6	6	0.99	0.99
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Urea mg/dL	0.99	7.22	8	0.99	0.98
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Carbonate %	0.99	0.019	2	0.99	0.94
Raw Milk	Cary 630 - Dial Path Accessory – 30 µm pathlength	Water %	0.99	0.36	9	0.99	1.02

Acknowledgement

The author thanks the following parties for their contribution to this project:

- Dr. Ashraf Ismail (McGill University) as the PI of this project.
- Dr. Jacqueline Sedman (McGill University) for her guidance on spectroscopy and quantitative analysis.
- PZL Industria Eletrônica Ltda (Londrina, PR, Brazil) for welcoming the author in their premises on two occasions, providing milk samples, chemicals and milk cryoscopy instrument.
- Breno Rachid, Maria Tereza-Piazzalunga, Rodrigo Teixeira de Andrade and Dilson Liukiti-Ito (PZL Industria Eletrônica Ltda) for their support during the two visits to Brazil.
- Agilent Technologies for providing portable FTIR spectrometer, Cary 630 (Agilent Technologies, Santa Clara, California, USA), and the DialPath sample introduction accessory (Agilent Technologies, Santa Clara, California, USA).
- Comline Elektronik Elektrotechnik GmbH (Wackersdorf, Germany) for building the lab-in-box analyzer.

Funding

This project was funded by two MITACS awards (Application Ref. IT06096 and IT10571).

Connecting Statement

In the previous chapter, principle component analysis (PCA) successfully isolated the spectral fingerprint from milk FTIR data that reflected the changes in milk chemical composition resulting from a systemic factor that altered milk composition, which was adulteration. In addition, PCA reduced the dimensionality of the milk FTIR spectral data and produced a new set of variables (i.e., principal components) that were used as predictors for classification models and as input variables to a model to test the significance of the systemic factor effect on milk spectral data. In this chapter, several multivariate analysis techniques, including PCA, will be evaluated as data mining tools to isolate the spectral fingerprint that reflects the effect of a housing treatment of dairy cattle on milk FTIR spectral data in the context of animal controlled-design trials.

Chapter 6: Evaluation of data analysis approaches of chemical milk composition and FTIR spectral data in the context of controlled-design trials testing dairy cattle housing welfare improvements

Abstract

Milk is a complex biological fluid whose components are synthesized from building blocks that are obtained from the blood plasma. These precursor molecules are either originated from diet or they are produced by metabolic processes in the cow's body. Several studies have demonstrated that milk composition reflects the concentrations of key blood plasma metabolites, the nutritional state of the cow and health conditions that might affect the cow's productivity. Some of these studies have relied on the numerical values of the concentrations of the studied milk components, others have directly relied on the FTIR spectrum of studied milk samples without using an intermediate prediction model to produce a specific number for a specific milk component.

The objective of this chapter is to evaluate the suitability of different data analysis approaches of milk composition data to detect the effect of tie rail (TR) configuration on milk composition. A TR is the pipe used as the attachment for the tie chain, which controls the forward location of each cow in her stall and facilitates or obstructs the cow movement in her stall while changing positions as well as cow access to feed; hence, affecting the cow's intake of key nutritional precursors required for the biosynthesis of different milk components. The evaluated data analysis approaches are categorized into two groups: univariate and multivariate approaches, which were applied to milk components' concentrations and milk FTIR spectral data. For the univariate approach, mixed modeling was employed, which is a powerful statistical tool that allows fitting mixed linear models. The multivariate approach employed principal component analysis (PCA) and hierarchical cluster analysis (HCA), which are unsupervised techniques, and partial least squares – discriminant analysis, (PLS-DA), which is a supervised technique.

The milk composition numerical data did not reveal significant treatment effect when analyzed by the mixed model, and PCA revealed that the predictions of minor milk components might have been biased by spectral contribution of major milk components. PLS-DA did not consider the repeated measurement structure of this study and it was limited to the milk components that were reported by the FTIR milk analyzer. On the other hand, PLS-DA analysis of the full milk FTIR

spectra revealed more details pertaining to the TR treatment effect on milk composition than PLS-DA analysis of the numerical dataset of milk composition. However, PLS-DA was not capable of testing the hypothesis or producing statistics regarding the significance of all the effects included in the statistical model of the study. In addition, unsupervised analysis of the full milk FTIR spectra revealed only the strongest effect on milk composition, which was the time effect (i.e., increased in cow days in milk and/or environmental changes such as ambient temperature as the trial was conducted over 20 weeks).

A hybrid approach was successfully applied to the FTIR spectral data, which retained the multivariate structure of the FTIR spectral data, while at the same time, accommodated the utilization of the mixed model to test fixed and random effects on the FTIR spectral data according to the statistical model that was defined by the experimental design of the trial with a repeated measurement structure and enabled hypothesis testing.

6.1 Introduction

Milk is a complex biological mixture of a solution of salts and carbohydrates, dispersed proteins, casein micelles and fat globules. Major milk components and some minor ones are synthesized in multiple organelles in the alveolar cells of the mammary glands. Milk proteins are synthesized in the rough endoplasmic reticulum from free amino acids or peptides that are absorbed from the bloodstream through the basolateral cell membrane of the alveolar cells. After that, milk proteins are transported to the Golgi apparatus where they undergo posttranslational modifications. The Golgi apparatus is also responsible for the synthesis of lactose from glucose, which is absorbed from the bloodstream. Synthesized milk proteins, lactose and other milk components, such as citrate, calcium and water, are packed into secretory vesicles that release these components through exocytosis into milk accumulating in the alveolar lumen. Fatty acids are the building blocks for milk triglycerides. Half of these fatty acids are derived from diet, which include most of C18 fatty acids (i.e., stearic, oleic and linoleic acids) and 30% of C16 fatty acids (i.e., palmitic acid). The other half includes shorter chain fatty acids, which are derived from *de novo* synthesis in the cytoplasm of the mammary glands using acetate and β -hydroxybutyric acid as precursors that are absorbed from the bloodstream. Milk fat droplets form near the rough endoplasmic reticulum, enlarge in size and make their way towards the apical membrane of the alveolar cell where they protrude from the cell into the alveolar lumen surrounded by portions of the cell membrane. Other milk components cross from the bloodstream through the basolateral membrane of the alveolar cells, traverse the cell and pass across the apical membrane into milk accumulated in the alveolar lumen. Such components include water, urea, glucose and some ions [97]. Several studies have demonstrated that milk composition reflects the concentrations of key blood plasma metabolites, such as non esterified fatty acids (NEFA) [98] and β -hydroxybutyric acid [99], the nutritional state of the cow [61, 100] and health conditions [15] that might affect the cow's productivity. Some of these studies have relied on the numerical values of the concentrations of the studied milk components that are determined by different analytical techniques, such as FTIR and gas chromatography, others have directly relied on the FTIR spectrum of studied milk samples without going through an intermediate prediction model to produce a specific number for a specific milk component.

The objective of this chapter is to evaluate the suitability of different data analysis approaches of milk composition data to detect the effect of tie rail (TR) configuration on milk composition. A

TR is the pipe used as the attachment for the tie chain, which controls the forward location of each cow in her stall and facilitates or obstructs the cow movement in her stall while changing positions as well as cow access to feed; hence, affecting the cow's intake of key nutritional precursors required for the biosynthesis of different milk components. Figure 6-1 summarizes the data analysis approaches of milk composition that will be covered in this chapter and that can be categorized into two groups: univariate and multivariate approaches, which will be applied to milk components' concentrations and milk FTIR spectral data. For the univariate approach, mixed modeling will be employed, which is a powerful statistical tool that allows fitting mixed linear models [101]. Mixed models are a generalization of the standard linear model. Mathematically, they are described as follows:

$$y = X\beta + Z\gamma + \epsilon$$

Where y is a vector of observed data, β is an unknown vector of fixed-effects parameters with known design matrix X , γ is an unknown vector of random-effects parameters with known design matrix Z and ϵ is an unknown random error vector.

Mixed models allow the modeling of the mean response that is described by fixed-effects parameters, in addition to the variance and covariance of the mean that are described by random-effects parameters. Fixed effects represent studied factors that are controlled in the trial with specific levels, such as different TR configurations. On the other hand, random effects represent factors that might affect the variability of the studied response but cannot be controlled during the trial, such as the cow. Fitting mixed models allows drawing statistical inferences and generating suitable statistics for hypothesis testing. However, mixed modeling can be applied to only one response at a time, while in fact, major and minor milk components are synthesized through biochemical pathways that are governed by the metabolic, nutritional and health state of the cow. For this reason, a multivariate approach might be more appropriate, in this case, because it can detect trends in the dataset based on changes in observed variables as a holistic unit.

The multivariate approach will employ principal component analysis (PCA) and hierarchical cluster analysis (HCA), which are unsupervised techniques, and partial least squares – discriminant analysis, (PLS-DA), which is a supervised technique. PCA assess the correlation among the studied variables by defining a smaller set of variables that are called principal components that eliminate unnecessary correlation between the original variables and that describe the unique

sources of variations in the original dataset. Hence, it can be used as a method to explore the structure in the relationships between the variables and to reduce the dimensionality of big data. HCA is an exploratory multivariate technique that aims at uncovering natural groupings of observations in a data set where the observations within each group are relatively homogeneous, yet the groups are unlike each other. However, both techniques do not provide suitable statistics for hypothesis testing and they do not take into consideration the statistical model of the experimental design under which the data is generated. PLS-DA is a supervised multivariate technique that aims at maximizing the separation of observations according to their class membership (i.e., treatment membership) to achieve maximum class separation. However, PLS-DA does not take into consideration the repeated measurement structure of the data collected from the same subjects over a specific period of time.

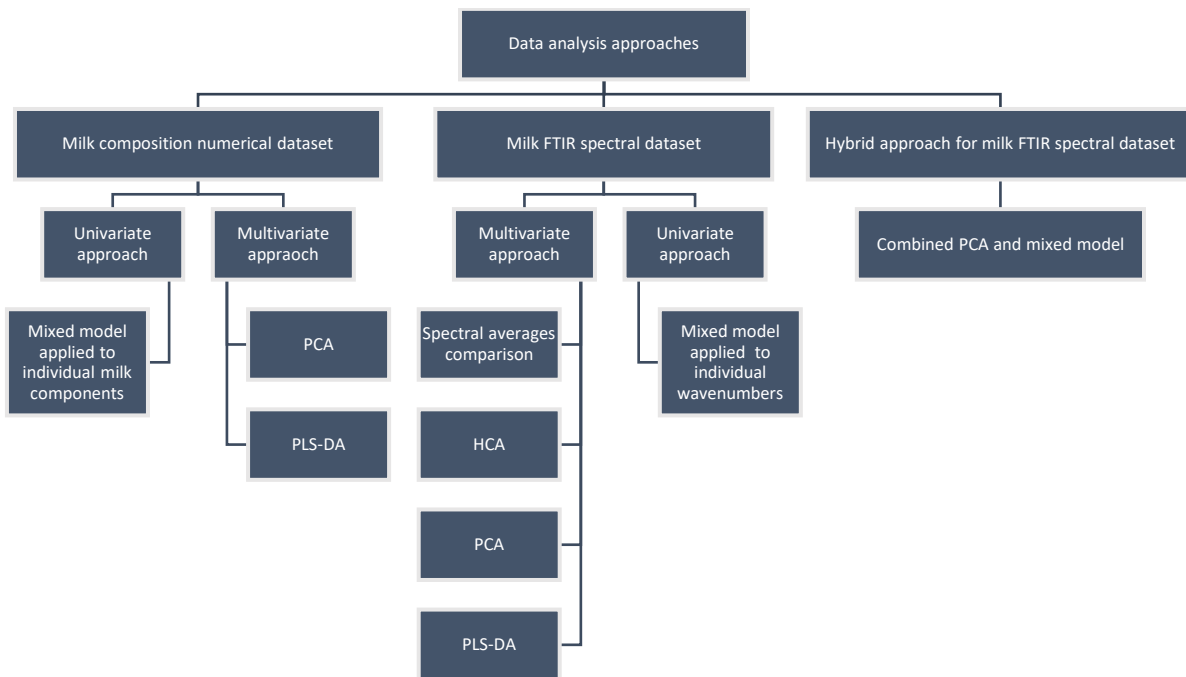


Figure 6-1 Data analysis approaches that are considered to detect the tie rail position effect on milk composition numerical and spectral data.

6.2 Materials and Methods

6.2.1 The data

The data used in this chapter was collected during an animal trial to test the effect of one element of stall configuration (i.e., tie rail) on cow's behaviour and welfare [102], which was conducted at the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). A tie rail (TR) is the pipe used as the attachment for the tie chain, which controls the forward location of each cow in her stall. In this study, 48 cows were assigned to 4 TR configurations, which were defined by the height and the forward position of TR. TR heights were 122, 122, 112 and 102 cm and forward positions were 18 cm, 36 cm, 18 cm and 36 cm for treatments T1, T2, T3 and T4, respectively. Treatments T1 and T2 are TR configurations that are recommended and commonly found on dairy farms, respectively. On the other hand, treatments T3 and T4 are new TR configurations designed to increase the opportunity of movements of the cow at her stall; hence, improve cow behavior and welfare. Cows were assigned to 6 different blocks to account for age of the cow, days in milk within current lactation and location in the barn effects. Half the cows underwent the trial during summer 2016 and the other half during fall 2016. Each period lasted 10 weeks (i.e., period 1: from July 25th to October 3rd, period 2: from October 10th to December 19th). The data of this trial was analyzed under the following statistical model:

$$Y_{ijklm} = \mu + trt_i + start_j + block_{kji} + cow_{lkji} + week_m + trt_i \times week_m + e_{ijklm}$$

Where Y_{ijklm} was the dependent variable; the outcome measure of the l^{th} cow from the k^{th} block (parity, DIM and location in the barn) and the j^{th} start date on the combination of the i^{th} tie-rail configuration and m^{th} week. trt_i was the fixed effect of the i^{th} tie-rail configuration. $start_j$ was the fixed effect of the j^{th} start date. $block_{kji}$ was the fixed effect of k^{th} parity, DIM and location in the barn from the j^{th} start date on the i^{th} tie-rail configuration treatment. cow_{lkji} was the random effect of the l^{th} cow from the j^{th} start date and the k^{th} block on the i^{th} tie-rail configuration treatment. $week_m$ was the fixed effect of the m^{th} week. $trt_i \times week_m$ is the interaction effect of the individual combination of the i^{th} tie-rail configuration treatment with the m^{th} week. e_{ijklm} was the random residual associated with the outcome measure of the l^{th} cow from j^{th} start date and k^{th} block on the combination of the i^{th} tie-rail configuration treatment and the m^{th} week [102].

Milk samples were collected weekly and they were analyzed by Valacta Inc. lab (Ste., Anne-de-Bellevue, QC, Canada) to determine major and minor milk components using commercial FTIR milk analyzer. A total of 19 milk components were included in this study. These components are fat, protein, lactose, total solids (TS), urea, β -hydroxybutyrate (BHB), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), short-chain fatty acids (SCFA), mid-chain fatty acids (MCFA), long-chain fatty acids (LCFA), saturated fatty acids (SFA), total unsaturated fatty acids (TUFA), mono-unsaturated fatty acids (MUFA), poly-unsaturated fatty acids (PUFA), trans fatty acids (TFA), free fatty acids (FFA) and fat-to- protein ratio (FP ratio). The total number of samples that were analyzed is 626. Two sets of data were received from Valacta. The first dataset comprised the concentrations of milk components that were determined by the commercial FTIR milk analyzer for individual milk samples. This dataset will be referred to as the numerical dataset. The second dataset contained FTIR spectra recorded for individual milk samples collected during the trial. This dataset will be referred to as the spectral dataset. Each FTIR spectrum consisted of 1060 spectral variables between 5008 cm^{-1} and 925 cm^{-1} .

6.2.2 Analysis of milk composition numerical dataset

Two approaches were applied to the analysis of the numerical dataset: a univariate and a multivariate approach. The Mixed procedure in SAS 9.4 (SAS Institute, Cary, NC, USA) was utilized to detect the treatment effect (i.e., TR configuration effect) on individual milk components. This procedure allows fitting of mixed linear models, which are a generalization of the standard linear model. A mixed linear model allows the modeling of variances and covariances in addition to the means of the dataset. The probability distribution of the data in a mixed linear model can be described by two sets of parameters. The fixed-effects parameters describe the mean of the model, while the random-effects parameters describe the variance-covariance of the model [101].

Alternatively, two multivariate techniques were used to analyse the numerical dataset. These techniques are principal component analysis (PCA), which is an unsupervised method, and partial least squares discriminant analysis (PLS-DA), which is a supervised method. PCA decomposes the variance in the dataset into a number of new variables that is equal to the number of observed variables (i.e., milk components) called principle components (PCs) [93]. These PCs are orthogonal to each other, which means that the correlation is 0 between any two PCs, and they describe the same variance structure as the original variables in the dataset. Each PC is a linear

combination of optimally weighted observed variables. However, only the few first PCs account for a significant portion of the variability in the original dataset. PCA loading plot reveals the main set of variables that drive variation in the dataset; on the other hand, PCA scores plot unveils tendencies toward clustering of samples.

PLS-DA is used for exploratory purpose in this study and not to build predictive models. It requires two matrices. The first is the X matrix that contains the observed variables, or the predictors, which are the concentrations of milk components for the observations. The second is the Y matrix that contains dummy variables that describe treatment membership of observations in the X matrix. PLS-DA performs a decomposition process on the X matrix (i.e., the numeric dataset) similar to that of PCA; however, this process is guided by observations' treatment membership described in the Y matrix [27]. The resulting new variables in PLS-DA are called latent variables or factors and they are calculated in a way to maximize observations' separation according to treatment. To construct a PLS-DA model, the X matrix data was centered and scaled. Statistically Inspired Modification of the PLS Method algorithm (SIMPLS) was used to build the PLS-DA model and it was validated by leave-one-out cross-validation approach. This algorithm was chosen because treatment has multiple levels [27]. Initially, all predictors were included in the PLS-DA model. The Variable Importance for the Projection (VIP) score was used to eliminate predictors that did not contribute to the modeling of the response (i.e., treatment membership) to comply with the parsimony principle. Only predictors with VIP score >0.8 were retained in the model. This process was repeated multiple times until a final model was obtained with the least number of predictors modeling the treatment membership of observations. JMP Pro 13.2.1 (SAS Institute, Cary, NC, USA) was used to perform PCA and PLS-DA on the numeric dataset.

6.2.3 Analysis of milk spectral dataset

Omnic 7.3 (Thermo Electron Corporation, Waltham, MA, USA) was used to visualize the milk FTIR spectra and to calculate average spectra by treatment and week. Only spectral regions containing information related to milk composition were retained for spectral analysis. These regions were $1612\text{--}925\text{ cm}^{-1}$, $1797\text{--}1681\text{ cm}^{-1}$ and $3061\text{--}2803\text{ cm}^{-1}$. The total number of spectral variables that were retained for analysis was 278 wavenumbers. MATLAB codes were written to calculate differential first derivative (FD) of the spectra with a derivative window of 1, to vector normalize (VN) the spectra [29] and to load individual spectra into a matrix for analysis by JMP

Pro 13.2.1. The FD will eliminate the offset of the spectral baseline, if present, and it will enhance small peaks that are especially present on the shoulder of larger peaks. VN will eliminate variability in the spectra that is not related to chemical composition of milk samples and that is originated from instrumentation, such as variability in the IR source intensity. Spectral analysis was performed on raw, VN raw, FD, VN FD spectral datasets. The four versions of the spectral dataset were analyzed by Hierarchical Cluster Analysis (HCA) and PCA. Forward search feature selection algorithm [103] was applied to the spectral datasets to pick up the relevant regions for classification according to the studied factors. This algorithm is implemented in DataAnalysis [89], which is an inhouse written software at the McGill IR group in the food science department. In addition, PLS-DA models were developed for the four versions of the milk spectral dataset using the same methodology that was used with the milk numeric dataset except that the models were validated by K-Fold (K=10) cross-validation approach for computational considerations. JMP Pro 13.2.1 (SAS Institute, Cary, NC, USA) was used to perform HCA, PCA and PLS-DA on the spectral datasets. On the other hand, a univariate approach was applied to the four versions of the milk FTIR spectral datasets. In this approach, the mixed model was applied to the individual spectral variables through a loop that iterated over these variables in the retained spectral regions using SAS 9.4 (SAS Institute, Cary, NC, USA). It must be noted, that Proc Mixed in SAS can test different effects on one response (i.e., the spectral variable) at a time.

6.2.4 Hybrid approach for milk spectral dataset analysis

A new approach has been developed to analyze the milk spectral dataset (Figure 6-15). This new approach combines multivariate analysis with mixed modeling to test specific treatment effect. This approach was first tested on milk samples spiked with lactose before applying it to the four versions of the spectral dataset of TR trial. For this purpose, skim milk packs were purchased from the local market and eleven samples were spiked with 1% lactose and another eleven samples were spiked with 5% lactose. Cary 630 (Agilent Technologies, Santa Clara, California, USA) spectrometer was used to record the FTIR spectra of these samples with a resolution of 16 cm^{-1} and 32 co-added scans. The spectral regions that were retained for spectral analysis were $1612\text{--}925\text{ cm}^{-1}$, $1797\text{--}1681\text{ cm}^{-1}$ and $3061\text{--}2803\text{ cm}^{-1}$ with a total of 278 spectral variables. Inhouse MATLAB codes were written to calculate differential FD of the spectra, to VN the spectra and to load individual spectra into a matrix for analysis by JMP Pro 13.2.1.

In JMP Pro, PCA was applied to the four versions of the spiked milk samples spectral dataset (i.e., raw, FD, VN raw, VN FD) as a dimension reduction method to produce a new set of orthogonal variables that explain the variance in the original dataset and to reduce the number of responses that will be tested by the mixed model [104]. PCs with eigenvalue ≥ 1 and that explained $\geq 1\%$ of the variation were considered meaningful and were retained for testing by the mixed model with a significance level $\alpha=0.05$. It must be noted that PCA was not used to model or to test any effect in this case.

In the mixed model, the effect of spiked lactose concentration was tested on retained PCs and labels “High” and “Low” were used to categorized samples with 5% and 1% spiked lactose, respectively. The PCA scores of the raw and FD spectral data were compared to understand the effect of FD on the PCA scores behaviour and their interpretation in each case. The same comparison was also applied to the means estimates of the mixed model for the retained PCs of the raw and FD spectral dataset to correctly interpret the differences in the means estimates in each case.

In case a PC revealed a significant treatment effect, then its loading spectrum was interpreted. A loading spectrum extracted from raw spectral dataset can be interpreted directly. On the other hand, loading spectrum extracted from a FD spectral dataset needs to be integrated to restore the actual peaks of the influential FTIR bands that became zeros when the first derivative was calculated. The cumulative trapezoidal numerical integration function in MATLAB was used to calculate the spectral integral for the loading spectrum in question. This function performs numerical integration via the trapezoidal method and preserves the intermediate integration values, which results in a spectrum that approximates the raw absorption spectrum prior to the FD calculation. The spectral integral will be missing the baseline, which is not an issue since the spectral integral will not be used for quantitative determinations. If the integrated loading spectrum produced wide humps with no clear peaks, the Peak Resolve feature in Omnic 7.3 (Thermo Electron Corporation, Waltham, MA, USA) was used to fit the integrated loading spectrum for probable peaks. To do so, the Voigt function [105] with low sensitivity was used and the baseline was set to none. The noise and the full width at half height (FWHH) of the narrowest peak in the region of interest were determined by the software. The fitting process was repeated several times until an acceptable residual spectrum was obtained.

The same approach was applied to the spectral datasets of the TR trial spectral dataset. The Mixed procedure in SAS 9.4 (SAS Institute, Cary, NC, USA) was utilized to test the TR treatment effect on the meaningful PCs extracted from the raw, FD, VN raw and VN FD spectral datasets. First, the dataset was sorted by treatment, start, block, cow and week. Then, a SAS macro was utilized to iterate over the meaningful PCs extracted from the spectral dataset in question and to test the treatment effect according to the statistical model that was defined by the experimental design. Finally, the least squares mean were tested for the PC that revealed a significant treatment effect.

6.3 Results and Discussion

6.3.1 Analysis of milk composition numerical dataset

The univariate approach that employed mixed modeling did not reveal significant treatment effect on any milk component that was determined by Valacta's FTIR milk analyzer (Table 6-1). The strongest significant effect was observed for week on almost all milk components followed by the block effect. The main drawback of this type of analysis is that it tests the treatment effect on each milk component separately, while in fact, major and minor milk components are synthesized through biochemical pathways that are governed by the metabolic, health and nutritional state of the cow. For example, lactose and minerals are the least variable milk components [106]. They maintain the osmolarity of milk at ~300 mOsm, which is equal to the osmolarity of blood, to facilitate secretion of milk. In addition, lactose synthesis in mammary glands consumes 85% of circulating glucose in the plasma, which is mainly provided by carbohydrate intake from feed [107]. Since the osmolarity of blood is stable in healthy cows, decreased lactose concentration in milk results from reduced plasma glucose concentration due to reduced carbohydrate intake [100], which can be an indicator of elevated body fat mobilization, or it can be an indicator of health issues, such as ketosis and mastitis [15]. In all cases, decreased lactose concentration is accompanied by changes in concentrations of other milk components. In the case of elevated body fat mobilization, increased fat metabolism is characterized by increased non-esterified fatty acids (NEFA), BHB and reduced glucose concentration in the plasma, and decreased protein and increased acetone, BHB, C18:1 and C16:0 in milk [98, 100]. Having said that, milk composition represents a snapshot of the metabolic and nutritional state of the cow, and changes to a single milk component cannot be considered as an indicator of the nutritional or metabolic state of the cow resulting from a specific treatment effect. This observation indicates that multivariate analysis might be a sound alternative approach to analyze this type of data. These techniques can detect trends in the data based on changes in observed variables as a holistic unit.

Table 6-1 *P* values of different effects on milk components for the TR configuration trial. No significant effect was detected for treatment. Week has the strongest significant effect on almost all milk components followed by Block.

Milk Component	Treatment	Start	Block (Start)	Week	Treatment*Week
Fat	0.3385	0.8329	0.1515	0.0691	0.5137
Protein	0.7165	0.2317	<.0001	<.0001	0.8060
Lactose	0.9320	0.0077	0.0031	<.0001	0.6973
TS	0.3143	0.2409	0.0164	<.0001	0.4894
Urea	0.5993	0.0254	0.5723	<.0001	0.9296
BHB	0.1614	0.1955	0.0707	<.0001	0.8236
C16_0	0.3812	0.2291	0.0483	0.0013	0.2456
C18_0	0.4320	0.1892	0.0191	0.0191	0.7989
C18_1	0.1873	0.0148	0.5218	0.0029	0.7582
SCFA	0.4203	0.0481	0.2833	0.0091	0.5678
MCFA	0.2875	0.9795	0.0205	<.0001	0.2449
LCFA	0.2099	0.9406	0.2570	0.1617	0.8289
SFA	0.3893	0.7455	0.1214	0.0417	0.3384
TUFA	0.4630	0.0895	0.6194	0.1189	0.8820
MUFA	0.3452	0.2299	0.5414	0.0093	0.7976
PUFA	0.4009	0.6416	0.9775	<.0001	0.5060
Trans FA	0.6415	0.3226	0.0486	<.0001	0.1055
FFA	0.5427	0.7243	0.0178	0.0007	0.9452
FP Ratio	0.5177	0.5930	0.5109	0.0086	0.5780

PCA revealed four meaningful PCs that explain 80.77% of the variation described by the observed variables, which are milk components' concentrations (Table 6-2). Two criteria were considered to extract meaningful PCs, which are eigenvalue ≥ 1 and explained variation $\geq 1\%$ [93]. PC1, PC2, PC3 and PC4 accounted for 50.46%, 16.53%, 7.42% and 6.53% of the total variation, respectively. The PCA scores plot (PC1 vs. PC2) shows that milk samples cluster at the origin of the PCA space and it did not reveal any clustering trends according to treatment, which suggests the absence of treatment effect (Figure 6-2). However, PCA is an unsupervised multivariate method that does not

take into consideration the statistical model under which the data was generated. This fact undermines the capabilities of PCA alone in detecting a treatment effect in the presence of other strong effects, such as week and block. On the other hand, the PCA loadings plot (PC1 vs. PC2) reveals an important observation (Figure 6-3). Milk components, which are determined by FTIR milk analyzer, are clustering into four groups that correspond to the spectral regions that are used to predict the concentrations of major milk components by PLS calibration models. The first group corresponds to the 1200-900 cm^{-1} spectral region in milk FTIR spectrum and it includes lactose and trans fatty acids, whose main FTIR bands in milk are centered at $\sim 1080 \text{ cm}^{-1}$ and $\sim 967 \text{ cm}^{-1}$, respectively. In this group, we also notice the presence of BHB, which has a minor absorption peak centered at $\sim 1080 \text{ cm}^{-1}$ in milk FTIR spectrum. However, the prominent absorption band of BHB is centered at $\sim 1405 \text{ cm}^{-1}$ in milk FTIR spectrum that is not affected by lactose absorption of IR energy, which means that changes in lactose and BHB might originate from the same source of variation. The second group corresponds to the C=O triglyceride ester linkage stretching band or Fat A region located at $\sim 1745\text{-}1725 \text{ cm}^{-1}$ and it includes C18:0, C18:1, TUFA, MUFA, PUFA, LCFA and FP ratio. The third group corresponds to the C–H stretching band or Fat B region located at $\sim 2980\text{--}2800 \text{ cm}^{-1}$ and it includes fat, TS, C16:0, SCFA, MCFA and SFA. The reports that were received from Valacta expressed fat as Fat B, based on this, the distinction between the two groups was made. The last group corresponds to the 1580-1200 cm^{-1} spectral region and it includes protein and urea that have characteristic bands in milk at $\sim 1565\text{--}1520 \text{ cm}^{-1}$ (i.e., Amide II) and $\sim 1465 \text{ cm}^{-1}$, respectively. It must be noted that urea and milk proteins share FTIR absorption band located at $\sim 1600\text{-}1700 \text{ cm}^{-1}$ that is overwhelmed by the immense water absorption of IR energy. This observation suggests that the concentrations of minor milk components that are predicted by the PLS calibration models of the FTIR milk analyzer might have been biased by the spectral contribution of major milk components, which undermines the potential of predicated minor milk components of exposing subtle differences related to changes in housing design (i.e., TR treatments) over a short period of time especially when a univariate approach is used to analyze this type of data.

Table 6-2 Principal components extracted from the numerical dataset of milk analysis for the TR configuration trial.

PC	Eigenvalue	Variation % Explained	Cum. Variation % Explained
PC1	9.5878	50.462	50.462
PC2	3.1071	16.353	66.815
PC3	1.4106	7.424	74.240
PC4	1.2406	6.529	80.769
PC5	0.9652	5.080	85.849
PC6	0.8242	4.338	90.187
PC7	0.7133	3.754	93.941
PC8	0.3734	1.965	95.907
PC9	0.3332	1.753	97.660
PC10	0.1819	0.957	98.617

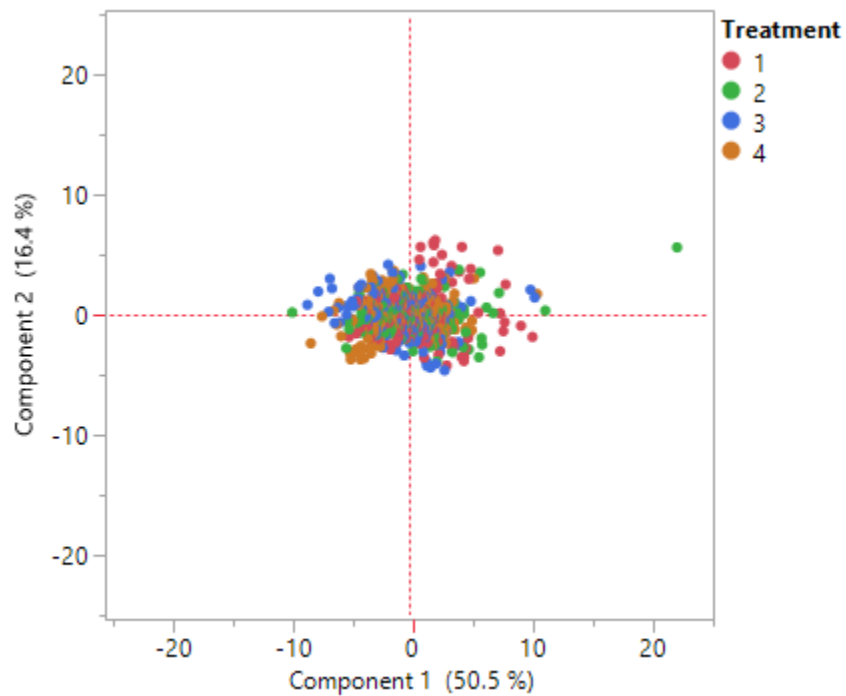


Figure 6-2 PCA scores plot projecting milk samples' scores for PC1 vs. PC2 for the TR configuration trial.

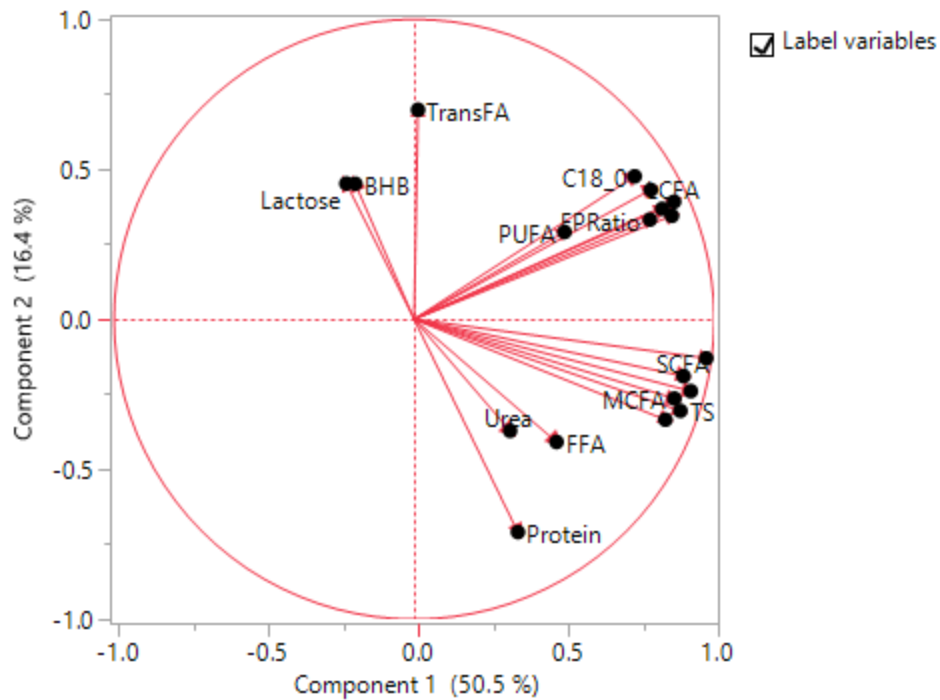


Figure 6-3 PCA loading plot (PC1 vs. PC2) projecting the main groups of milk components that explains the majority of the variation in the milk numerical dataset of the TR configuration trial.

PLS-DA was the second multivariate technique that was used to analyze the milk numerical dataset. A parsimonious PLS-DA model was obtained after eliminating predictors that did not contribute to the modeling of treatment membership of observations. The eliminated predictors were fat, protein, TS, C18_0, LCFA, SFA, TUFA, FFA and FP ratio. Only 10 milk components had a VIP score greater than 0.8 (Table 6-3). Inspection of the coefficients of these milk components for the classification of observations according to treatment reveals that only T3 shows a pattern that might be in agreement with elevated body fat mobilization, which includes increased BHB and C16:0, which is a long chain fatty acid, and decreased lactose (Figure 6-4) [98, 100]. However, this PLS-DA model explained only 10.79% of the variation related to treatment membership of observations, which suggests that the treatment effect on milk composition was negligible or the period of exposure to treatment was not sufficient to produce a significant treatment effect on milk composition, which explains the lack of effect on other milk components, such as protein and C18:1. PLS-DA was the only technique that provided some insight into the effect of treatment on milk composition. However, the conclusions drawn from PLS-DA might be undermined by the fact that it does not consider the repeated measurement structure of this study

and the effects other than treatment that are included in the statistical model under which the data was generated.

Table 6-3 VIP scores of predictors of the parsimonious PLS-DA classification model according to treatment.

Milk Component	VIP Score
C18_1	1.2096
SCFA	1.1330
BHB	1.0890
MUFA	1.0364
Urea	1.0016
PUFA	0.9785
MCFA	0.9590
Trans FA	0.8626
Lactose	0.8310
C16_0	0.8237

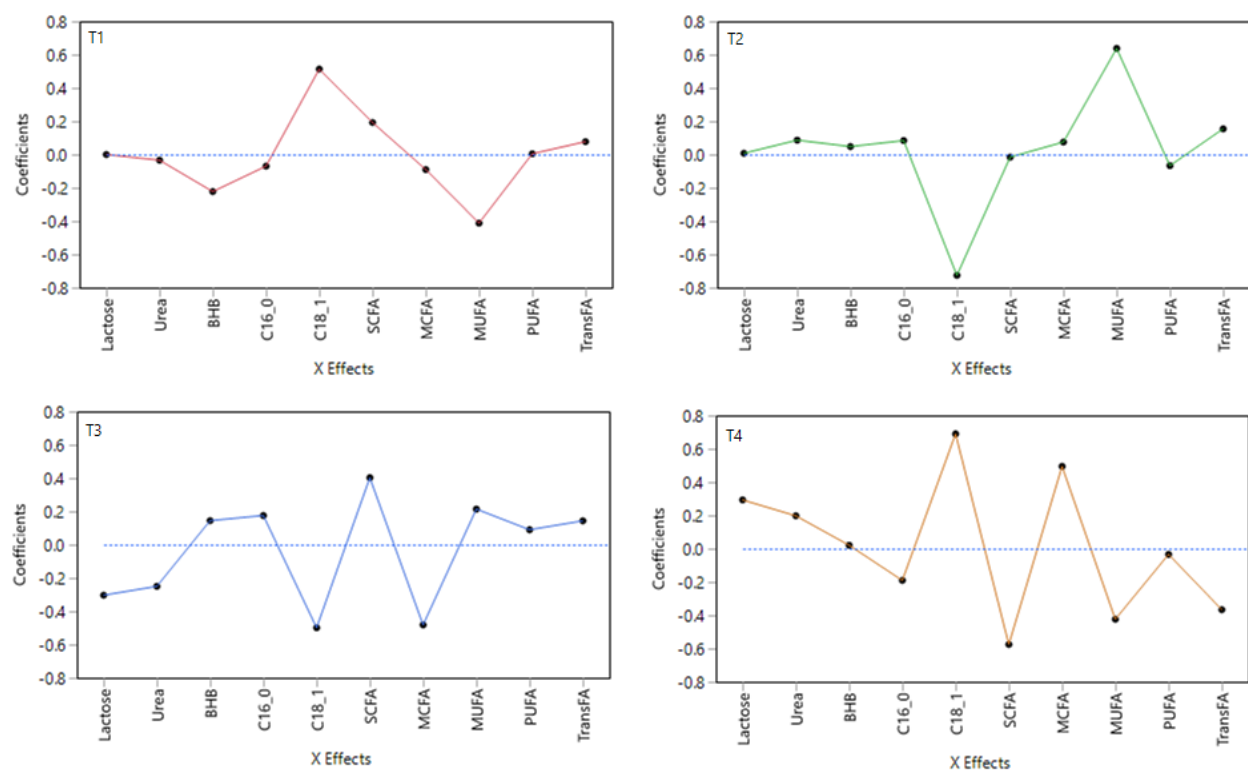


Figure 6-4 Coefficients of important milk components for classification of observations according to treatment by the parsimonious PLS-DA model. T3 shows signs of elevated body fat mobilization similar to what has been reported in the literature, such as increased BHB and C16:0, a long chain fatty acid, and decreased lactose.

6.3.2 Analysis of milk Spectral dataset

The analysis of the milk composition numerical dataset was limited to milk components that were provided by Valacta. However, there are other minor milk components that have been reported in the literature as biomarkers for the cow's metabolic state, such as citrate, acetone and acetoacetate [59], which are not included in this dataset and which might have provided better insight about the treatment effect on milk composition. On the other hand, milk FTIR spectra have information about all molecules that contain covalent bonds present in milk, which means that the milk spectral dataset contains more details about milk composition of samples collected during this study. Having said that, several approaches were evaluated for analysing the spectral dataset of milk. The first approach was to visually inspect the average treatment spectra calculated for week 3 and 10. For example, Figure 6-5 shows that the average raw spectra for week 3 and 10 of T1 during season 1 of the trial have more intense absorption bands than those of T3 at $\sim 1745 \text{ cm}^{-1}$, or Fat A region, which suggests that milk samples of cows assigned to T1 had higher fat content than milk samples of cows assigned to T3 during the first season of the trial. In addition, it can be noticed that the fat

content decreased and increased for T1 and T3, respectively, between weeks 3 and 10 in the first season of the trial. By visually inspecting the average raw spectra, only general remarks could have been stated about changes to the major milk components' FTIR bands. This method is incapable of testing any treatment effect; hence, visualizing spectra is not an appropriate approach to detect or test a treatment effect on milk FTIR spectra.

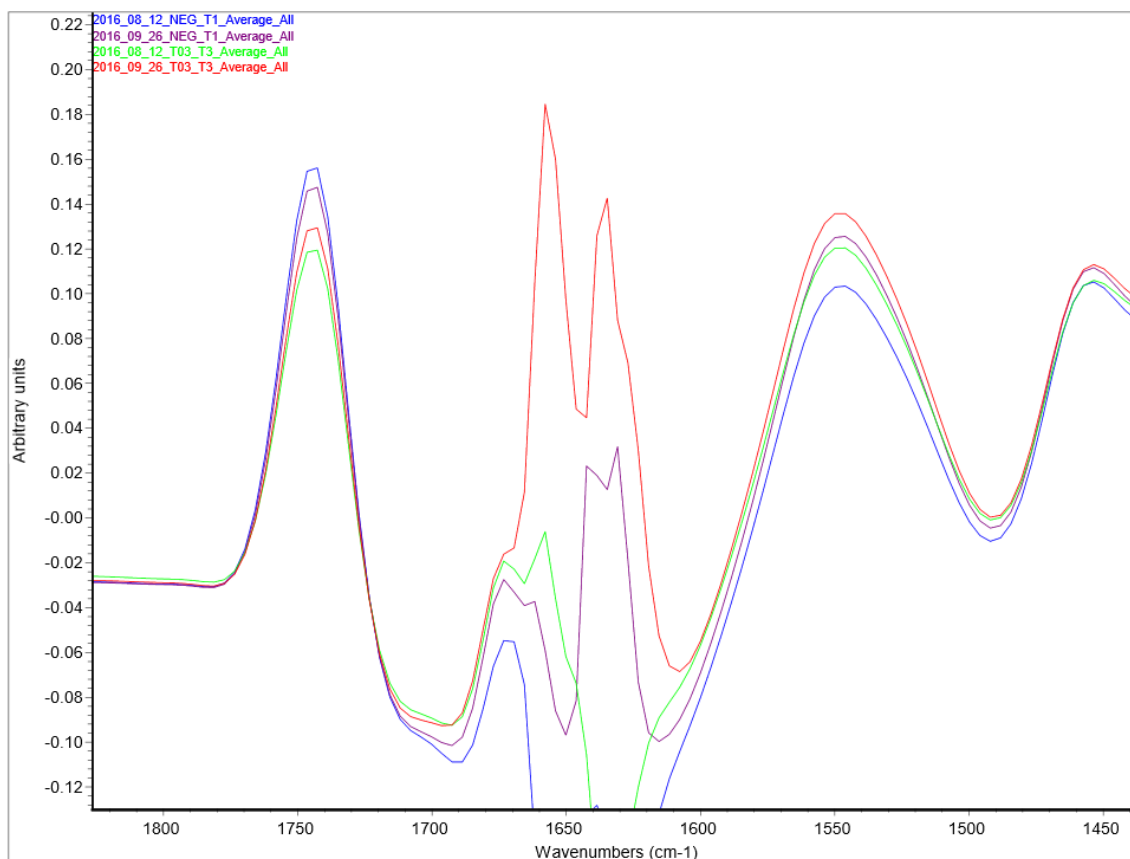


Figure 6-5 Spectral averages comparison. Blue: average spectrum of T1 week 3, purple: average spectrum of T1 week 10, red: average spectrum of T3 week 10, green: average spectrum of T3 week 3. T1 has more intense absorption band than T3 at $\sim 1745 \text{ cm}^{-1}$. In addition, the intensity of the band decreased and increased for T1 and T3, respectively, between weeks 3 and 10 in the first season of the trial.

The second approach was to apply HCA to the four versions of the milk spectral dataset (i.e. raw, FD, VN raw, VN FD). HCA is an unsupervised classification method that aims at uncovering of natural groups or clusters that are present in a dataset where observations are relatively homogenous within a cluster and different from other observations in other clusters. To achieve this goal, a specific distance measure, such as Euclidian distance, is calculated between pairs of all observations, then a linkage algorithm is used to determine the similarity between observations

and eventually their cluster membership. The result of this analysis is projected in a dendrogram, which is a tree like graph that visualizes the similarities between different clusters and observations [94]. In this case, the HCA dendrogram revealed that spectra of milk samples tend to cluster according to season. The trend was most notable in the raw spectral dataset (Figure 6-6). This observation suggests that season has the strongest effect on the spectral data, which agrees with the findings of the univariate analysis of the mixed procedure of milk components in the previous section. The tendency to cluster by season became more evident after applying the forward search feature selection algorithm to the raw spectral dataset, which aims at selecting the significant features or wavenumbers that are responsible for the main sources of variation in the spectral dataset. This algorithm found that regions $1134\text{-}1095\text{ cm}^{-1}$ and $1330\text{-}1292\text{ cm}^{-1}$ were the most significant for clustering according to season (Figure 6-7). In the dendrogram, most of the samples with the red color, which represent season 1, fell in the upper arm. On the other hand, most of the samples with the blue color, which represent season 2, fell in the lower arm. Since HCA is an unsupervised classification method, the data did not show any tendency to cluster according to any other studied factor, which renders HCA functionality limited in this context of spectral analysis.

The findings of the third approach, PCA, were consistent with those of HCA. The PCA score plot shows that samples tend to cluster according to season in the raw spectral dataset (Figure 6-8) when all spectral regions that contain information related to milk chemical composition were used (i.e., $3061\text{-}2803\text{ cm}^{-1}$, $1797\text{-}1681\text{ cm}^{-1}$ and $1612\text{-}925\text{ cm}^{-1}$). This trend became more evident when PCA was restricted to regions $1134\text{-}1095\text{ cm}^{-1}$ and $1330\text{-}1292\text{ cm}^{-1}$, which were detected by the forward search feature selection algorithm. In the first case, seven meaningful PCs, whose eigenvalue is ≥ 1 and the percentage of explained variation is $\geq 1\%$, were observed and they explained 98.40% of the variation in the raw spectral dataset. PC1, PC2 and PC3 explained 47.33%, 35.46% and 11.19% of the variation in the raw spectral dataset, respectively. In the second case, only two meaningful PCs were observed that explained 99.02% of the variation in the raw spectral dataset. PC1 and PC2 explained 76.22% and 22.80% of the variation in the raw spectral dataset, respectively. This low number of meaningful PCs suggests that the trend that have been observed in the raw spectral dataset (i.e., clustering according to time) might be an artifact resulting from variability of instruments or the application of the analytical procedure at Valacta. When contacted about this regard, Valacta confirmed that all milk samples were analyzed by the same analyzer (i.e., line E), which was frequently standardized according to the standard operational

procedure (SOP) in effect. Nevertheless, trends observed only in the raw spectral dataset should be interpreted with caution since they might be a result of procedural systematic factors that are not controlled during the trial. However, if the signal, which is detected at regions 1134-1095 cm^{-1} and 1330-1292 cm^{-1} by the forward search feature selection algorithm, is true then it can be assigned to one or more of the following IR bands: CH_3 wagging of fatty acids at $\sim 1123 \text{ cm}^{-1}$, $\text{N}-\text{H}$ wagging of the Amide III band of proteins at $\sim 1300 \text{ cm}^{-1}$, bending of symmetric bond (HCH) and of CH_2OH of carbohydrates between 1500 cm^{-1} and 1200 cm^{-1} , or C–O bond and C–C bond stretching of carbohydrates between 1200 cm^{-1} and 950 cm^{-1} [108]. No specific assignment can be made in this case due to the fact that the forward search feature selection algorithm does not provide a loading spectrum as PCA does; hence, the exact position of the significant FTIR bands will not be known. The only conclusion that can be drawn in this case is that fat, protein and lactose concentrations might have been changing over the trial period, which does not prove any treatment effect. This conclusion is consistent with the findings of the spectral visual inspection.

PLS-DA was the fourth approach that was used to analyze the milk spectral dataset. The VN FD spectral dataset resulted in a parsimonious model with 54 significant spectral features that explained 96.56% and 20.49% of the variation in the X and Y matrices, respectively (Table 6-4). The spectral dataset explains more variation related to the treatment membership (i.e., variation in the Y matrix) than the milk composition numerical dataset. The spectral and the numerical datasets explain 20.49% and 10.79% of the variation in the Y matrix, respectively. This observation confirms that the full milk FTIR spectrum contains more details related to effects of TR configuration treatments on milk composition than numbers reported by FTIR analyzers. Hence, analysis of full milk FTIR spectrum is preferred over analysis of few selected milk components that are reported in this type of studies.

The variable importance plot for the PLS-DA model for the VN FD spectral dataset reveals that the most influential spectral variables that explain the treatment membership of milk samples are 1400 cm^{-1} , 1076 cm^{-1} , 1053 cm^{-1} , 2822 cm^{-1} , 2926 cm^{-1} and 1280 cm^{-1} (Figure 6-9). It must be noted that these wavenumbers represent the band limits and not the actual peaks of the significant FTIR bands. In the case of first derivative, the band peak becomes zero and band limits become maxima and minima. In addition, the fact that the VN FD spectral dataset was more efficient in modeling the treatment membership with a smaller number of spectral variables than the VN raw

spectral dataset indicates that the differences in milk composition between treatments are more probably reflected in minor milk components. For example, the wavenumber 1400 cm^{-1} could be the limit for the 1370 cm^{-1} band of acetone in milk FTIR spectrum [109] that can be assigned to the bending vibration of the C-H bond [108]. Acetone is a biomarker related to elevated body fat mobilization [100]. This observation is consistent with the findings of the PLS-DA analysis of the milk numerical dataset which concluded that minor milk components mostly reflect the treatment effect. However, no further conclusions could have been drawn from this analysis, such as which treatment was significantly different from the others in terms of acetone concentration, which keeps the main question of this study unanswered.

None of the previous approaches could incorporate the statistical model of the experimental design of this study to test the TR treatment effect on the spectral dataset. For this reason, a univariate approach was applied to the four versions of the milk FTIR spectral dataset. It could test the different effects defined in the statistical model and provided *P* values for each tested spectral variable (Table 6-5). This approach revealed a significant season, time and block effects on multiple spectral variables, but no significant treatment effect was detected. Four spectral variables in the FD spectral dataset revealed significant interactions between time and treatment. These spectral variables were 1558 cm^{-1} , 1554 cm^{-1} , 1361 cm^{-1} and 1357 cm^{-1} . However, it was not possible to interpret these results from a spectroscopic point of view because this approach does not test the studied effects on the spectrum as a wholistic entity.

To summarize, different approaches have been evaluated to detect TR configuration effect on milk composition. None of them proved to be adequate in answering the questions of the current problem. As a result, a new approach had to be developed to test a TR treatment effect on milk composition that will address the following points:

- 1- The new approach must be a multivariate one because treatments affecting milk composition will probably affect multiple components in milk.
- 2- The analysis must be performed on the FTIR spectral data of milk because this data contains more details about milk composition than numbers reported by a commercial analyzer. The PLS-DA models that were previously discussed showed that the spectral dataset explained more variation in the Y matrix (i.e., the treatment membership) than the

numerical dataset. The spectral and the numerical datasets explained 20.49% and 10.79% of the variation in the Y matrix, respectively.

- 3- The new multivariate approach must be tweaked to accommodate powerful univariate statistical tools, such as mixed modeling, that will test the effects of different factors and that will produce indicators about the significance of these factors (e.g., *P* values) while retaining the multivariate structure of the spectral data.

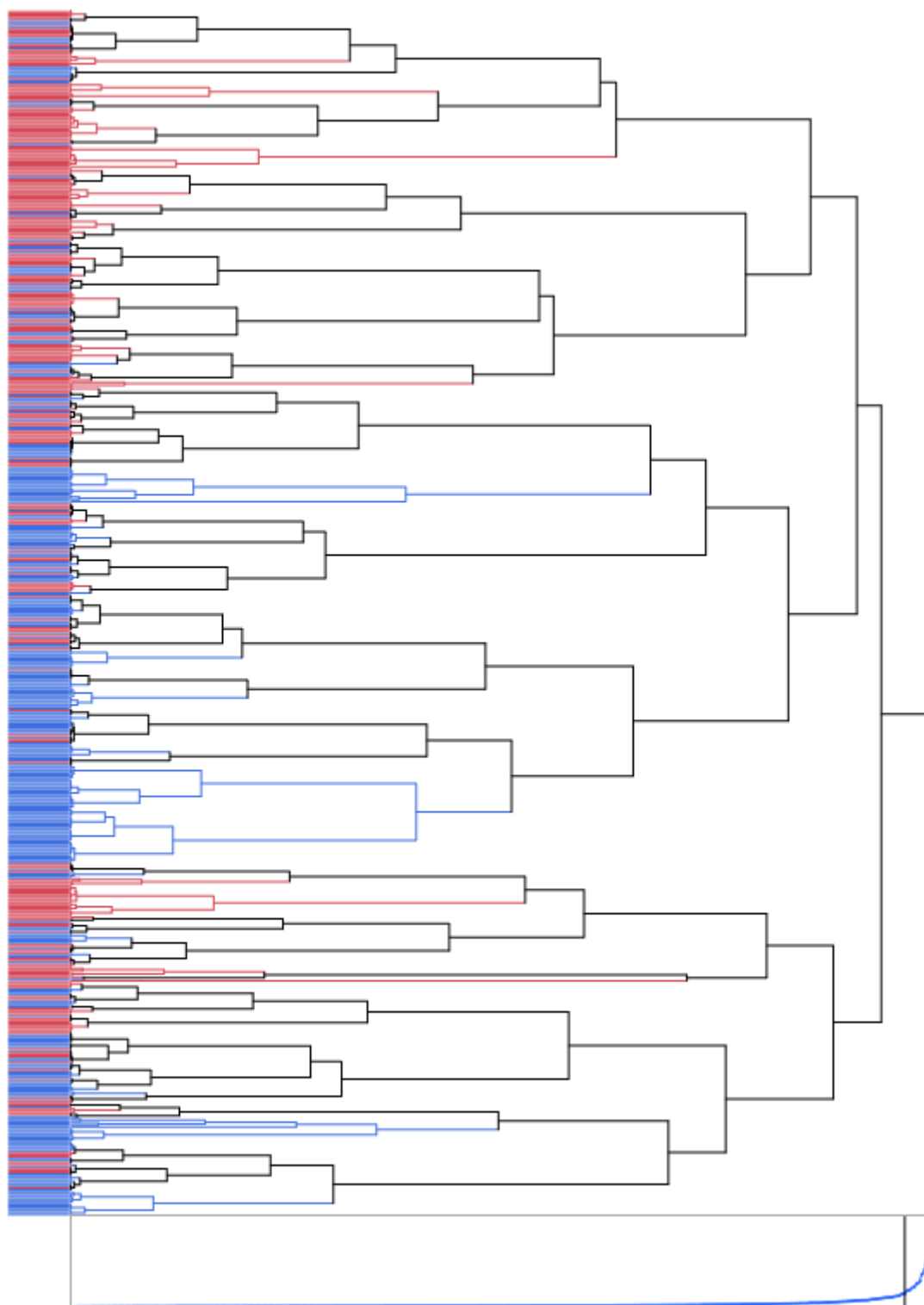


Figure 6-6 HCA dendrogram of raw spectral data. Samples tend to cluster according to season red is season 1, blue is season 2.

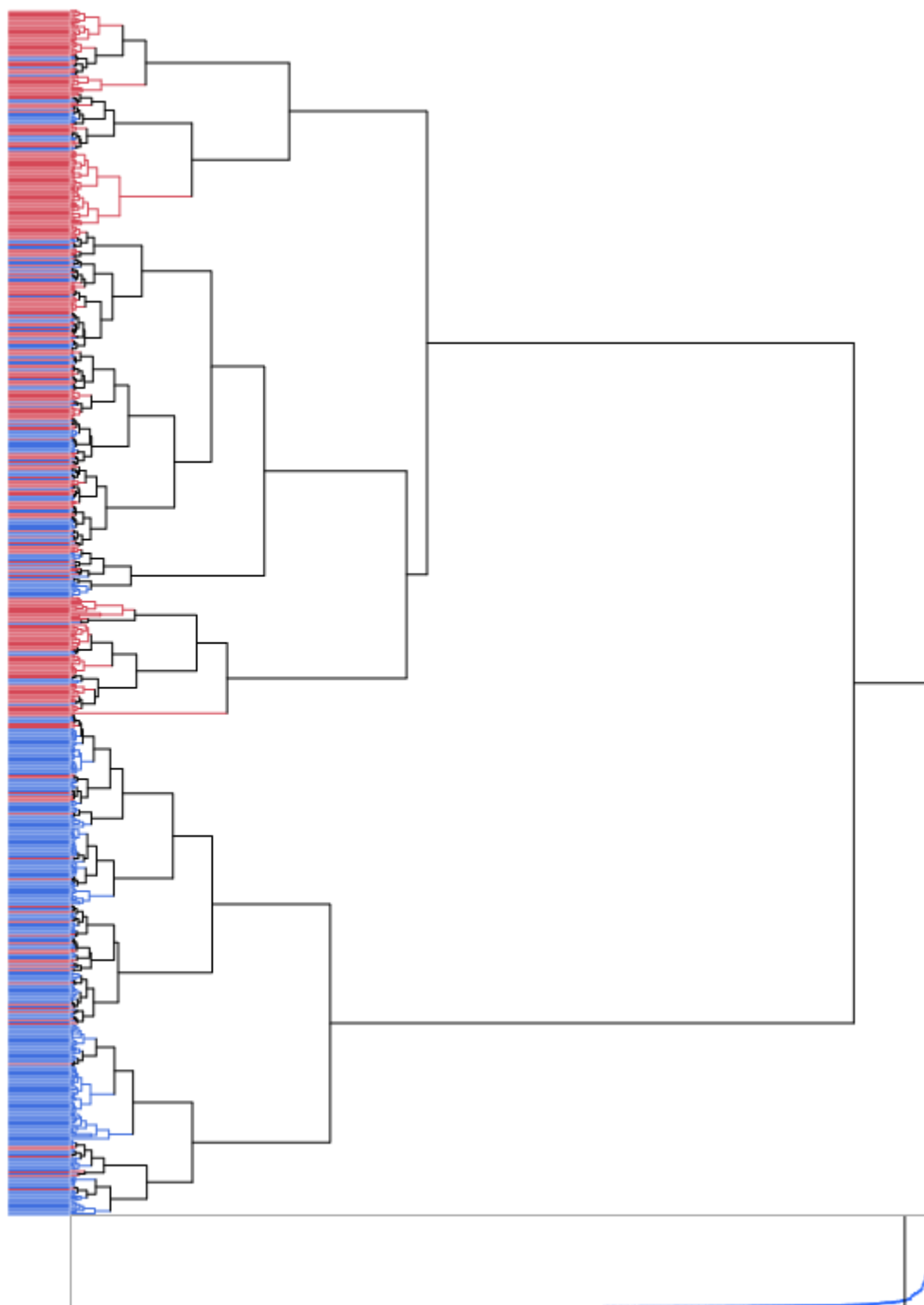


Figure 6-7 HCA dendrogram of raw spectral data after forward feature selection. The spectral regions that are used are 1134-1095 cm^{-1} and 1330-1292 cm^{-1} . Clustering according to season becomes more evident. Upper arm is season 1 (red), lower arm is season 2 (blue).

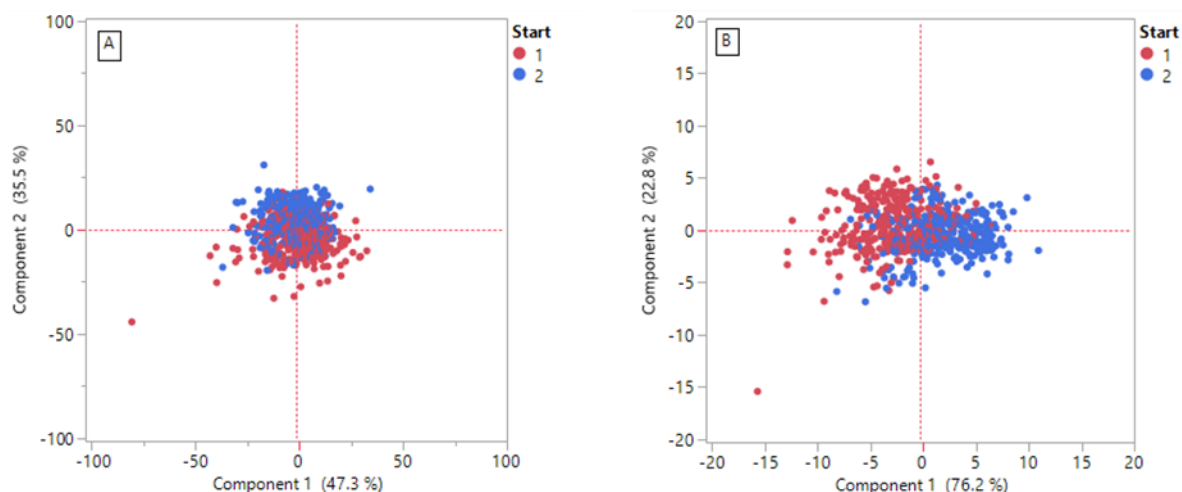


Figure 6-8 PCA scores of the raw spectral dataset. Samples tend to cluster according to season. A: the spectral regions used are $3061\text{--}2803\text{ cm}^{-1}$, $1797\text{--}1681\text{ cm}^{-1}$, $1612\text{--}925\text{ cm}^{-1}$, B: the separation became more evident when PCA was restricted to spectral regions detected by forward search feature selection algorithm $1134\text{--}1095\text{ cm}^{-1}$ and $1330\text{--}1292\text{ cm}^{-1}$.

Table 6-4 Model comparison summary of PLS-DA models developed for the four spectral dataset of milk samples. The vector normalized first derivative spectral dataset resulted in the most parsimonious model.

Spectral Dataset	Model	Number of factors	Percent variation explained for X matrix	Percent variation explained for Y matrix	Number of VIP>0.8
Raw	Initial	14	99.48	21.24	179
	Parsimonious	14	99.63	20.88	118
FD	Initial	14	94.05	20.96	155
	Parsimonious	11	96.78	21.13	83
VN Raw	Initial	13	99.04	21.87	148
	Parsimonious	12	99.35	19.83	94
VN FD	Initial	13	92.03	22.10	150
	Parsimonious	10	96.56	20.49	54

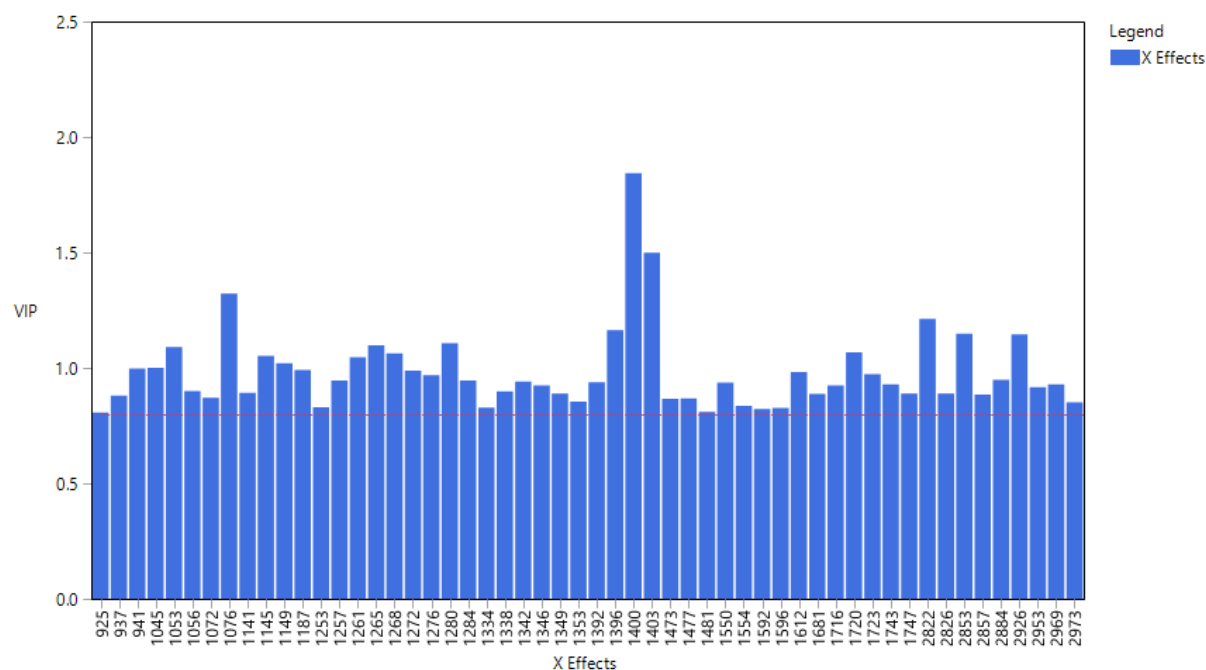


Figure 6-9 Variable importance plot for the PLS-DA model for the vector normalized first derivative spectra dataset. The most influential variables that explain the treatment membership for milk samples are centered at $\sim 1400\text{ cm}^{-1}$, 1076 cm^{-1} , 1053 cm^{-1} , 2822 cm^{-1} , 2926 cm^{-1} and 1280 cm^{-1} . It must be noted that these wavenumbers represent the band limits in the case of first derivative and not the actual peaks of the IR bands.

Table 6-5 P values obtained from the mixed procedure that was applied to individual spectral variables in the raw, FD, VN raw and VN FD spectral datasets of milk.

		Treatment	Start	Block (Start)	Week	Treatment*Week
Raw	Min. P value	0.2082	< 0.0001	0.0001	< 0.0001	0.4267
	Max. P value	0.9730	0.9849	0.7882	0.6444	0.9994
FD	Min. P value	0.0797	< 0.0001	0.0001	< 0.0001	0.0231
	Max. P value	0.9995	0.9868	0.9010	0.8877	0.9993
VN Raw	Min P value	0.1533	0.0004	< 0.0001	< 0.0001	0.2901
	Max P value	0.9955	0.9686	0.7298	0.6991	0.9831
VN FD	Min P value	0.0829	< 0.0001	< 0.0001	< 0.0001	0.0528
	Max P value	0.9965	0.9979	0.9983	0.7778	0.9937

6.3.3 Hybrid approach for milk spectral dataset analysis

Before applying the hybrid approach to the TR trial's spectral data, milk samples were spiked with lactose at two levels, 5% and 1%, to test the feasibility of this approach in detecting a treatment's effect on milk composition FTIR spectral data.

PCA was applied to the spectral regions $3061\text{--}2803\text{ cm}^{-1}$, $1797\text{--}1681\text{ cm}^{-1}$ and $1612\text{--}925\text{ cm}^{-1}$ of the spiked milk samples. The VN raw and VN FD spectral datasets produced 6 and 9 meaningful PCs (i.e., eigenvalue ≥ 1 and percent of explained variation $\geq 1\%$) that explained 94.91% and 95.91% of the variation in the dataset, respectively (Table 6-6). For the VN raw spectral data, PC1 and PC2 explained 79.15% and 9.10% of the variation, respectively. For the VN FD spectral data, PC1 and PC2 explained 73.80% and 8.83% of the variation, respectively.

First, the effect of the spiked lactose concentration was tested by the mixed model that was applied to the meaningful PCs that were isolated from each spectral dataset. The spiked lactose concentration, which is considered as the treatment, showed significant effect on PC1 and PC2 in VN raw and VN FD spectral datasets, respectively. This step provided a clear answer to one of the questions in this problem, which is “Does the treatment have a significant effect on the FTIR spectral data of milk composition?”. Another question that needed to be answered is “Which treatment level did significantly change milk components' concentrations?”. The answer to this question lies in the estimates of the least squares means of the significant PC scores for each treatment level that are produced by the mixed model. Table 6-7 shows that these estimates increase with increased lactose concentration in the VN raw spectral dataset, on the other hand, they decrease with increased lactose concentration in the VN FD spectral dataset. The same experiment was repeated with three levels of lactose in aqueous solutions, 1% 2% and 5%, and the same observation was revealed. In this case, the increased lactose concentration will linearly increase the intensities of the spectral variables between 1200 cm^{-1} and 950 cm^{-1} , which will produce PCA scores with increasing and decreasing values in the VN raw and VN FD spectral dataset, respectively. In addition, the differences of the least squares means will reveal the significant differences among different treatment levels (Table 6-8).

Table 6-6 Meaningful principal components isolated from the VN raw and VN FD spectral datasets of milk samples spiked with two levels of lactose. The spiked lactose concentration yielded a significant effect on PC1 and PC2 for VN raw and VN FD spectral datasets, respectively, when tested with the mixed procedure.

PC	Eigenvalue		Explained Variation %		P value	
	VN Raw	VN FD	VN Raw	VN FD	VN Raw	VN FD
PC1	111.6042	104.0581	79.152	73.800	<0.0001*	0.1367
PC2	12.8309	12.4435	9.100	8.825	0.8326	<0.0001*
PC3	3.3335	3.8386	2.364	2.722	0.9660	0.9894
PC4	2.3771	3.5680	1.686	2.531	0.9825	0.9928
PC5	1.9246	3.1382	1.365	2.226	0.9960	0.7790
PC6	1.7584	2.8542	1.247	2.024	0.9932	0.8690
PC7	-	2.1694	-	1.539	-	0.7830
PC8	-	1.7054	-	1.209	-	0.9790
PC9	-	1.4517	-	1.030	-	0.7728

Table 6-7 Estimates of the least squares means of the significant PC scores produced by the mixed model for lactose spiked milk samples and aqueous solutions for the VN raw and VN FD spectral datasets. These means increase and decrease with increased lactose concentration in the VN raw and VN FD spectral datasets, respectively.

Sample	Spectral dataset	Significant PC	Category	Estimate	Standard Error	P Value
Milk	VN Raw	PC1	High	12.3584	0.3117	<.0001
			Low	-8.5558	0.2593	<.0001
	VN FD	PC2	High	-3.7315	0.5230	<.0001
			Low	2.5833	0.4352	<.0001
Water	VN Raw	PC1	High	14.5585	0.09661	<.0001
			Mid	-4.7313	0.09661	<.0001
			Low	-9.8272	0.09661	<.0001
	VN FD	PC1	High	-13.4298	0.1958	<.0001
			Mid	4.4075	0.1958	<.0001
			Low	9.0223	0.1958	<.0001

Table 6-8 Differences of least squares means of PC scores produced by the mixed model for lactose spiked milk samples and aqueous solutions for the VN raw and VN FD spectral datasets. Significant differences among treatment levels are revealed by *P* values.

Sample	Spectral dataset	Significant PC	Category	Estimate	Standard Error	<i>P</i> Value
Milk	VN Raw	PC1	High vs. Low	20.9142	0.4055	<.0001
	VN FD	PC2	High vs. Low	-6.3148	0.6804	<.0001
Water	VN Raw	PC1	High vs. Low	24.3857	0.1366	<.0001
			High vs. Mid	19.2899	0.1366	<.0001
			Low vs. Mid	-5.0958	0.1366	<.0001
	VN FD	PC1	High vs. Low	-22.4521	0.2769	<.0001
			High vs. Mid	-17.8373	0.2769	<.0001
			Low vs. Mid	4.6148	0.2769	<.0001

Second, differences in the behaviour of the PCA scores had to be understood in cases of raw and FD spectral data. Inspection of the PCA score plots (Figure 6-10) reveals that samples with high lactose concentration load positively on the PC that was extracted from the VN raw spectral dataset and that showed significant treatment effect (i.e., PC1). On the other hand, the PCA scores of samples with high lactose concentrations load negatively on the PC that was extracted from the VN FD spectral dataset and that showed significant treatment effect (i.e., PC2). This observation proves that PCA scores of the FD spectral data behave in an opposite way to PCA scores obtained from raw spectral data.

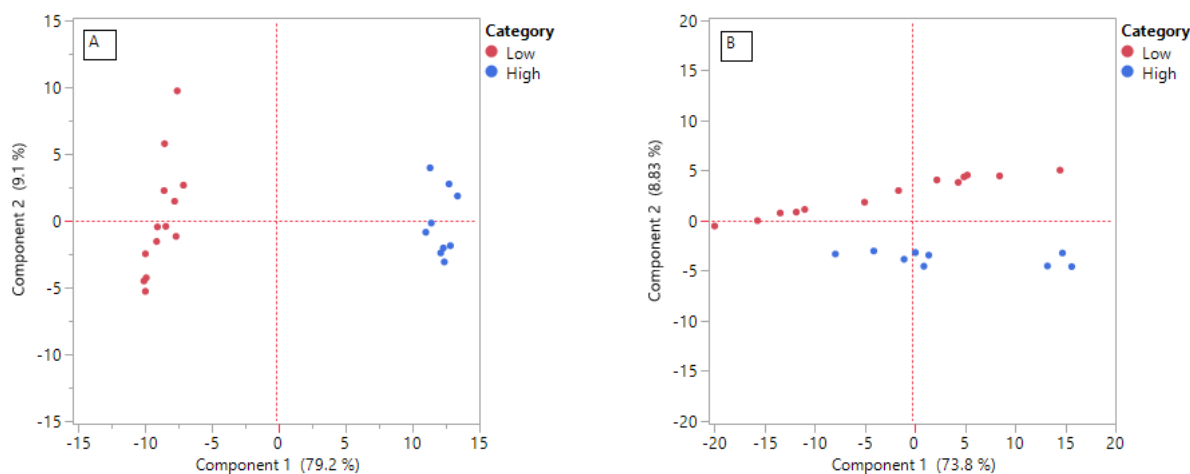


Figure 6-10 PCA scores of milk samples spiked with two levels of lactose. A: samples spiked with high lactose concentration are loading positively on PC1 that showed significant effect for spike level in the VN raw spectral dataset. B: samples spiked with high lactose concentration are loading negatively on PC2 that showed significant effect for spike level in the VN FD spectral dataset.

The third question that this approach answers is “What are the spectral variables that were affected by changing milk components’ concentrations?”. This question can be answered by inspecting the loading spectrum of the PC that showed a treatment significant effect. The loadings represent the weights that linearly combines the spectral variables to calculate the scores for every sample for this specific PC, in other words, they represent the contribution of each spectral variable to this specific PC scores. In case the loading spectrum was obtained from raw spectral dataset, then it can be interpreted directly and conclusion about influential spectral variables can be drawn. For example, the loading spectrum of PC1, which was extracted from VN raw spectral dataset and that showed significant treatment effect, positively correlates with wavenumbers that had increased spectral intensities as a result of lactose spiking (Figure 6-11). These wavenumbers are located between 1466 cm^{-1} and 989 cm^{-1} . The strongest correlation was observed between 1175 cm^{-1} and 989 cm^{-1} , which is dominated by lactose spectral contribution in milk FTIR spectrum. On the other hand, this loading spectrum negatively correlates with spectral variables whose intensities decreased with lactose spiking, especially in the spectral region from 1582 cm^{-1} to 1511 cm^{-1} , which is dominated by milk protein spectral contribution in milk FTIR spectrum.

In case the loading spectrum was obtained from FD spectral dataset (i.e., PC2), it cannot be interpreted directly because the maximum of an FTIR band peak becomes zero when FD is calculated (Figure 6-12). Hence, the spectral integral of the loading spectrum obtained from FD spectral dataset must be calculated before interpretation to restore back the influential spectral

peaks in the loading spectrum. The cumulative trapezoidal numerical integration function in MATLAB was used to calculate the spectral integral of the loading spectrum obtained from FD spectral dataset. The spectral integral reveals clear positive correlations with spectral variables from 1200 cm^{-1} to 950 cm^{-1} which are dominated by lactose spectral contribution (Figure 6-13). It must be noted that when calculating the FD, the baseline is lost, and integration will never restore it back. This issue is not problematic because the loading spectrum is only used to detect increased or decreased correlations with wavenumbers and to locate peaks that corresponds to influential spectral variables for the PC in question. The loading spectrum will never be used for quantitative determinations. The integration process of the loading spectrum was followed by peak fitting to determine the spectral variables that represent the centers of influential FTIR bands. This process located the most prominent peak in the positively correlating portion of the integrated PC2 loading spectrum at 1086 cm^{-1} (Table 6-9), which had the greatest peak height and peak area. The band centered around 1086 cm^{-1} has been reported in the literature as the band with the most intense absorption for D-(+)-lactose [110]. In addition, all detected peaks, 1086 cm^{-1} and 1139 cm^{-1} , fall within the region of lactose absorption from 1200 cm^{-1} to 950 cm^{-1} .

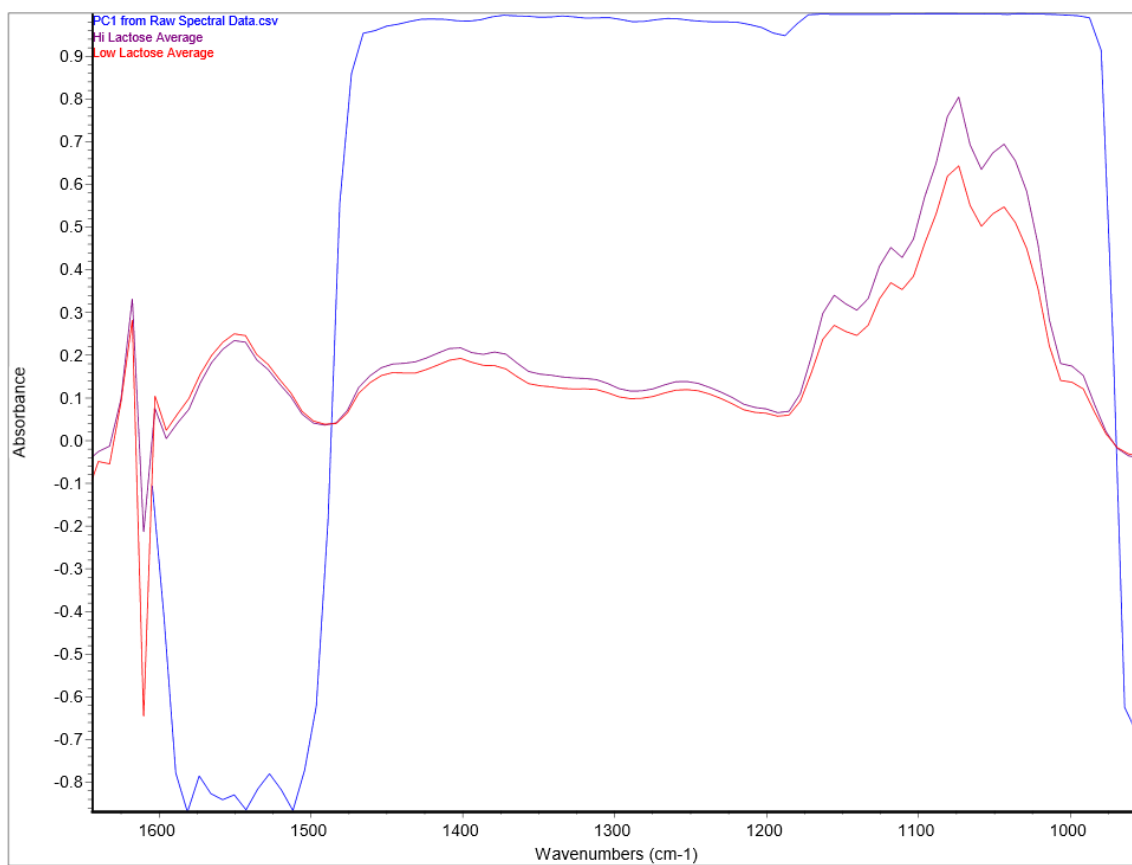


Figure 6-11 Comparison between FTIR spectra of milk samples spiked with lactose and the principal component extracted from raw spectra that revealed significant effect for lactose concentration. Red: FTIR spectrum of milk sample spiked with 1% lactose, purple: FTIR spectrum of milk spectrum spiked with 5% lactose, blue: loading spectrum of PC1, which is the principle component that showed significant treatment effect (i.e., lactose spiking). Spectral variables with increased spectral intensities correlate positively with PC1 and vice versa.

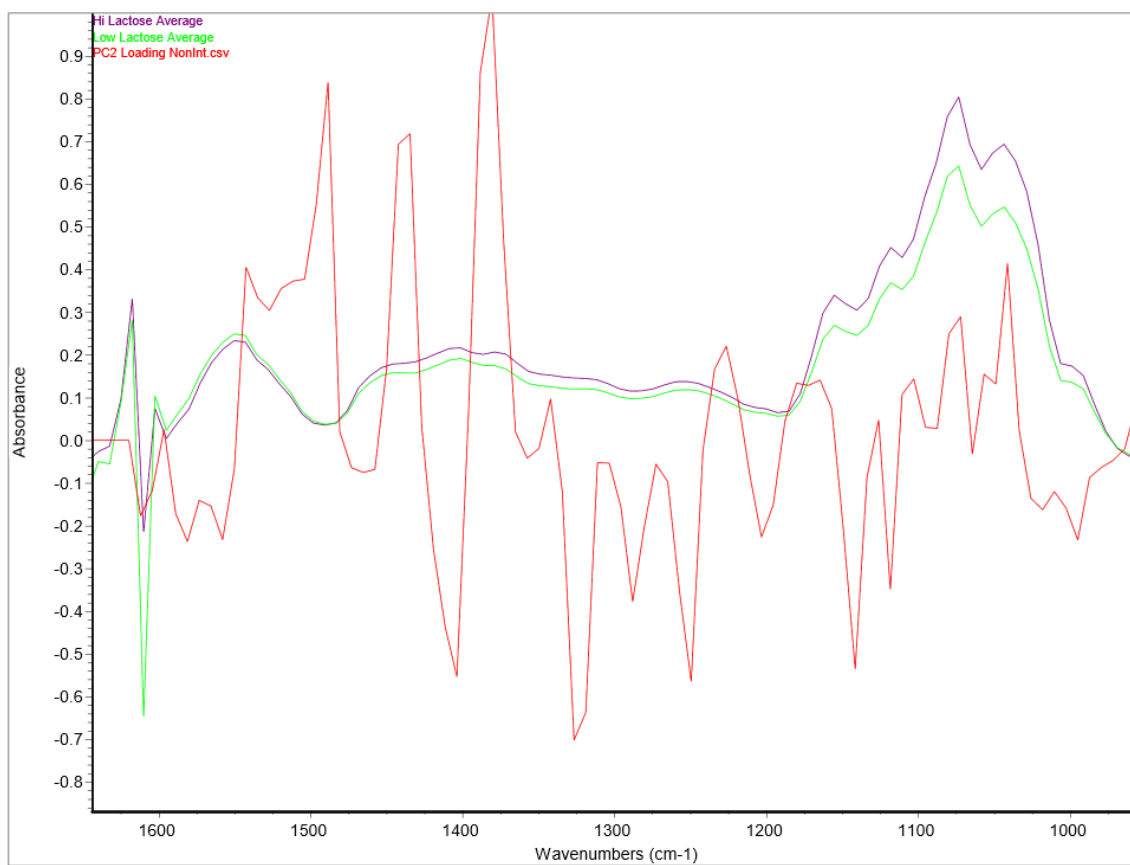


Figure 6-12 Comparison between FTIR spectra of milk samples spiked with lactose and the principal component extracted from first derivative spectra that revealed significant effect for lactose concentration. Green: FTIR spectrum of milk sample spiked with 1% lactose, purple: FTIR spectrum of milk spectrum spiked with 5% lactose, red: loading spectrum of PC2 isolated from VN FD spectral dataset of milk samples spiked with lactose before integration. No conclusions regarding influential spectral variables can be drawn in this case.

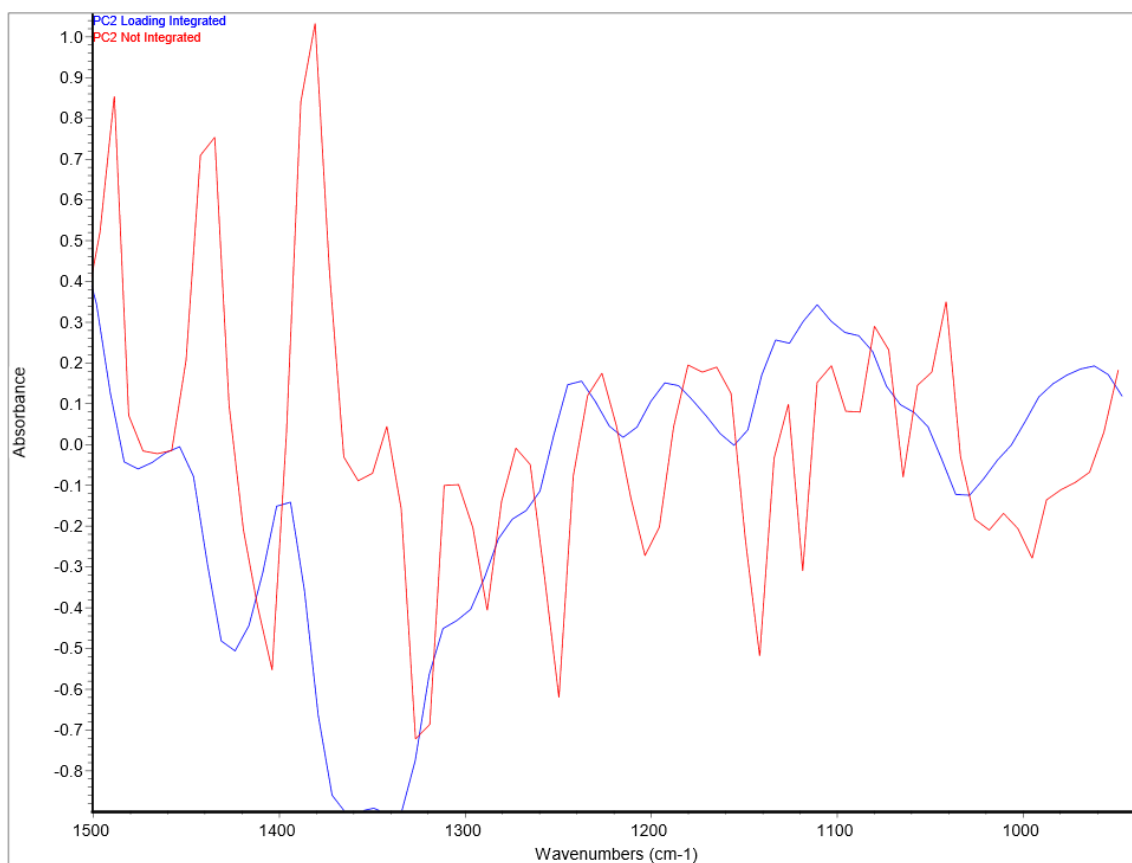


Figure 6-13 Comparison of loading spectrum of PC2 obtained from VN FD spectral dataset of milk samples spiked with different levels of lactose (red) and the spectral integral of loading spectrum of PC2 (blue). The spectral integral shows clear positive correlation with spectral variables from 1200 cm^{-1} to 950 cm^{-1} which are dominated by lactose spectral contribution.

Table 6-9 Peak fitting results for PC2 integrated loading spectrum that was extracted from VN FD spectral dataset of milk samples spiked with lactose. The most prominent peak is at 1086 cm^{-1} , which is very close to lactose peak at 1076 cm^{-1} .

Peak #	Peak Type	Center X	Height	FWHH	Area
1	Voigt	1039.88	0	23.4975	0
2	Voigt	1086.787	4.7302	23.0564	271.6144
3	Voigt	1113.078	0	28.4228	0
4	Voigt	1139.808	0.9093	25.6215	50.244
5	Voigt	1184.314	0	28.0263	0
6	Voigt	1238.561	0	26.3452	0

The final question that this approach answers is “What molecules in milk can be assigned to the detected peaks in the loading spectrum?”. To answer this question, several approaches will be employed. First, milk samples will be spiked with minor milk components, such as urea, citrate and acetone, to detect their peaks in milk FTIR spectrum. Second, the literature will be reviewed for any milk FTIR peaks that will be detected by the hybrid approach and that will not be assigned to any molecule by the spiking experiments. Third, FTIR reference spectra of minor milk components will be reviewed for possible candidate molecules that can be assigned to any unknown peaks in milk FTIR spectrum. These reference spectra will be downloaded from the National Institute of Standards and Technology (NIST) [111]. It must be noted that peaks extracted from a loading spectrum will not be treated as band centers because the hybrid approach approximate the influential FTIR peaks in a loading spectrum that are related to a treatment effect. If they fall within FTIR band of a specific bond of a molecule then they will be assigned to that molecule. The detected peaks in the loading spectrum need not be an exact match of FTIR band center of a specific bond in a molecule in milk.

When the hybrid approach was applied to the spectral datasets of the TR trial, PCA yielded five, ten, six and twelve meaningful PCs from the raw, FD, VN raw and VN FD spectral datasets that explained 97.13%, 93.72%, 97.07% and 93.97% of the variation in the spectral dataset, respectively (Table 6-10). Only PC5 isolated from the VN raw spectral dataset revealed a significant treatment effect ($P = 0.0478$). The loading spectrum of this PC shows strong correlations at 2833 cm^{-1} , 1775 cm^{-1} , 1720 cm^{-1} and 1206 cm^{-1} (Figure 6-14). These wavenumbers fall on the limits of the FTIR bands of Fat B, Fat A, carboxylic group and Amide III bands in the FTIR spectra of the trial’s milk samples, which might imply that the treatment influenced fat, fatty acids and protein levels in the collected milk samples. However, PC5 also reveals significant effects for the season ($P = 0.0481$), the block ($P = 0.0030$) and the week ($P < 0.0001$) effects. In addition, the least squares means output of the mixed procedure shows significant interactions between some weeks and treatments (Table 6-11). These observations undermine the significance of the treatment effect; hence, the loading spectrum will not be interpreted any further.

To summarize, the hybrid approach was successfully applied to the FTIR spectral data of milk samples collected during the TR trial, which retained the multivariate structure of the FTIR spectral data and allowed the utilization of the mixed model as a powerful tool to test the treatment effect on the FTIR spectral data of collected milk samples according to the statistical model that was defined by the experimental design of the trial with a repeated measurement structure that included fixed and random effects. In addition, the hybrid approach highlighted the need to alter the statistical model in order to eliminate the strong week effect, which might overshadow any treatment effect. In the next chapter, the statistical model will be altered to eliminate the week effect and the hybrid approach, that was developed in this chapter, will be applied to the FTIR data of the TR trial.

Table 6-10 The hybrid analysis approach applied to meaningful PCs extracted from raw, FD, VN raw and VN FD spectral datasets. PC5 extracted from VN raw dataset reveals significant treatment effect ($P = 0.0478$) in addition to a strong significant week effect ($P < 0.0001$).

Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values				
					Treatment	Start	Block (Start)	Week	Treatment*Week
Raw	PC1	131.58	47.33	47.33	0.3650	0.4918	0.1384	0.0129	0.5465
	PC2	98.57	35.46	82.79	0.7025	<0.0001	0.0063	<0.0001	0.9963
	PC3	31.11	11.19	93.98	0.9527	0.0003	0.0001	<0.0001	0.7137
	PC4	5.24	1.89	95.86	0.5947	0.4105	0.3527	<0.0001	0.5595
	PC5	3.53	1.27	97.13	0.1990	0.1050	0.1891	<0.0001	0.2675
FD	PC1	147.73	53.14	53.14	0.3516	0.3079	0.0778	0.0006	0.4919
	PC2	43.03	15.48	68.62	0.9454	0.0077	0.0012	<0.0001	0.9664
	PC3	29.38	10.57	79.18	0.5440	0.0002	0.0104	0.0008	0.1017
	PC4	12.94	4.66	83.84	0.1885	0.0172	0.0212	0.0421	0.6272
	PC5	6.60	2.37	86.21	0.9179	0.0565	0.1188	<0.0001	0.6114
	PC6	5.84	2.10	88.31	0.4658	0.1925	0.0411	<0.0001	0.6197
	PC7	4.55	1.64	89.95	0.7516	0.0003	0.0002	<0.0001	0.7027
	PC8	4.03	1.45	91.40	0.5454	0.5727	0.0614	<0.0001	0.9560
	PC9	3.38	1.22	92.62	0.9286	0.0020	0.6008	<0.0001	0.5458
	PC10	3.07	1.11	93.72	0.8729	0.3596	0.0344	<0.0001	0.5593
VN Raw	PC1	188.05	67.64	67.64	0.4479	0.5917	0.2798	0.2537	0.6130
	PC2	55.76	20.06	87.70	0.9279	0.7324	<0.0001	<0.0001	0.9827
	PC3	14.19	5.11	92.80	0.5406	0.0009	0.0445	<0.0001	0.8577
	PC4	5.72	2.06	94.86	0.1824	0.1428	0.2127	<0.0001	0.5590
	PC5	3.54	1.27	96.14	0.0478	0.0481	0.0030	<0.0001	0.6112
	PC6	2.59	0.93	97.07	0.2151	0.0121	0.0316	<0.0001	0.6463
VN FD	PC1	158.87	57.15	57.15	0.4527	0.8100	0.3782	0.4379	0.5152
	PC2	45.17	16.25	73.40	0.7708	0.8727	0.0002	<0.0001	0.3095
	PC3	13.96	5.02	78.42	0.1848	0.3740	0.0326	0.0019	0.5701
	PC4	8.89	3.20	81.62	0.9806	0.0002	0.0897	<0.0001	0.5731
	PC5	6.41	2.31	83.93	0.5798	0.0007	0.0362	<0.0001	0.5042
	PC6	5.00	1.80	85.73	0.8833	0.0016	0.0041	<0.0001	0.8177
	PC7	4.61	1.66	87.38	0.1836	0.1060	0.0037	<0.0001	0.6829
	PC8	4.36	1.57	88.95	0.8591	0.0044	0.0032	<0.0001	0.7012
	PC9	4.06	1.46	90.41	0.3142	0.1344	0.0915	0.0001	0.7577
	PC10	3.73	1.34	91.75	0.3109	0.0474	0.0137	<0.0001	0.7254
	PC11	3.26	1.17	92.92	0.4412	0.6319	0.0768	<0.0001	0.4111
	PC12	2.90	1.04	93.97	0.1575	0.7255	0.1740	0.0009	0.1174

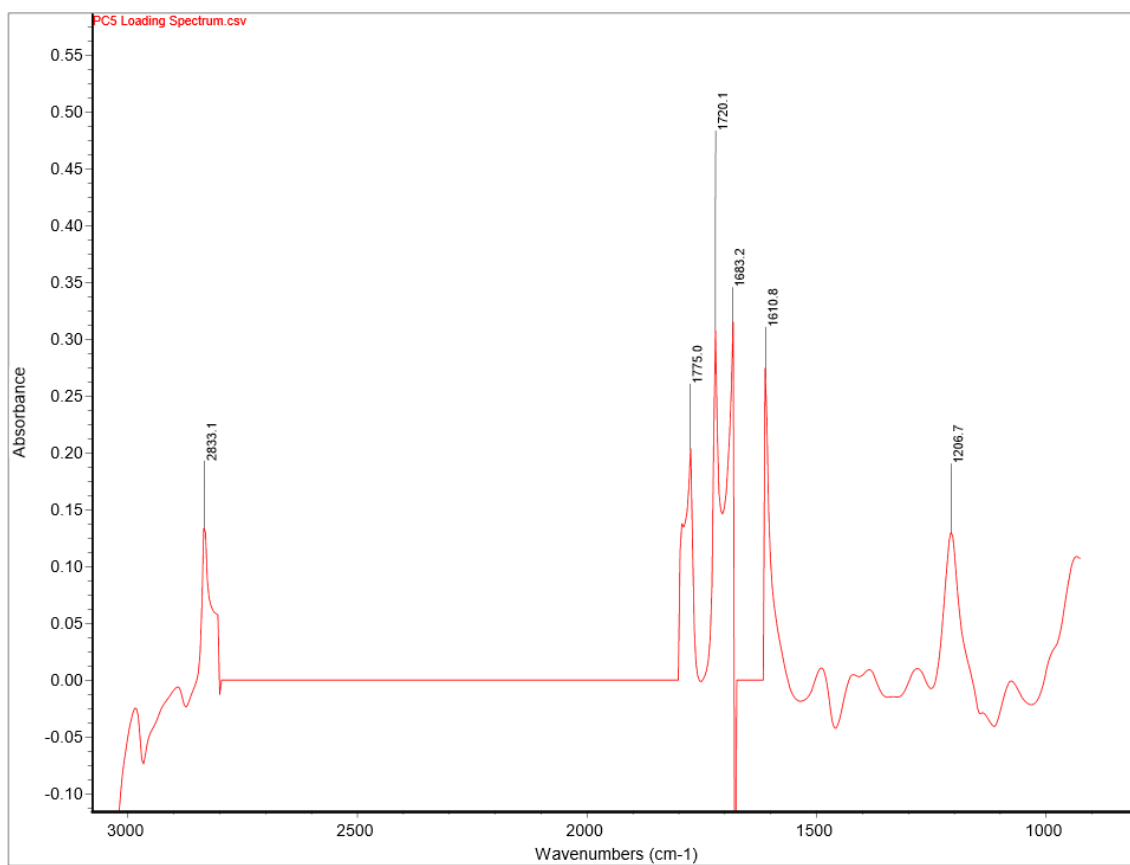


Figure 6-14 The loading spectrum of PC5 that was extracted from the VN raw spectral dataset of the TR trial. This PC reveals significant treatment effect ($P = 0.0478$) and it shows strong correlations at 2833 cm^{-1} , 1775 cm^{-1} , 1720 cm^{-1} and 1206 cm^{-1}

Table 6-11 The least squares means output of the mixed procedure that was applied to PC5 of the VN raw spectral dataset of the TR trial. It reveals significant interactions between some weeks and treatments.

Effect	Treatment	Week	Estimate	STD Error	DF	t Value	P Value
Treatment	1		0.234	0.1661	41.8	1.41	0.1662
Treatment	2		0.3784	0.1757	41.5	2.15	0.0371
Treatment	3		0.0614	0.1562	42.6	0.39	0.6962
Treatment	4		-0.267	0.1562	42.6	-1.71	0.0946
Week		1	-0.7108	0.1561	331	-4.55	<.0001
Week		2	0.2929	0.1567	336	1.87	0.0624
Week		3	0.212	0.2648	564	0.8	0.4238
Week		4	0.8318	0.2648	564	3.14	0.0018
Week		5	-0.5412	0.2648	564	-2.04	0.0415
Week		6	0.9179	0.2621	563	3.5	0.0005
Week		7	-0.00878	0.2648	564	-0.03	0.9735
Week		8	0.1348	0.2648	564	0.51	0.6109
Week		9	0.4452	0.2648	564	1.68	0.0932
Week		10	-0.5568	0.2648	564	-2.1	0.0359
Treatment*Week	1	1	-0.9092	0.3333	354	-2.73	0.0067
Treatment*Week	1	2	0.2851	0.3277	344	0.87	0.385
Treatment*Week	1	3	0.5123	0.5339	565	0.96	0.3377
Treatment*Week	1	4	0.9947	0.5339	565	1.86	0.063
Treatment*Week	1	5	-0.1088	0.5339	565	-0.2	0.8386
Treatment*Week	1	6	1.6543	0.512	557	3.23	0.0013
Treatment*Week	1	7	0.1829	0.5339	565	0.34	0.7321
Treatment*Week	1	8	-0.3702	0.5339	565	-0.69	0.4883
Treatment*Week	1	9	0.5806	0.5339	565	1.09	0.2773
Treatment*Week	1	10	-0.4819	0.5339	565	-0.9	0.3671
Treatment*Week	2	1	-0.5497	0.318	304	-1.73	0.0849
Treatment*Week	2	2	0.1528	0.3278	329	0.47	0.6414
Treatment*Week	2	3	0.8766	0.5605	564	1.56	0.1184
Treatment*Week	2	4	1.046	0.5605	564	1.87	0.0625
Treatment*Week	2	5	0.1597	0.5605	564	0.28	0.7758
Treatment*Week	2	6	1.0781	0.5605	564	1.92	0.0549
Treatment*Week	2	7	-0.1244	0.5605	564	-0.22	0.8245
Treatment*Week	2	8	1.0727	0.5605	564	1.91	0.0562
Treatment*Week	2	9	0.5122	0.5605	564	0.91	0.3612
Treatment*Week	2	10	-0.4395	0.5605	564	-0.78	0.4333
Treatment*Week	3	1	-0.7019	0.2969	347	-2.36	0.0186
Treatment*Week	3	2	0.1299	0.2969	347	0.44	0.6619
Treatment*Week	3	3	0.331	0.5103	567	0.65	0.5168
Treatment*Week	3	4	0.4616	0.5103	567	0.9	0.3661
Treatment*Week	3	5	-1.3314	0.5103	567	-2.61	0.0093
Treatment*Week	3	6	0.614	0.5103	567	1.2	0.2294
Treatment*Week	3	7	0.147	0.5103	567	0.29	0.7735
Treatment*Week	3	8	0.08745	0.5103	567	0.17	0.864
Treatment*Week	3	9	0.8144	0.5103	567	1.6	0.1111
Treatment*Week	3	10	0.06197	0.5103	567	0.12	0.9034
Treatment*Week	4	1	-0.6825	0.2969	347	-2.3	0.0221
Treatment*Week	4	2	0.604	0.2969	347	2.03	0.0427
Treatment*Week	4	3	-0.8721	0.5103	567	-1.71	0.088
Treatment*Week	4	4	0.825	0.5103	567	1.62	0.1065
Treatment*Week	4	5	-0.8841	0.5103	567	-1.73	0.0837
Treatment*Week	4	6	0.325	0.5103	567	0.64	0.5245
Treatment*Week	4	7	-0.2406	0.5103	567	-0.47	0.6375
Treatment*Week	4	8	-0.2507	0.5103	567	-0.49	0.6234
Treatment*Week	4	9	-0.1262	0.5103	567	-0.25	0.8048
Treatment*Week	4	10	-1.3679	0.5103	567	-2.68	0.0076

6.4 Conclusion

Several approaches have been evaluated to detect the effect of TR configuration on milk composition. The numerical milk composition data did not reveal significant treatment effect when analyzed by the mixed procedure, which is a univariate hypothesis testing approach. On the other hand, PCA revealed that the predictions of the minor milk components might have been biased by the spectral contributions of the major milk components and it did not provide any insight on the treatment effect on milk composition. Only PLS-DA provided such insight; however, the conclusions drawn from this analysis were undermined by the fact that it does not take into account the repeated measurement structure of this study and the effects other than the treatment (e.g., block, week, season) that are included in the statistical model under which the data was generated. In addition, this analysis was limited to milk components that were reported by the FTIR milk analyzer, which did not include all important biomarkers in dairy cows, such as citrate, acetone and acetoacetate.

Unsupervised analysis of the full milk FTIR spectrum revealed only the strongest effect on milk composition, which was the time effect. Techniques, such as HCA and PCA, could not test or reveal the TR treatment effect in the presence of an overwhelming effect, such as the week or the season effect. On the other hand, supervised analysis of the full milk FTIR spectrum by PLS-DA revealed more details pertaining to the TR treatment effect on milk composition than PLS-DA analysis of the numerical dataset of milk composition. However, PLS-DA was not capable of answering the main questions of the current problem.

A new hybrid approach was developed for assessing a treatment effect on FTIR milk spectral data that could answer all the questions related to the current problem. These questions are:

- Does a treatment have a significant effect on the FTIR spectral data of milk composition?
- Which treatment level did significantly change milk components' concentrations?
- What are the spectral variables that were affected by changing milk components' concentrations?
- What molecules in milk can be assigned to FTIR peaks that are significantly affected by the treatment?

This hybrid approach was successfully applied to the FTIR spectral data of milk samples collected during the TR trial, which retained the multivariate structure of the FTIR spectral data, while at the same time, accommodated the utilization of the mixed model as a powerful tool to test fixed and random effects on the FTIR spectral data of collected milk samples according to the statistical model that was defined by the experimental design of the trial with a repeated measurement structure and enabled hypothesis testing.

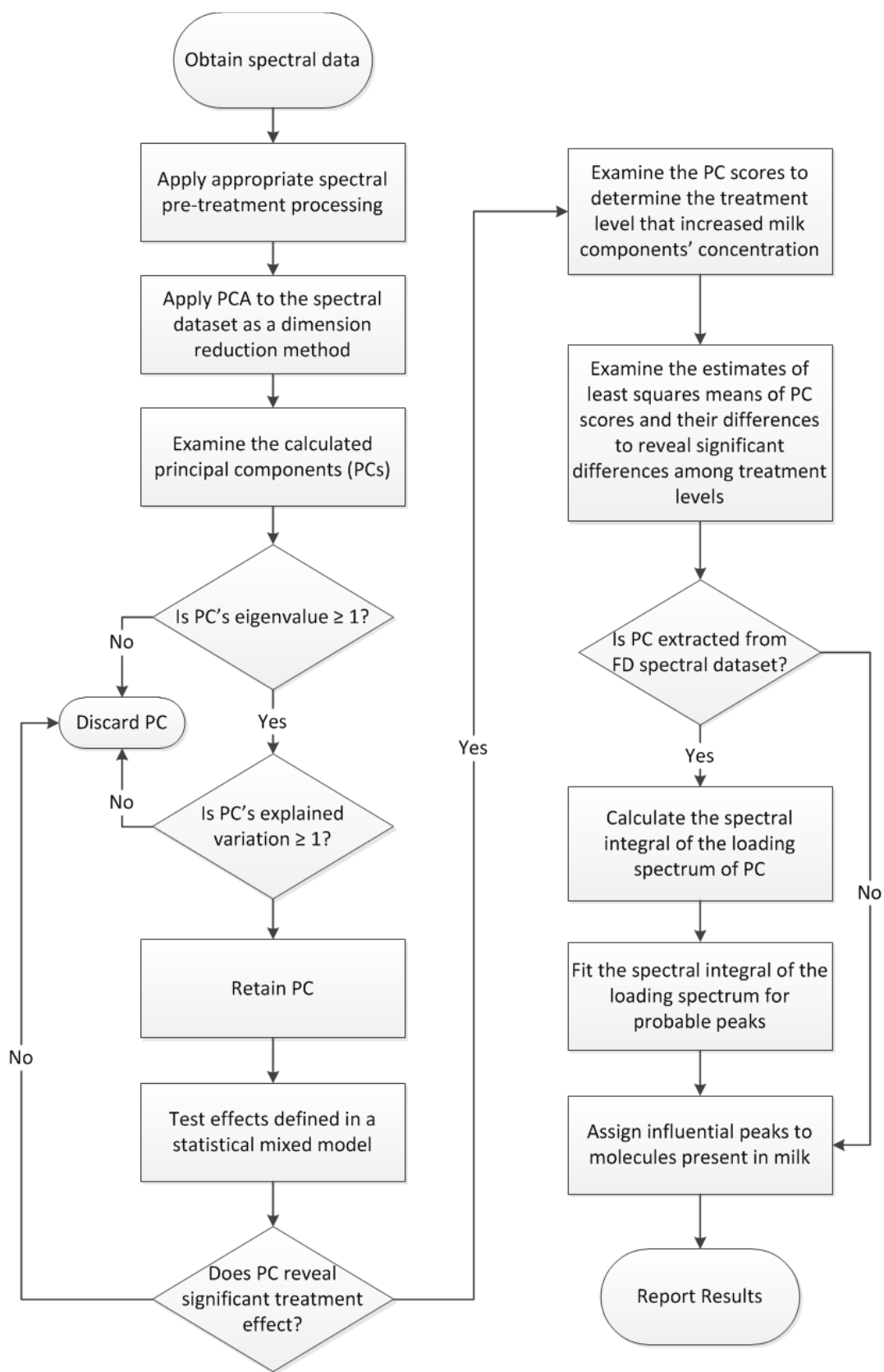


Figure 6-15 Workflow of the hybrid data analysis approach of FTIR milk spectral data to detect a treatment effect.

Acknowledgment

The author thanks the following people for their contribution to this project:

- Dr. Elsa Vasseur (McGill University) as the PI of this project.
- Dr. Roger Cue (McGill University) for his guidance on statistical analysis by mixed models.
- Jessica St John (McGill University) conducted the tie rail animal trial and developed the SAS code to detect the treatment effect on milk components by the mixed procedure as part of her M. Sc. Thesis [102].
- Tania Wolfe (McGill University) for her guidance on the trial and its statistical model.
- Valacta Inc. for analyzing milk samples collected for this project.
- Dr. Débora Santschi (Valacta) for converting the spectral files of the FOSS analyzer from the FSS proprietary format to the TXT format.

Funding

The animal trial was funded by Vasseur's NSERC-Novalait-Dairy Farmers of Canada-Valacta/Lactanet Industrial Research Chair on Sustainable Life of Dairy Cattle. Complementary funding for the study (physiological indicators) was provided by NSERC Collaborative Research Development Grant (Project no. CRDPJ 499520 16) & FRQNT-CRIBIQ-Novalait through Programme de recherche en partenariat pour l'innovation en production et transformation laitières (Project no. EPI 2016-138) and Valacta/Lactanet (Principal Investigator: Vasseur).

Connecting statement

In the previous chapter, a hybrid analysis approach of milk FTIR data was developed by combining principle component analysis (PCA) and mixed modeling as a data mining tool to isolate the spectral fingerprint that reflects the effect of a housing treatment of dairy cattle on milk FTIR spectral data in the context of controlled-design trials without relying on predictions of milk components from the FTIR spectral data. This hybrid approach will push milk FTIR spectroscopy beyond the paradigm of predicting specific milk components by PLS regression models to make inferences about the metabolic state of the animal. In this chapter, the hybrid analysis approach will be applied to milk FTIR spectra collected for animal trials that aim at studying the effects of tie rail position, chain length, stall width, stall length and manger wall height on animal welfare in the tie stall dairy farming system. The spectral fingerprint that represents the changes in milk composition related to the housing treatment effect will be interpreted in light of behavioural and welfare data collected for the trials' subjects. The hybrid approach will provide a novel angle to study animal welfare.

Chapter 7: Data mining of milk FTIR spectral data by combining mixed modeling and multivariate analysis to study the effects of housing treatments on cow welfare

Abstract

Principal component analysis and mixed modeling were successfully combined in a hybrid approach to analyze milk FTIR spectral data. This approach was applied to milk FTIR spectral data of 4 animal trials designed to study the effects of different housing treatments on the chemical composition of milk. In the tie rail trial, PC7 extracted from long-term VN FD spectral average dataset revealed significant effect on milk composition ($P = 0.0106$) for T3, which is the treatment that had the tie rail height at 112 cm and its forward position at 18 cm. The loading spectrum of this principal component revealed features that could be attributed to molecules in milk associated with elevated body fat mobilization, which indicated that T3 tie rail configuration was probably obstructing the cow access to feed. In the chain length trial, PC6 ($P = 0.0323$) extracted from long-term VN FD spectral average dataset revealed significant effect of the tie chain length on milk composition. The loading spectrum of this principal component revealed features that could be attributed to molecules in milk associated with increased incidents of acidotic rumen insults. The findings suggest that cows enrolled in the longer chain treatment had increased saliva production that stabilized the ruminal pH. In the stall width trial, PC5 extracted from long-term VN Raw spectral dataset revealed significant effect ($P = 0.0423$) for the single width treatment and a significant block effect ($P = 0.0008$). The loading spectrum of this principle component revealed milk samples collected from cows enrolled in the single width treatment (i.e., T1) had higher milk fat triglycerides, milk fatty acids and BHB. However, this effect was the result of a missing milk sample for the double width treatment (i.e., T2) that rendered the median parity of cows of T2 greater than that of T1, which are 3 and 2, respectively. Hence, the cows of T2 were older and one lactation higher than those of T1. In the manger wall height and stall length trial, PC6 extracted from long-term FD spectral dataset revealed significant stall length effect ($P = 0.0355$) on milk composition, which was due to higher incidents of mastitis in cow enrolled in the shorter stall length. The SCC average for stall length treatments confirmed this observation.

7.1 Introduction

Milk analysis by FTIR spectroscopy is currently being used to determine major milk components, milk fatty acids and minor milk components, such as urea, BHB and acetone. This technology can be implemented as a precision dairy monitoring technology that can identify individual cows with health or physiological issues through exception reporting of parameters that are significantly deviating from regular baselines recorded for individual animals [112]. Several milk components that are determined by FTIR milk analyzers are used to detect health and wellbeing disorders in dairy cows. For example, LeBlanc *et al.*, (2005) found a strong association between displaced abomasum incidence and BHB levels in milk $>200 \mu\text{mol/L}$ [113]. Toni *et al.*, (2011) reported that increased fat-to-protein ratio is an indicator of elevated lipid mobilization and increased risk of ketosis, displaced abomasum, lameness and mastitis [113]. Geishauser *et al.*, (2001) reported thresholds for the following molecules in milk for diagnosis of subclinical ketosis: $100 \mu\text{mol/L}$ of BHB, $100 \mu\text{mol/L}$ of acetoacetate or $250 \mu\text{mol/L}$ of acetone [113]. Enemark (2008) reported depressed milk fat percentage as an indicator of increased risk of subacute ruminal acidosis (SARA) incidence in dairy cows; however, the author also acknowledge the difficulty in relying on this indicator alone to diagnose SARA due to high variability in milk fat overtime [113]. Forsbäck *et al.*, (2010) observed little day-to-day variations in milk components and milk somatic cell count (SCC) in healthy lactating cows compared with cows diagnosed with mastitis or subclinical mastitis [113]. Nielsen *et al.*, (2005) reported consistent lactose concentration of 4.7% or higher was associated with lower SCC and reduced risk of intramammary infections [113]. However, by the time the animal exhibits clinical signs of stress or illness, it might be too late to intervene [112]. On the other hand, the above-mentioned health issues are also associated with behavioural changes that precede the clinical diagnosis. For example, cows that were diagnosed with a left displaced abomasum showed increased step activity in comparison to healthy animals during the week prior to the clinical diagnosis [113]. Goldhawk *et al.*, (2009) observed that cows at risk of subclinical ketosis had fewer feeding visits and 18% less dry matter intake in comparison with healthy controls in the week leading up to calving. The authors also acknowledged that measuring feed intake is difficult on commercial farms and that changes in feeding behavior, milk composition and milk yield are affected by multitude of factors; hence, undermining their reliability in diagnosing diseases and disorders in dairy cows [113]. Yeiser (2011) observed that rest time was greater and average daily steps were significantly lower five days before cows were

diagnosed with mastitis [113]. He also noticed that 2 days before onset rest time decreased compared with that of healthy cows [113].

One of the multiple benefits of milk analysis by FTIR is that its implementation does not require an invasive procedure or extensive labor. Increased frequency of the acquisition of milk mid-infrared spectra by existing dairy control laboratories has the potential to provide more information regarding the efficiency of feed management, health, reproduction and wellbeing of individual cows [112]. This approach will allow the producers to adopt proactive practices and help them intervene at the early stages of probable issues when actual milk components diverge from expected levels for individual animals [112]. However, commercial milk FTIR analyzers rely on partial least squares (PLS) models to predict milk components' levels from milk mid-infrared spectra. The accuracy of these PLS models is a function of the accuracy of the reference method that was used to determine the actual levels of the analyte of interest in milk samples that were used to calibrate the PLS model.

The objective of this chapter is to evaluate the potential of combined mixed modeling and multivariate analysis of milk mid-IR spectra in detecting changes in milk chemical composition resulting from physiological conditions that can be related to different housing treatments intended to improve the level of comfort and welfare of individual cows, without relying on determining intermediate values for specific milk components. In other words, the capability of the hybrid analysis approach will be evaluated for its potential to capture the spectral fingerprint of specific trend of changing milk components that can later be used to build prediction models of the health and welfare condition of individual cows. Animal behavioural data will also be used to support any trends detected by the hybrid analysis approach of spectral data.

7.2 Materials and Methods

7.2.1 The data

For each trial, milk samples were collected weekly and they were analyzed in Valacta Inc. lab (Ste., Anne-de-Bellevue, QC, Canada) to determine major and minor milk components using commercial FTIR milk analyzer. Two sets of data were received from Valacta. The first dataset comprised the concentrations of milk components that were determined by the commercial FTIR milk analyzer for individual milk samples. This dataset will be referred to as the numerical dataset. The second dataset contained FTIR spectra recorded for individual milk samples collected during the trial. This dataset will be referred to as the spectral dataset. Each FTIR spectrum consisted of 1060 spectral variables between 5008 cm^{-1} and 925 cm^{-1} .

A total of 19 milk components were included in the numerical dataset for the tie rail, chain length and stall width trials. For the manger wall and stall length, a total of 22 milk components were included. For all the trials, milk numeric data included concentrations of the following milk components: fat, protein, lactose, urea, beta-hydroxybutyrate (BHB), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), short-chain fatty acids (SCFA), mid-chain fatty acids (MCFA), long-chain fatty acids (LCFA), saturated fatty acids (SFA), total unsaturated fatty acids (TUFA), mono-unsaturated fatty acids (MUFA), poly-unsaturated fatty acids (PUFA), trans fatty acids (TFA), free fatty acids (FFA). All trials' milk numeric data included total solids (TS) except for the manger wall and stall length trial; on the other hand, all trials' milk numeric data included myristic acid (C14:0) except for the tie rail trial. In addition, the tie rail trial's milk numeric data included fat-to-protein ratio and the manger wall and stall length trial's milk numeric data included *de novo* fatty acids, mixed fatty acids, preformed fatty acids and true protein. These numerical datasets will be used to confirm observations that were concluded from the application of the hybrid spectral analysis, when possible. The total number of samples that were analyzed were 626, 355, 175, 476 for the tie rail, chain length, stall width and the manger wall and stall length trials, respectively.

7.2.2 Experimental design and statistical model

7.2.2.1 Tie rail trial

The objective of this trial was to study the effect of the tie rail height and forward position on cow's behaviour and welfare [102], which was conducted at the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). A tie rail (TR) is the pipe used as the attachment for the tie chain, which controls the forward location of each cow in her stall. In this study, 48 cows were assigned to 4 TR configurations, which were defined by the height and the forward position of TR. TR heights were 122, 122, 112 and 102 cm and forward positions were 18 cm, 36 cm, 18 cm and 36 cm for treatments T1, T2, T3 and T4, respectively. Treatments T1 and T2 are TR configurations that are recommended and commonly found on dairy farms, respectively. On the other hand, treatments T3 and T4 are new TR configurations designed to increase the opportunity of movements of the cow at her stall; hence, improve cow behavior and welfare. Cows were assigned to 6 different blocks to account for age of the cow, days in milk within current lactation and location in the barn effects. Half the cows underwent the trial during summer 2016 and the other half during fall 2016. Each period lasted 10 weeks (i.e., period 1: from July 25th to October 3rd, period 2: from October 10th to December 19th). The original statistical model of this trial was:

$$Y_{ijklm} = \mu + trt_i + start_j + block_{kji} + cow_{lkji} + week_m + trt_i \times week_m + e_{ijklm}$$

Where Y_{ijklm} was the dependent variable; the outcome measure of the l^{th} cow from the k^{th} block (parity, DIM and location in the barn) and the j^{th} start date on the combination of the i^{th} tie-rail configuration and m^{th} week. trt_i was the fixed effect of the i^{th} tie-rail configuration. $start_j$ was the fixed effect of the j^{th} start date. $block_{kji}$ was the fixed effect of k^{th} parity, DIM and location in the barn from the j^{th} start date on the i^{th} tie-rail configuration treatment. cow_{lkji} was the random effect of the l^{th} cow from the j^{th} start date and the k^{th} block on the i^{th} tie-rail configuration treatment. $week_m$ was the fixed effect of the m^{th} week. $trt_i \times week_m$ is the interaction effect of the individual combination of the i^{th} tie-rail configuration treatment with the m^{th} week. e_{ijklm} was the random residual associated with the outcome measure of the l^{th} cow from j^{th} start date and k^{th} block on the combination of the i^{th} tie-rail configuration treatment and the m^{th} week [102].

To eliminate the week effect, short-term and long-term averages were calculated for each cow for FTIR spectra and milk composition data that included samples collected from week 1 to week 3 and from week 8 to week 10, respectively. As a result, the statistical model was modified to become as follows:

$$Y_{ijk} = \mu + trt_i + start_j + block_{kj} + e_{ijk}$$

Where Y_{ijk} was the dependent variable; the outcome measure of the cow of the i^{th} tie-rail configuration from the k^{th} block (parity, DIM and location in the barn) and the j^{th} start date. trt_i was the fixed effect of the i^{th} tie-rail configuration. $start_j$ was the fixed effect of the j^{th} start date. $block_{kj}$ was the fixed effect of k^{th} parity, DIM and location in the barn from the j^{th} start date. e_{ijk} was the random residual associated with the outcome measure of the cow from j^{th} start date and k^{th} block on the i^{th} tie-rail configuration treatment.

7.2.2.2 Chain length trial

The objective of this animal trial was to study the effect of the tie chain length (TCL) on cow's behaviour and welfare [114]. The trial was conducted at the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). A tie chain confines a cow to her stall space and allows for ease of lunging, resting in the head back position, or grooming. In this study, 24 cows were assigned to 2 TCL treatments. T1, which is the control, had a chain length of 1 m. On the other hand, T2, which is a suggested treatment, had a chain length of 1.4 m that intends to increase the cow's movement ability in her stall. Cows were assigned to 12 different blocks to account for age of the cow (i.e., parity and stage of lactation) and days in milk within current lactation, they were placed evenly into two rows within the barn (i.e., row 1 and 4). The trial lasted for 10 weeks from February 20th, 2017 to May 1st, 2017. The original statistical model of this trial was:

$$Y_{ijkl} = \mu + trt_i + block_j + cow_{ij} + row_k + week_l + trt_i \times week_l + e_{ijkl}$$

Where: Y_{ijkl} was the dependant variable; the outcome measure of the cow from the j^{th} block (parity and lactation stage) in the k^{th} row on the combination of the i^{th} chain length and the l^{th} week; trt_i is the fixed effect of the i^{th} chain length; $block_j$ is the fixed effect of the j^{th} parity and lactation stage combination; cow_{ij} is the random effect of the cow from the j^{th} block on the i^{th}

chain length; row_k is the random effect of the k^{th} row in the barn; $week_l$ is the fixed effect of the l^{th} week; $trt_i \times week_l$ is the fixed effect of the interaction, the specific effect of the combination of the i^{th} chain length and the l^{th} week; e_{ijkl} is the random residual associated with the outcome measure of the cow from the j^{th} block in the k^{th} row on the combination of the i^{th} chain length and the l^{th} week [114].

To eliminate the week effect, short-term and long-term averages were calculated for each cow for FTIR spectra and milk composition data that included samples collected from week 1 to week 3 and from week 8 to week 10, respectively. As a result, the statistical model was modified to become as follows:

$$Y_{ijk} = \mu + trt_i + block_j + row_k + e_{ijk}$$

Where: Y_{ijkl} was the dependant variable; the outcome measure of the cow from the j^{th} block (parity and lactation stage) in the k^{th} row on the i^{th} chain length; trt_i is the fixed effect of the i^{th} chain length; $block_j$ is the fixed effect of the j^{th} parity and lactation stage combination; row_k is the random effect of the k^{th} row in the barn; e_{ijk} is the random residual associated with the outcome measure of the cow from the j^{th} block in the k^{th} row on the i^{th} chain length.

7.2.2.3 Stall width trial

The objective of this animal trial was to study the effect of the stall width (SW) on cow's behaviour and welfare [114]. The stall width allows the cows to rest in wider positions. The trial was conducted at the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). In this study, 16 cows were assigned to 2 SW treatments. T1, which is the control, had a single width stall that was calculated as follows: $2 \times (\text{width of cow at hips}) + 2$ inches. On the other hand, T2, which is a suggested treatment, had a double width stall that intends to increase the cow's comfort in her stall. The stall width for this treatment was calculated as follows: $2 \times (2 \times (\text{width of cow at hips}) + 2 \text{ inches})$. Cows were assigned to 8 different blocks to account for age of the cow (i.e., parity and stage of lactation) and days in milk within current lactation and they were placed evenly into two rows within the barn (i.e., row 1 and 4). The trial lasted for 6 weeks from June 5th, 2017 to July 14th, 2017. The original statistical model of this trial was:

$$Y_{ijkl} = \mu + trt_i + block_j + cow_{ij} + row_k + week_l + trt_i \times week_l + e_{ijkl}$$

Where: Y_{ijkl} was the dependant variable; the outcome measure of the cow from the j^{th} block (parity and lactation stage) in the k^{th} row on the combination of the i^{th} stall width and the l^{th} week; trt_i is the fixed effect of the i^{th} stall width; $block_j$ is the fixed effect of the j^{th} parity and lactation stage combination; cow_{ij} is the random effect of the cow from the j^{th} block on the i^{th} stall width; row_k is the random effect of the k^{th} row in the barn; $week_l$ is the fixed effect of the l^{th} week; $trt_i \times week_l$ is the fixed effect of the interaction, the specific effect of the combination of the i^{th} stall width and the l^{th} week; e_{ijkl} is the random residual associated with the outcome measure of the cow from the j^{th} block in the k^{th} row on the combination of the i^{th} stall width and the l^{th} week [114].

To eliminate the week effect, data of weeks 1 and 6 were used as short-term and long-term datasets for data analysis. As a result, the statistical model was modified to become as follows:

$$Y_{ijk} = \mu + trt_i + block_j + row_k + e_{ijk}$$

Where: Y_{ijk} was the dependant variable; the outcome measure of the cow from the j^{th} block (parity and lactation stage) in the k^{th} row on the i^{th} stall width; trt_i is the fixed effect of the i^{th} stall width; $block_j$ is the fixed effect of the j^{th} parity and lactation stage combination; row_k is the random effect of the k^{th} row in the barn; e_{ijk} is the random residual associated with the outcome measure of the cow from the j^{th} block in the k^{th} row on the i^{th} stall width.

7.2.2.4 Manger wall and stall length trial

The objective of this animal trial was to study the combined effect of the stall length and manger wall height on cow's behaviour and welfare [115]. The trial was conducted at the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). The length of two rows in the barn were modified as follows: row 1 was 178 cm (i.e., short or L1), which is commonly found on Quebec's farm, and row 4 was 188 cm (i.e., long or L2), which is a suggested modification. Two manger wall heights were applied randomly to stalls within each row: 20 cm (i.e., regular or T1), which is the upper limit of recommendation and 5 cm (i.e., low or T2). In this study, 24 cows were randomly divided into 4 groups. Two groups were assigned to each row and subjected to both manger wall treatments in a crossover design (1 week for

habituation, 6 weeks of data collection/treatment). Cows were assigned to 6 different blocks to account for age of the cow (i.e., parity and stage of lactation) and days in milk within current lactation. Increasing the stall length and reducing the manger wall will increase the space available for the cow, which will ease its movement and lying; thus, reducing its susceptibility to injuries. The trial lasted for 14 weeks from February 26th, 2018 to June 4th, 2018. The original statistical model of this trial was:

$$Y_{ijkmnpq} = \mu + length_i + seq_{ij} + block_k + cow_{ijkm} + period_n + week_p + trt_{iq} + trt_{iq} \times week_p + e_{ijkmnpq}$$

Where: $Y_{ijkmnpq}$ was the dependent variable; the outcome measure of the m^{th} cow of the k^{th} block in the j^{th} seq of the i^{th} length, the q^{th} trt of the i^{th} length, and the n^{th} period and p^{th} week; $length_i$ is the fixed effect of the i^{th} stall length; seq_{ij} is the fixed effect of the j^{th} sequence on the i^{th} stall length; $block_k$ is the fixed effect of the k^{th} parity and stage of lactation combination; cow_{ijkm} is the random effect of the cow from the k^{th} block on the j^{th} sequence of the i^{th} stall length; $period_n$ is the fixed effect of the n^{th} period; $week_p$ is the fixed effect of the p^{th} week; trt_{iq} is the fixed effect of the q^{th} manger wall height treatment on the i^{th} stall length treatment; $trt_{iq} \times week_p$ is the fixed effect of the interaction, the specific effect of the combination of the p^{th} week and the q^{th} manger wall height on the i^{th} stall length; and $e_{ijkmnpq}$ is the random residual associated with the outcome measure of the cow from the k^{th} block on the j^{th} seq of the i^{th} length, the q^{th} manger wall height (on the i^{th} stall length) and on the p^{th} week of the n^{th} period [115].

To eliminate the week effect, data of weeks 1 and 6 were used as short-term and long-term datasets for data analysis. As a result, the statistical model was modified to become as follows:

$$Y_{ijkmnq} = \mu + length_i + seq_{ij} + block_k + cow_{ijkm} + period_n + trt_{iq} + e_{ijkmnq}$$

Where: Y_{ijkmnq} was the dependent variable; the outcome measure of the m^{th} cow of the k^{th} block in the j^{th} seq of the i^{th} length, the q^{th} trt of the i^{th} length, and the n^{th} period; $length_i$ is the fixed effect of the i^{th} stall length; seq_{ij} is the fixed effect of the j^{th} sequence on the i^{th} stall length; $block_k$ is the fixed effect of the k^{th} parity and stage of lactation combination; cow_{ijkm} is the random effect of the cow from the k^{th} block on the j^{th} sequence of the i^{th} stall length; $period_n$ is the fixed effect of the n^{th} period; trt_{iq} is the fixed effect of the q^{th} manger wall height treatment on the i^{th} stall length treatment; and e_{ijkmnq} is the random residual associated with the outcome measure of

the cow from the k^{th} block on the j^{th} seq of the i^{th} length, the q^{th} manger wall height (on the i^{th} stall length) and of the n^{th} period.

7.2.3 Spectral data analysis

The hybrid approach that was developed in the previous chapter was applied to the short-term and long-term spectral datasets that were calculated for each trial. The spectral regions that were retained for spectral analysis were $1612\text{-}925\text{ cm}^{-1}$, $1797\text{-}1681\text{ cm}^{-1}$ and $3061\text{-}2803\text{ cm}^{-1}$ with a total of 278 spectral variables. These spectral regions were considered because they contain the information related to milk chemical composition. Inhouse MATLAB codes were written to calculate differential FD of the spectra, to VN the spectra and to load individual spectra into a matrix. These spectral pre-treatments were applied to spectra of individual samples prior to calculating the short-term and long-term averages. In JMP Pro 13.2.1., PCA was applied to the four versions of the short-term and long-term spectral datasets (i.e., raw, FD, VN raw, VN FD) as a dimension reduction method to reduce the number of responses to be tested by the mixed model. PCs with eigenvalue ≥ 1 and that explained 1% of the variation or more were considered meaningful and were retained for testing by the mixed model. The Mixed procedure in SAS 9.4 (SAS Institute, Cary, NC, USA) was utilized to test for the treatment effect at significance level $\alpha=0.05$. If a PC revealed a significant treatment effect, then the least squares means of its scores were examined to determine the treatment levels that are significantly different from the other levels using a Scheffé adjustment for multiple comparisons. The influential spectral features can be directly extracted from the loading spectrum of the PC that revealed the significant treatment effect, if it was extracted from raw spectral dataset. If this PC was obtained from FD spectral dataset, then the spectral integral of the PC's loading spectrum must be calculated before extracting the influential spectral features. The cumulative trapezoidal numerical integration function in MATLAB was used to calculate the spectral integral for the loading spectrum in question. If the integrated loading spectrum had produced wide humps with no clear peaks, the Peak Resolve feature in Omnic 7.3 (Thermo Electron Corporation, Waltham, MA, USA) was used to fit the integrated loading spectrum for probable peaks. To do so, the Voigt function with low or high sensitivity was used and the baseline was set to none. The noise and the full width at half height (FWHH) of the narrowest peak in the region of interest was determined by the software. The fitting process was repeated several times until an acceptable residual spectrum was obtained. Milk

numerical dataset will be used to confirm observations from the spectral analysis related to determined milk components.

In addition, parsimonious PLS-DA model was developed for the tie rail trial for the averaged numerical milk composition data for exploratory purposes according to the method described in the previous chapter.

7.2.4 Interpretation of spectral features in loading spectra

Bulk tank raw milk was obtained from the Dairy Research Complex, Macdonald Campus, McGill University (Ste. Anne-de-Bellevue, QC, Canada). Milk samples of 35 mL were spiked with minor milk components and aqueous solutions of these chemicals were prepared. Their FTIR spectra were recorded at Valacta by their milk analyzers. Urea, β -hydroxybutyric acid (BHBA or BHB) and acetone were chosen because they are routinely determined by dairy control laboratories. Linoleic acid was chosen as an example of unsaturated fatty acid. It must be noted that different fatty acids do not produce distinct FTIR signals from each other, specially when they are present in a mixture of fatty acids [42]. In addition to urea, other nonprotein nitrogen (NPN) compounds that are present in milk were used to spike milk samples. These compounds are ammonium, creatine, histamine, orotic acid and hippuric acid [84]. In addition to BHB and acetone, citrate and acetate are also markers for energy intake related issues in dairy cows [59]. In addition, phosphate, lactose, glucose and galactose were also chosen. For compounds that do not have solubility limits in water, 1% (w/v) and 0.5% (w/v) were used for sample and solution preparations; otherwise, solubility in water was used as the maximum concentration for the preparations (Table 7-1). Additional milk samples were prepared by spiking them with the mean quantity found in milk for compounds that are naturally present in low concentrations in milk [84]. The following chemicals were obtained from Sigma-Aldrich: BHB (i.e., DL- β -Hydroxybutyric acid sodium salt ~98%), citrate (i.e., citric acid trisodium salt dihydrate), phosphate (i.e., sodium phosphate monobasic monohydrate $H_2NaO_4P \cdot H_2O$), α -D glucose, D-Galactose, α -D-Lactose monohydrate, linoleic acid (i.e., linoleic acid sodium salt 99%), creatine anhydrous, histamine, orotic acid monohydrate (i.e., 6-Carboxy-2,4-dihydropyrimidine) and hippuric acid (i.e., benzoylaminoacetic acid sodium salt 99%). Urea and acetone were obtained from Fisher scientific. Acetate (i.e., sodium acetate anhydrous $NaC_2H_3O_2$) and ammonium chloride were obtained from Mallinckrodt and VWR, respectively.

Omic 7.3 (Thermo Electron Corporation, Waltham, MA, USA) was used to calculate variance spectra and the second derivative for the collected spectra, and the Find Peaks functionality was used to determine IR peaks' positions in milk and in the aqueous solution for each compound.

Table 7-1 Concentrations of spiked milk samples and aqueous solutions of milk minor components

Compound	Concentrations (w/v)
Urea	1%, 0.5%
BHB	1%, 0.5%
Acetone	1%, 0.5%
Citrate	1%, 0.5%
Acetate	1%, 0.5%
Phosphate	1%, 0.5%
Glucose	1%, 0.5%
Galactose	1%, 0.5%
Lactose	1%, 0.5%
Ammonium	1%, 0.5%, 0.002% (milk mean)
Linoleic acid	0.14%
Creatine	1%, 0.5%, 0.002% (milk mean)
Histamine	0.28%
Orotic acid	0.14% (solubility), 0.002% (milk mean)
Hippuric acid	0.3% (solubility), 0.0006% (milk mean)

7.3 Results and Discussion

7.3.1 Interpretation of spectral features in loading spectra

Table 7-2 summarizes the spectral features of minor milk components observed in the second derivative of FTIR spectra of milk samples spiked with these molecules and their respective aqueous solutions. These observed features will be used to identify the influential features and their respective molecules in the loading spectra of principal components that will reveal significant treatment effect in the studies covered in this chapter.

Urea is a primary amide (Figure 7-7) that has a carbonyl ($C=O$) and two amino ($-NH_2$) functional groups. The spectra of milk samples spiked with increasing amounts of urea (Figure 7-8) reveal increased absorbance intensity at peaks centered around ~ 1457 and 1156 cm^{-1} . These peaks are assigned to the $C-N$ stretching band and $C-N$ stretching coupled with the stretching of adjacent bond in the molecule, respectively [116]. The amide I band at 1650 cm^{-1} that results from $C=O$ stretching is not observable in aqueous samples because it overlaps with the $O-H$ bending vibration band of water, which is considered as a noisy region.

β -hydroxybutyric acid is a β -hydroxy acid (Figure 7-9) that has carboxyl and hydroxy functional groups separated by two carbon atoms, in addition to methyl and methylene groups. The spectra of milk samples spiked with increasing amounts of β -hydroxybutyric acid (Figure 7-10) reveal increased absorbance intensity at peaks centered around ~ 2926 , 1554 , 1405 , 1316 , 1077 cm^{-1} . The peak at 2926 is assigned to the asymmetrical stretching ($\nu_{as}CH_2$) of the methylene group, which overlaps with the signal of the $C-H$ stretching of milk fat or the Fat B region [116]. The peak at 1316 cm^{-1} is assigned to the $C-O$ stretching [116]. On the other hand, the peak at 1077 cm^{-1} is assigned to and $C-H$ bending vibrations [117], and the $C-O/C-C$ bond stretching [110], which overlaps with the signal coming from lactose. The carboxylate ion gives rise to two bands: a strong asymmetrical stretching at 1554 cm^{-1} and a weaker symmetrical stretching at 1405 cm^{-1} . This band can also be assigned to the $C-O-H$ bending [116].

Acetone is a ketone (Figure 7-11) that has a carbonyl functional group ($C=O$) and two methyl (CH_3) groups. The spectra of milk samples spiked with increasing amounts of acetone (Figure 7-12) reveal increased absorbance intensity at peaks centered around ~ 1690 , 1414 , 1373 and 1239 cm^{-1} . The peak at 1690 cm^{-1} is assigned to the carbonyl group ($C=O$) stretching [116]. Normally,

carbonyl group exhibit absorbance band at 1715 cm^{-1} , but in this case the position of the carbonyl band has shifted into a reduced wavenumber because of the polarity of water in milk. The peaks at 1414 and 1373 cm^{-1} are assigned to the asymmetrical ($\delta_{as}CH_3$) and symmetrical (δ_sCH_3) bending vibration of the methyl group, respectively [116]. The peak at 1239 cm^{-1} is assigned to the stretching of $C - C - C$ group and the bending of $C - C(=O) - C$ in the $C - C - C$ group [116].

Citrate is a derivative of citric acid (Figure 7-13), which is a weak organic acid. It has three carboxyl function groups and one hydroxy functional group. The spectra of milk samples spiked with increasing amounts of citrate (Figure 7-14) reveal increased absorbance intensity at peaks centered around $\sim 2926, 1557, 1394, 1248, 1078\text{ cm}^{-1}$. The peak at 2926 is assigned to the asymmetrical stretching ($\nu_{as}CH_2$) of the methylene group, which overlaps with the signal of the $C - H$ stretching of milk fat or the Fat B region [116]. The peaks at $1557, 1394$ and 1248 cm^{-1} are assigned to the strong symmetrical stretching of the carboxylate ion, the $C - O - H$ bending and the $C - O$ stretching [116]. The peak at 1078 cm^{-1} is assigned to the $C - H$ bending vibrations [117] and the $C - O/C - C$ bond stretching [110], which overlaps with the signal coming from lactose.

Acetate is a derivative of acetic acid (Figure 7-15), which is a weak organic acid. Acetate has one carboxyl and one methyl functional group. The spectra of milk samples spiked with increasing amounts of acetate (Figure 7-16) reveal increased absorbance intensity at peaks centered around ~ 1551 and 1414 cm^{-1} . The peaks at 1551 and 1414 cm^{-1} are assigned to the strong symmetrical stretching of the carboxylate ion [116] and $C - O$ stretching [118], respectively.

Phosphate is a derivative of phosphoric acid (Figure 7-17), which is a weak acid. It contains one phosphoryl group. The spectra of milk samples spiked with increasing amounts of phosphate (Figure 7-18) reveal increased absorbance intensity at peaks centered around $\sim 1156, 1077$ and 940 cm^{-1} . The peaks at 1156 and 1077 cm^{-1} are assigned to the asymmetric and symmetric stretching of the phosphoryl group ($P = O$) [119]. The peak at 940 cm^{-1} is assigned to the stretching of $P - O - H$ [116, 119].

Glucose, galactose and lactose are carbohydrates (Figure 7-19). Glucose and galactose are classified as aldohexoses monosaccharides. They consist of six carbon atoms backbone and an

aldehyde functional group on carbon atom number 1. Galactose is an epimer of glucose. They differ in their stereochemistry at carbon atom number 4. Lactose is a disaccharide that consists of galactose and glucose subunits linked by a β 1 $\rightarrow$ 4 glycosidic bond. The spectra of milk samples spiked with increasing amounts of glucose, galactose and lactose (Figure 7-20) reveal increased absorbance intensity at a peak centered around $\sim 1076\text{ cm}^{-1}$. In addition, a distinct increased absorbance intensity is observed at ~ 1043 and 1154 cm^{-1} for glucose and galactose, respectively. In an aqueous solution of glucose, the distinct peak appears at 1033 cm^{-1} , which has been reported as an IR marker peak for glucose in the literature [110, 120]. All peaks observed in the region $1200\text{-}800\text{ cm}^{-1}$ are assigned to the $C - O/C - C$ bond stretching [110].

Ammonium results from the protonation of ammonia, which is an azane that consists of a single nitrogen atom covalently bonded to three hydrogen atoms (Figure 7-21). The spectra of milk samples spiked with increasing amounts of ammonium chloride (Figure 7-22) reveal increased absorbance intensity at peak centered around $\sim 1457\text{ cm}^{-1}$. This broad band is assigned to the $N - H$ bending of the NH_4^+ ion [116].

Linoleic acid is an octadecadienoic acid with long aliphatic unsaturated chain and carboxylic functional group in which the two double bonds are at positions 9 and 12 and have *cis* stereochemistry (Figure 7-23) [121]. The spectra of milk samples spiked with linoleic acid (Figure 7-24) reveal increased absorbance intensity at peaks centered around $\sim 3012, 2927, 2857, 1705, 1581, 1554, 1458, 1408$ and 987 cm^{-1} . The peak at 3012 cm^{-1} is assigned to the $C - H$ stretching in the alkene (olefinic) bond ($C = C - H$) [116]. The peaks at 2926 and 2857 cm^{-1} are assigned to the asymmetrical stretching ($\nu_{as}CH_2$) and symmetrical stretching (ν_sCH_2) of the methylene group, which overlaps with the signal of the $C - H$ stretching of milk fat or the Fat B region [116]. The peak at 1705 cm^{-1} is assigned to the $C = O$ stretching vibration in the carboxyl functional group [116]. The carboxylate ion gives rise to two bands: a strong asymmetrical stretching at 1554 cm^{-1} and a weaker symmetrical stretching at 1408 cm^{-1} . The peak at 1408 cm^{-1} also results from the $C - O - H$ bending in the protonated carboxylic functional group [116]. The peak at 1458 cm^{-1} can be assigned either to the asymmetrical bending vibration of the methyl group ($\delta_{as}CH_3$) or to the scissoring band of the methylene group (δ_sCH_2) [116]. The peak at 987 cm^{-1} is assigned to the out-of-plane alkene $C - H$ bending vibration, which is a characteristic band for alkenes [116].

Creatine is an organic compound that is derived from the amino acid glycine. It contains carboxylic, methyl and amidine ($-C(=NH)NH_2$) functional groups (Figure 7-25) [122]. The spectra of milk samples spiked with increasing amounts of creatine (Figure 7-26) reveal increased absorbance intensity at peaks centered around $\sim 1538, 1396, 1311, 1106, 980\text{ cm}^{-1}$. The peak centered at 1396 cm^{-1} can be assigned to the $C - O - H$ bending band of the carboxylic functional group, to the symmetrical stretching of the carboxylate ion or to the $C - N$ stretching band [116]. The peaks centered at $1538, 1311, 1106$ and 980 cm^{-1} are assigned to the deformation vibration band of $(=NH_2)$ in the amidine functional group [123], to the $C - O$ stretching band of the carboxylic functional group, to the $C - N$ stretching band coupled with stretching of adjacent bonds in the molecule and to the out-of-plane bending of the bonded $O - H$ in dimeric carboxylic acids, respectively [116].

Histamine is an organic compound that is derived from the amino acid histidine. It is an 1H-imidazole substituted at position C-4 by a 2-aminoethyl group (Figure 7-27) [124]. The spectra of milk samples spiked with increasing amounts of histamine (Figure 7-28) reveal increased absorbance intensity at peaks centered around $\sim 3012, 2857, 1581, 1457, 1315, 1033, 987\text{ cm}^{-1}$. These peaks are assigned to NH_3^+ stretching (νNH_3^+), CH_2 symmetrical stretching ($\nu_s CH_2$) of the methylene group of the side chain, stretching of the imidazole ring which appears in the aqueous solution at 1573 and 1488 cm^{-1} , NH_3^+ symmetrical bending ($\delta_s NH_3^+$), CH_2 wagging (ωCH_2), $C - N$ stretching (νCN) and $C - H$ out-of-plane bending (γCH), respectively [125].

Orotic acid is a pyrimidine monocarboxylic acid that is uracil bearing a carboxy substituent at position C-6 (Figure 7-29) [126]. The spectra of milk samples spiked with increasing amounts of orotic acid (Figure 7-30) reveal increased absorbance intensity at peaks centered around $\sim 1700, 1500, 1377$ and 1033 cm^{-1} . The peak at 1700 cm^{-1} is assigned to $C = O$ stretching [127]. The peaks at 1500 and 1377 cm^{-1} are assigned to the heteroaromatic ring stretching vibration [116]. The peak at 1033 cm^{-1} is assigned to the ring in-plane deformation and stretching [127].

Hippuric acid is an N-acylglycine in which the acyl group is specified as benzoyl (Figure 7-31) [128]. The spectra of milk samples spiked with increasing amounts of hippuric acid (Figure 7-32) reveal increased absorbance intensity at peaks centered around $\sim 1581, 1400$ and 1307 cm^{-1} . These peaks are assigned to the bending vibration (δNH) of $N - H$, in-plane bending (δCH_2) of the methylene group, and the bending vibration (δCH_2) of the methylene group [129], respectively.

Table 7-2 IR peaks of minor milk components detected in FTIR spectra of spiked milk samples and aqueous solutions of the respective molecules.

Molecule	Peaks in milk cm^{-1} (2 nd derivative)	Peaks in water cm^{-1} (2 nd derivative)
Urea	1457, 1156	1461, 1160
β -Hydroxybutyric acid (BHB)	2926, 1554, 1405, 1316, 1077	2981, 1559, 1404, 1311, 1269, 1207, 1130, 1060, 948
Acetone	1690, 1414, 1373, 1239	1689, 1424, 1370, 1239, 1096
Citrate	2926, 1557, 1394, 1248, 1078	2923, 1581-1566, 1390, 1288, 1093
Acetate	1551, 1414	1554, 1416, 1348, 1060, 1021, 933
Phosphate	1156, 1077, 940	1261, 1236, 1160, 1077, 941
Ammonium chloride	1457	1454
Linoleic acid (fatty acid)	3012, 2927, 2857, 1705, 1581, 1554, 1458, 1408, 987	3011, 2929, 2861, 1597, 1554, 1458, 1405
Creatine	1538, 1396, 1311, 1106, 980	2950, 2835, 1538, 1431, 1396, 1307, 1168, 1107, 1049, 976
Histamine	3012, 2857, 1581, 1457, 1315, 1033, 987	3008, 2888, 1573, 1488, 1310, 1033, 987, 941
Orotic acid	1700, 1500, 1377, 1033	1700, 1497, 1377, 1014
Hippuric acid	1581, 1400, 1307	1584, 1489, 1396, 1301

7.3.2 Tie rail trial

Table 7-3 and Table 7-4 summarize the meaningful PCs that were extracted from the raw, FD, VN raw and VN FD spectral datasets of the long-term and short-term milk samples spectral averages of the tie rail trial, respectively. PCA yielded five, seven, four and nine meaningful PCs from the raw, FD, VN raw and VN FD long-term spectral average datasets that explained 97.35%, 94.92%, 96.99% and 95.44% of the variation in the spectral datasets, respectively. These PCs, whose eigenvalue and percentage of explained variation ≥ 1 , represent the sources of variation in their respective spectral datasets that were separated from the noise and that were tested for the treatment, start (i.e. season) and block effects by the SAS Mixed procedure. PC6 ($P = 0.0371$), PC4 ($P = 0.0462$) and PC7 ($P = 0.0106$) extracted from long-term FD, VN raw, VN FD spectral average datasets, respectively, revealed significant treatment effect. The calculation of the spectral averages was successful in eliminating the strong week effect that was noticed in all analysis approaches evaluated in the previous chapter. Among the three PCs that revealed significant treatment effect, PC7 ($P = 0.0106$) isolated from the long-term VN FD spectral dataset revealed the strongest treatment effect, and unlike the other two PCs, start ($P = 0.5590$) and block ($P =$

0.0600) had insignificant effects on it. This observation can be explained by the fact that FD exposed more details in the spectral dataset and VN eliminated procedural variability not related to chemical composition of milk samples, which facilitated the isolation of the treatment effect from the other two studied effects. For these reasons, PC7 isolated from the long-term VN FD spectral dataset will be subjected to further analysis to determine the treatment levels that are significantly different from each other and the spectral features that are responsible for these differences. PC 7 explains 1.37% of the variation in its respective dataset, which suggests that the treatment effect is limited, and it was detected in an early stage.

Table 7-5 and Table 7-6 summarize the least squares means and their differences produced by the Mixed procedure for the scores of PC7 extracted from long-term VN FD spectral average dataset that revealed a significant treatment effect. These tables show that T3 PC7 scores were significantly different from the scores of other treatments ($P = 0.0038$) and the Scheffé adjusted P value shows that T3 is significantly different from T1 ($P = 0.0332$).

Inspection of the integral of PC7 loading spectrum (Figure 7-1) revealed clear peaks at the following wavenumbers: 3008, 2919, 2851, 1716, and 1407, which can be assigned to the following IR bands: the $C - H$ stretching in the alkene (olefinic) bond ($C = C - H$) in unsaturated fatty acids, the asymmetrical stretching ($\nu_{as}CH_2$) of the methylene group in fatty acids, symmetrical stretching (ν_sCH_2) of the methylene group in fatty acids, the $C = O$ stretching vibration in the carboxyl functional group in free fatty acids and the symmetrical stretching of the carboxylate ion or the $C - O - H$ bending in BHB, respectively (Table 7-2).

Table 7-7 summarizes the results of the peak fitting process for regions 1250-1180 cm^{-1} , 1390-1250 cm^{-1} and 1618-1424 cm^{-1} , which did not show any clear peaks. In the first region, a peak was detected at 1237 cm^{-1} , which can be assigned to the stretching of $C - C - C$ group and the bending of $C - C(=O) - C$ in the $C - C - C$ group in acetone. In the second region, peaks were detected at 1287, 1317 and 1372 cm^{-1} , which can be assigned to the $C - O$ stretching that appears in the FTIR spectrum of the aqueous solution of citrate, the $C - O$ stretching in BHB and the symmetrical (δ_sCH_3) bending vibration of the methyl group in acetone, respectively. In the third region, peaks were detected at 1460 and 1541 cm^{-1} . The 1460 cm^{-1} peak can be assigned to the asymmetrical bending vibration of the methyl group ($\delta_{as}CH_3$) or to the scissoring band of the methylene group

($\delta_s CH_2$) in fatty acids. On the other hand, the 1541 cm^{-1} peak can be assigned to the symmetrical stretching of the carboxylate ion that is found in citrate, BHB, free fatty acids and acetate.

In addition, an upward trend is noticed between $\sim 950\text{ cm}^{-1}$ and $\sim 1040\text{ cm}^{-1}$, with a small bump at $\sim 967\text{ cm}^{-1}$. Subsequently, a downward trend is noticed between $\sim 1040\text{ cm}^{-1}$ and $\sim 1100\text{ cm}^{-1}$ followed by another upward trend that peaks at 1130 cm^{-1} , which corresponds to a band present in the FTIR spectrum of the aqueous solution of BHB. These trends can be interpreted that milk samples collected from cows assigned to T3 in the last 3 weeks of the trial had a higher content of trans fatty acids and BHB and lower content of lactose in comparison to samples collected from cows assigned to T1 during the same period. Trans fatty acids, lactose and BHB have absorption IR bands centered at $\sim 967\text{ cm}^{-1}$ [130], $\sim 1076\text{ cm}^{-1}$ [130] and 1130 cm^{-1} , respectively. It can be noticed that the PCA decomposition process of variation in the spectral dataset allowed the detection of minor milk components' IR bands that are overwhelmed by major milk components' absorption bands.

The spectral analysis concludes that milk samples collected from cows assigned to T3 had higher levels of BHB, acetone, citrate, acetate and trans fatty acids and lower levels of lactose during the last three weeks of the trial. All these molecules, except for trans fatty acids, are considered markers of elevated body fat mobilization or negative energy balance [15, 59]. Milk composition numerical data only included levels of BHB, lactose and trans fatty acids. Inspection of milk composition numerical data confirms the observations of the spectral analysis (Table 7-9 and Table 7-10). During the last 3 weeks, average lactose content was 4.62% and 4.60% for T1 and T3, respectively, and average BHB content was 0.05 mmol/L and 0.06% mmol/L for T1 and T3, respectively. For week 9, average lactose content was 4.63% and 4.59% for T1 and T3, respectively. For week 10, average BHB content was 0.05 mmol/L and 0.07 mmol/L for T1 and T3, respectively. Reduced lactose concentrations and increased BHB in milk suggest that cows assigned to T3 might have been experiencing higher level of body fat mobilization in comparison to cows assigned to T1 during the last 3 weeks of the trial [100]. Reduced protein concentration in milk is another indicator of increased body fat mobilization [100]. The long-term average protein concentration was 3.44% and 3.34% for T1 and T3, respectively, and a downward trend between 1550 cm^{-1} and 1600 cm^{-1} was noticed in the integral of PC7 loading spectrum. In addition, the

average trans fatty acids content during the last 3 weeks was 0.09% and 0.10% for T1 and T3, respectively, which agrees with the findings of the spectral analysis.

A parsimonious PLS-DA model of the milk composition numerical dataset for weeks 8-10 to classify milk samples according to treatment revealed that BHB was the most significant variable to discriminate milk samples by treatment (Table 7-8). In addition, it revealed that lactose and protein were not significant variables for such discrimination. This observation might explain the absence of a clear peak at 1076 cm^{-1} and around the Amide II or Amide III bands in the integral of PC7 loading spectrum. The VIP scores also show that differences in milk composition according to treatment are reflected in minor milk component rather than major ones, which explains why PC7 extracted from VN FD spectral dataset revealed a strong treatment effect. It must be noted that this PLS-DA model explained only 13.68% of the variation related to treatment membership of milk samples, which implies that the treatment effect on milk composition was limited and it agrees with the fact that PC7 explained 1.3% of the variation in the long-term VN FD averaged spectral dataset.

The conclusion that cows assigned to T3 might have been experiencing increased body fat mobilization in comparison to cows assigned to T1 could be somehow corroborated by the neck injuries that were recorded during the trial. While both T1 and T3 show increased injuries on the proximal area of the cow's neck (higher portion, closest to the body), T3 (i.e., Neckline1 in the original thesis) was the only treatment out of the 4 tested to show increased injury on 2 locations of the cow's neck; indeed, T3 had an additional increase in injuries on the medial area of the cow's neck (lower portion, closest to the head). These injuries results from the cows putting pressure on their neck through repeated contact with the tie-rail, while transitioning from lying to standing positions and, possibly in an attempt to reach feed [102]. However, no difference in eating-rumination time was found between tie-rail treatments at any time point of the trial (51.5, 50, 45.8, 49.6 % of time eating-ruminating per h in the long-term for T1, T2, T3 and T4, respectively) [102]. Appearance of neck injuries at two locations of the neck indicates that T3 tie rail configuration was probably obstructing the cow access to feed, which may have resulted in possible reduced feed intake (not measured in the trial) that led to an elevated body fat mobilization. It must be noted that the effect was limited, and the milk numerical data did not indicate any clinical issues

with the cows assigned to T3, which means that the hybrid spectral analysis approach could detect the trend in an early stage before it becomes problematic.

Table 7-3 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral averages for the tie rail trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment, start and block effects that are tested in this trial.

Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Long-term			
				Cumulative Explained Variation %	P Values		
					Treatment	Start	Block (Start)
Raw	PC1	144.01	51.61	51.62	0.2897	0.0768	0.0081
	PC2	83.97	30.10	81.71	0.9753	0.2931	0.1052
	PC3	38.26	13.71	95.43	0.7750	0.0013	0.0001
	PC4	5.36	1.92	97.35	0.0836	0.7465	0.2169
	PC5	2.70	0.97	98.31	0.3495	0.1821	0.0859
FD	PC1	161.77	58.19	58.19	0.3120	0.1375	0.0091
	PC2	39.03	14.04	72.23	0.4568	0.0130	0.0010
	PC3	33.63	12.10	84.33	0.7935	0.0307	0.0388
	PC4	12.46	4.48	88.81	0.0519	0.0817	0.0071
	PC5	7.31	2.63	91.44	0.6068	0.0001	0.4963
	PC6	5.39	1.94	93.38	0.0371	0.0238	0.0464
	PC7	4.27	1.54	94.92	0.8602	0.7468	0.7219
VN Raw	PC1	189.23	68.07	68.07	0.4486	0.3247	0.0941
	PC2	62.68	22.55	90.62	0.9549	0.1570	0.0003
	PC3	11.23	4.04	94.66	0.6285	0.6290	0.1392
	PC4	6.49	2.34	96.99	0.0462	0.0695	0.0223
VN FD	PC1	161.38	58.05	58.05	0.4429	0.2729	0.1311
	PC2	52.76	18.98	77.30	0.3412	0.0794	0.0001
	PC3	17.70	6.37	83.40	0.0698	0.2443	0.0145
	PC4	10.96	3.94	87.34	0.4485	0.0001	0.0513
	PC5	7.23	2.60	89.94	0.1883	0.0031	0.1370
	PC6	5.10	1.84	91.77	0.6687	0.2201	0.8147
	PC7	3.82	1.37	93.15	0.0106	0.5590	0.0600
	PC8	3.32	1.20	94.34	0.1827	0.1467	0.3407
	PC9	3.05	1.10	95.44	0.5853	0.9014	0.3648

Table 7-4 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral averages for the tie rail trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment, start and block effects that are tested in this trial.

Short-term							
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values		
					Treatment	Start	Block (Start)
Raw	PC1	120.20	43.24	43.24	0.8027	0.2834	0.3742
	PC2	116.59	41.94	85.17	0.4505	<0.0001	0.0024
	PC3	28.43	10.23	95.40	0.9673	0.0022	0.0005
	PC4	4.70	1.69	97.09	0.7538	0.5885	0.2833
	PC5	3.42	1.23	98.32	0.4750	0.0667	0.1672
FD	PC1	143.72	51.70	51.70	0.7183	0.8642	0.2918
	PC2	53.42	19.22	70.916	0.8662	<0.0001	0.0028
	PC3	30.73	11.05	81.97	0.7711	0.0011	0.0509
	PC4	14.51	5.22	87.19	0.5312	0.0157	0.0699
	PC5	8.54	3.07	90.26	0.4634	0.0012	0.0001
	PC6	7.12	2.56	92.82	0.9830	0.4151	0.0185
	PC7	4.03	1.45	94.27	0.6678	0.1466	0.0445
	PC8	3.38	1.22	95.49	0.9792	0.2750	0.7527
VN Raw	PC1	184.33	66.305	66.305	0.8430	0.0835	0.4132
	PC2	59.13	21.27	87.57	0.8334	0.2707	0.0002
	PC3	18.65	6.70	94.28	0.3532	0.0001	0.0252
	PC4	6.05	2.18	96.46	0.3044	0.0524	0.2503
	PC5	3.00	1.08	97.54	0.5695	0.0047	0.1282
	PC6	2.14	0.77	98.31	0.3011	0.1517	0.0000
	PC7	1.42	0.51	98.82	0.2854	0.0331	0.2038
VN FD	PC1	157.19	56.54	56.54	0.8203	0.1957	0.5475
	PC2	53.57	19.26	75.80	0.9078	0.4261	0.0057
	PC3	15.28	5.50	81.30	0.6044	0.0272	0.0968
	PC4	11.74	4.22	85.52	0.3441	<0.0001	0.0070
	PC5	7.86	2.83	88.35	0.3957	0.3243	<0.0001
	PC6	6.25	2.25	90.60	0.6979	0.0064	0.1319
	PC7	5.70	2.05	92.64	0.7479	0.9305	0.9583
	PC8	4.20	1.51	94.15	0.9162	0.5622	0.9332
	PC9	2.99	1.08	95.23	0.2639	0.8408	0.2663
	PC10	2.89	1.04	96.27	0.6651	0.2540	0.3726

Table 7-5 Least squares means produced by the Mixed procedure for the scores of PC7 extracted from long-term VN FD spectral average dataset and that revealed a significant treatment effect. This table shows that T3 scores were significantly different from the other treatments.

Treatment	Estimate	Standard Error	DF	t Value	P Value
1	0.6846	0.4959	30	1.38	0.1776
2	0.5001	0.5267	30	0.95	0.3499
3	-1.4596	0.4652	30	-3.14	0.0038
4	0.4188	0.4652	30	0.9	0.3752

Table 7-6 Differences of least squares means for PC7's scores. The Scheffé adjusted P values show that T3 is significantly different from T1.

Treatment	Treatment	Estimate	Standard Error	DF	t Value	P Value	Scheffé Adj. P Value
1	2	0.1845	0.7097	30	0.26	0.7967	0.9953
1	3	2.1442	0.68	30	3.15	0.0037	0.0332
1	4	0.2658	0.68	30	0.39	0.6986	0.9845
2	3	1.9597	0.7027	30	2.79	0.0091	0.0711
2	4	0.08133	0.7027	30	0.12	0.9086	0.9996
3	4	-1.8784	0.6579	30	-2.86	0.0077	0.0622

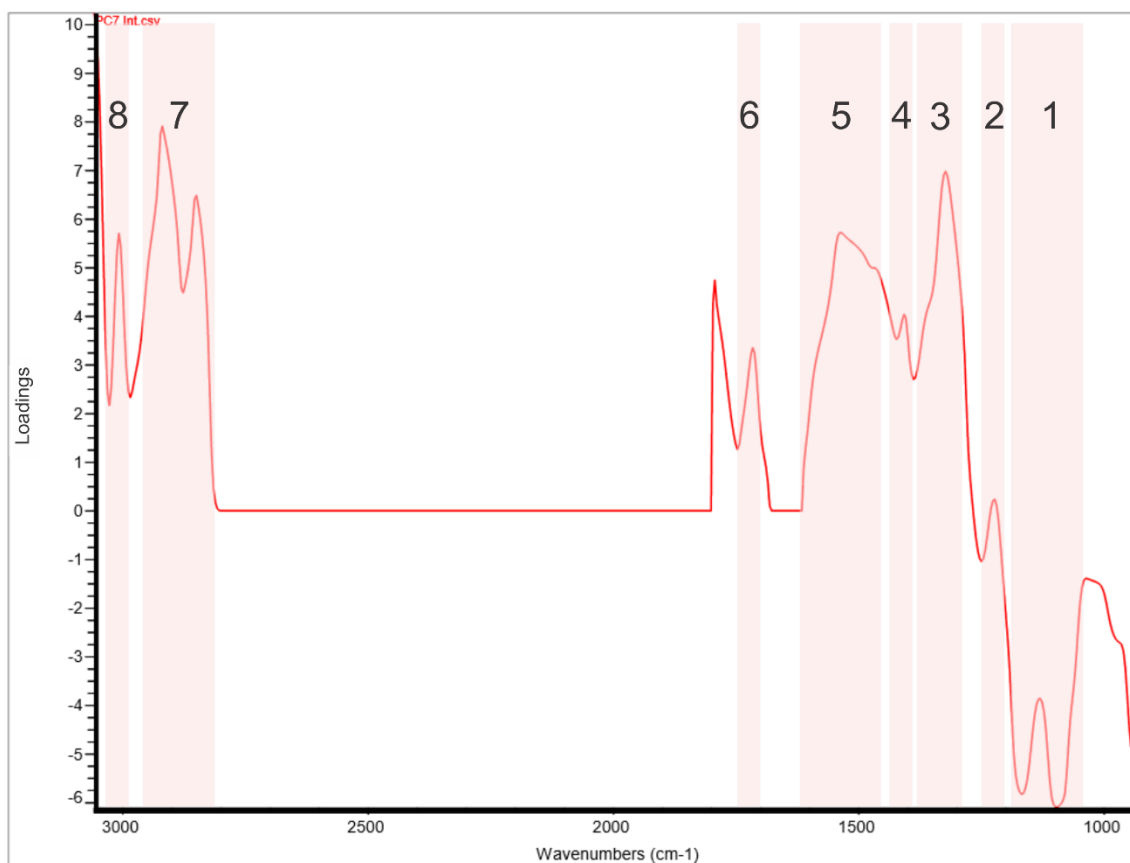


Figure 7-1 The spectral integral of PC7 loading spectrum extracted from long-term VN FD spectral average dataset for the TR trial. Shaded regions can be assigned to: 1) lactose 1200-1000 cm^{-1} , 2) acetone $\sim 1237 \text{ cm}^{-1}$, 3) citrate, BHB and acetone 1390-1250 cm^{-1} , 4) BHB $\sim 1404 \text{ cm}^{-1}$, 5) fatty acids and carboxylate ion in citrate, BHB, free fatty acids and acetate 1618-1424 cm^{-1} , 6) Carboxylic group of free fatty acids $\sim 1716 \text{ cm}^{-1}$, 7) CH stretching of fatty acids 3000-2800 cm^{-1} , 8) =C-H stretching of fatty acids $\sim 3008 \text{ cm}^{-1}$.

Table 7-7 Results of peak fitting of lumpy regions in the integral of the loading spectrum of PC7 and probable molecules that can be assigned to the resulting peaks.

Range cm^{-1}	Peak Type	Center X	Height	Area	Probable molecule
1250-1180	Voigt	1237.178	0.0228	1.6199	Acetone
	Voigt	1287.064	0.2825	26.7635	Citrate
1390 -1250	Voigt	1317.459	6.6918	738.564	BHB
	Voigt	1372.16	1.5591	204.3213	Acetone
1618-1424	Voigt	1460.422	4.1669	674.6262	Fatty acids
	Voigt	1541.455	4.2489	490.2986	Carboxylate ion in citrate, BHB, free fatty acids and acetate

Table 7-8 Variable importance for the projection scores of significant milk components obtained from parsimonious PLS-DA model for milk composition numerical data for weeks 8-10 to classify milk samples by tie rail treatment.

Milk Component	VIP Score
BHB	1.1430
C16_0	1.0903
FP Ratio	1.0418
Trans FA	1.0128
Urea	1.0006
C18_1	0.9324
MUFA	0.8762
PUFA	0.8690

Table 7-9 Milk composition data $\pm$ SD for weeks 8-10 for the tie rail trial with T1 and T2 averages by treatment.

Treatment	T1				T2			
Week	8	9	10	Avg.	8	9	10	Avg.
Fat %	4.22 $\pm$ 0.70	4.16 $\pm$ 0.71	4.10 $\pm$ 0.46	4.16 $\pm$ 0.62	4.20 $\pm$ 0.57	4.28 $\pm$ 0.63	4.16 $\pm$ 0.64	4.21 $\pm$ 0.59
Protein %	3.44 $\pm$ 0.32	3.44 $\pm$ 0.27	3.44 $\pm$ 0.26	3.44 $\pm$ 0.27	3.37 $\pm$ 0.33	3.38 $\pm$ 0.31	3.44 $\pm$ 0.31	3.39 $\pm$ 0.31
Lactose %	4.62 $\pm$ 0.12	4.63 $\pm$ 0.17	4.61 $\pm$ 0.18	4.62 $\pm$ 0.15	4.65 $\pm$ 0.16	4.65 $\pm$ 0.16	4.61 $\pm$ 0.19	4.64 $\pm$ 0.17
TS %	13.26 $\pm$ 0.77	13.20 $\pm$ 0.76	13.12 $\pm$ 0.58	13.19 $\pm$ 0.69	13.19 $\pm$ 0.70	13.29 $\pm$ 0.70	13.18 $\pm$ 0.79	13.22 $\pm$ 0.71
Urea mg/dL	14.29 $\pm$ 2.46	14.71 $\pm$ 2.90	13.90 $\pm$ 2.72	14.30 $\pm$ 2.63	14.03 $\pm$ 2.07	14.50 $\pm$ 3.50	13.54 $\pm$ 2.57	14.02 $\pm$ 2.71
BHB mmol/L	0.05 $\pm$ 0.03	0.05 $\pm$ 0.03	0.05 $\pm$ 0.03	0.05 $\pm$ 0.03	0.07 $\pm$ 0.03	0.06 $\pm$ 0.04	0.08 $\pm$ 0.03	0.07 $\pm$ 0.03
C16:0 %	1.105 $\pm$ 0.198	1.108 $\pm$ 0.197	1.081 $\pm$ 0.147	1.098 $\pm$ 0.177	1.118 $\pm$ 0.164	1.157 $\pm$ 0.231	1.105 $\pm$ 0.198	1.127 $\pm$ 0.194
C18:0 %	0.406 $\pm$ 0.071	0.379 $\pm$ 0.064	0.389 $\pm$ 0.045	0.391 $\pm$ 0.060	0.391 $\pm$ 0.046	0.392 $\pm$ 0.039	0.391 $\pm$ 0.055	0.391 $\pm$ 0.045
C18:1 %	0.912 $\pm$ 0.144	0.932 $\pm$ 0.150	0.930 $\pm$ 0.093	0.925 $\pm$ 0.127	0.864 $\pm$ 0.123	0.907 $\pm$ 0.095	0.905 $\pm$ 0.156	0.892 $\pm$ 0.124
SCFA %	0.508 $\pm$ 0.121	0.500 $\pm$ 0.127	0.485 $\pm$ 0.099	0.497 $\pm$ 0.113	0.487 $\pm$ 0.085	0.482 $\pm$ 0.082	0.477 $\pm$ 0.093	0.482 $\pm$ 0.084
MCFA %	2.013 $\pm$ 0.459	1.920 $\pm$ 0.505	1.912 $\pm$ 0.339	1.948 $\pm$ 0.428	2.062 $\pm$ 0.467	2.047 $\pm$ 0.621	2.021 $\pm$ 0.530	2.043 $\pm$ 0.524
LCFA %	1.565 $\pm$ 0.247	1.525 $\pm$ 0.236	1.542 $\pm$ 0.138	1.544 $\pm$ 0.206	1.541 $\pm$ 0.170	1.552 $\pm$ 0.141	1.550 $\pm$ 0.231	1.548 $\pm$ 0.178
SFA %	2.807 $\pm$ 0.503	2.736 $\pm$ 0.504	2.656 $\pm$ 0.330	2.733 $\pm$ 0.443	2.795 $\pm$ 0.421	2.797 $\pm$ 0.491	2.702 $\pm$ 0.434	2.764 $\pm$ 0.436
TUFA %	1.236 $\pm$ 0.180	1.209 $\pm$ 0.185	1.196 $\pm$ 0.122	1.214 $\pm$ 0.160	1.207 $\pm$ 0.146	1.214 $\pm$ 0.121	1.206 $\pm$ 0.189	1.209 $\pm$ 0.149
MUFA %	1.040 $\pm$ 0.143	1.035 $\pm$ 0.152	1.026 $\pm$ 0.110	1.033 $\pm$ 0.132	1.004 $\pm$ 0.136	1.034 $\pm$ 0.106	1.032 $\pm$ 0.158	1.023 $\pm$ 0.131
PUFA %	0.158 $\pm$ 0.031	0.150 $\pm$ 0.043	0.163 $\pm$ 0.027	0.157 $\pm$ 0.034	0.132 $\pm$ 0.040	0.144 $\pm$ 0.045	0.147 $\pm$ 0.032	0.141 $\pm$ 0.039
Trans FA %	0.094 $\pm$ 0.015	0.092 $\pm$ 0.018	0.098 $\pm$ 0.020	0.095 $\pm$ 0.017	0.104 $\pm$ 0.017	0.102 $\pm$ 0.019	0.102 $\pm$ 0.017	0.103 $\pm$ 0.017
FFA %	6.983 $\pm$ 1.744	6.678 $\pm$ 1.666	7.162 $\pm$ 1.418	6.941 $\pm$ 1.577	6.869 $\pm$ 1.374	7.076 $\pm$ 2.111	6.788 $\pm$ 1.673	6.911 $\pm$ 1.689
FP Ratio	1.23 $\pm$ 0.20	1.21 $\pm$ 0.17	1.19 $\pm$ 0.10	1.21 $\pm$ 0.16	1.25 $\pm$ 0.16	1.27 $\pm$ 0.17	1.21 $\pm$ 0.14	1.24 $\pm$ 0.15

Table 7-10 Milk composition data $\pm$ SD for weeks 8-10 for the tie rail trial with T3 and T4 averages by treatment

Treatment	T3				T4			
Week	8	9	10	Avg.	8	9	10	Avg.
Fat %	3.89 $\pm$ 0.45	3.94 $\pm$ 0.54	3.92 $\pm$ 0.49	3.91 $\pm$ 0.48	3.73 $\pm$ 0.64	3.90 $\pm$ 0.60	4.11 $\pm$ 0.56	3.91 $\pm$ 0.60
Protein %	3.34 $\pm$ 0.27	3.33 $\pm$ 0.29	3.36 $\pm$ 0.31	3.34 $\pm$ 0.28	3.41 $\pm$ 0.29	3.39 $\pm$ 0.30	3.42 $\pm$ 0.31	3.41 $\pm$ 0.29
Lactose %	4.63 $\pm$ 0.16	4.59 $\pm$ 0.20	4.59 $\pm$ 0.21	4.60 $\pm$ 0.18	4.61 $\pm$ 0.10	4.60 $\pm$ 0.11	4.59 $\pm$ 0.13	4.60 $\pm$ 0.11
TS %	12.83 $\pm$ 0.59	12.83 $\pm$ 0.70	12.85 $\pm$ 0.63	12.84 $\pm$ 0.62	12.73 $\pm$ 0.83	12.86 $\pm$ 0.84	13.09 $\pm$ 0.80	12.90 $\pm$ 0.82
Urea mg/dL	12.84 $\pm$ 2.77	13.05 $\pm$ 2.15	12.59 $\pm$ 2.13	12.83 $\pm$ 2.31	13.35 $\pm$ 1.74	14.48 $\pm$ 2.37	14.57 $\pm$ 2.34	14.13 $\pm$ 2.18
BHB mmol/L	0.06 $\pm$ 0.02	0.06 $\pm$ 0.02	0.07 $\pm$ 0.02	0.06 $\pm$ 0.02	0.06 $\pm$ 0.02	0.05 $\pm$ 0.03	0.07 $\pm$ 0.02	0.06 $\pm$ 0.03
C16:0 %	1.015 $\pm$ 0.139	1.037 $\pm$ 0.188	1.027 $\pm$ 0.174	1.026 $\pm$ 0.164	0.965 $\pm$ 0.152	1.001 $\pm$ 0.165	1.053 $\pm$ 0.167	1.006 $\pm$ 0.161
C18:0 %	0.371 $\pm$ 0.044	0.369 $\pm$ 0.051	0.378 $\pm$ 0.045	0.373 $\pm$ 0.046	0.357 $\pm$ 0.071	0.372 $\pm$ 0.068	0.404 $\pm$ 0.066	0.378 $\pm$ 0.069
C18:1 %	0.846 $\pm$ 0.074	0.896 $\pm$ 0.069	0.858 $\pm$ 0.080	0.867 $\pm$ 0.075	0.841 $\pm$ 0.140	0.935 $\pm$ 0.138	0.966 $\pm$ 0.117	0.914 $\pm$ 0.139
SCFA %	0.447 $\pm$ 0.086	0.449 $\pm$ 0.099	0.454 $\pm$ 0.083	0.450 $\pm$ 0.087	0.423 $\pm$ 0.112	0.436 $\pm$ 0.102	0.466 $\pm$ 0.099	0.442 $\pm$ 0.103
MCFA %	1.783 $\pm$ 0.374	1.786 $\pm$ 0.436	1.818 $\pm$ 0.362	1.796 $\pm$ 0.381	1.724 $\pm$ 0.517	1.710 $\pm$ 0.461	1.930 $\pm$ 0.426	1.788 $\pm$ 0.467
LCFA %	1.455 $\pm$ 0.144	1.460 $\pm$ 0.144	1.447 $\pm$ 0.148	1.454 $\pm$ 0.141	1.421 $\pm$ 0.259	1.511 $\pm$ 0.238	1.603 $\pm$ 0.214	1.512 $\pm$ 0.243
SFA %	2.570 $\pm$ 0.347	2.554 $\pm$ 0.409	2.559 $\pm$ 0.357	2.561 $\pm$ 0.361	2.465 $\pm$ 0.448	2.487 $\pm$ 0.419	2.627 $\pm$ 0.380	2.526 $\pm$ 0.411
TUFA %	1.161 $\pm$ 0.091	1.168 $\pm$ 0.102	1.156 $\pm$ 0.095	1.162 $\pm$ 0.093	1.143 $\pm$ 0.194	1.210 $\pm$ 0.169	1.261 $\pm$ 0.163	1.205 $\pm$ 0.177
MUFA %	0.962 $\pm$ 0.076	1.003 $\pm$ 0.093	0.966 $\pm$ 0.089	0.977 $\pm$ 0.086	0.958 $\pm$ 0.167	1.048 $\pm$ 0.169	1.074 $\pm$ 0.143	1.027 $\pm$ 0.163
PUFA %	0.139 $\pm$ 0.022	0.152 $\pm$ 0.038	0.142 $\pm$ 0.023	0.144 $\pm$ 0.029	0.129 $\pm$ 0.016	0.157 $\pm$ 0.037	0.156 $\pm$ 0.029	0.147 $\pm$ 0.031
Trans FA %	0.099 $\pm$ 0.016	0.096 $\pm$ 0.023	0.097 $\pm$ 0.023	0.097 $\pm$ 0.020	0.088 $\pm$ 0.020	0.092 $\pm$ 0.020	0.094 $\pm$ 0.019	0.091 $\pm$ 0.019
FFA %	6.665 $\pm$ 1.275	6.514 $\pm$ 1.171	6.527 $\pm$ 0.939	6.569 $\pm$ 1.106	6.234 $\pm$ 1.371	6.841 $\pm$ 1.831	7.279 $\pm$ 1.311	6.785 $\pm$ 1.541
FP Ratio	1.17 $\pm$ 0.16	1.18 $\pm$ 0.14	1.17 $\pm$ 0.17	1.17 $\pm$ 0.15	1.09 $\pm$ 0.17	1.15 $\pm$ 0.13	1.20 $\pm$ 0.14	1.15 $\pm$ 0.15

7.3.3 Tie chain length trial

Table 7-11 and Table 7-12 summarize the meaningful PCs that were extracted from the raw, FD, VN raw and VN FD spectral datasets of the long-term and short-term milk samples spectral averages of the tie chain length trial, respectively. PCA yielded four, seven, four and eight meaningful PCs from the raw, FD, VN raw and VN FD long-term spectral average datasets that explained 98.76%, 97.53%, 98.47% and 96.74% of the variation in the spectral datasets, respectively. These PCs, whose eigenvalue and percentage of explained variation ≥ 1 , represent the sources of variation in their respective spectral datasets that were separated from the noise and that were tested for the treatment and block effects by the SAS Mixed procedure. PC6 ($P = 0.0337$) and PC6 ($P = 0.0323$) extracted from long-term FD, and VN FD spectral average datasets, respectively, revealed significant treatment effect. The later PC will be interpreted because it yielded a smaller P value and it was extracted from VN spectra. VN eliminates procedural variability not related to chemical composition of milk samples. PC 6 extracted from the long-term VN FD spectral average dataset explains 1.70% of the variation in its respective dataset, which suggests that the treatment effect is limited, and it was detected in an early stage. In addition, this PC did not reveal a significant block effect ($P = 0.0882$).

Table 7-13 and Table 7-14 summarize the least squares means and their differences produced by the Mixed procedure for the scores of PC6 extracted from long-term VN FD spectral average dataset that revealed a significant treatment effect. These tables show that the estimates of the mean scores of PC6 for T1 and T2 are negative and positive values, respectively, which means that the influential spectral features in the loading spectrum of this PC have more intense absorbance in spectra collected for milk samples of cows assigned to T1. It was proven in the previous chapter that spectral features with more intense absorbances yield greater negative scores when PCA is applied to the FD of the spectral data. This observation means that molecules affected by the treatment are present in higher concentrations in milk samples collected from cows assigned to T1, which is the control treatment (i.e., the shorter chain length).

Inspection of the integral of PC6 loading spectrum (Figure 7-2) revealed several peaks at the following wavenumbers: 2919, 2851, 1715, 1576, 1541, 1461, 1419 and 968 cm^{-1} . The peaks 2919, 2851, 1715, 1541, 1419 and 968 cm^{-1} can be assigned to the following IR bands: the asymmetrical stretching ($\nu_{as}\text{CH}_2$) of the methylene group in fatty acids, the symmetrical stretching ($\nu_s\text{CH}_2$) of

the methylene group in fatty acids and histamine, the $C = O$ stretching vibration in the carboxyl functional group in free fatty acids, the deformation vibration band of ($= NH_2$) in the amidine functional group in creatine, the $C - O$ stretching band in acetate, the alkene trans double bond in trans fatty acids, respectively. The peak at 1576 cm^{-1} can be assigned to: the stretching of the imidazole ring in histamine, the bending vibration (δNH) of $N - H$ in hippuric acid and the symmetrical stretching of the carboxylate ion in citrate and fatty acids. The peak at 1461 cm^{-1} can be assigned to: the $C - N$ stretching band in urea, the $N - H$ bending of the NH_4^+ ion and NH_3^+ symmetrical bending ($\delta_s NH_3^+$) in histamine. Table 7-15 summarizes the results of the peak fitting process for regions $1365\text{-}1160\text{ cm}^{-1}$, which revealed peaks at 1347 , 1299 and 1249 cm^{-1} , which can be assigned to the CH_3 deformation band [118] that appears in the FTIR spectrum of the aqueous solution of acetate, the bending vibration (δCH_2) of the methylene group in hippuric acid and the $C - O$ stretching band in citrate, respectively.

The spectral analysis concludes that milk samples collected from cows assigned to T1, which is the control treatment, had higher levels of trans fatty acids, acetate, citrate and non protein nitrogen compounds (NPNs), which are urea, ammonium, hippuric acid, creatine and histamine, during the last three weeks of the trial. All these molecules, and histamine in specific, are markers of decreased pH in the rumen [17-20, 131]. This observation suggests that cows assigned to the longer chain length treatment had more stable ruminal pH and less incidences of acidotic insults in the rumen.

Acute and subacute ruminal acidosis (SARA) are associated with the accumulation of lactic acid and volatile fatty acids (VFAs), respectively, in the rumen and the subsequent decrease in the ruminal pH for several hours per day [17] that results from feeding high grain diets that are low in fiber to high yielding dairy cows under intensive livestock production systems that are adapted to digesting forage diets [18, 19]. Grain in the diet increases ammonia, VFAs, acetate, butyrate, propionate, and valerate concentrations in the rumen [17]. The decrease of the ruminal pH induces changes in the microbial flora profile of the rumen. Microorganisms that synthesize lactic acid (i.e., *Streptococcus bovis* and lactobacilli) outnumber those that utilize lactic acid (*Megasphaera elsdenii* and *Selenomonas ruminantium*) [20]. In addition, high concentrations of short-chain fatty acids (SCFA), mainly acetate, propionate, and butyrate, damage the epithelial barrier of the rumen during an acidotic insult. This damage is not immediately reversible upon the removal of the

acidotic insult, which leads to leakage of rumen metabolites to the blood serum [131]. In addition, these conditions liberates *Fusiformis necrophorus* [20] into the portal circulation, predisposing the cow to liver abscesses. The degeneration process associated with this chain of events cause release of histamine. Histamine, which has been hypothesized to have an important role in acidosis, is also produced by the decarboxylation of the amino acid histidine at low ruminal pH by the bacteria *Allisonella histaminiiformans* [17]. In addition, histamine and endotoxins are released during the decline of the ruminal pH as a result of bacteriolysis and tissue degradation [20].

Changes to milk composition are reported in cases of reduced ruminal pH. These changes include depressed milk fat [18], reduced protein and increased milk NPNs [20], such as ammonium, urea, histamine, creatine and hippuric acid [84]. In addition, decreased ruminal pH alters the biohydrogenation pathway of linoleic acid and increases the production of *trans-10* C18:1 fatty acid. Thus, more trans fatty acids are absorbed, even if the intake of unsaturated fatty acids is not necessarily high [18, 21]. In fact, the milk numerical dataset (Table 7-16) shows that long-term average of trans fatty acids for T1, which is 0.06%, is higher than the average for T2, which is 0.05%. The difference between the two averages revealed a significant treatment effect ($P = 0.0445$) when tested by the Mixed procedure. We also notice an increase and a decrease in the long-term average of urea and protein for T1 and T2, respectively. These averages were 12.28 mg/dL and 12.14 mg/dL for urea, and 3.28% and 3.34% for protein for T1 and T2, respectively. Milk fat depression is also noticed for T1 milk samples during the last week of the trial. The average fat levels for T1 and T2 were 4.01% and 4.32%, respectively, during that period. In addition, the long-term averages of SCFA were 0.35% and 0.33% for T1 and T2, respectively. The increase in SCFA, which are *de novo* synthesized in the mammary glands, might require increased amounts of the cofactor nicotinamide adenine dinucleotide phosphate (NADPH) that results from the oxidation of isocitrate to α -ketoglutaric acid in the Krebs cycle, which might explain an increase in citrate in milk samples collected from cows assigned to T1 [97].

Ruminal pH is a result of the balance between acid production from carbohydrates in the rumen and the buffering action of sodium bicarbonate that are present in saliva, which are produced during chewing or rumination [18-20]. Hence, enhanced chewing or rumination activity by the cow reduces the risk of an acidotic insult. In the tie chain length trial, the time spent outside of the stall by the withers of the cows assigned to the longer chain treatment was significantly greater in

comparison to cows assigned to the recommended length (11 ± 1.1 vs 7 ± 1.1 % of daily time; $P = 0.05$) [114], and it significantly increased between short- and long-term measurements ($+ 3$ % of daily time; $P \leq 0.05$) [114]. This measurement indicates that cows assigned to longer chains were spending more time outside the stall perimeter in the manger area. In addition, the distance outside of stall perimeter for the withers, which represents the average distance outside of the stall in the manger area, increased significantly between the short- and long-term ($+ 0.9$ cm; $P \leq 0.05$) for both treatments [114]; however, this measurement did not differ between treatments. These observations suggest that cows assigned to the longer chain treatment were spending more time at the manger. Since feeding is the assumed usage of that area, we may assume that cows assigned to longer chain treatment might have been chewing more; hence, they were producing more saliva to balance the ruminal pH, which reduces the risk of an acidotic insult. It must be noted that no detailed measures for the manger space usage were collected for this trial. In addition, the combined “eating/rumination time” measurement did not reveal significant differences among the treatments. However, the device that was used for this measurement was not capable of recording the times of the two actions separately. Currently, the only diagnostic test for subclinical acidosis is ruminal pH [20]. In this trial, milk FTIR spectroscopy showed promising results in rapidly detecting metabolites in milk related to rumen acidotic insults in their early stages.

Table 7-11 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral averages for the tie chain length trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.

Long-term						
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values	
					Treatment	Block
Raw	PC1	158.97	57.19	57.19	0.9253	0.0955
	PC2	80.81	29.07	86.25	0.1965	0.2271
	PC3	32.06	11.53	97.78	0.4082	0.5837
	PC4	2.70	0.97	98.76	0.3985	0.0461
FD	PC1	188.03	67.64	67.64	0.9370	0.0788
	PC2	33.77	12.15	79.79	0.3728	0.5928
	PC3	28.94	10.41	90.20	0.2021	0.1266
	PC4	8.67	3.19	93.31	0.9431	0.6860
	PC5	5.63	2.03	95.34	0.3739	0.0198
	PC6	3.76	1.35	96.69	0.0337	0.0679
	PC7	2.33	0.84	97.53	0.6143	0.2491
VN Raw	PC1	210.88	75.86	75.86	0.7279	0.1006
	PC2	51.15	18.40	94.26	0.6815	0.3665
	PC3	8.54	3.07	97.33	0.3112	0.3810
	PC4	3.18	1.14	98.47	0.4405	0.1283
VN FD	PC1	188.40	67.77	67.77	0.8630	0.1848
	PC2	43.33	15.59	83.36	0.6391	0.2985
	PC3	12.88	4.63	87.99	0.3567	0.1941
	PC4	8.05	2.90	90.89	0.9436	0.0205
	PC5	5.17	1.86	92.75	0.7679	0.3388
	PC6	4.75	1.70	94.46	0.0323	0.0882
	PC7	3.40	1.22	95.68	0.2885	0.1233
	PC8	2.96	1.06	96.74	0.8024	0.4088

Table 7-12 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral averages for the tie chain length trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.

Short-term						
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values	
					Treatment	Block
Raw	PC1	149.34	53.72	53.72	0.3465	0.1303
	PC2	88.36	31.79	85.50	0.7148	0.0335
	PC3	34.31	12.34	97.84	0.3370	0.4137
	PC4	3.05	1.20	98.94	0.5802	0.5166
FD	PC1	179.27	64.49	64.49	0.3585	0.0959
	PC2	36.08	12.98	77.47	0.2820	0.4186
	PC3	35.44	12.75	90.21	0.7909	0.0741
	PC4	10.16	3.65	93.87	0.8219	0.9010
	PC5	4.47	1.61	95.48	0.9004	0.4182
	PC6	2.83	1.02	96.50	0.0731	0.5845
	PC7	2.70	0.97	97.47	0.1404	0.5390
VN Raw	PC1	204.06	73.67	73.67	0.4274	0.2042
	PC2	60.05	21.68	95.35	0.3757	0.0510
	PC3	6.13	2.22	97.56	0.4904	0.2932
VN FD	PC1	184.41	66.34	66.33	0.5209	0.2876
	PC2	52.02	18.71	85.05	0.2414	0.0613
	PC3	9.26	3.33	88.37	0.1178	0.2708
	PC4	7.50	2.70	91.07	0.3463	0.2496
	PC5	5.37	1.93	93.00	0.9029	0.9686
	PC6	4.09	1.47	94.48	0.0543	0.7157
	PC7	3.33	1.20	95.67	0.5632	0.2725
	PC8	3.03	1.09	96.76	0.4896	0.0189

Table 7-13 Least squares means produced by the Mixed procedure for the scores of PC6 extracted from long-term VN FD spectral average dataset and that revealed a significant treatment effect.

Treatment	Estimate	Standard Error	DF	t Value	P Value
1	-0.8393	0.4293	9	-1.96	0.0823
2	0.8426	0.5079	9	1.66	0.1315

Table 7-14 Differences of least squares means for PC6's scores. The Scheffé adjusted P value shows that T1 is significantly different from T2.

Treatment	Treatment	Estimate	Standard Error	DF	t Value	P Value	Scheffé Adj. P Value
1	2	-1.6819	0.6650	9	-2.53	0.0323	0.0323

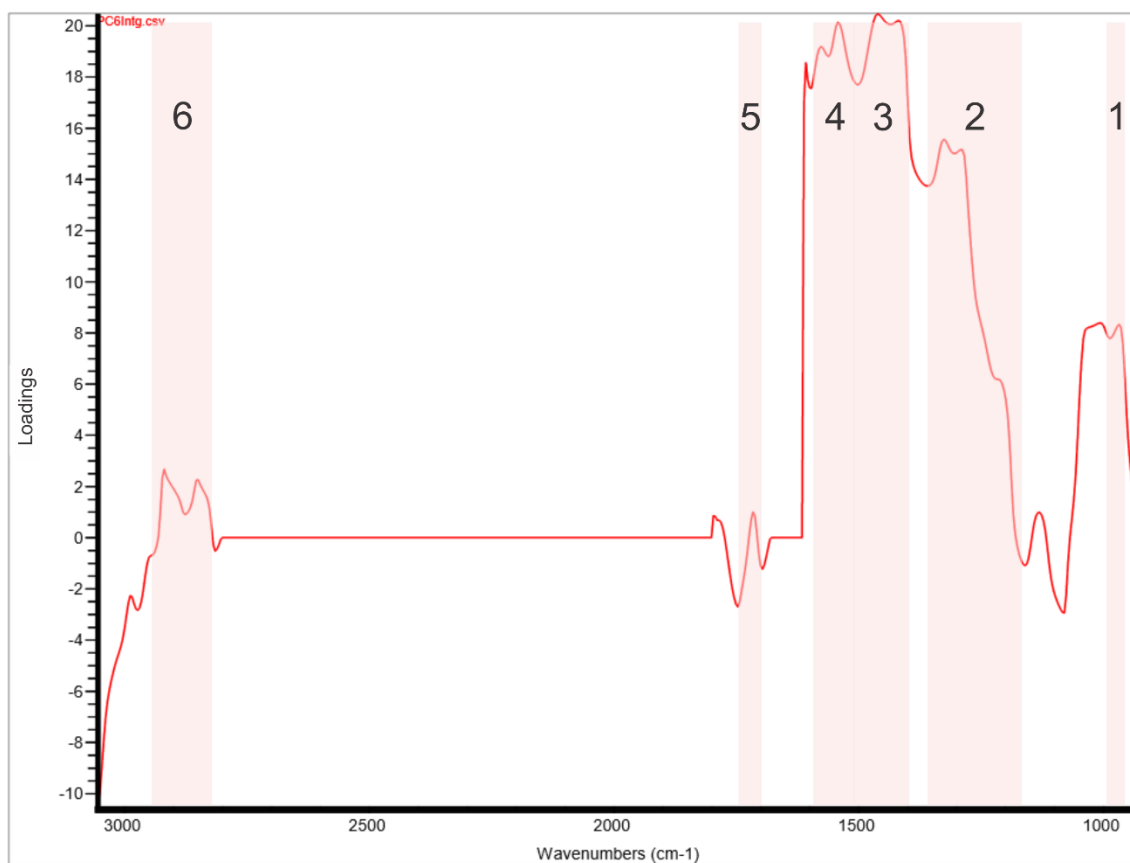


Figure 7-2 The spectral integral of PC6 loading spectrum extracted from long-term VN FD spectral average dataset for the TCL trial. Shaded regions can be assigned to: 1) trans fatty acids 968 cm^{-1} , 2) citrate, hippuric acid and acetate 1365-1160 cm^{-1} , 3) acetate 1419 cm^{-1} , fatty acids and NPN (i.e., urea, ammonium and histamine) 1461 cm^{-1} , 4) creatine 1541 cm^{-1} , histamine, hippuric acid, citrate and fatty acids 1576 cm^{-1} , 5) Carboxylic group of free fatty acids ~ 1715 cm^{-1} , 6) CH stretching of fatty acids 3000-2800 cm^{-1} .

Table 7-15 Results of peak fitting of region 1365-1160 cm^{-1} in the integral of the loading spectrum of PC6 and probable molecules that can be assigned to the resulting peaks.

Range cm^{-1}	Peak Type	Center X	Height	Area	Probable molecule
1365-1160	Voigt	1249.745	3.5993	50.9682	Citrate
	Voigt	1299.439	7.107	60.9523	Hippuric acid
	Voigt	1347.338	10.3636	65.2514	Acetate

Table 7-16 Milk composition data $\pm$ SD for weeks 8-10 for the tie chain trial with long-term averages by treatment.

Treatment	T1				T2			
Week	8	9	10	Avg.	8	9	10	Avg.
Fat %	3.70 $\pm$ 0.68	3.88 $\pm$ 0.71	4.01 $\pm$ 0.56	3.86 $\pm$ 0.65	3.69 $\pm$ 0.61	3.70 $\pm$ 0.64	4.32 $\pm$ 1.56	3.87 $\pm$ 0.99
Protein %	3.25 $\pm$ 0.36	3.33 $\pm$ 0.36	3.27 $\pm$ 0.38	3.28 $\pm$ 0.36	3.31 $\pm$ 0.29	3.34 $\pm$ 0.32	3.36 $\pm$ 0.48	3.34 $\pm$ 0.35
Lactose %	4.61 $\pm$ 0.15	4.63 $\pm$ 0.16	4.64 $\pm$ 0.18	4.62 $\pm$ 0.16	4.58 $\pm$ 0.21	4.59 $\pm$ 0.19	4.54 $\pm$ 0.20	4.57 $\pm$ 0.19
TS %	12.53 $\pm$ 0.93	12.81 $\pm$ 0.99	12.89 $\pm$ 0.84	12.75 $\pm$ 0.91	12.56 $\pm$ 0.90	12.61 $\pm$ 0.96	13.19 $\pm$ 1.92	12.76 $\pm$ 1.27
Urea mg/dL	11.96 $\pm$ 2.35	12.62 $\pm$ 1.73	12.27 $\pm$ 1.76	12.28 $\pm$ 1.93	11.72 $\pm$ 1.47	12.24 $\pm$ 2.02	12.53 $\pm$ 2.40	12.14 $\pm$ 1.92
BHB mmol/L	0.11 $\pm$ 0.01	0.08 $\pm$ 0.02	0.09 $\pm$ 0.02	0.09 $\pm$ 0.02	0.12 $\pm$ 0.02	0.10 $\pm$ 0.02	0.13 $\pm$ 0.09	0.11 $\pm$ 0.05
C14:0 %	0.431 $\pm$ 0.094	0.463 $\pm$ 0.104	0.457 $\pm$ 0.078	0.450 $\pm$ 0.091	0.434 $\pm$ 0.097	0.455 $\pm$ 0.095	0.469 $\pm$ 0.144	0.451 $\pm$ 0.109
C16:0 %	1.321 $\pm$ 0.287	1.406 $\pm$ 0.329	1.430 $\pm$ 0.253	1.386 $\pm$ 0.287	1.357 $\pm$ 0.301	1.379 $\pm$ 0.295	1.545 $\pm$ 0.562	1.419 $\pm$ 0.384
C18:0 %	0.366 $\pm$ 0.045	0.384 $\pm$ 0.044	0.414 $\pm$ 0.042	0.388 $\pm$ 0.047	0.362 $\pm$ 0.041	0.361 $\pm$ 0.060	0.448 $\pm$ 0.170	0.386 $\pm$ 0.104
C18:1 %	0.697 $\pm$ 0.108	0.749 $\pm$ 0.086	0.788 $\pm$ 0.101	0.745 $\pm$ 0.103	0.672 $\pm$ 0.093	0.686 $\pm$ 0.091	0.916 $\pm$ 0.488	0.747 $\pm$ 0.282
SCFA %	0.341 $\pm$ 0.100	0.349 $\pm$ 0.105	0.357 $\pm$ 0.090	0.349 $\pm$ 0.096	0.327 $\pm$ 0.087	0.327 $\pm$ 0.094	0.330 $\pm$ 0.111	0.328 $\pm$ 0.093
MCFA %	1.673 $\pm$ 0.447	1.726 $\pm$ 0.460	1.753 $\pm$ 0.374	1.717 $\pm$ 0.418	1.706 $\pm$ 0.358	1.646 $\pm$ 0.405	1.949 $\pm$ 0.943	1.754 $\pm$ 0.587
LCFA %	1.269 $\pm$ 0.197	1.341 $\pm$ 0.169	1.429 $\pm$ 0.176	1.346 $\pm$ 0.188	1.233 $\pm$ 0.143	1.214 $\pm$ 0.167	1.598 $\pm$ 0.763	1.330 $\pm$ 0.444
Saturated FA %	2.522 $\pm$ 0.493	2.614 $\pm$ 0.552	2.667 $\pm$ 0.429	2.601 $\pm$ 0.484	2.506 $\pm$ 0.474	2.511 $\pm$ 0.475	2.746 $\pm$ 0.828	2.576 $\pm$ 0.583
Total Unsaturated FA %	0.828 $\pm$ 0.137	0.854 $\pm$ 0.120	0.911 $\pm$ 0.127	0.864 $\pm$ 0.129	0.821 $\pm$ 0.109	0.789 $\pm$ 0.116	1.065 $\pm$ 0.553	0.879 $\pm$ 0.320
Monounsaturated FA %	0.684 $\pm$ 0.114	0.720 $\pm$ 0.097	0.757 $\pm$ 0.109	0.720 $\pm$ 0.108	0.663 $\pm$ 0.096	0.657 $\pm$ 0.098	0.850 $\pm$ 0.424	0.715 $\pm$ 0.246
Polyunsaturated FA %	0.086 $\pm$ 0.028	0.118 $\pm$ 0.023	0.124 $\pm$ 0.022	0.110 $\pm$ 0.029	0.084 $\pm$ 0.027	0.115 $\pm$ 0.030	0.121 $\pm$ 0.047	0.106 $\pm$ 0.037
Trans FA %	0.058 $\pm$ 0.014	0.059 $\pm$ 0.010	0.064 $\pm$ 0.014	0.060 $\pm$ 0.013	0.049 $\pm$ 0.010	0.050 $\pm$ 0.012	0.048 $\pm$ 0.020	0.049 $\pm$ 0.014
FFA %	5.958 $\pm$ 1.217	7.563 $\pm$ 1.432	7.555 $\pm$ 1.067	7.025 $\pm$ 1.433	8.254 $\pm$ 3.089	9.138 $\pm$ 3.327	12.835 $\pm$ 10.514	9.879 $\pm$ 6.269

7.3.4 Stall width trial

Table 7-17 and Table 7-18 summarize the meaningful PCs that were extracted from the raw, FD, VN raw and VN FD spectral datasets of the long-term and short-term milk samples spectral datasets of the stall width trial, respectively. PCA yielded five, nine, five and nine meaningful PCs from the raw, FD, VN raw and VN FD long-term spectral datasets that explained 98.45%, 98.21, 98.28% and 98.08% of the variation in the spectral datasets, respectively. These PCs, whose eigenvalue and percentage of explained variation ≥ 1 , represent the sources of variation in their respective spectral datasets that were separated from the noise and that were tested for the treatment and block effects by the SAS Mixed procedure. PC5 ($P = 0.0423$) extracted from long-term VN Raw spectral dataset revealed significant treatment effect; however, it also revealed a significant block effect ($P = 0.0008$). For this reason, the interpretation of the loading spectrum of this PC should take into consideration all factors accounted for by the blocking variable, which are parity and days in milking (DIM). PC 5 extracted from the long-term VN Raw spectral dataset explains 1.13% of the variation in its respective dataset, which suggests that the treatment and the block effects are limited. In addition, this PC will be interpreted directly without being integrated because it was extracted from a raw spectral dataset. This fact implies that the treatment and the block effects are mostly related to changes in major milk components.

Table 7-20 and Table 7-21 summarize the least squares means and their differences produced by the Mixed procedure for the scores of PC5 extracted from long-term VN Raw spectral dataset that revealed significant treatment and block effects. These tables show that the estimates of the mean scores of PC5 for T1 and T2 are positive and negative values, respectively, which means that the influential spectral features in the loading spectrum of this PC have more intense absorbance in spectra collected for milk samples of cows assigned to T1. It was proven in the previous chapter that spectral features with more intense absorbances yield greater positive scores when PCA is applied to the raw spectral data. This observation means that molecules affected by the treatment and block effects are present in higher concentrations in milk samples collected from cows assigned to T1, which is the control treatment (i.e., the single width stall).

Inspection of PC5 loading spectrum (Figure 7-3) revealed peaks at the following wavenumbers: 2920, 2855, 1721 and 1205 cm^{-1} . The peaks 2920, 2855 and 1721 cm^{-1} are assigned to the asymmetrical stretching ($\nu_{as}CH_2$) of the methylene group in fatty acids of milk fat, the

symmetrical stretching ($\nu_s CH_2$) of the methylene group in fatty acids of milk fat and the $C = O$ stretching vibration of ester linkage in milk fat, respectively. The region that shows the greatest loadings is between $\sim 1250\text{ cm}^{-1}$ and $\sim 1140\text{ cm}^{-1}$ with a peak centered at $\sim 1205\text{ cm}^{-1}$. This region can be assigned to the methylene twisting and wagging vibrations of fatty acids, esters [116] and BHB, whose FTIR spectrum of the aqueous solution shows an absorption band centered at 1207 cm^{-1} (Figure 7-4). In addition, the $C - O$ stretching vibration of the $C - C(=O) - O$ band of saturated esters shows strong absorption in the 1210 cm^{-1} to 1163 cm^{-1} region [116], which is observed in the spectra of milk samples with increasing fat content (Figure 7-5). Milk fat is predominantly triglycerides, which are esters derived from glycerol and three fatty acids.

Table 7-22 summarizes the results of the peak fitting process for region $1100\text{-}940\text{ cm}^{-1}$, which revealed peaks at 1063 and 940 cm^{-1} that appears in the FTIR spectrum of the aqueous solution of BHB (Table 7-2).

The spectral analysis concludes that milk samples collected from cows assigned to T1, which is the control treatment, had higher levels of milk fat triglycerides, milk fatty acids and BHB, which is a precursor for the *de novo* synthesis of fatty acids in the mammary glands [97]. In this case, BHB is not considered as an indicator of negative energy balance due to the absence of features in PC5 loading spectrum that can be assigned to other markers of this metabolic issue. This observation suggests that cows assigned to T1 were synthesizing more fatty acids and milk fat over the long-term of the trial. This conclusion is supported by the numerical data set of milk components during the last week of the trial (Table 7-23). The average of all reported milk components was higher for T1 during that period except for urea and trans fatty acids.

In addition to nutrition and herd management practices, several factors affect milk composition. These factors are region, season, breed, individuality, age, disease, diurnal rhythm, stage of lactation and parity [132]. The trial was conducted in the same premises during one season, all cows were of the same breed, which was Holstein, and collected milk samples combined portions of morning and evening milking to eliminate the variations in milk composition related to diurnal rhythm. Blocking was implemented in the experimental design of the trial to account for variations related to the remaining factors, which are stage of lactation, parity and age (Table 7-24). However, milk sample of cow 2057 from block 8 in T2 was missing on week 6 of the trial. This cow was the only primiparous one in T2. In addition, T2 had a cow in its seventh parity (i.e., cow 7097); while

the greatest parity in T1 was the fifth (i.e., cow 419). The missed milk sample of cow 2057 rendered the median parity of cows of T2 greater than that of T1, which are 3 and 2, respectively. Hence, cows of T2 were older and one lactation higher than those of T1. Milk fat and protein decline as the animal becomes older. Milk fat falls about 0.2% each year from the first to fifth lactation as a result of higher production and more udder infections; while protein decreases 0.02 to 0.05% each lactation as animals age [133]. This fact is reflected in the numerical milk data where the averages of fat were 3.96% and 3.76% and the averages of protein were 3.34% and 3.28% for T1 and T2, respectively. In fact, the loading spectrum of PC5 extracted from long-term VN Raw spectral dataset shows increased loadings between 1565 cm^{-1} and 1520 cm^{-1} which corresponds to the Amide II band of proteins in milk FTIR spectrum [22]. This observation explains the strong block effect ($P = 0.0008$) and the treatment effect ($P = 0.0423$) on PC5. In this case, the treatment effect might have originated from the imbalance of age and parity in blocks of T1 and T2 that resulted from the missing milk sample of cow 2057 from the long-term dataset. The effect of a greater age and parity in T2 might have overshadowed any effect on milk composition related to the stall width. The hybrid spectral analysis was applied to the spectral datasets of week 5 and none of the extracted PCs revealed a significant treatment effect (Table 7-19). In this trial, the hybrid spectral analysis could detect effects of multiple factors on the FTIR spectra of milk samples and provided explanation related to specific experimental details of the trial.

Table 7-17 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the long-term milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.

Long-term						
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values	
					Treatment	Block
Raw	PC1	146.50	52.70	52.70	0.6103	0.3639
	PC2	76.97	27.69	80.39	0.5711	0.2751
	PC3	40.09	14.42	94.81	0.5856	0.1811
	PC4	6.78	2.44	97.25	0.7496	0.7710
	PC5	3.36	1.21	98.45	0.5696	0.3811
FD	PC1	148.33	53.36	53.36	0.4548	0.3200
	PC2	46.01	16.55	69.91	0.5975	0.1332
	PC3	30.35	10.92	80.83	0.6819	0.6276
	PC4	20.25	7.29	88.11	0.9316	0.5952
	PC5	10.42	3.75	91.86	0.6152	0.0134
	PC6	6.76	2.43	94.29	0.2942	0.8353
	PC7	4.72	1.70	95.99	0.2918	0.6883
	PC8	3.56	1.28	97.27	0.1399	0.5155
	PC9	2.62	0.94	98.21	0.6917	0.7075
VN Raw	PC1	199.72	71.84	71.84	0.6957	0.4815
	PC2	50.71	18.24	90.08	0.5269	0.2258
	PC3	14.78	5.32	95.40	0.4743	0.4272
	PC4	4.87	1.75	97.15	0.4372	0.2245
	PC5	3.13	1.13	98.28	0.0423	0.0008
VN FD	PC1	175.90	63.27	63.27	0.6769	0.4643
	PC2	38.64	13.90	77.17	0.6043	0.1012
	PC3	21.69	7.80	84.98	0.6936	0.7748
	PC4	13.90	5.00	89.98	0.1414	0.0035
	PC5	6.340	2.28	92.26	0.2300	0.6657
	PC6	5.20	1.87	94.13	0.9493	0.9771
	PC7	4.64	1.67	95.80	0.0916	0.5014
	PC8	3.61	1.30	97.09	0.6612	0.1740
	PC9	2.74	0.99	98.08	0.3364	0.2808

Table 7-18 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial.

Short-term						
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values	
					Treatment	Block
Raw	PC1	150.38	54.09	54.09	0.3546	0.0176
	PC2	92.18	33.16	87.25	0.8075	0.2866
	PC3	23.83	8.57	95.82	0.3476	0.1208
	PC4	6.54	2.35	98.18	0.3555	0.3405
	PC5	3.13	1.13	99.30	0.4765	0.9212
FD	PC1	167.78	60.35	60.35	0.4435	0.0196
	PC2	35.34	12.71	73.07	0.5139	0.2447
	PC3	30.83	11.09	84.16	0.3807	0.4828
	PC4	18.35	6.60	90.76	0.2811	0.3418
	PC5	8.10	2.91	93.67	0.7191	0.6465
	PC6	5.33	1.91	95.59	0.1686	0.8037
	PC7	4.09	1.47	97.06	0.1488	0.1938
VN Raw	PC1	213.76	76.89	76.89	0.4997	0.0614
	PC2	37.90	13.63	90.53	0.4937	0.2029
	PC3	15.46	5.56	96.09	0.3052	0.2461
	PC4	4.74	1.70	97.79	0.0743	0.0171
	PC5	3.27	1.18	98.97	0.3028	0.9457
VN FD	PC1	191.97	69.05	69.05	0.4677	0.0563
	PC2	30.13	10.84	79.89	0.1613	0.3021
	PC3	18.44	6.63	86.52	0.8666	0.3497
	PC4	11.05	3.98	90.50	0.3685	0.2558
	PC5	7.58	2.73	93.23	0.2124	0.8537
	PC6	5.48	1.97	95.20	0.2010	0.5612
	PC7	3.33	1.20	96.40	0.7711	0.8458
	PC8	2.78	1.00	97.40	0.1893	0.8713

Table 7-19 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of week 5 milk spectral datasets for the stall width trial. The table also lists P values obtained from the SAS Mixed Procedure for the treatment and block effects that are tested in this trial

Week 5						
Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Cumulative Explained Variation %	P Values	
					Treatment	Block
Raw	PC1	177.50	63.85	63.85	0.2285	0.0800
	PC2	57.88	20.82	84.67	0.1899	0.0630
	PC3	34.32	12.35	97.01	0.2589	0.1503
	PC4	5.26	1.89	98.91	0.9269	0.7909
FD	PC1	167.09	60.10	60.10	0.4310	0.0898
	PC2	44.69	16.08	76.18	0.2321	0.2646
	PC3	29.51	10.61	86.79	0.2594	0.2298
	PC4	17.97	6.47	93.26	0.6512	0.7352
	PC5	4.58	1.65	94.91	0.1013	0.5006
	PC6	3.49	1.25	96.16	0.9563	0.5699
	PC7	2.92	1.05	97.21	0.8166	0.7843
VN Raw	PC1	224.92	80.91	80.91	0.2122	0.0704
	PC2	36.82	13.24	94.15	0.2888	0.2928
	PC3	7.45	2.68	96.82	0.9588	0.5065
	PC4	4.39	1.58	98.41	0.9383	0.4886
VN FD	PC1	195.07	70.17	70.17	0.2110	0.0756
	PC2	35.07	12.61	82.78	0.2931	0.3356
	PC3	16.10	5.79	88.57	0.9989	0.5778
	PC4	7.311	2.63	91.20	0.8035	0.8637
	PC5	6.98	2.51	93.72	0.1216	0.0818
	PC6	5.07	1.82	95.54	0.1011	0.3773
	PC7	3.29	1.18	96.72	0.0645	0.7708
	PC8	2.75	0.99	97.71	0.2212	0.5114

Table 7-20 Least squares means produced by the Mixed procedure for the scores of PC5 extracted from long-term VN Raw spectral average dataset and that revealed a significant treatment effect.

Treatment	Estimate	Standard Error	DF	t Value	P Value
1	0.2842	1.1704	1	0.24	0.8483
2	-0.3179	1.1740	1	-0.27	0.8317

Table 7-21 Differences of least squares means for PC5's scores. The Scheffé adjusted P value shows that T1 is significantly different from T2.

Treatment	Treatment	Estimate	Standard Error	DF	t Value	P Value	Scheffé Adj. P Value
1	2	0.6021	0.2224	5.01	2.71	0.0423	0.0423

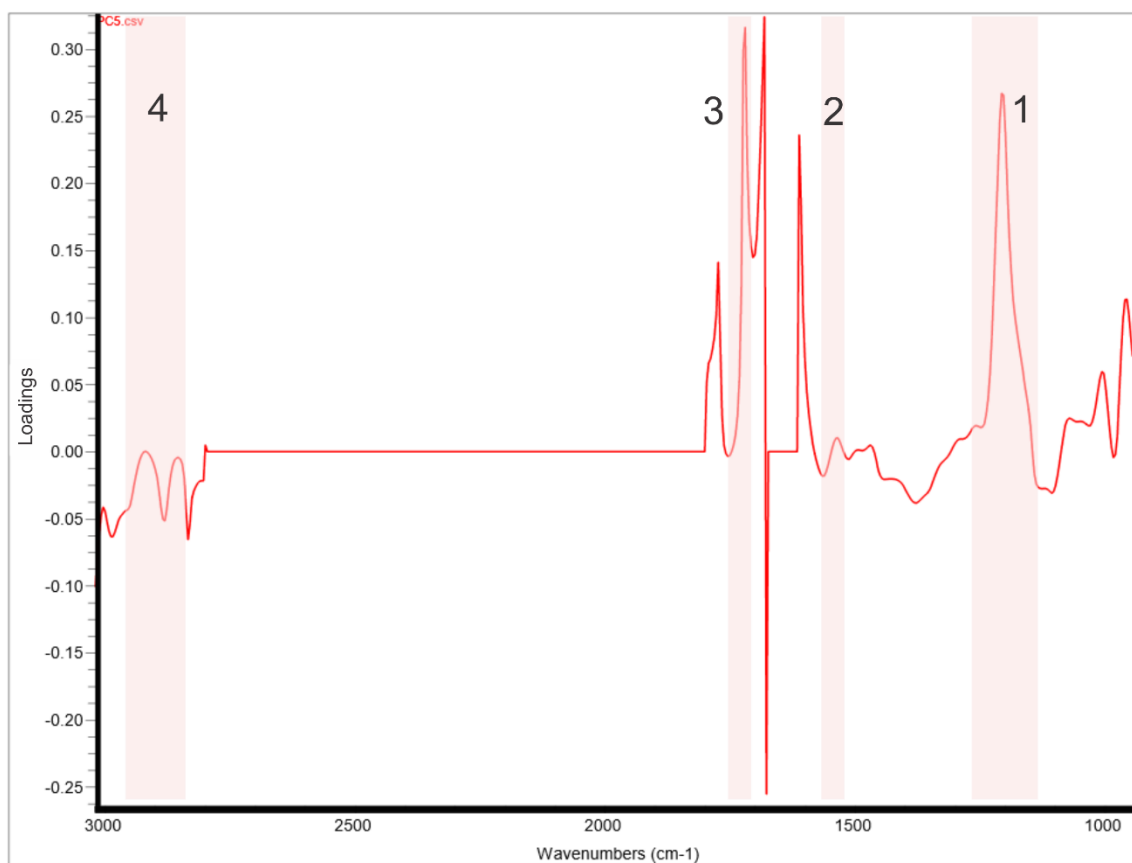


Figure 7-3 The loading spectrum of PC5 extracted from long-term VN Raw spectral dataset for the stall width trial. Shaded regions can be assigned as follows: 1) 1250-1140 cm^{-1} to the methylene twisting and wagging vibrations of fatty acids, esters and BHB, and to the C – O stretching vibration of the C – C(=O) – O band of saturated esters, 2) 1565-1520 cm^{-1} to the Amide II band of milk proteins, 3) 1721 cm^{-1} to the C = O stretching vibration of ester linkage in milk fat, 4) 3000-2840 cm^{-1} to the asymmetrical stretching ($\nu_{\text{as}}\text{CH}_2$) and the symmetrical stretching ($\nu_{\text{s}}\text{CH}_2$) of the methylene group in fatty acids of milk fat.

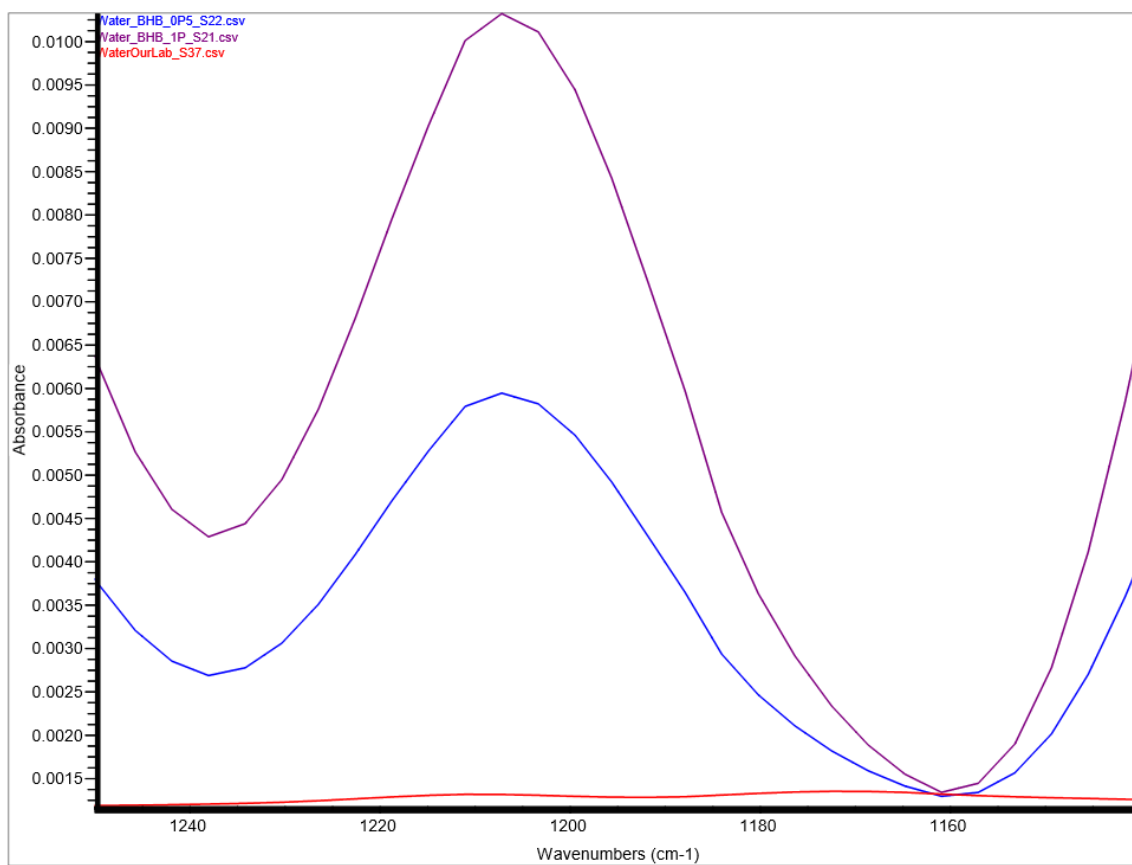


Figure 7-4 FTIR spectra of BHB aqueous solutions. Red: water, blue: 0.5% aqueous solution of BHB, purple: 1% aqueous solution of BHB. The spectra show an absorption band centred at 1207 cm⁻¹, which can be assigned to the methylene twisting and wagging vibrations.

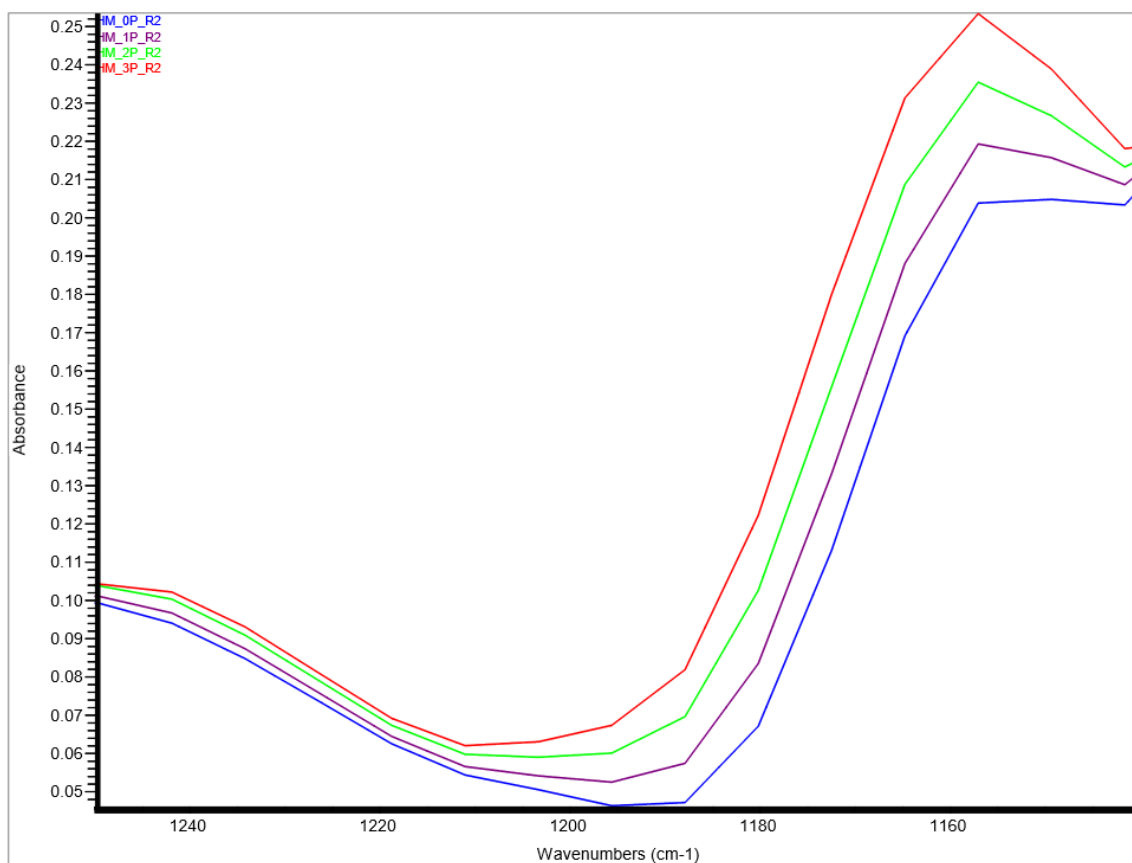


Figure 7-5 FTIR spectra of Milk samples with different fat concentrations. Blue: skim milk, purple: 1% fat, green: 2% fat, red: 3.25% fat. The C – O stretching vibration of the C – C(=O) – O band of saturated esters shows strong absorption in the 1210 cm^{-1} to 1163 cm^{-1} region.

Table 7-22 Results of peak fitting of region 1100-940 cm^{-1} in the loading spectrum of PC5 and probable molecules that can be assigned to the resulting peaks.

Range cm^{-1}	Peak Type	Center X	Height	Area	Probable molecule
1100-940	Voigt	944.001	0.0773	11.105	BHB
	Voigt	1063.754	0.0058	0.8025	BHB

Table 7-23 Average $\pm$ SD milk composition data for week 6 for the stall width trial.

Treatment	T1	T2
Fat %	3.96 $\pm$ 0.54	3.76 $\pm$ 0.37
Protein %	3.34 $\pm$ 0.22	3.28 $\pm$ 0.18
Lactose %	4.57 $\pm$ 0.11	4.56 $\pm$ 0.19
TS %	12.85 $\pm$ 0.64	12.57 $\pm$ 0.43
Urea mg/dL	12.85 $\pm$ 1.22	13.07 $\pm$ 1.76
BHB mmol/L	0.13 $\pm$ 0.02	0.13 $\pm$ 0.04
C14:0 %	0.450 $\pm$ 0.084	0.417 $\pm$ 0.083
C16:0 %	1.074 $\pm$ 0.183	1.005 $\pm$ 0.159
C18:0 %	0.377 $\pm$ 0.054	0.357 $\pm$ 0.016
C18:1 %	0.805 $\pm$ 0.134	0.804 $\pm$ 0.064
SCFA %	0.349 $\pm$ 0.054	0.314 $\pm$ 0.040
MCFA %	1.860 $\pm$ 0.219	1.700 $\pm$ 0.152
LCFA %	1.086 $\pm$ 0.156	1.061 $\pm$ 0.066
Saturated FA %	2.577 $\pm$ 0.407	2.410 $\pm$ 0.309
Total unsaturated FA %	1.135 $\pm$ 0.148	1.095 $\pm$ 0.082
Monounsaturated FA %	0.944 $\pm$ 0.142	0.894 $\pm$ 0.076
Polyunsaturated FA %	0.157 $\pm$ 0.023	0.149 $\pm$ 0.025
Trans FA %	0.056 $\pm$ 0.017	0.058 $\pm$ 0.012
FFA %	8.725 $\pm$ 2.402	8.366 $\pm$ 2.800

Table 7-24 Block design on week 6 of the stall width trial. E: early lactation stage 0-100 days in milk, M: mid lactation stage 100-200 days in milk. ^a Oldest cow among subjects of T2. ^b cow with a missing milk sample.

Treatment	Block	Cow ID	Lactation	Lactation stage	Median lactations per treatment
T1	1	5322	2	E	2.00
	2	5300	2	M	
	3	419	5	M	
	4	5275	3	M	
	5	5294	2	M	
	6	6294	4	M	
	7	5308	2	M	
	8	2063	1	E	
T2	1	2029	2	E	3.00
	2	5296	2	M	
	3	7097^a	7	M	
	4	5327	2	M	
	5	5283	3	M	
	6	5258	3	M	
	7	5279	3	M	
	8	2057^b	1	E	

7.3.5 Manger wall and stall length trial

The manger wall and stall length trial was conducted under a crossover design in which each subject was exposed to the two treatment levels within each row (i.e., T1 and T2), which means each subject contributed two observations to the collected data. However, this fact did not prevent the application of PCA as a dimensionality reduction method to the collected data in crossover experimental design because PCA will not be used to model any effects. There have been published studies in which PCA was applied as a dimensionality reduction method to experimental data collected with a crossover experimental design [134, 135]. In this trial, the hybrid spectral analysis approach that was developed in the previous chapter proved to be compatible with different experimental designs that can be analyzed by mixed models that include randomized complete

block designs, repeated measures and crossover designs, which makes this approach a versatile tool for spectral analysis in controlled-design trials.

Table 7-25 and Table 7-26 summarize the meaningful PCs that were extracted from the raw, FD, VN raw and VN FD spectral datasets of the long-term and short-term milk samples spectral datasets of the stall length and manger wall trial, respectively. PCA yielded five, eight, four and eight meaningful PCs from the raw, FD, VN raw and VN FD long-term spectral datasets that explained 98.47%, 95.53%, 97.01% and 94.24% of the variation in the spectral datasets, respectively. These PCs, whose eigenvalue and percentage of explained variation ≥ 1 , represent the sources of variation in their respective spectral datasets that were separated from the noise and that were tested for the stall length, sequence, block, manger wall height treatment on the stall length treatment and period effects by the SAS Mixed procedure. The manger wall height treatment and its combined effect with the stall length did not reveal any significant effect on the short-term and long-term spectral datasets, which means that the manger wall and stall length treatment did not affect milk composition during the trial. On the other hand, PC6 ($P = 0.0355$) extracted from long-term FD spectral dataset revealed significant length effect, which means that the stall length had a significant effect on milk composition over the long-term of the trial. This PC explains 1.77% of the variation in its respective dataset, which suggests that the length effect is limited. The loading spectrum of PC6 had to be integrated before interpretation because it was isolated from FD spectral dataset. In the following discussion, we will refer to the short and long stalls as L1 and L2, respectively, to distinguish them from the two treatments (i.e., T1 and T2), which represent the combined effect of the stall length and manger wall height.

Table 7-27 and Table 7-28 summarize the least squares means and their differences produced by the Mixed procedure for the scores of PC6 extracted from long-term FD spectral dataset that revealed a significant length effect. These tables show that the estimates of the mean scores of PC6 for L1 and L2 are negative and positive values, respectively, which means that the influential spectral features in the loading spectrum of this PC have more intense absorbance in spectra collected for milk samples of cows assigned to L1. It was proven in the previous chapter that spectral features with more intense absorbances yield greater negative scores when PCA is applied to the FD of the spectral data. This observation means that molecules affected by the stall length

are present in higher concentrations in milk samples collected from cows assigned to L1, which is the shorter stall length.

Inspection of the spectral integral of PC6 loading spectrum (Figure 7-6) revealed peaks at the following wavenumbers: 1575, 1460, 1408, 1033, 980 and 946 cm^{-1} . The peak at $\sim 1575 \text{ cm}^{-1}$ is observed in the FTIR spectrum of histamine aqueous solution and it can be assigned to the stretching of the imidazole ring. The peak at $\sim 1460 \text{ cm}^{-1}$ is observed in the FTIR spectra of milk samples spiked with urea and ammonium; hence, it can be assigned to the $C - N$ stretching band in urea or to the $N - H$ bending band in ammonium. The peak at $\sim 1408 \text{ cm}^{-1}$ is observed in milk samples spiked with BHB and it can be assigned to the $C - O - H$ bending band or the symmetrical stretching of carboxylate ion. The peaks at ~ 1033 and 980 cm^{-1} are observed in the FTIR spectra of milk samples spiked with histamine. The peak at $\sim 946 \text{ cm}^{-1}$ is observed in FTIR spectra of aqueous solutions of histamine and BHB. In addition, 980 cm^{-1} and 946 cm^{-1} has already been assigned to bound Ca and organic P in milk [58].

The spectral analysis concludes that milk samples collected from cows assigned to L1 (i.e., the short length row) had higher levels of BHB, urea, ammonium and histamine during the long-term of each period of the trial. The last three compounds are considered milk NPNs. Milk numeric data supports the conclusion regarding BHB and urea. Over the long-term, the averages for period 1 of BHB content were 0.06 mmol/L and 0.05 mmol/L for L1 and L2, respectively, and 0.05 mmol/L and 0.04 mmol/L for L1 and L2, respectively, for period 2. The averages for period 1 of urea content were 13.02 mg/dL and 11.79 mg/dL for L1 and L2, respectively, and 14.01 mg/dL and 12.43 mg/dL for L1 and L2, respectively, for period 2.

Histamine in milk originates from the blood serum [84] and its elevated levels may be attributed to the corium tissue breakdown (i.e., resulting in skin lesions or injuries) or stress [20]. Injury severity decreased at several different locations on the cows over time, regardless of treatment. It has been found that cows had 4-8 times less contacts with the tie-rail while they were rising in long stalls regardless of the manger height but only while comparing to short stalls with low manger, which may have led to possible reduction of injuries on the cow's neck [115]. However, reduction in injuries on the cow's neck was greater for cows in short stalls with high manger while compared to low manger. The key finding on the outcome measures of welfare comparing long stalls to short stalls is the drastic increase of 1 h per day in lying time [115]. Increased lying time

of 1 h per day for cows assigned to L2 (i.e., the longer stalls) indicates a more comfortable environment; thus, may explain the reduced histamine levels in the blood.

Another source of elevated histamine in the blood is protein degradation that is associated with necrotic diseases such as mastitis and metritis [20]. The averages of somatic cell count (SCC) for L1 and L2 were 281,435 and 133,136 cells/mL, respectively, in milk samples collected during the long-term of both periods. In Canadian Holstein cows, SCC greater than 200,000 cells/mL is considered a sign of mastitis for cows that are more than 30 DIM [15]. In addition, the average SCC for L1 (i.e., 281,435 cells/mL) exceeds the geometric mean of SCC for the province of Quebec, which is 215,000 cells/mL [136]. If these numbers are broken down by period, the averages of SCC for L1 and L2 become 528,273 cells/mL and 203,700 cells/mL, respectively, for period 2. This observation indicates that cows assigned to shorter stalls were releasing more histamine due to udder inflammation into the blood than those assigned to longer stalls, which corroborates the findings of the spectral analysis. Nevertheless, no solid conclusion can be made regarding the relationship between the stall length, milk composition and susceptibility to mastitis in this trial. In fact, due to a strict maintenance of stall and bedding, no difference in stall cleanliness or dryness, or in udder health were found between treatments and overtime [115]. While the literature reports experimental studies that find that longer stalls increase udder dirtiness [137], observational ones did not confirm this relationship [138-140]. The dynamics of infectious microorganisms' proliferation on dairy farms is complex and the stall length and manger wall trial was not designed to consider factors that have significant effect on the proliferation of mastitis, such as management practices [137].

Table 7-25 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of long-term milk spectral datasets for the stall length and manger wall trial. The table also lists P values obtained from the SAS Mixed Procedure for the effects that were tested in this trial.

Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Long-term					
				Cumulative Explained Variation %	P Values				
					length	seq(length)	block	trt(length)	period
Raw	PC1	158.84	57.14	57.14	0.4466	0.9078	0.1115	0.7793	0.1256
	PC2	71.90	25.86	83.00	0.4065	0.6420	0.0131	0.4777	<0.0001
	PC3	35.59	12.80	95.80	0.5914	0.7760	0.0934	0.4070	0.0004
	PC4	4.03	1.45	97.25	0.6814	0.2743	0.5259	0.5997	0.0001
	PC5	3.39	1.22	98.47	0.0590	0.1678	0.1449	0.5110	<0.0001
FD	PC1	172.42	62.02	62.02	0.3954	0.9197	0.1148	0.8642	0.5032
	PC2	46.40	16.69	78.71	0.3227	0.5131	0.0241	0.1711	<0.0001
	PC3	22.43	8.07	86.78	0.7148	0.9572	0.0801	0.7467	<0.0001
	PC4	7.87	2.83	89.61	0.7260	0.1368	0.1738	0.8487	<0.0001
	PC5	5.60	2.01	91.62	0.6060	0.4096	0.9410	0.7316	0.0119
	PC6	4.92	1.77	93.39	0.0355	0.0375	0.3748	0.2756	0.9115
	PC7	3.10	1.12	94.51	0.4702	0.8323	0.9112	0.6305	0.5355
	PC8	2.84	1.02	95.53	0.8848	0.0693	0.0389	0.1293	0.8755
VN Raw	PC1	210.37	75.67	75.67	0.3895	0.9458	0.0986	0.9735	0.6202
	PC2	45.96	16.53	92.20	0.6187	0.5421	0.0906	0.2465	<0.0001
	PC3	9.65	3.47	95.68	0.2230	0.4455	0.3281	0.5177	<0.0001
	PC4	3.70	1.33	97.01	0.1589	0.1703	0.6535	0.7369	0.0004
VN FD	PC1	187.51	67.45	67.45	0.4192	0.9179	0.1040	0.9781	0.0414
	PC2	35.97	12.94	80.39	0.3945	0.6005	0.0570	0.6406	0.0001
	PC3	11.32	4.07	84.46	0.1020	0.2641	0.0631	0.8256	<0.0001
	PC4	8.02	2.89	87.34	0.5462	0.4502	0.1249	0.4358	0.5410
	PC5	6.69	2.41	89.75	0.7541	0.0830	0.9953	0.6990	0.2240
	PC6	5.28	1.90	91.65	0.3356	0.8266	0.7761	0.2235	0.1408
	PC7	3.84	1.38	93.03	0.5208	0.5644	0.9576	0.3690	0.2010
	PC8	3.35	1.21	94.24	0.1591	0.0027	0.0480	0.0581	0.5281

Table 7-26 Principal components extracted from the raw, FD, VN raw, VN FD spectral datasets of the short-term milk spectral datasets for the stall length and manger wall trial. The table also lists P values obtained from the SAS Mixed Procedure for the effects that were tested in this trial.

Spectral Dataset	Meaningful PC	Eigenvalue	Explained Variation %	Short-term					
				Cumulative Explained Variation %	P Values				
					length	seq(length)	block	trt(length)	period
Raw	PC1	144.33	51.92	51.92	0.2579	0.9280	0.1675	0.8666	0.0023
	PC2	92.20	33.17	85.08	0.5433	0.9861	0.0220	0.4867	<0.0001
	PC3	33.65	12.11	97.19	0.2322	0.7154	0.0460	0.9156	0.0001
	PC4	2.81	1.01	98.20	0.5377	0.0816	0.7792	0.2083	0.5328
FD	PC1	166.18	59.78	59.78	0.2150	0.9438	0.1940	0.8597	0.0435
	PC2	48.58	17.47	77.25	0.0822	0.5589	0.0026	0.8058	<0.0001
	PC3	32.09	11.54	88.79	0.4677	0.8917	0.0544	0.4873	0.0112
	PC4	8.40	3.02	91.82	0.2212	0.3661	0.8604	0.9767	0.6371
	PC5	6.43	2.31	94.13	0.6158	0.2777	0.0890	0.8287	<0.0001
	PC6	3.77	1.36	95.49	0.0894	0.9774	0.4045	0.6288	0.6860
VN Raw	PC1	204.74	73.65	73.65	0.2443	0.9570	0.2631	0.5677	0.1116
	PC2	56.44	20.30	93.95	0.2362	0.7380	0.0399	0.8788	0.0041
	PC3	6.43	2.31	96.27	0.9240	0.1253	0.9148	0.0894	<0.0001
	PC4	3.64	1.31	97.57	0.2230	0.1684	0.1441	0.5018	<0.0001
VN FD	PC1	179.34	64.51	64.51	0.2171	0.9720	0.2728	0.4520	0.0042
	PC2	45.02	16.19	80.70	0.1509	0.8364	0.0400	0.7609	0.0857
	PC3	16.05	5.77	86.48	0.9141	0.2355	0.4883	0.9741	<0.0001
	PC4	8.68	3.12	89.60	0.2601	0.6865	0.8022	0.9715	0.1621
	PC5	6.10	2.19	91.80	0.2197	0.9643	0.1640	0.4886	0.8574
	PC6	4.62	1.66	93.46	0.1530	0.0150	0.0799	0.9530	0.0932
	PC7	3.23	1.16	94.62	0.1905	0.0663	0.2297	0.7606	0.2904
	PC8	2.91	1.05	95.67	0.9910	0.4457	0.2886	0.2040	0.3811

Table 7-27 Least squares means produced by the Mixed procedure for the scores of PC6 extracted from long-term FD spectral average dataset and that revealed a significant length effect.

Length	Estimate	Standard Error	DF	t Value	P Value
L1	-0.6226	0.4187	33	-1.49	0.1465
L2	0.6924	0.4339	33	1.60	0.1201

Table 7-28 Differences of least squares means for PC6's scores. The Scheffé adjusted P value shows that L1 is significantly different from L2

Length	Length	Estimate	Standard Error	DF	t Value	P Value	Scheffé Adj. P Value
L1	L2	-1.3150	0.5998	33	-2.19	0.0355	0.0355

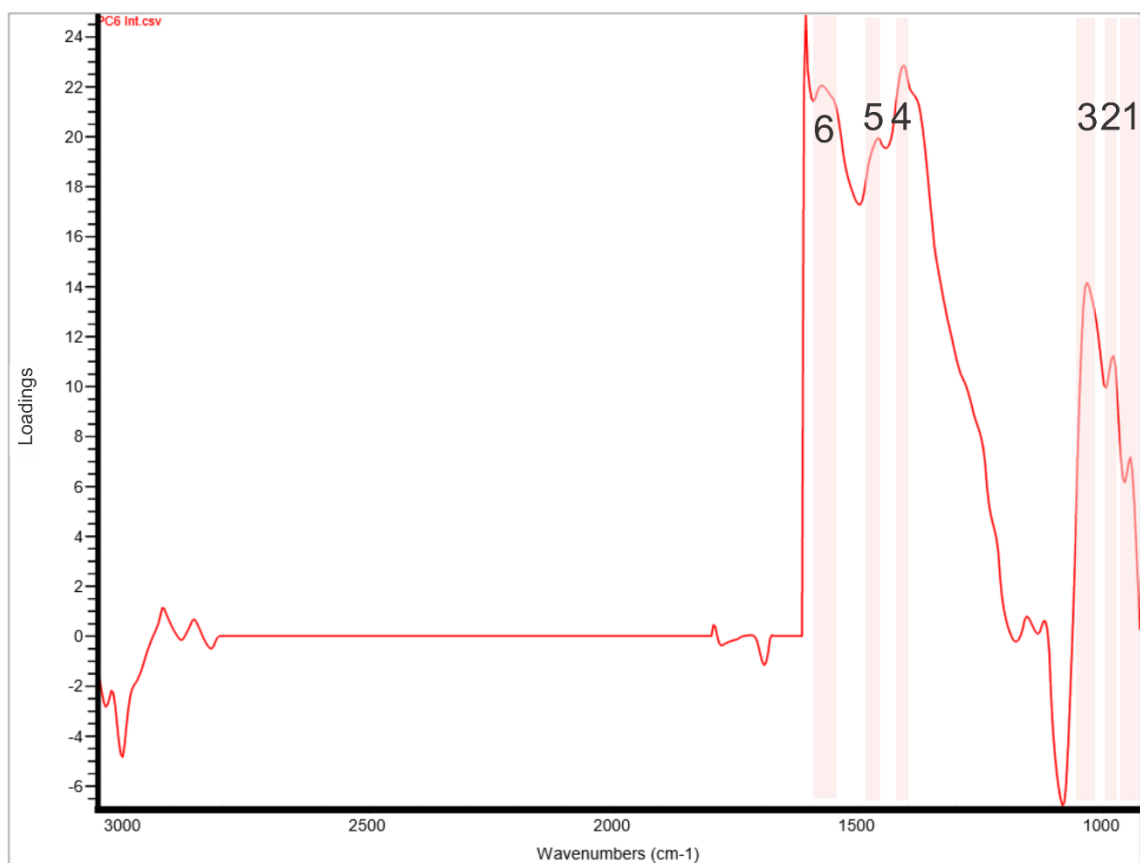


Figure 7-6 The loading spectrum of PC6 extracted from long-term FD spectral dataset for the stall length and manger wall trial. Shaded regions can be assigned as follows: 1-3) peaks at ~ 1033 , 980 and 946 cm^{-1} are found in FTIR spectra of milk samples spiked with histamine and BHB and their aqueous solutions, 4) peak at $\sim 1408\text{ cm}^{-1}$ is observed in FTIR spectra of milk samples spiked with BHB, 5) peak at $\sim 1460\text{ cm}^{-1}$ is observed in FTIR spectra of milk samples spiked with urea and ammonium, 6) peak at $\sim 1575\text{ cm}^{-1}$ is observed in FTIR spectrum of histamine aqueous solution.

Table 7-29 Milk composition average $\pm$ SD data for week 6 for the stall length and manger wall trial.

Row	L1		L2		Average by length	
Period	P1	P2	P1	P2	L1	L2
Fat %	3.85 $\pm$ 0.87	3.92 $\pm$ 0.71	4.06 $\pm$ 0.75	4.44 $\pm$ 0.48	3.89 $\pm$ 0.78	4.23 $\pm$ 0.66
Protein %	3.34 $\pm$ 0.29	3.45 $\pm$ 0.26	3.45 $\pm$ 0.28	3.65 $\pm$ 0.22	3.39 $\pm$ 0.27	3.54 $\pm$ 0.27
Lactose %	4.67 $\pm$ 0.11	4.65 $\pm$ 0.19	4.61 $\pm$ 0.18	4.60 $\pm$ 0.21	4.66 $\pm$ 0.15	4.60 $\pm$ 0.19
Urea mg/dL	13.03 $\pm$ 1.72	14.01 $\pm$ 2.21	11.79 $\pm$ 3.32	12.43 $\pm$ 2.46	13.50 $\pm$ 1.99	12.08 $\pm$ 2.91
BHB mmol/L	0.06 $\pm$ 0.02	0.05 $\pm$ 0.01	0.05 $\pm$ 0.02	0.04 $\pm$ 0.02	0.06 $\pm$ 0.02	0.05 $\pm$ 0.02
C14:0 %	0.444 $\pm$ 0.132	0.464 $\pm$ 0.109	0.468 $\pm$ 0.099	0.515 $\pm$ 0.053	0.453 $\pm$ 0.119	0.489 $\pm$ 0.083
C16:0 %	1.114 $\pm$ 0.349	1.224 $\pm$ 0.254	1.236 $\pm$ 0.288	1.401 $\pm$ 0.170	1.167 $\pm$ 0.305	1.311 $\pm$ 0.251
C18:0 %	0.299 $\pm$ 0.054	0.263 $\pm$ 0.035	0.309 $\pm$ 0.054	0.294 $\pm$ 0.049	0.281 $\pm$ 0.048	0.302 $\pm$ 0.051
C18:1 %	0.791 $\pm$ 0.093	0.701 $\pm$ 0.106	0.811 $\pm$ 0.144	0.800 $\pm$ 0.126	0.748 $\pm$ 0.107	0.806 $\pm$ 0.133
SCFA %	0.377 $\pm$ 0.110	0.354 $\pm$ 0.073	0.377 $\pm$ 0.079	0.392 $\pm$ 0.059	0.366 $\pm$ 0.093	0.384 $\pm$ 0.069
MCFA %	1.736 $\pm$ 0.540	1.909 $\pm$ 0.441	1.925 $\pm$ 0.411	2.171 $\pm$ 0.252	1.819 $\pm$ 0.492	2.037 $\pm$ 0.362
LCFA %	1.109 $\pm$ 0.148	1.049 $\pm$ 0.159	1.125 $\pm$ 0.163	1.154 $\pm$ 0.145	1.080 $\pm$ 0.153	1.138 $\pm$ 0.152
SFA %	2.611 $\pm$ 0.689	2.664 $\pm$ 0.511	2.752 $\pm$ 0.533	2.978 $\pm$ 0.319	2.636 $\pm$ 0.597	2.855 $\pm$ 0.453
UFA %	1.051 $\pm$ 0.151	1.059 $\pm$ 0.189	1.102 $\pm$ 0.213	1.235 $\pm$ 0.169	1.055 $\pm$ 0.166	1.163 $\pm$ 0.202
MUFA %	0.897 $\pm$ 0.120	0.810 $\pm$ 0.127	0.920 $\pm$ 0.155	0.933 $\pm$ 0.134	0.855 $\pm$ 0.128	0.926 $\pm$ 0.143
PUFA %	0.149 $\pm$ 0.026	0.167 $\pm$ 0.015	0.149 $\pm$ 0.025	0.178 $\pm$ 0.022	0.158 $\pm$ 0.023	0.162 $\pm$ 0.027
TFA %	0.027 $\pm$ 0.020	0.086 $\pm$ 0.031	0.023 $\pm$ 0.026	0.094 $\pm$ 0.025	0.055 $\pm$ 0.039	0.055 $\pm$ 0.044
FFA %	5.912 $\pm$ 1.742	4.618 $\pm$ 1.565	5.572 $\pm$ 1.068	4.770 $\pm$ 1.070	5.293 $\pm$ 1.751	5.207 $\pm$ 1.120
De novo FA %	1.029 $\pm$ 0.304	1.028 $\pm$ 0.219	1.076 $\pm$ 0.237	1.173 $\pm$ 0.142	1.029 $\pm$ 0.260	1.120 $\pm$ 0.201
Mixed FA %	1.308 $\pm$ 0.379	1.354 $\pm$ 0.271	1.404 $\pm$ 0.334	1.525 $\pm$ 0.185	1.330 $\pm$ 0.326	1.459 $\pm$ 0.277
Preformed FA %	1.158 $\pm$ 0.151	0.940 $\pm$ 0.174	1.190 $\pm$ 0.183	1.091 $\pm$ 0.163	1.054 $\pm$ 0.194	1.145 $\pm$ 0.177
True Pro %	3.12 $\pm$ 0.29	3.25 $\pm$ 0.26	3.24 $\pm$ 0.29	3.45 $\pm$ 0.21	3.18 $\pm$ 0.28	3.33 $\pm$ 0.27

7.4 Conclusion

A hybrid spectral analysis approach was developed to be applied to milk FTIR spectral data that combined PCA and mixed modeling to detect a treatment effect in the context of controlled-design dairy cattle welfare trials. This approach retained the multivariate structure of the FTIR data and was applicable to different experimental designs that can be analyzed by mixed models, such as randomized complete block designs, repeated measures and crossover designs. In the tie-rail and chain length trials, the experimental design isolated the treatment effect (i.e., housing modification) from the effects of other factors in the trial and the hybrid spectral analysis approach determined the sources of variation in milk composition that related to treatment effect. In the tie rail and chain length trials, PCs explaining 1.37% and 1.70% of the variation in milk FTIR data, respectively, revealed significant treatment effects. The loading spectra of these PCs revealed spectral features of molecules related to elevated body fat mobilization and reduced ruminal pH in the tie rail and chain length trials, respectively. These findings were corroborated by the data that was collected during the trials (i.e., behavioural data) and by the milk numeric data, which did not include all the molecules revealed by the spectral features in the loading spectrum of the PCs in question. In the stall width trial, a missing milk sample had a detrimental effect on the balance of the blocking factor, which rendered the median parity of cows of the double width stall treatment greater by one lactation cycle than that of the single width treatment. This imbalance in the blocking factor was mainly reflected on milk fat and other milk components in milk numerical data and the loading spectrum of the PC that revealed the block and treatment effect. In this case, the hybrid spectral analysis approach could detect effects of multiple factors on the FTIR spectra of milk samples and provided explanation related to specific experimental details of the trial. The stall length and manger wall trial had a crossover experimental design. The hybrid approach revealed that cows on longer stalls had less histamine in their milk samples which resulted from an increase in lying time which states a great improvement in cow comfort at her stall. It can be concluded that the hybrid approach for spectral analysis provides an innovative tool for studying animal welfare from a novel prospective that focuses on the repercussions of animal comfort on its biochemical process and their effects on milk synthesis and its fine composition.

Acknowledgment

The author thanks the following people for their contribution to this project:

- Dr. Elsa Vasseur (McGill University) as the PI of this project.
- Dr. Roger Cue (McGill University) for his guidance on statistical analysis by mixed models.
- Jessica St John (McGill University) conducted the tie rail animal trial and developed the SAS code to detect the treatment effect on individual milk components by the mixed procedure as part of her M. Sc. Thesis [102].
- Véronique Boyer (McGill University) conducted the chain length and stall width animal trials and developed the SAS codes to detect the treatment effect on individual milk components by the mixed procedure as part of her M. Sc. Thesis [114].
- Sarah McPherson (McGill University) conducted the manger wall and stall length animal trial and developed the SAS codes to detect the treatment effect on individual milk components by the mixed procedure as part of her M. Sc. Thesis [115].
- Tania Wolfe (McGill University) for her guidance on the trials and their statistical model.
- Dr. Ashraf Ismail (McGill University), Dr. Salwa Karboune (McGill University) and Dr. Raj Duggavathi (McGill University) for providing different chemical standards.
- Dr. Jacqueline Sedman (McGill University) reviewed the band assignment of chemical standards reported in this study.
- Valacta Inc. for analyzing milk samples collected for this project.
- Dr. Débora Santschi (Valacta) for converting the spectral files of the FOSS analyzer from the FSS proprietary format to the TXT format.

Funding

The animal trial was funded by Vasseur's NSERC-Novalait-Dairy Farmers of Canada-Valacta/Lactanet Industrial Research Chair on Sustainable Life of Dairy Cattle. Complementary funding for the study (physiological indicators) was provided by NSERC Collaborative Research Development Grant (Project no. CRDPJ 499520 16) & FRQNT-CRIBIQ-Novalait through Programme de recherche en partenariat pour l'innovation en production et transformation laitières (Project no. EPI 2016-138) and Valacta/Lactanet (Principal Investigator: Vasseur).

Appendix

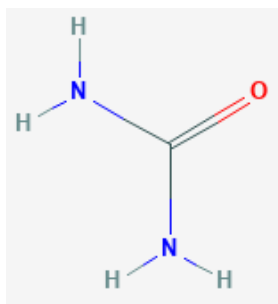


Figure 7-7 Urea [141]

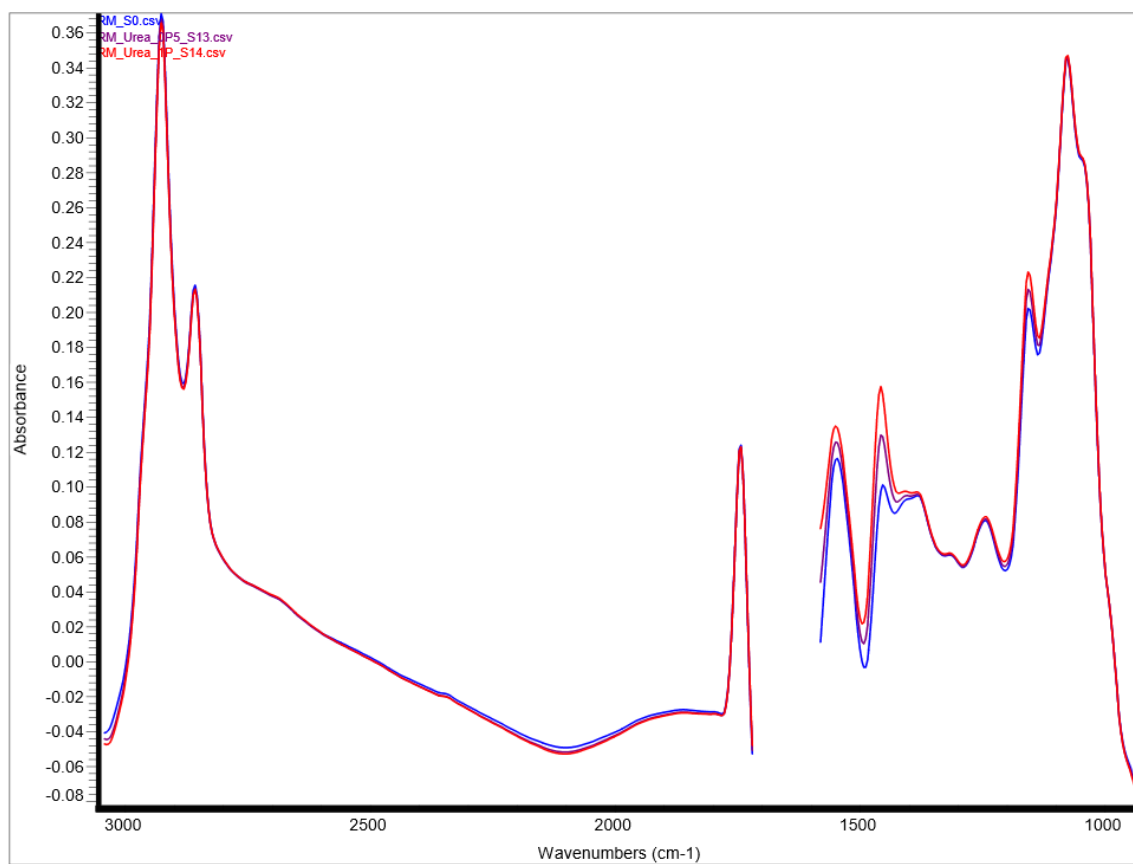


Figure 7-8 Raw milk spiked with urea. Blue: raw milk, purple: raw milk spiked with 0.5% urea, red: raw milk spiked with 1% urea.

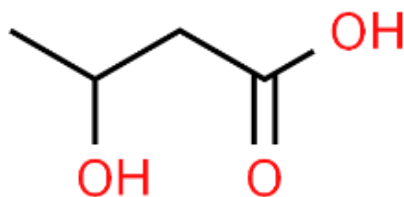


Figure 7-9 β -Hydroxybutyric acid (BHBA or BHB) [142]

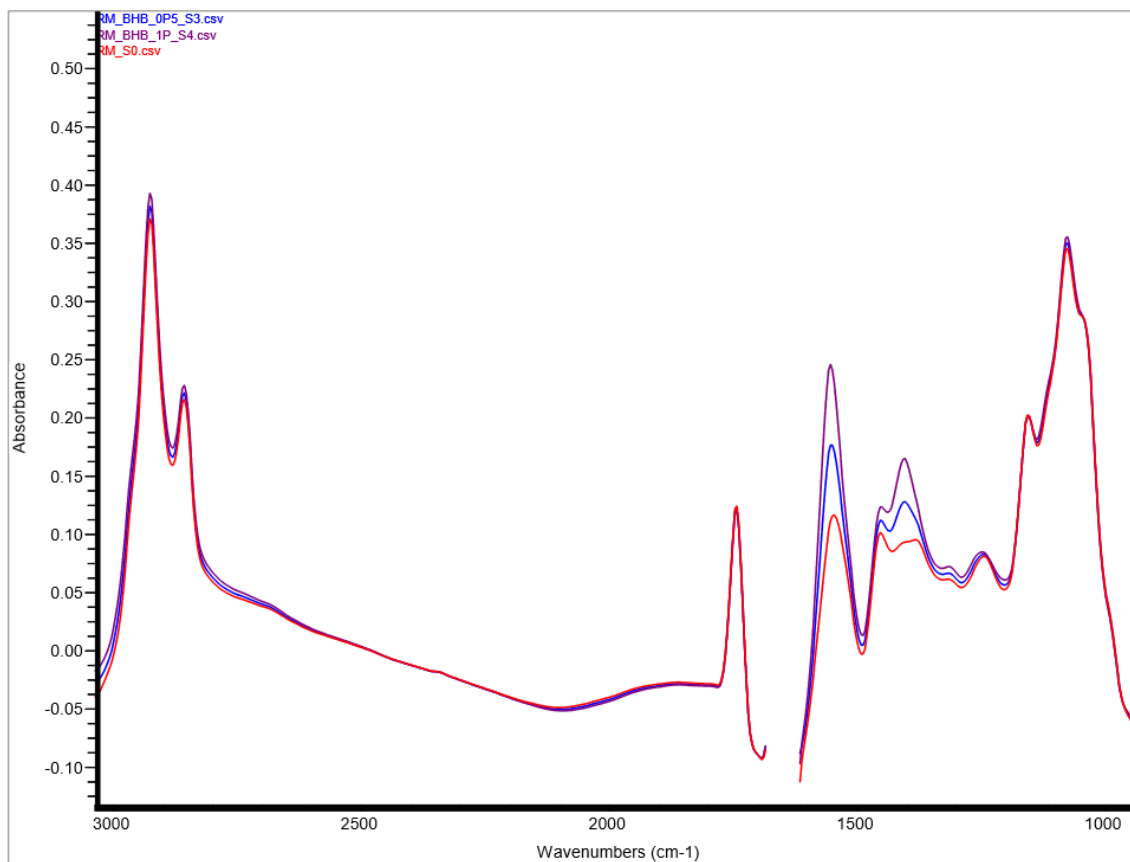


Figure 7-10 Raw milk spiked with BHB standard. Red: raw milk, blue: raw milk spiked with 0.5% BHB, purple: raw milk spiked with 1% BHB.

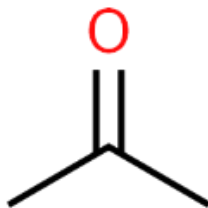


Figure 7-11 Acetone [143]

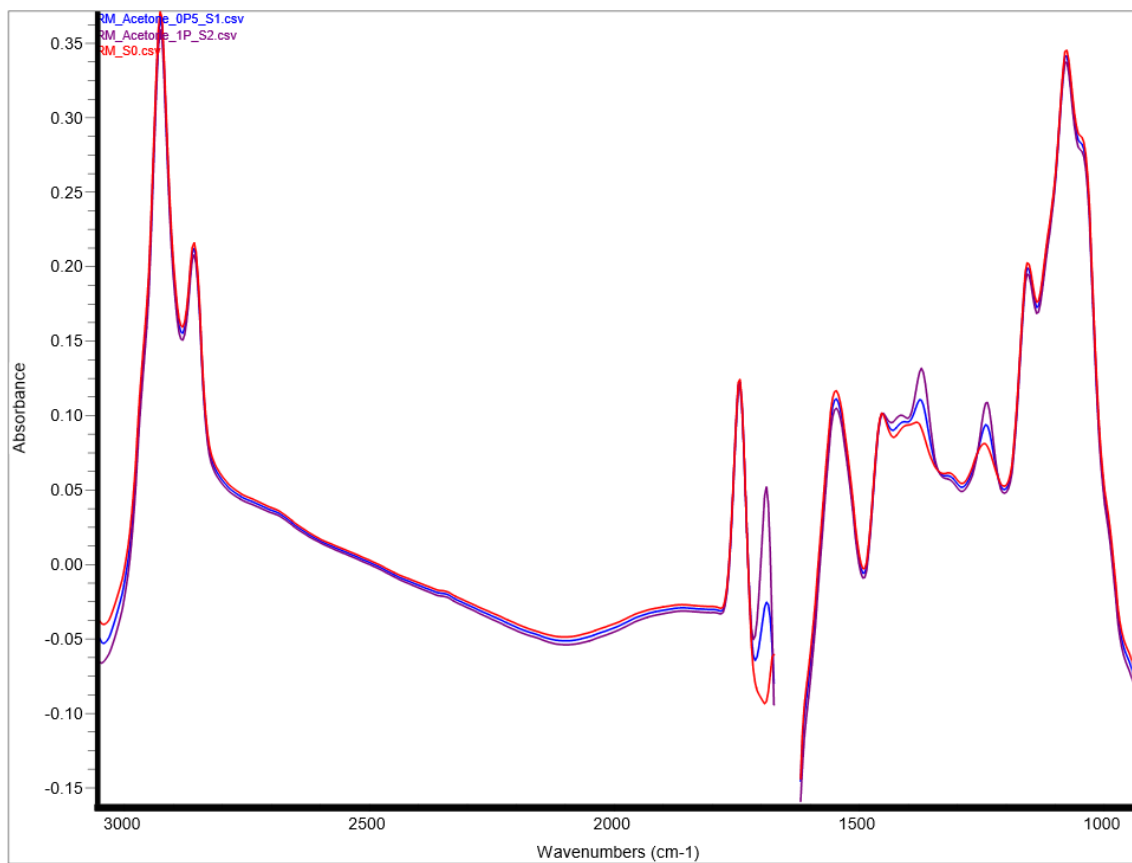


Figure 7-12 Raw milk spiked with acetone. Red: raw milk, blue: raw milk spiked with 0.5% acetone, purple: raw milk spiked with 1% acetone.

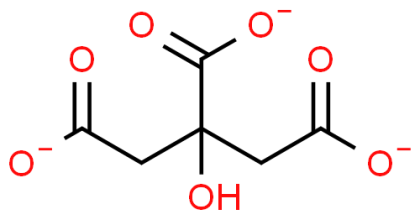


Figure 7-13 Citrate [144]

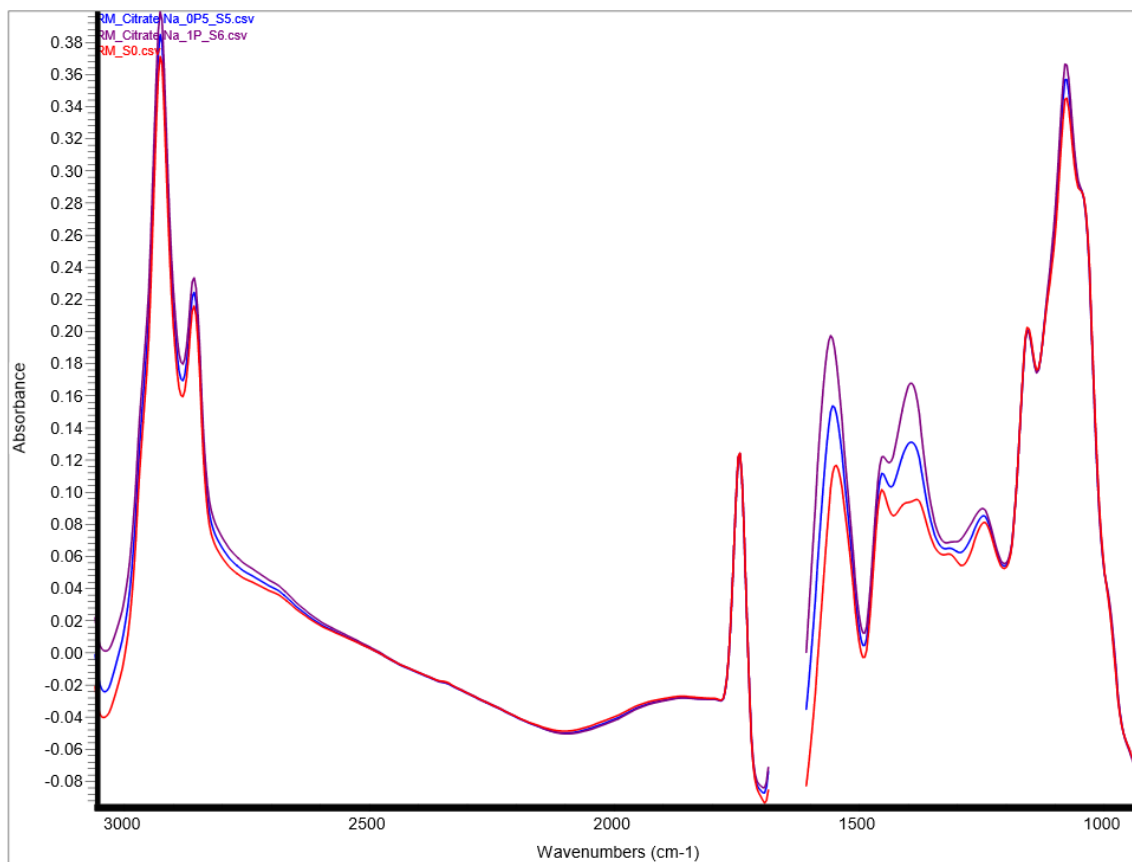


Figure 7-14 Raw milk spiked with citrate. Red: raw milk, blue: raw milk spiked with 0.5% citrate, purple: raw milk spiked with 1% citrate.

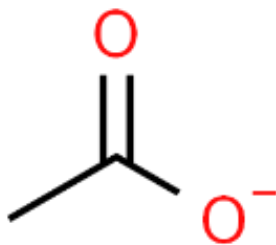


Figure 7-15 Acetate [145]

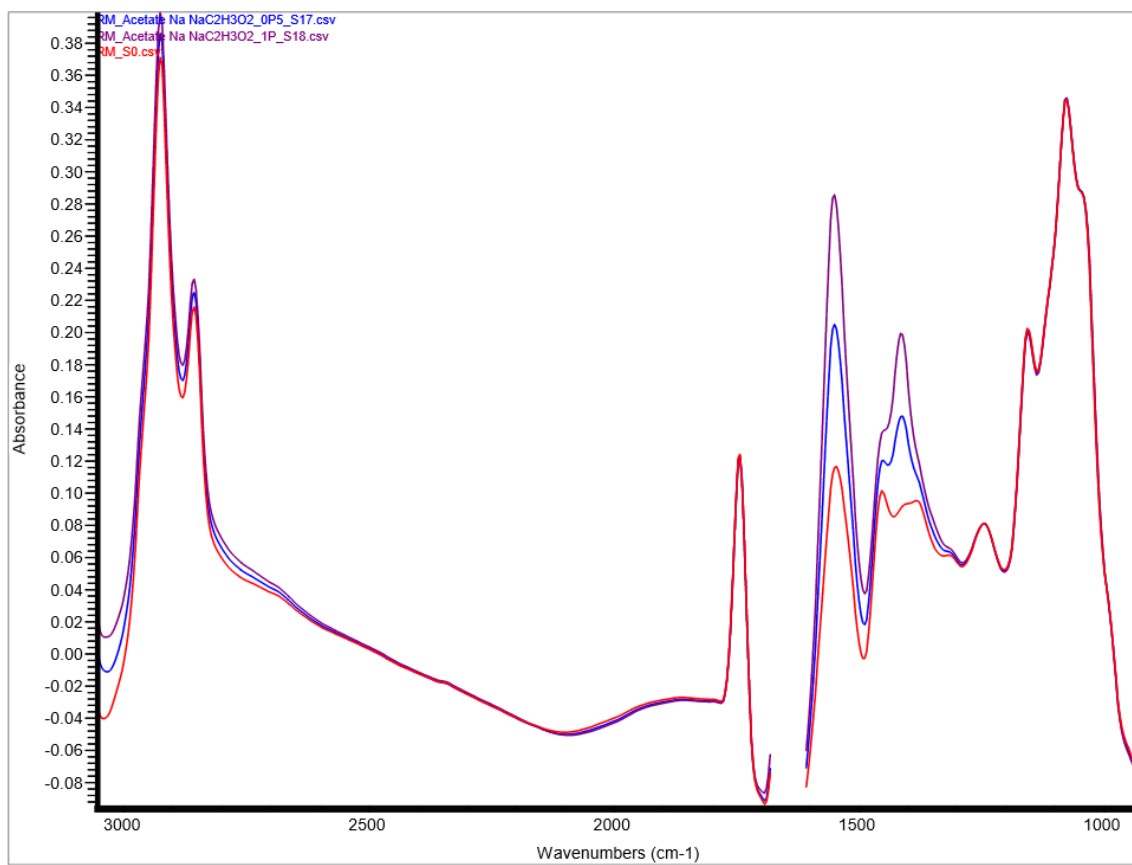


Figure 7-16 Raw milk spiked with acetate. Red: raw milk, blue: raw milk spiked with 0.5% acetate, purple: raw milk spiked with 1% acetate.

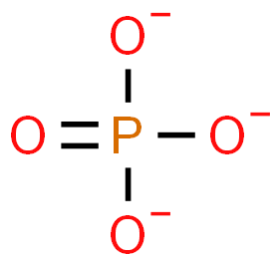


Figure 7-17 Phosphate [146]

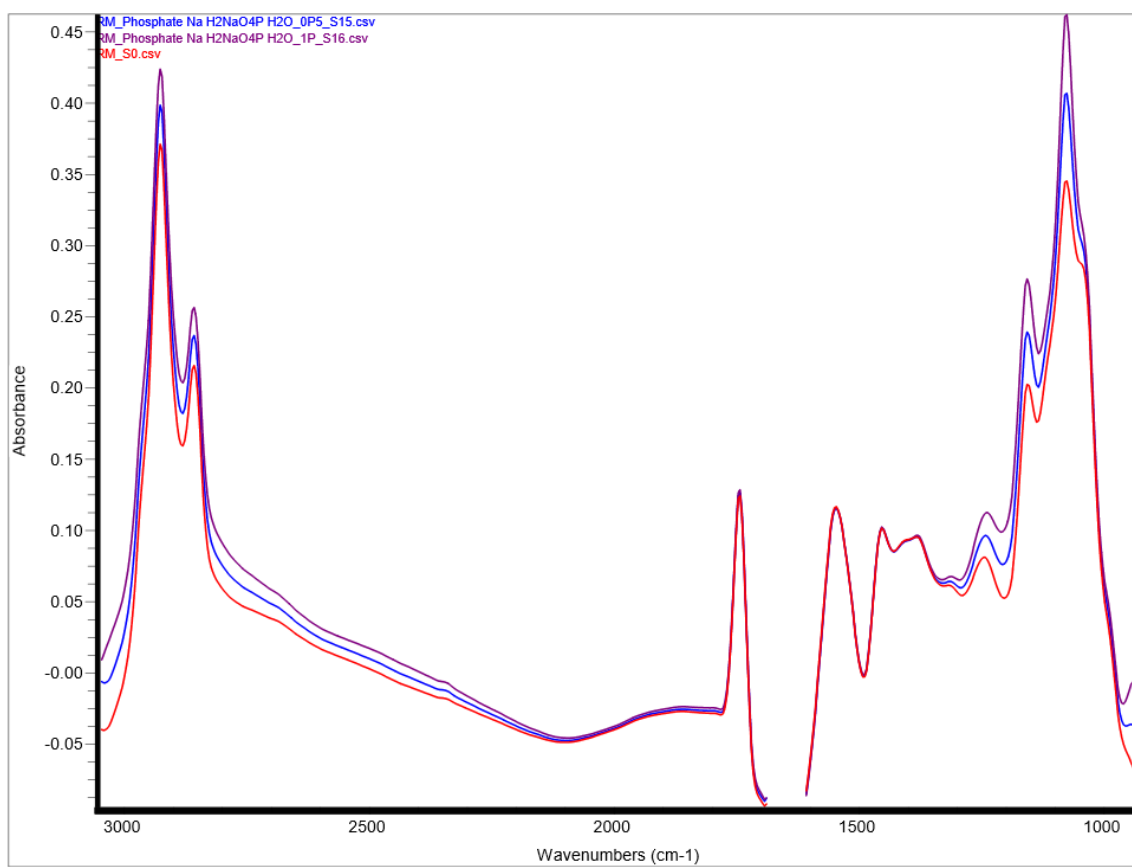


Figure 7-18 Raw milk spiked with phosphate. Red: raw milk, blue: raw milk spiked with 0.5% phosphate, purple: raw milk spiked with 1% phosphate.

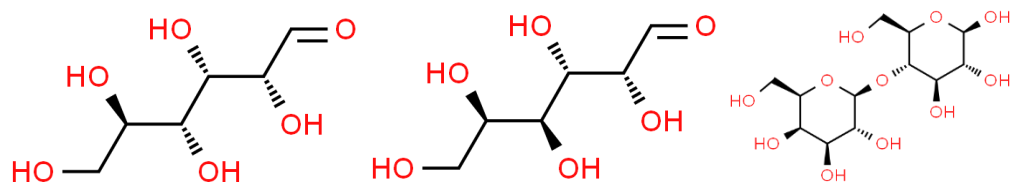


Figure 7-19 From left to right: D-(+)-glucose, D-(+)-galactose, β -lactose [147-149]

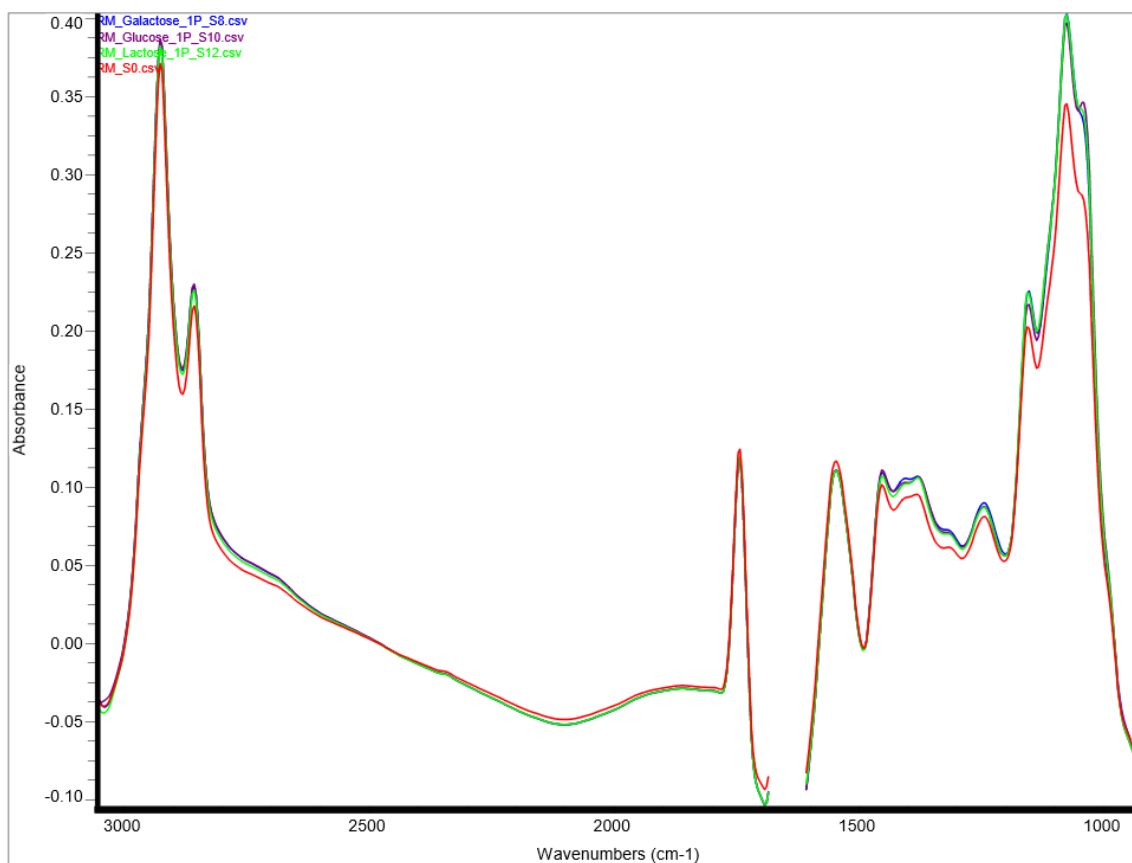


Figure 7-20 Red: raw milk, blue: raw milk spiked with 1% galactose, purple: raw milk spiked with 1% glucose, green: raw milk spiked with 1% lactose.



Figure 7-21 Ammonium ion [150]

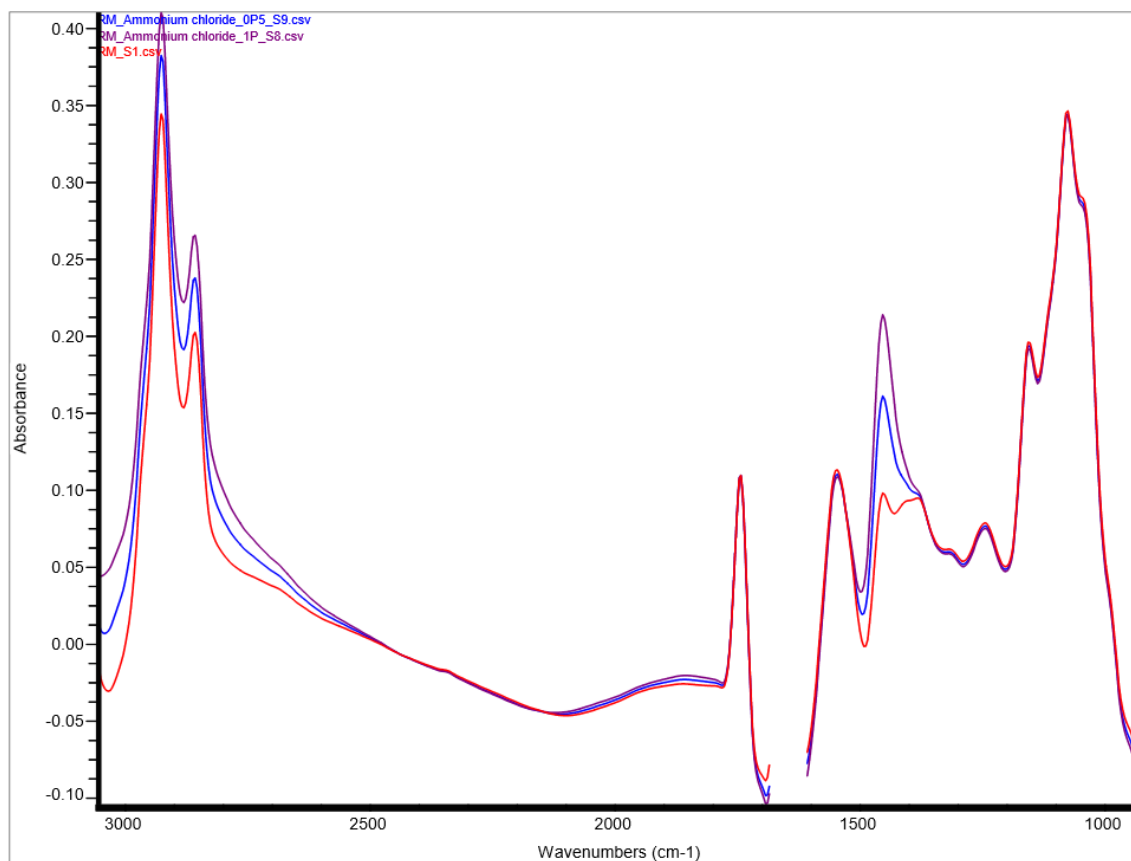


Figure 7-22 Raw milk spiked with ammonium chloride. Red: raw milk, blue: raw milk spiked with 0.5% ammonium chloride, purple: raw milk spiked with 1% ammonium chloride.

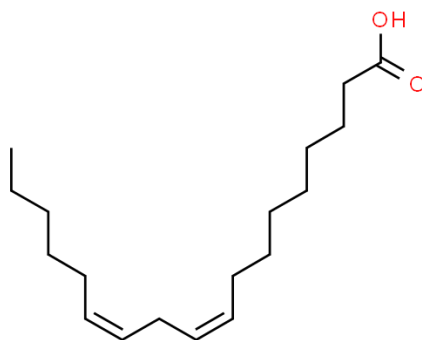


Figure 7-23 Linoleic acid C18:2 cis 9,12 [121]

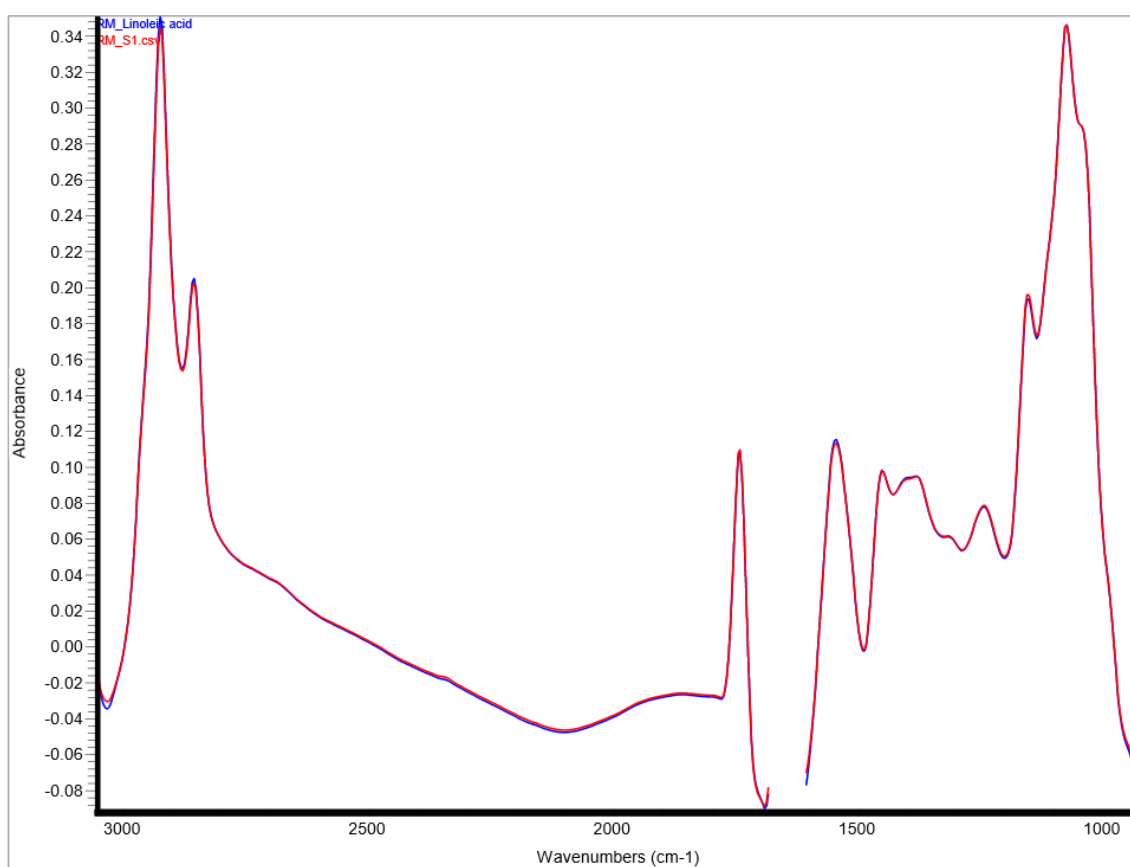


Figure 7-24 Raw milk spiked with linoleic acid. Red: raw milk, blue: raw milk spiked with linoleic acid 0.14%.

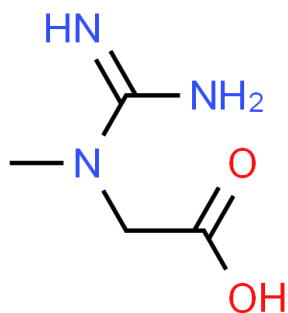


Figure 7-25 Creatine [122]

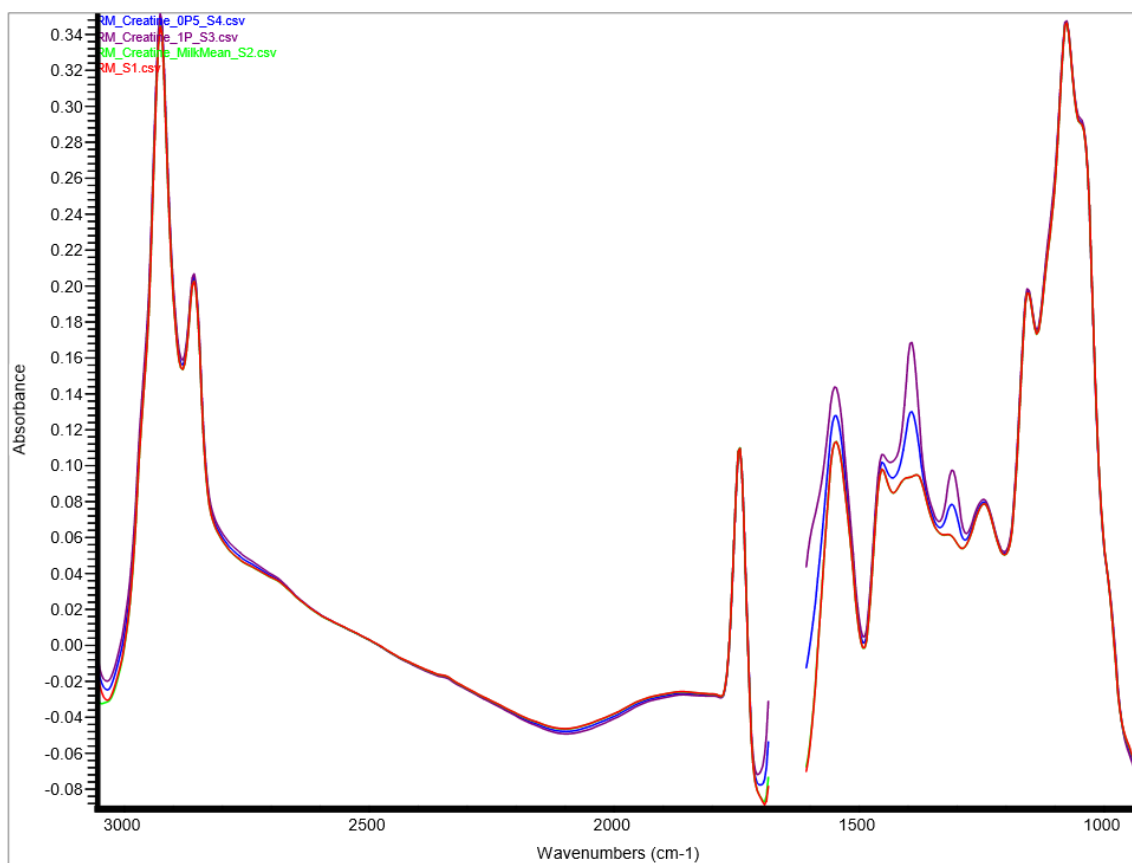


Figure 7-26 Raw milk spiked with creatine. Red: raw milk, blue: raw milk spiked with 0.5% creatine, purple: raw milk spiked with 1% creatine, green: raw milk spiked with 0.002% (milk mean) creatine.

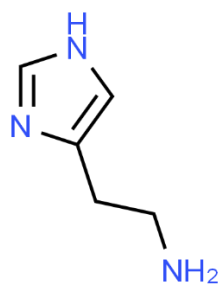


Figure 7-27 Histamine [124]

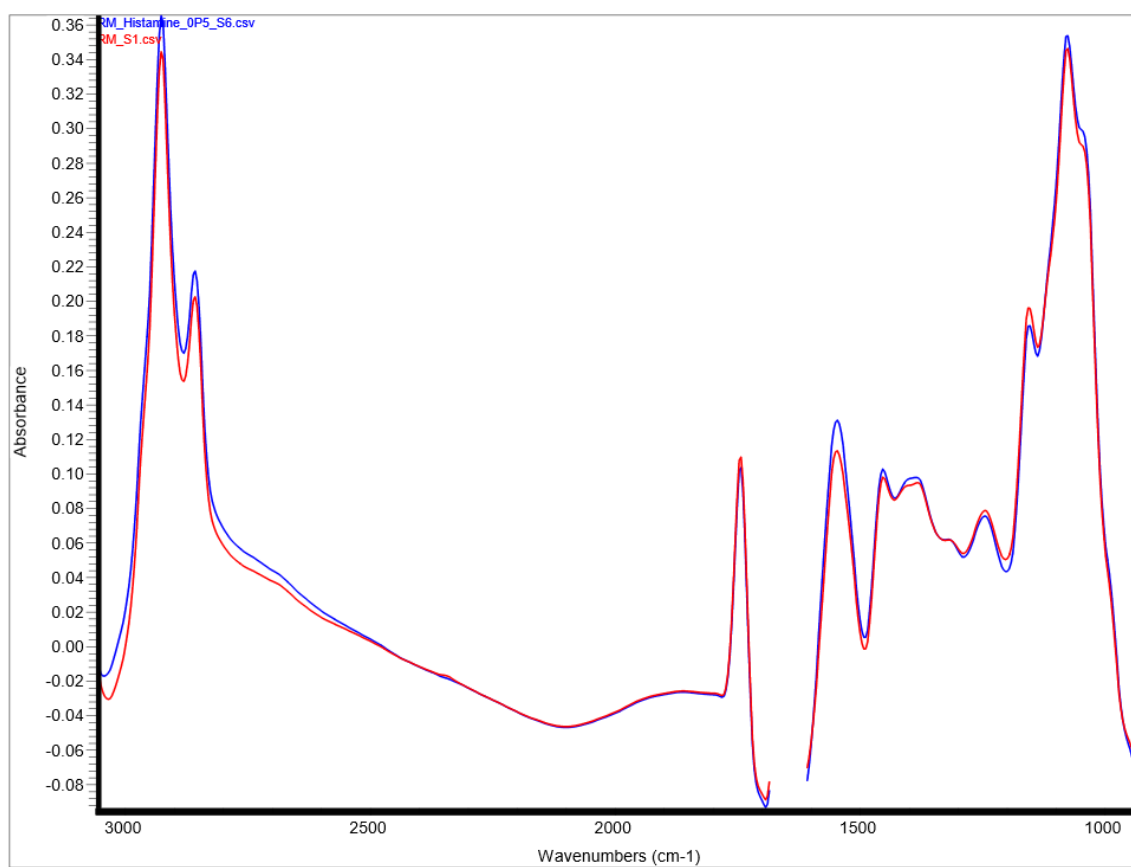


Figure 7-28 Raw milk spiked with histamine. Red: raw milk, blue: raw milk spiked with histamine 0.28%.

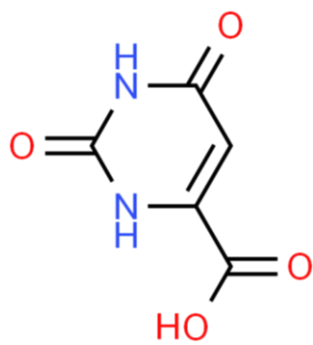


Figure 7-29 Orotic acid [126]

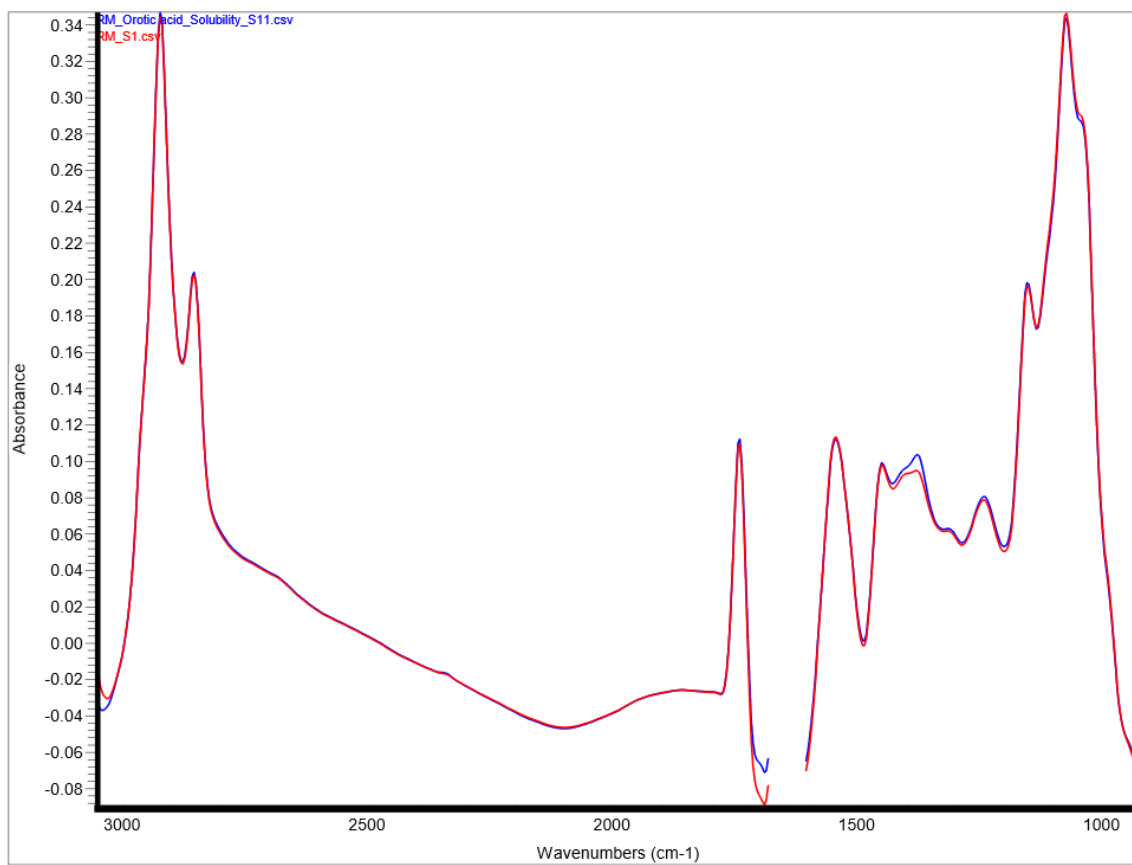


Figure 7-30 Raw milk spiked with orotic acid. Red: raw milk, blue: raw milk spiked with orotic acid 0.14%.

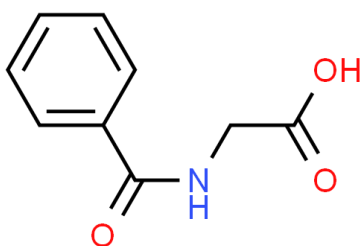


Figure 7-31 Hippuric acid [128]

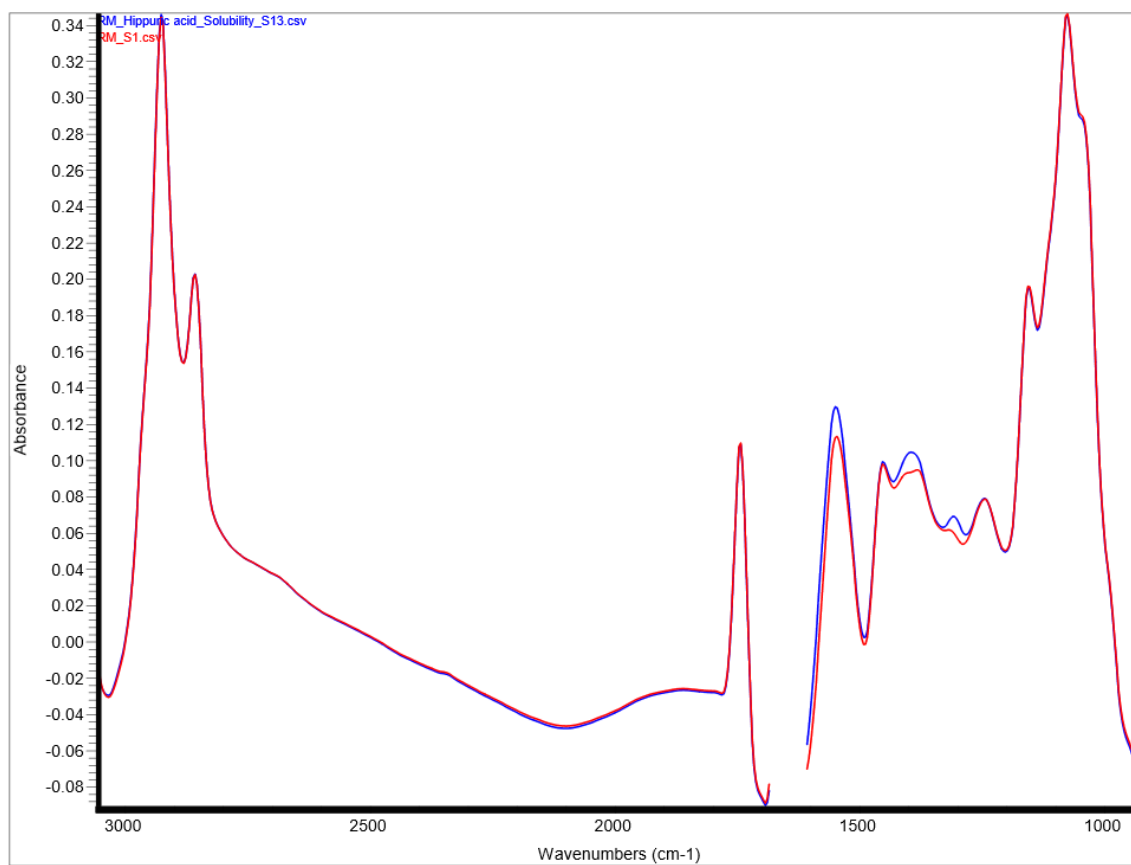


Figure 7-32 Raw milk spiked with hippuric acid. Red: raw milk, blue: raw milk spiked with hippuric acid 0.3%.

Chapter 8: General discussion and conclusion

8.1 Discussion

In this thesis, two propositions were argued for expanding the capabilities of the milk recording system. The first was on-site milk analysis. The current system relies on collecting milk samples from farms enrolled in the milk recording system and then shipping those samples to the central dairy laboratory of a dairy herd improvement agency (DHI), such as Valacta. The current system stipulates 10 analyses within 12 months period, which results in an estimated 3 million samples being shipped annually to the central dairy laboratory. In addition, monitoring milk composition has been proved to be an effective tool in managing the production process on dairy farms to monitor animal health or nutritional status [14, 15], as it was presented in the literature review in chapter 2. Modern milk analyzers are state-of-the-art instruments that are renowned for their throughput, accuracy and precision [5]. However, those instruments are expensive, and they are designed to operate under specific conditions, which makes on-site analysis of milk infeasible. If a dairy farmer wishes to adopt a proactive approach and follow a self monitoring program of milk composition, then there are currently two available options. The first option is to ship more samples from the dairy farm to the central dairy laboratory; hence, increasing the carbon footprint of dairy farming, which is already a concern. The second option is to rely on analyzers of milking robots; however, those analyzers use near infrared (NIR) detectors that produce milk spectra with broad, overlapping and low intensity bands [11]. Electromagnetic radiation in the NIR region (i.e., 2,500 to 750 nm) has a shorter wavelength than the mid-IR region (i.e., 10 to 2.5 μm), which causes greater light scattering that makes predictions of milk components less accurate than those produced by analyzers built with mid-IR detectors, especially for minor milk components. For these reasons, we consider that having an on-site milk analyzer based on mid-IR sensor will be a valuable tool for dairy farmers who wish to self monitor milk composition of their cows more frequently to adopt a proactive decision-making process rather than being reactive to issues that have already become problematic.

For this purpose, two types of infrared (IR) spectrometers have been evaluated for on-site milk analysis in combination with two sample introduction methods. In chapter 3, the first combination was a portable Fourier transform infrared (FTIR) spectrometer with a transmission cell as a sample introduction method. This combination is the basis of the official method for milk analysis by IR

spectroscopy [4] and it is already implemented in commercial milk analyzers [5]. This combination requires milk fat to be homogenized to reduce light scattering; hence, increasing the accuracy of the analysis, and to facilitate the passage of milk through the transmission cell with the micrometric optical path. Ultrasonic processing proved to be successful in achieving this goal. Applying 3000 joules of ultrasonic energy to 5 mL milk sample for 120s yielded a consistent particle distribution profile of milk fat that was similar to that of industrially homogenized milk with a two-stage high pressure homogenizer. In addition, the ultrasonic treatment degassed the milk sample and heated it to a temperature appropriate for the analysis. Partial least squares (PLS) regression models were developed to predict major milk components and some minor ones from FTIR milk spectra. The external validation study of the final prototype that was evaluated in this study yielded mean differences (MD) values that were ≤ 0.05 for fat, protein and lactose, which comply with the stipulations of the AOACI official method 972.16, 33.2.31 [37]. This proof of concept lay the foundations for the design of a dedicated FTIR on-site milk analyzer. Nevertheless, it is recommended to investigate the potential of implementing microfluidics as a sample delivery system. A microfluidic system will reduce the volume of milk sample required for the analysis, which will reduce the amount of ultrasonic energy and time required to homogenize milk fat. In addition, it will reduce the amount of energy that is needed to drive the milk sample through the analyzer.

In chapter 4, attenuated total reflectance (ATR) was evaluated as an alternative sample introduction method for the analysis of raw milk. It goes without saying that the accuracy of the analysis will not be as accurate as it is with homogenized milk. ATR will eliminate the need for a pumping accessory to deliver liquid milk for analysis and it will avoid all mechanical issues that might result from the clogging of the micrometric path length of the transmission cell by raw milk. Contrary to what has been reported in the literature so far, the study relied on large sample size of producer raw milk, which helped model the chemical information related to milk composition in a more realistic manner that lead to reduced prediction errors, especially for milk protein. The correlation coefficients values of cross validation of PLS regression models were 0.95, 0.98 and 0.98, and RMSECV values were 0.06%, 0.07% and 0.06% for lactose, protein and non-fatty solids, respectively. As for milk fat, the correlation coefficient and RMSECV for cross validation were 0.75 and 0.37%, respectively. Milk fat is present in the form of globules that range in diameters from <0.2 to >15 μm . The small fat globules represent 80% of the total number of fat globules but

they contain only 3% of the mass of fat. On the other hand, large globules represent only 2% of the total number of fat globules but they contain 95% of the mass of fat [5]. Taking into consideration that the path length of the evanescent wave at the ATR surface is only 2-3 μm [7], most of the fat in milk deposited on the ATR surface is not probed by the evanescent wave. This observation proves that ATR can be used as a sample introduction method for milk analysis for applications where milk fat is not required to be determined. However, most herd management decisions rely heavily on fat levels in milk, which renders ATR infeasible for such applications.

In addition, a novel type of IR spectrometers was evaluated for milk analysis in chapter 4, which relies on a linear variable filter (LVF) as a dispersive element mounted on top of a pyroelectric line sensor to resolve different wavelength in the IR spectrum rather than interferometry, which means it does not contain any moving parts. This type of spectrometer is significantly cheaper than its FTIR counterpart, which will reduce the cost of an on-site analyzer. It is also equipped with an ATR sample introduction method, which will eliminate the need for pumping accessory. This spectrometer yielded an acceptable analytical performance for the determination of lactose, protein and water in milk, and prediction errors were improved when milk was homogenized with an ultrasonic probe. However, this spectrometer did not reveal acceptable analytical performance with milk fat. Milk spectra obtained by this spectrometer had uneven data spacing and the number of data points in the Fat A region was too low to adequately capture the chemical information related to milk fat, which makes it suitable for applications that do not require the determination of milk fat.

In chapter 5, a case study was presented in which on-site milk analysis by IR spectroscopy provided a solution for a serious issue that faces the dairy industry in Brazil. There, water addition to milk is a rampant practice that can be detected by milk cryoscopy [79], which is the official method in Brazil to prove that milk is not watered-down. However, milk cryoscopy readings of freezing point depression of milk can be manipulated and restored to its legal value by the addition of some chemicals with the added water. To overcome this problem, a lab-in-box instrument was developed using the LVF spectrometer, which was tested in the previous chapter. This detector could differentiate genuine milk samples from watered-down ones, and it could predict the percentage of added water regardless of the chemical that was added to restore the freezing point depression of milk to its legal value. In addition, an IR-based solution was developed that relied

on portable FTIR spectrometer combined with two tiers classification model and PLS regression models to differentiate watered-down milk samples from genuine ones and to identify and quantify the added chemical. In this case, the crucial decision was dependent on the outcome of the first classification model, which was a principal component based quadratic discriminant analysis model that determines the authenticity of the milk sample, rather than a PLS regression model that played a secondary role in this application. At the moment, PLS regression is the prevailing algorithm that is implemented with milk analyzers. This case study demonstrated that an on-site milk analyzer based on mid-IR spectroscopy combined with multivariate classification algorithms can be a valuable tool to support the dairy industry in combating problems related to milk production.

The second proposition that was discussed in this thesis to expand the capabilities of the milk recording system is the exploitation of milk FTIR spectra beyond the paradigm of predicting specific milk components by PLS regression models that will be used in the decision-making process on dairy farms. DHI agencies provide services that do not necessarily rely on data generated by the milk recording system, such as the assessment of animal welfare. For this purpose, trained technicians visit dairy farms to evaluate animal injuries, quality of the cow's lifts and sets, body condition and lameness. However, some conditions related to animal welfare, such as laminitis, cause physiological changes in the animal's body that might lead to increased concentrations of some metabolites in the blood serum [20]. Since precursors of milk components are obtained from the blood, we hypothesize that the animal welfare state might affect milk composition, which might be reflected in the milk FTIR spectrum. It might be difficult to detect such changes in milk components that are currently determined by central dairy laboratories because they might not be affected by the animal welfare state. Alternatively, data mining techniques were applied to milk FTIR spectra to determine the spectral fingerprint associated with the animal welfare state.

In chapter 6, several multivariate data analysis algorithms were evaluated as data mining tools to detect the spectral fingerprint in milk FTIR spectra that reflects changes in milk composition associated with the effects of different housing treatments that aim at improving animal welfare in animal trials with controlled experimental design. Data mining is broadly defined as the process of finding patterns in data to improve the decision-making process in a specific domain [93]. To

achieve this goal, it combines tools from statistics, artificial intelligence (AI) and machine learning, which include hypothesis testing, clustering, classification and regression algorithms. In chapter 6, hierarchical cluster analysis (HCA), principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were evaluated as tools for data mining of milk FTIR spectra. The major drawback of the three algorithms was their inability of hypothesis testing for fixed and random effects that had been defined in the statistical model that was used to generate the data in the animal trial with the controlled experimental design. To overcome this obstacle, PCA was used as a dimensionality reduction method for the spectral data that created a new dataset with smaller number of variables. These variables were used as input for hypothesis testing by mixed modeling to test the different effects that were considered in the statistical model of the animal trial. The “PCA-mixed modeling” hybrid approach maintained the multivariate structure of milk FTIR data and accommodated a powerful hypothesis testing tool that proved to be successful in answering the following questions:

- Does a treatment have a significant effect on the FTIR spectral data of milk composition?
- Which treatment level did significantly change milk components’ concentrations?
- What are the spectral variables that were affected by changing milk components’ concentrations?
- What molecules in milk can be assigned to FTIR peaks that are significantly affected by the treatment?

In chapter 7, the “PCA-mixed modeling” hybrid approach was successfully applied to milk FTIR data collected during animal trials designed to study the effects of tie rail position, chain length, stall width, stall length and manger wall height on dairy cows’ welfare state in the tie stall dairy system. For each trial, the isolated spectral fingerprint represented changes to milk composition associated with the significant treatment effect and it was interpreted in light of the behavioural data that was collected during the trial. This hybrid approach bridged the gap between two different domains, which are milk FTIR spectroscopy and behavioral animal science, and provided a tool to study animal welfare from a novel angle. The hybrid approach detected the trend of subtle changes in milk composition before the appearance of any problematic clinical signs on the trials’ subjects. On the other hand, testing of individual milk components, which were predicted by PLS regression models, for treatment effect by mixed modeling did not reveal any trends in the changes

of milk composition. This observation proves that milk FTIR data can be used directly in applications beyond the paradigm of predicting specific milk components by PLS regression models. For future studies, it is recommended to add a milk metabolomics component to better understand the changes in milk chemical composition as a function of the animal welfare state. This metabolic component will verify the findings of the hybrid spectral analysis.

8.2 Conclusion

Milk recording system is a valuable tool for enforcing the standards of the dairy industry on milk producers and for providing support for dairy farmers to constantly improve their herd management practices. In this thesis, we presented two propositions to expand the capabilities of the milk recording system. The first was on-site analysis, which can be achieved by having portable transmission based FTIR milk analyzer. It was proved in this thesis that portable FTIR spectrometer equipped with a transmission cell and combined with PLS regression can accurately predict milk components, which included major ones and some minor ones. Such a tool will enable dairy farmers to implement a self monitoring program of milk composition without increasing the number of samples that need to be shipped to central dairy laboratories for analysis; hence, avoid increasing the carbon footprint of this process. On the other hand, the novel LVF-IR spectrometer proved to be effective in determining milk water, lactose and protein, which made it a good candidate for on-site analysis for applications that do not rely on determining milk fat, such as differentiation of watered-down milk from genuine one. An IR-based solution for detecting this practice was highly appreciated by dairy industry stakeholders in countries where this practice is a serious problem.

The second proposition was to exploit milk FTIR data directly in data mining process to detect trends of changes in milk composition without relying on prediction of specific milk components by PLS regression models. Combining PCA and mixed modeling proved to be a successful strategy for achieving this goal that helped bridge the gap between two different scientific domains, which are milk FTIR spectroscopy and animal behaviour science. This strategy will open the door to study animal welfare from a novel angle, which will eventually help DHI agencies to provide new services for dairy farmers in the field of animal welfare based on milk FTIR data that is routinely collected.

References

1. Goulden, J.D.S., *Analysis of milk by infrared absorption*. J Dairy Res 1964. **31**(3): p. 273-284.
2. Biggs, D.A., *Milk Analysis with the Infrared Milk Analyzer*. Journal of Dairy Science, 1967. **50**(5): p. 799-803.
3. Voort, F.R.V.d., et al., *Assessment of Fourier Transform Infrared Analysis of Milk*. Journal of AOAC International, 1992. **75**(5).
4. AOAC, *AOAC Official Methods of Analysis, Method 990.21*. Association of Official Analytical Chemists International. 2000, Gaithersburg, Maryland.
5. Andersen, S.K., P.W. Hansen, and H.V. Andersen, *Vibrational spectroscopy in the analysis of dairy products and wine*. Handbook of vibrational spectroscopy, 2002.
6. Skoog, D.A., et al., *Fundamentals of analytical chemistry*. 2013: Nelson Education.
7. Eastwood, C.R., D.F. Chapman, and M.S. Paine, *Networks of practice for co-construction of agricultural decision support systems: Case studies of precision dairy farms in Australia*. AGSY Agricultural Systems, 2012. **108**: p. 10-18.
8. Bewley, J. *Precision dairy farming: Advanced analysis solutions for future profitability*. in *Proceedings of the first North American conference on precision dairy management, Toronto, Canada*. 2010.
9. *Canadian dairy industry procedures for herd recording service standards and disciplinary sanctions*. 2011.
10. *Dairy herd recording service standards for herds qualifying for publishable lactations and/or genetic evaluations in Canada*. 2011.
11. Agelet, L.E. and C.R. Hurburgh, Jr., *A Tutorial on Near Infrared Spectroscopy and Its Calibration*. Critical Reviews in Analytical Chemistry, 2010. **40**(4): p. 246-260.
12. Borchers, M.R., *An Evaluation of Precision Dairy Farming Technology Adoption, Perception, Effectiveness, and Use*. 2015.
13. Borchers, M.R. and J.M. Bewley, *An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness*. Journal of dairy science, 2015. **98**(6): p. 4198-205.
14. Gengler, N., et al., *Capitalizing on fine milk composition for breeding and management of dairy cows*. Journal of dairy science, 2016. **99**(5): p. 4071-4079.
15. Arnould, V.M.-R., et al., *Review: milk composition as management tool of sustainability*. Biotechnologie, Agronomie, Société et Environnement, 2013. **17**(4): p. 613-621.
16. Geishauser, T., et al., *Evaluation of five cowside tests for use with milk to detect subclinical ketosis in dairy cows*. Journal of dairy science, 1998. **81**(2): p. 438-43.
17. Golder, H.M., et al., *Effects of grain, fructose, and histidine on ruminal pH and fermentation products during an induced subacute acidosis protocol*. Journal of dairy science, 2012. **95**(4): p. 1971-82.
18. Abdela, N., *Sub-acute Ruminal Acidosis (SARA) and its Consequence in Dairy Cattle: A Review of Past and Recent Research at Global Prospective*. Achievements in the Life Sciences, 2016. **10**(2): p. 187-196.
19. Humer, E., et al., *Invited review: Practical feeding management recommendations to mitigate the risk of subacute ruminal acidosis in dairy cattle*. Journal of Dairy Science, 2018. **101**(2): p. 872-888.

20. Nocek, J.E., *Bovine Acidosis: Implications on Laminitis*. Journal of Dairy Science, 1997. **80**(5): p. 1005-1028.
21. Bauman, D., et al., *Major advances associated with the biosynthesis of milk*. Journal of dairy science, 2006. **89**(4): p. 1235-1243.
22. van de Voort, F.R. and A.A. Ismail, *Proximate analysis of foods by mid-FTIR spectroscopy*. Trends in Food Science & Technology, 1991. **2**: p. 13-17.
23. Smith, B.C., *Fundamentals of Fourier transform infrared spectroscopy*. 2011: CRC press.
24. Li-Chan, E.C.Y., et al., *Vibrational Spectroscopy of Food and Food Products*, in *Handbook of Vibrational Spectroscopy*. 2006, John Wiley & Sons, Ltd.
25. Etzion, Y., et al., *Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy*. JODS Journal of Dairy Science, 2004. **87**(9): p. 2779-2788.
26. Luinge, H.J., et al., *Determination of the fat, protein and lactose content of milk using Fourier transform infrared spectrometry*. *ACA* Analytica Chimica Acta, 1993. **284**(2): p. 419-433.
27. Cox, I., M. Gaudard, and S.A.S. Institute, *Discovering partial least squares with JMP*. 2013, SAS Institute: Cary, N.C.
28. Wojciechowski, K.L. and D.M. Barbano, *Prediction of fatty acid chain length and unsaturation of milk fat by mid-infrared milk analysis*. JODS Journal of Dairy Science, 2016. **99**(11): p. 8561-8570.
29. Esbensen, K., et al., *Multivariate data analysis : an introduction to multivariate analysis, process analytical technology and quality by design*. 6th edition. ed. 2018, Oslo, Norway: CAMO.
30. Franca, A.S. and L.M.L. Nollet, *Chapter 10: Multivariate Statistical Analysis and Chemometrics*, in *Spectroscopic methods in food analysis*. 2017, CRC Press, Taylor & Francis Group: Boca Raton, FL.
31. Lynch, J.M., et al., *Precalibration Evaluation Procedures for Mid-Infrared Milk Analyzers1*. JODS Journal of Dairy Science, 2006. **89**(7): p. 2761-2774.
32. Horwitz, W., *Official methods of analysis of the AOAC International*. Vol. 18. 2000: The Association.
33. Kaylegian, K.E., et al., *Impact of fatty acid composition on the accuracy of mid-infrared fat analysis of farm milks1*. JODS Journal of Dairy Science, 2009. **92**(6): p. 2502-2513.
34. Kaylegian, K.E., et al., *Calibration of Infrared Milk Analyzers: Modified Milk Versus Producer Milk1*. JODS Journal of Dairy Science, 2006. **89**(8): p. 2817-2832.
35. Barbano, D., K. Wojciechowski, and J. Lynch, *Effect of preservatives on the accuracy of mid-infrared milk component testing*. Journal of dairy science, 2010. **93**(12): p. 6000-6011.
36. Kaylegian, K.E., et al., *Lipolysis and Proteolysis of Modified and Producer Milks Used for Calibration of Mid-Infrared Milk Analyzers1*. JODS Journal of Dairy Science, 2007. **90**(2): p. 602-615.
37. Kaylegian, K.E., et al., *Modified versus producer milk calibration: mid-infrared analyzer performance validation*. Journal of dairy science, 2006. **89**(8): p. 2833-45.
38. Soyeurt, H., et al., *Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry*. JODS Journal of Dairy Science, 2006. **89**(9): p. 3690-3695.
39. Soyeurt, H., et al., *Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries*. Journal of dairy science, 2011. **94**(4): p. 1657-67.

40. Rutten, M.J., et al., *Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer*. Journal of dairy science, 2009. **92**(12): p. 6202-9.
41. Afseth, N.K., et al., *Predicting the fatty acid composition of milk: A comparison of two Fourier transform infrared sampling techniques*. Applied spectroscopy, 2010. **64**(7): p. 700-707.
42. Eskildsen, C.E., et al., *Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: understanding predictions of highly collinear reference variables*. Journal of dairy science, 2014. **97**(12): p. 7940-51.
43. Bonfatti, V., et al., *Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental*. JODS Journal of Dairy Science, 2016. **99**(10): p. 8216-8221.
44. Kaylegian, K.E., et al., *Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis*. Journal of dairy science, 2009. **92**(6): p. 2485-501.
45. Fox, P.F. and P.L.H. McSweeney, *Advanced dairy chemistry. Volume 2, Volume 2*. 2006.
46. Di Marzo, L., P. Cree, and D.M. Barbano, *Prediction of fat globule particle size in homogenized milk using Fourier transform mid-infrared spectra*. Journal of Dairy Science, 2016. **99**(11): p. 8549-8560.
47. Bonfatti, V., G. Di Martino, and P. Carnier, *Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows*. Journal of Dairy Science Journal of Dairy Science, 2011. **94**(12): p. 5776-5785.
48. Lindmark-Månsson, H., et al., *Relationship between somatic cell count, individual leukocyte populations and milk components in bovine udder quarter milk*. International Dairy Journal, 2006. **16**(7): p. 717-727.
49. Soyeurt, H., et al., *Mid-infrared prediction of lactoferrin content in bovine milk: potential indicator of mastitis*. Animal : an international journal of animal bioscience, 2012. **6**(11): p. 1830-8.
50. De Marchi, M., et al., *Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy*. JODS Journal of Dairy Science, 2009. **92**(1): p. 423-432.
51. Dal Zotto, R., et al., *Reproducibility and repeatability of measures of milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy*. Journal of dairy science, 2008. **91**(10): p. 4103-12.
52. De Marchi, M., et al. *Prediction of milk coagulation properties by Fourier transform mid-infrared spectroscopy (FTMIR) for genetic purposes, herd management and dairy profitability*. in *Proceedings of the 38th International Committee for Animal Recording (ICAR) Meeting; Cork, Ireland*. 2012.
53. Chladek, G., et al., *Effect of season and herd on rennet coagulation time and other parameters of milk technological quality in holstein dairy cows*. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2011. **59**(5): p. 113-118.
54. De Marchi, M., et al., *Prediction of coagulating and noncoagulating milk samples using mid-infrared spectroscopy*. Journal of dairy science, 2013. **96**(7): p. 4707-15.
55. Cecchinato, A., et al., *Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk*. Journal of dairy science, 2009. **92**(10): p. 5304-13.

56. Wray, N. and P. Visscher, *Estimating trait heritability*. Nature Education, 2008. **1**(1): p. 29.
57. Neale, M. and L. Cardon, *Methodology for genetic studies of twins and families*. Vol. 67. 2013: Springer Science & Business Media.
58. Soyeurt, H., et al., *Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry*. JODS Journal of Dairy Science, 2009. **92**(6): p. 2444-2454.
59. Grelet, C., et al., *Development of Fourier transform mid-infrared calibrations to predict acetone, b-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network*. Journal of dairy science, 2016. **99**(6): p. 4816-25.
60. McParland, S., et al., *The use of mid-infrared spectrometry to predict body energy status of Holstein cows*. Journal of dairy science, 2011. **94**(7): p. 3651-61.
61. McParland, S., et al., *Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows*. JODS Journal of Dairy Science, 2014. **97**(9): p. 5863-5871.
62. Vanlierde, A., et al., *Hot topic: Innovative lactation-stage-dependent prediction of methane emissions from milk mid-infrared spectra*. JODS Journal of Dairy Science, 2015. **98**(8): p. 5740-5747.
63. Dehareng, F., et al., *Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows*. Animal : an international journal of animal bioscience, 2012. **6**(10): p. 1694-701.
64. Laine, A., et al., *How to use mid-infrared spectral information from milk recording system to detect the pregnancy status of dairy cows*. Commun. Agric. Appl. Biol. Sci., 2014. **79**: p. 33-38.
65. Capuano, E., et al., *Verification of fresh grass feeding, pasture grazing and organic farming by FTIR spectroscopy analysis of bovine milk*. Food Research International Food Research International, 2014. **60**: p. 59-65.
66. De Marchi, M., et al., *Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits*. JODS Journal of Dairy Science, 2014. **97**(3): p. 1171-1186.
67. *Omnic FTIR Software User's Guide*. 1993, Nicolet Analytical Instruments - Technical Publications Department.
68. Chittur, K.K., *FTIR/ATR for protein adsorption to biomaterial surfaces*. JBMT</cja:jid> Biomaterials, 1998. **19**(4): p. 357-369.
69. Wiesent, B., D. Dorigo, and A. Koch, *A miniaturized MID-IR-Spectrometer based on a linear variable filter and pyroelectric line array—Monitoring oil condition*. Proceedings IRS² 2013, 2013: p. 59-64.
70. Wang, T.D., et al., *Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(40): p. 15864-9.
71. Ertugay, M.F., M. sengul, and M. sengul, *Effect of Ultrasound Treatment on Milk Homogenisation and Particle Size Distribution of Fat*. Türk veterinerlik ve hayvancilik dergisi = Turkish journal of veterinary & animal sciences., 2004. **28**: p. 303-308.
72. Iñón, F.A., S. Garrigues, and M. de la Guardia, *Nutritional parameters of commercially available milk samples by FTIR and chemometric techniques*. ACA</cja:jid> Analytica Chimica Acta, 2004. **513**(2): p. 401-412.

73. Linker, R. and Y. Etzion, *Potential and limitation of mid-infrared attenuated total reflectance spectroscopy for real time analysis of raw milk in milking lines*. Journal of Dairy Research, 2009. **76**(1): p. 42-48.
74. Bassbasi, M., et al., *FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods*. FOCH Food Chemistry, 2014. **146**: p. 250-254.
75. FAO, *Dairy Market Review*. 2019, Food and Agriculture Organization: Rome.
76. Johnson, R., *Food fraud and economically motivated adulteration of food and food ingredients*. 2014, Congressional Research Service Washington, DC.
77. Finete, V.d.L.M., et al., *Is it possible to screen for milk or whey protein adulteration with melamine, urea and ammonium sulphate, combining Kjeldahl and classical spectrophotometric methods?* Food chemistry, 2013. **141**(4): p. 3649-3655.
78. Hortvet, J., *The Cryoscopy of Milk*. Industrial & Engineering Chemistry, 1921. **13**(3): p. 198-208.
79. Harding, F., *Milk adulteration-freezing point depression*. Journal of the Society of Dairy Technology, 1990. **43**(3).
80. Chen, P., X.D. Chen, and K.W. Free, *Measurement and data interpretation of the freezing point depression of milks*. Journal of Food Engineering, 1996. **30**(1-2): p. 239-253.
81. Henno, M., et al., *Factors affecting the freezing point stability of milk from individual cows*. International Dairy Journal, 2008. **18**(2): p. 210-215.
82. Rangappa, K., *Cryoscopy and refraction in milk*. Biochimica et biophysica acta, 1948. **2**: p. 207-209.
83. Singh, P. and N. Gandhi, *Milk Preservatives and Adulterants: Processing, Regulatory and Safety Issues*. Food Rev. Int. Food Reviews International, 2015. **31**(3): p. 236-261.
84. Jensen, R.G., *Handbook of milk composition*. 1995, Academic Press: San Diego.
85. Cassoli, L.D., et al., *An assessment of Fourier transform infrared spectroscopy to identify adulterated raw milk in Brazil*. International Journal of Dairy Technology, 2011. **64**(4): p. 480-485.
86. Santos, P.M., E.R. Pereira-Filho, and L.E. Rodriguez-Saona, *Application of hand-held and portable infrared spectrometers in bovine milk analysis*. Journal of agricultural and food chemistry, 2013. **61**(6): p. 1205-11.
87. Jha, S.N., et al., *Detection and quantification of urea in milk using attenuated total reflectance-Fourier transform infrared spectroscopy*. Food and Bioprocess Technology, 2015. **8**(4): p. 926-933.
88. Botelho, B.G., et al., *Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA*. Food chemistry, 2015. **181**: p. 31-37.
89. Lam, L.M., et al., *Reagent-Free Identification of Clinical Yeasts by Use of Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy*. Journal of clinical microbiology, 2019. **57**(5): p. e01739-18.
90. Nijpels, H., et al., *Application of cryoscopy for the measurement of enzymatic hydrolysis of lactose*. Journal of Food Science, 1980. **45**(6): p. 1684-1687.
91. Ministério da Agricultura, P.e.A., Gabinete do Ministro, Brazil, *Instrução Normativa Nº 62, de 29 de Dezembro de 2011*, P.e.A. Ministério da Agricultura, Editor. 2011: Brazil.
92. Sun, D.-W., *Multivariate Classification for Qualitative Analysis*, in *Infrared spectroscopy for food quality analysis and control*. 2009, Academic Press/Elsevier: Amsterdam; Boston.

93. Klimberg, R.K., B.D. McCullough, and S.A.S. Institute, *Fundamentals of predictive analytics with JMP*. 2016, SAS Institute: Cary, N.C.
94. Bihl, T. and S.A.S. Institute, *Biostatistics using JMP : a practical guide*. 2017, SAS Institute: Cary, NC.
95. Schäfer, J. and K. Strimmer, *A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics*. Statistical applications in genetics and molecular biology, 2005. **4**(1).
96. Poonia, A., et al., *Detection of adulteration in milk: A review*. International journal of dairy technology, 2017. **70**(1): p. 23-42.
97. Akers, R.M., *Milk component biosynthesis*, in *Lactation and the mammary gland*. 2002, Blackwell Publishing: Ames, Iowa. p. 199-218.
98. Jorjong, S., et al., *Milk fatty acids as possible biomarkers to early diagnose elevated concentrations of blood plasma nonesterified fatty acids in dairy cows*. Journal of Dairy Science, 2014. **97**(11): p. 7054-7064.
99. Pralle, R.S., K.W. Weigel, and H.M. White, *Predicting blood β -hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network*. Journal of Dairy Science, 2018. **101**(5): p. 4378-4387.
100. Weber, C., et al., *Variation in fat mobilization during early lactation differently affects feed intake, body condition, and lipid and glucose metabolism in high-yielding dairy cows*. Journal of Dairy Science, 2013. **96**(1): p. 165-180.
101. Institute, S., *Sas/stat 9.2 User's Guide: The MIXED Procedure (book Excerpt)*. 2008: Sas Inst.
102. St.John, J., *Effect of Positioning the Tie-rail to Follow the Natural Neck Line of Cows when Eating and Rising on the Welfare of Dairy Cows Housed in Tie-stall Barns*, in *Animal Science Department*. 2018, McGill University: Montreal.
103. Xiaobo, Z., et al., *Variables selection methods in near-infrared spectroscopy*. Analytica Chimica Acta, 2010. **667**(1): p. 14-32.
104. Lehman, A., et al., *JMP for basic univariate and multivariate statistics: methods for researchers and social scientists*. 2013: Sas Institute.
105. Bradley, M., *Curve fitting in Raman and IR spectroscopy: basic theory of line shapes and applications*. Thermo Fisher Scientific, Madison, USA, Application Note, 2007. **50733**.
106. Akers, R.M., *Biochemical properties of mammary secretions*, in *Lactation and the mammary gland*. 2002, Blackwell Publishing: Ames, Iowa. p. 199-218.
107. Akers, R.M., *Functional development of the mammary gland*, in *Lactation and the mammary gland*. 2002, Blackwell Publishing: Ames, Iowa. p. 66-87.
108. Sun, D.-W., *Infrared spectroscopy for food quality analysis and control*. 2009, Academic Press/Elsevier: Amsterdam; Boston.
109. Heuer, C., et al., *Determination of acetone in cow milk by Fourier transform infrared spectroscopy for the detection of subclinical ketosis*. Journal of dairy science, 2001. **84**(3): p. 575-582.
110. Wiercigroch, E., et al., *Raman and infrared spectroscopy of carbohydrates: A review*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017. **185**: p. 317-335.
111. *NIST Chemistry WebBook*. 2018; Available from: <https://webbook.nist.gov/chemistry/#>.

112. Bewley, J., et al., *Precision dairy monitoring technology implementation opportunities and challenges*, in *Pages 1251–1261 in Large Dairy Herd Management*. 2017, Am. Dairy Sci. Assoc., Champaign, IL.
113. Petersson-Wolfe, C., et al., *Opportunities for identifying animal health and well-being disorders using precision technologies*, in *Pages 1279–1292 in Large Dairy Herd Management*. 2017, Am. Dairy Sci. Assoc., Champaign, IL.
114. Boyer, V., *How can we change stalls to better meet our cows' needs? Increasing chain length and stall width to enhance dairy cows' ease of movement and ability to rest in tie-stalls*, in *Animal Science Department*. 2019, McGill University: Montreal.
115. McPherson, S., *Making stall beds more comfortable: The effect of longitudinal space and deep-bedding on the comfort and welfare of tie-stall-housed dairy cows*, in *Animal Science Department*. 2019, McGill University: Montreal.
116. Silverstein, R.M., F.X. Webster, and D.J. Kiemle, *Spectrometric identification of organic compounds*. 8th ed. / ed. 2015, Hoboken, NJ: John Wiley & Sons.
117. Vasko, P.D., J. Blackwell, and J. Koenig, *Infrared and raman spectroscopy of carbohydrates: Part I: Identification of O-H and C-H related vibrational modes for D-glucose, maltose, cellobiose, and dextran by deuterium-substitution methods*. Carbohydrate Research, 1971. **19**(3): p. 297-310.
118. Ito, K. and H.J. Bernstein *The vibrational spectra of the formate, acetate, and oxalate ions*. Canadian Journal of Chemistry, 1956. **34**, 170-178 DOI: 10.1139/v56-021.
119. Klähn, M., et al., *IR spectra of phosphate ions in aqueous solution: predictions of a DFT/MM approach compared with observations*. The Journal of Physical Chemistry A, 2004. **108**(29): p. 6186-6194.
120. Petibois, C., et al., *Determination of glucose in dried serum samples by Fourier-transform infrared spectroscopy*. Clinical chemistry, 1999. **45**(9): p. 1530-5.
121. Linoleic acid CSID:4444105, <http://www.chemspider.com/Chemical-Structure.4444105.html> (accessed 16:37, Jul 16, 2019).
122. Creatine CSID:566, <http://www.chemspider.com/Chemical-Structure.566.html> (accessed 16:04, Jul 16, 2019).
123. Socrates, G., *Infrared and Raman characteristic group frequencies: tables and charts*. 2004: John Wiley & Sons.
124. Histamine CSID:753, <http://www.chemspider.com/Chemical-Structure.753.html> (accessed 16:14, Jul 16, 2019).
125. Collado, J.A. and F.J. Ramírez, *Infrared and Raman spectra of histamine-NH₄ and histamine-Nd₄ monohydrochlorides*. Journal of Raman Spectroscopy, 1999. **30**(5): p. 391-397.
126. Orotic acid CSID:942, <http://www.chemspider.com/Chemical-Structure.942.html> (accessed 16:21, Jul 16, 2019).
127. Hernanz, A., et al., *Vibrational analysis and spectra of orotic acid*. Biopolymers, 2000. **57**(3): p. 187-198.
128. Hippuric acid CSID:451, <http://www.chemspider.com/Chemical-Structure.451.html> (accessed 16:25, Jul 16, 2019).
129. Karabacak, M., M. Cinar, and M. Kurt, *Molecular structure and vibrational assignments of hippuric acid: A detailed density functional theoretical study*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2009. **74**(5): p. 1197-1203.
130. Sun, D.-W., *Infrared spectroscopy for food quality analysis and control*. 2009.

131. Meissner, S., et al., *Key role of short-chain fatty acids in epithelial barrier failure during ruminal acidosis*. Journal of dairy science, 2017. **100**(8): p. 6662-6675.
132. Jensen, R.G., *Miscellaneous Factors Affecting Composition and Volume of Human and Bovine Milks*, in *Handbook of milk composition*. 1995, Academic Press: San Diego.
133. Heinrichs, J., C. Jones, and K. Bailey, *Milk components: Understanding the causes and importance of milk fat and protein variation in your dairy herd*. Dairy Anim. Sci, 1997. **5**: p. 1e-8e.
134. Bischoff, P., et al., *Topography of clonidine-induced electroencephalographic changes evaluated by principal component analysis*. Anesthesiology, 2000. **92**(6): p. 1545-52.
135. Milot, M.H., et al., *A crossover pilot study evaluating the functional outcomes of two different types of robotic movement training in chronic stroke survivors using the arm exoskeleton BONES*. Journal of neuroengineering and rehabilitation, 2013. **10**: p. 112.
136. Olde Riekerink, R.G.M., et al., *Incidence Rate of Clinical Mastitis on Canadian Dairy Farms*. Journal of Dairy Science, 2008. **91**(4): p. 1366-1377.
137. Plesch, G. and U. Knierim, *Effects of housing and management conditions on teat cleanliness of dairy cows in cubicle systems taking into account body dimensions of the cows*. Animal : an international journal of animal bioscience, 2012. **6**(8): p. 1360-8.
138. Veissier, I., J. Capdeville, and E. Delval, *Cubicle housing systems for cattle: Comfort of dairy cows depends on cubicle adjustment*. Journal of animal science, 2004. **82**(11): p. 3321-37.
139. Martiskainen, P., T. Koistinen, and J. Mononen. *Cubicle dimensions affect resting-related behaviour, injuries and dirtiness of loose-housed dairy cows*. in *Proceedings of the XIII (13th) International Congress in Animal Hygiene, Tartu, Estonia, 175pp*. 2007.
140. Lombard, J.E., et al., *Associations between cow hygiene, hock injuries, and free stall usage on US dairy farms*. Journal of Dairy Science, 2010. **93**(10): p. 4668-4676.
141. Urea National Center for Biotechnology Information. PubChem Database. Urea, CID=1176, <https://pubchem.ncbi.nlm.nih.gov/compound/Urea> (accessed on July 6, 2019).
142. β -Hydroxybutyric acid CSID:428, <http://www.chemspider.com/Chemical-Structure.428.html> (accessed 18:20, Jul 6, 2019)
143. Acetone CSID:175, <http://www.chemspider.com/Chemical-Structure.175.html> (accessed 19:23, Jul 6, 2019).
144. Citrate CSID:29081, <http://www.chemspider.com/Chemical-Structure.29081.html> (accessed 20:04, Jul 6, 2019).
145. Acetate CSID:170, <http://www.chemspider.com/Chemical-Structure.170.html> (accessed 20:28, Jul 6, 2019).
146. Phosphate CSID:1032, <http://www.chemspider.com/Chemical-Structure.1032.html> (accessed 00:21, Jul 7, 2019).
147. Glucose CSID:96749, <http://www.chemspider.com/Chemical-Structure.96749.html> (accessed 01:19, Jul 27, 2019).
148. Galactose CSID:2301265, <http://www.chemspider.com/Chemical-Structure.2301265.html> (accessed 01:19, Jul 27, 2019).
149. Lactose CSID:5904, <http://www.chemspider.com/Chemical-Structure.5904.html> (accessed 02:02, Jul 27, 2019).
150. Ammonium ion CSID:218, <http://www.chemspider.com/Chemical-Structure.218.html> (accessed 20:34, Jul 31, 2019).