

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Identification of key mid-infrared wavelengths associated with red deer milk composition

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Animal Science

at Massey University, Palmerston North, New Zealand



Leith Anderson

2026

Abstract

Mid-infrared (MIR) spectroscopy has been applied to predict composition of milk from different animal species, but no study has reported the key mid-infrared wavelengths associated with the concentrations of main milk components. The objective of this study was to identify the MIR wavelengths most strongly associated with each milk component to improve understanding of the spectral regions contributing to prediction accuracy in deer milk. A total of 202 milk samples collected from lactating red deer hinds at the Aratiatia Deer Milk Farm, Wairakei, Taupō, New Zealand, were analysed using a MilkoScan 4000 MIR spectrometer. Partial Least Squares Regression (PLSR) models were developed using MIR spectral data to perform an external validation of the equations already integrated in the MilkoScan 4000 MIR spectrometer machine. Model performance was assessed using calibration, 10-fold cross-validation, and independent hold-out validation datasets based on the coefficient of determination (R^2), root mean square error of prediction ($RMSEP$), and residual predictive deviation (RPD). PLSR coefficient profiles were examined to identify the most influential spectral regions contributing to prediction accuracy. The PLSR models achieved excellent agreement with the FOSS-derived reference values for all major milk components. Validation models produced R^2 values of 0.9994 for fat, 0.9975 for protein, and 0.9953 for lactose, with corresponding $RMSEP$ values of 0.0405%, 0.0282%, and 0.0281%, respectively, demonstrating close agreement between predicted and reference values. The influential spectral regions corresponded closely with established absorption bands associated with lipid, protein, and carbohydrate functional groups, indicating that the models captured chemically meaningful information rather than random statistical relationships. This study provides a greater understanding of the spectral basis of these predictions supporting the development of herd-testing systems, genetic improvement programmes, and the emerging deer dairy industry.

Acknowledgements

I would like to express gratitude to my supervisor, Dr Nicolas Lopez-Villalobos, for his guidance, support, and expertise throughout my Master's. His knowledge and encouragement were invaluable during the completion of this thesis. With each check-in and feedback helping me gain confidence and advance my research skills.

I would like to thank Penny Back for helping me build confidence in myself and for providing opportunities to gain practical, hands-on experience. Those opportunities played an important role in developing both my confidence and desire to complete a master's degree. I would also like to acknowledge the wider Animal Science Department at Massey University for their support and contribution throughout both my undergraduate and postgraduate studies.

I gratefully acknowledge the financial support provided through the Hudson and McGowan Scholarship (2026), the LIC Brian Aspen Scholarship (2026), and the Akers Postgraduate Scholarship (2025). Being selected as the scholarship recipient is a significant honour, and it eases the financial burden, allowing me to fully dedicate my time and energy to my research. I would also like to thank Livestock Improvement Corporation for providing the milk sample MIR data required to complete this research, providing the opportunity to gain valuable knowledge in a research gap.

I would also like to thank my friends for their constant support throughout this journey. From the geopolitical discussions and puzzle breaks to the shared baking, you all made me look forward to working in the postgraduate office. Thanks to those who shared the highs and lows of postgraduate life, the support made this experience a happy one to remember, and I truly could not have done it without you all.

I would also like to thank my partner for always listening to my endless rants and for providing a space away from university where I could truly switch off, relax, and enjoy the moment. Thank you for encouraging me to prioritise myself when needed and teaching me that being more selfish with my time and energy is sometimes necessary to achieve my goals.

To my sisters, thank you for your constant encouragement and support throughout this degree. Thank you for always listening when I needed to vent, making me laugh when I needed it most,

and reminding me how far I have come whenever I doubted myself or felt behind. Saving the best for last, the biggest thanks goes to my father, who never gave up. Thank you for all your hard work and sacrifices to ensure I could live the best life possible and strive for higher education. Any problem, big or small, you were always the first person I would call. I truly value your advice and wisdom. Whether I actually took your advice is another story, but know that I was always listening and that your guidance meant more to me than you probably realise.

Table of contents

Abstract.....	i
Acknowledgements.....	iii
Table of contents.....	vi
List of tables.....	viii
List of figures.....	x
List of abbreviations.....	xii
List of appendices.....	xv
Chapter 1	
General introduction.....	1
Chapter 2	
Literature review.....	6
Chapter 3	
Material and methods.....	31
Chapter 4	
Results.....	40
Chapter 5	
Discussion and conclusions.....	52
List of references.....	88

List of tables

Table 2.1.	Reported ranges of deer milk composition from different deer species.....	17
Table 4.1.	Mean, standard deviation (SD) and number (N) for red deer milk composition traits (fat, protein, and lactose).....	42
Table 4.2.	Calibration, 10-fold cross-validation, and validation statistics for the PLSR models predicting fat, protein, and lactose concentrations in red deer milk.....	42

List of figures

Figure 4.1.	Observed versus cross-validated predicted red deer milk fat concentration (%) for the 13-component PLSR model.....	43
Figure 4.2.	Observed versus predicted red deer milk fat concentration (%) for the validation dataset using the 13-component PLSR model...	44
Figure 4.3.	Observed versus cross-validated predicted red deer milk protein concentration (%) for the 16-component PLSR model.....	45
Figure 4.4.	Observed versus cross-validated predicted red deer milk protein concentration (%) for the validation dataset using the 16-component PLSR model.....	45
Figure 4.5.	Observed versus cross-validated predicted red deer milk lactose concentration (%) for the 21-component PLSR model.....	46
Figure 4.6.	Observed versus cross-validated predicted red deer milk lactose concentration (%) for the validation dataset using the 21-component PLSR model.....	47
Figure 4.7.	Manhattan plot of the absolute PLSR regression coefficients for the 13-component fat model, highlighting the 80 most influential spectral pins.....	48
Figure 4.8.	Manhattan plot of the absolute PLSR regression coefficients for the 16-component protein model, highlighting the 80 most influential spectral pins.....	49
Figure 4.9.	Manhattan plot of the absolute PLSR regression coefficients for the 21-component lactose model, highlighting the 80 most influential spectral pins.....	50

List of abbreviations

AIC	Akaike Information Criterion
BCFA	Branched-Chain Fatty Acid
BW	Breeding Worth
CLA	Conjugated Linoleic Acid
DINZ	Deer Industry New Zealand
DMNZ	Deer Milking New Zealand
FA	Fatty Acid
FTIR	Fourier Transform Infrared Spectroscopy
GWAS	Genome-Wide Association Study
MIR	Mid-Infrared Spectroscopy
MLR	Multiple Linear Regression
MUFA	Monounsaturated Fatty Acid
PLSR	Partial Least Squares Regression
PW	Production Worth
QTL	Quantitative Trait Locus
R^2	Coefficient of Determination
R^2_{CV}	Coefficient of Determination of Cross-Validation
R^2_v	Coefficient of Determination of Validation
R^2_{ExV}	Coefficient of Determination of External Validation
RCT	Rennet Coagulation Properties
RER	Ratio of Error Range
RMSE	Root Mean Square Error
RMSE _C	Root Mean Square Error of Calibration
RMSE _{CV}	Root Mean Square Error of Cross-Validation
RMSE _P	Root Mean Square Error of Prediction
RPD	Residual Predictive Deviation
RPE	Relative Prediction Error
SE _{CV}	Standard Error of Cross-Validation

SEP	Standard Error of Prediction
SFA	Saturated Fatty Acid

List of appendices

Appendix 1	Fat, protein, lactose spectral pins retained following exploratory backward stepwise regression	66
Appendix 2	Top 80 influential spectral pins identified by PLSR coefficients for predicting milk fat concentration	70
Appendix 3	Top 80 influential spectral pins identified by PLSR coefficients for predicting milk protein concentration.....	76
Appendix 4	Top 80 influential spectral pins identified by PLSR coefficients for predicting milk lactose concentration.....	82

Chapter 1: General Introduction

The New Zealand deer industry is an established contributor to the agricultural sector, generating value primarily through venison, velvet, and other high-value products (Hall, 2023). In recent years, there has been increasing interest in the development of a commercial deer milk industry in New Zealand, driven by the distinctive composition and potential premium value of deer milk. Red deer milk has been reported to contain relatively high concentrations of fat and protein, while lactose concentrations are generally lower than those of bovine milk (Landete-Castillejos et al., 2000; Malacarne et al., 2015; Li et al., 2023). These characteristics have contributed to interest in deer milk as a source of premium nutritional products and specialised dairy ingredients, and other high-value food applications.

Although the commercial deer dairy industry is relatively small, deer milk products are already being produced and marketed in New Zealand. Commercial operations have demonstrated opportunities for processing deer milk into high-value products, including milk powders and other value-added products (Pāmu n.d.; Piddock 2023). However, the development of the industry is constrained by the relatively low milk yield of red deer and the seasonal nature of lactation. Red deer hinds are typically milked for a comparatively short period, and milk production is substantially lower than that of conventional dairy species (NZIPIM 2020; Piddock 2023). These characteristics increase the importance of efficient milk collection, processing, and analytical methods for supporting the development of the industry.

The successful application of MIR spectroscopy depends on the development of calibration equations using reference data that adequately represent the population and conditions in which the equations will subsequently be applied. Variation in factors such as animal population, breed, nutrition, management, stage of lactation, and production system can influence milk composition and, consequently, the relationship between MIR spectra and milk constituents. Therefore, calibration equations developed using one population or production system may not necessarily maintain the same predictive performance when applied to an independent population. External validation is consequently important for assessing the predictive performance and generalisability of existing MIR equations when they are applied to new populations (Ho et al., 2021; Ní Fhuaráin et al. 2024).

Despite the established application of MIR spectroscopy in bovine, ovine, and caprine milk, little research has investigated its application to red deer milk (Cellesi et al. 2019; Dagnachew et al., 2013; De Marchi et al. 2011). In particular, there is limited information regarding the performance of existing MIR calibration equations when applied to New Zealand red deer milk

and the spectral regions contributing to the prediction of its major milk components. This represents an important knowledge gap given the increasing commercial interest in deer milk and the need for rapid and reliable methods of assessing milk composition.

The present study therefore aimed to evaluate the predictive performance of an existing species-specific FOSS MilkoScan equations for estimating fat, protein, and lactose concentrations in New Zealand red deer milk. A second objective was to identify the MIR wavelengths most strongly associated with each milk component, providing greater insight into the spectral regions contributing to the existing predictions. Together, these objectives provide an initial assessment of the applicability of MIR spectroscopy to red deer milk and establish a foundation for future development and independent validation of deer-specific MIR calibration equations.

Chapter 2: Literature Review

2.1 Deer history

2.1.1 Introduction of deer to New Zealand

Deer were introduced to New Zealand primarily for recreational hunting. While specific dates are debatable, it is estimated that introductions occurred between 1851 and 1871 (Hall 2023; Nixon 2004). Seven types of deer were brought to New Zealand. The most abundant were red deer, with a sizable release near Nelson in 1854 (Figgins and Holland 2012; Hall 2023; Nixon 2004). Initially, individual farmers were responsible for these imports, but with the support of the government in the 1860s, Acclimatisation Societies took over (Figgins and Holland 2012; Nixon 2004). New Zealand's temperate climate, abundant vegetation, mild winters, and absence of natural predators, pests, and diseases created ideal conditions for deer survival. Allowing them to achieve higher weights than their counterparts in the Northern Hemisphere (Nixon 2004). Additionally, their early maturity and initial protection against hunting, such as fines for out-of-season killing, contributed to rapid population growth (Figgins and Holland 2012; Nixon 2004).

2.1.2 Shift from game to pest

The rising deer population quickly led to significant environmental degradation, shifting their status from prized game to serious pests (Figgins and Holland 2012; Hall 2023; Nixon 2004). Public concerns about the damage deer posed to New Zealand's forests were raised as early as 1892. By 1919, it was estimated that the red deer population was 300,000, growing at a concerning 25% annually. The explosion in numbers also meant intense competition for grazing land with domestic livestock (Figgins and Holland 2012). The mounting pressure from farmers and conservationists peaked at the 1930 "Deer Menace" conference, where all remaining legal protection for deer was lifted, officially declaring them as pests in need of eradication (Nixon 2004).

2.1.3 Establishment of deer farming

A government-managed culling effort began in the 1960s to reduce wild deer numbers, which led to substantial export of wild-shot venison (Fennessy 1995; Figgins and Holland 2012; Nixon 2004). However, the intensive hunting, particularly with the widespread use of helicopters from 1964, led to incredibly high kill rates, reaching 145,000 carcasses in 1973

(Fennessy 1995; Nixon 2004). This rate of culling was unsustainable, as wild deer numbers dropped dramatically, making a consistent supply difficult (Fennessy 1995). Wild deer also became more elusive, changing their feeding habits and retreating into dense bush to avoid helicopters, further complicating hunting efforts (Nixon 2004).

Due to existing demand for venison in Germany and emerging markets in Asia for velvet antlers and other co-products used in traditional oriental medicine, the New Zealand government introduced legislation allowing deer farming in 1969. The first license to formally farm deer was issued in 1970 (Fennessy 1995; Nixon 2004; Pearse 2022). This legal change enabled commercial hunters to transition from culling to capturing live deer to establish farms, with approximately 70,000 wild deer transferred to pastoral properties (Figgins and Holland 2012). New Zealanders adapted their experience from ovine and bovine farming to deer, learning about their biological needs and developing management systems for grazing them on pasture (Nixon 2004; Pearse 2022).

2.2 Deer industry

New Zealand is now the largest exporter of deer meat and products globally, with the industry generating an average of \$280 million in export earnings annually (Hall 2023). As of 2024, there are 709,000 deer farmed over 1,200 farms (DINZ 2025; Stats NZ 2024). The Deer Industry New Zealand (DINZ) is the overarching industry body responsible for promoting and assisting the development of the industry (Pearse 2022). DINZ has released the “Thrive 2035” plan that aims to overcome stalled growth and historical market volatility, targeting an annual growth rate of 5% to 10% for stakeholders. In the hope of doubling the industry size and achieving projected revenues exceeding \$660 million by 2035 (DINZ 2025).

2.2.1 Venison industry

Venison comprises the majority of the total volume of deer products exported (91%) and accounts for over 58% of the total value. In 2022, one-third of the venison exported to the United States was due to the Hotel, Restaurant, and Café (HORECA) industry. The success in this market is partly due to the perception of venison as a lean, tender meat high in iron, appealing to nutrition and health-centric customers (Hall 2023). In 2024, the venison export value was \$205 million, with 15,319 tonnes exported at the average price of \$8.90 per kg (DINZ n.d.).

Germany has been a main and historically well-established market for New Zealand venison, exerting significant control over the market and dictating prices. To gain better returns, New Zealand processors have increasingly sought to bypass wholesalers and establish direct relationships with supermarket chains (Nixon 2004). This strategy aims to put processors more in touch with consumer trends and reduce price volatility.

Other European states consume approximately 30% of New Zealand's venison exports. The demand is driven by traditional consumption patterns, particularly around Christmas, when game meat is viewed as a treat (Nixon 2004). European markets, especially for chilled venison, show a seasonal peak in demand in October, coinciding with the approach of winter and increased consumption of venison stews (Hall 2023). New Zealand's ability to supply young 9–12-month-old males during the southern hemisphere's spring and early summer allows it to meet this peak chilled season demand, with prices almost double that of frozen venison. China is an emerging market for New Zealand venison, though it faced import issues and confusion over venison's status as game meat during the COVID-19 pandemic (Hall 2023; Pearse 2022). Efforts are ongoing to increase market foothold in China and other Asian countries.

2.2.2 Velvet industry

China has recently become the largest customer for New Zealand velvet, followed by South Korea, with emerging interest also seen in Taiwan. In 2024, the velvet export value was \$104 million, with 1,043 tonnes exported at the average price of \$105.00 per kg. New Zealand velvet is marketed for use in traditional medicine, where it is promoted for a range of purported health benefits including wound healing, stress reduction and immune support. However, robust clinical evidence supporting many of these claims remains limited (DINZ n.d.; Hall 2023; Pearse 2022). DINZ focuses on premium positioning and sustainable on-farm value. Signing Free Trade Agreements (FTAs) with China (2008) and South Korea (2015) has significantly improved market access. The industry strategically advertises towards new consumers, 30-50-year-old mothers in South Korea. Promoting velvet as a healthy functional food ingredient beyond traditional medicine (Pearse 2022).

2.2.3 By-product industry

The New Zealand deer industry generates significant value not only from venison and velvet but also from a variety of other by-products that were historically considered waste (Hall 2023).

Now they contribute to the export revenue. China and South Korea are the largest consumers of pizzlies, tails, sinews, and dried blood products, which are high-value medicine ingredients for vitality and remedies. They are sometimes boiled, used in soups or consumed with alcohol (Hall 2023; Nixon 2004; Pearse 2022). China imported 80% of New Zealand's total exports of these by-products in 2022, with an export value of \$14 million in 2023/24, excluding velvet (DINZ n.d.; Hall 2023). Hong Kong, Thailand, and Malaysia are also significant markets for oriental medicine products. For skins and hides, which are made into leather products, the main markets include Germany, America, and Japan (Hall 2023; Nixon 2004). Edible offal is popular in European and Scandinavian markets, with the United States being a consistent major importer, primarily for the pet food industry. China and Hong Kong also import offal for pet food (Nixon 2004).

2.3 Deer milk industry

2.3.1 History of deer milking

Reindeer milking emerged thousands of years ago in northern Eurasia, with intensive practice evolving at least 2,000 years ago around Lake Baikal in eastern Siberia, and later spread to nearby ethnic groups along the borders of Russia, Mongolia, and China (Holand et al. 2017; Wang et al. 2017). There are records of the Sámi in northern Fennoscandia also milking reindeer, influenced by Nordic milking husbandry. Traditionally, the deer were maintained in small, tame, multipurpose herds for milk, as well as hunting and gathering. The milk was consumed fresh, in tea, or processed into cheese, butter, and sour cream, also serving medicinal purposes (Holand et al. 2017; Wang et al. 2017). The practice faded out in the early 1900s, however, attempts at commercial reindeer dairying in the Soviet Union occurred in the late 1920s. Initial reports were enthusiastic about the possibility of large-scale production, but the late 1930s shut the idea down due to the change in the economy. As collectivisation was introduced and advancements in technology favoured extensive meat production (Fondahl 1989).

2.3.2 Deer milking industry in New Zealand

New Zealand has three known commercial deer milking operations. These include Aratiatia Farm, located north of Taupō, which is owned and operated by Pāmu (Piddock 2023). The second is the fifth-generation McIntyre family's Benio Farm in Southland, which has worked

with Pāmu since 2016, supplying milk for the company's deer milk brand (Pāmu n.d.). The third, Deer Milking New Zealand (DMNZ), was established by Mark Faulks and Graham Carr at Lincoln Hills Farm in South Canterbury in 2016 (NZIPIM 2020). On these farms, the red deer hinds are typically milked once a day, with the milking season running from early summer (December/January) to autumn (April/May). The milking season averages 140 days, with approximately 1 litre of milk produced per hind per day. (NZIPIM 2020; Piddock 2023). Since the milking season is so short and production levels are low when compared to other animals, the high prices of deer milk products reflect its scarcity.

Pāmu Deer Milk is marketed as a premium nutrition product, naturally high in energy, protein, calcium, phosphorus, and vitamin A, with lower lactose than bovine milk (Pāmu n.d.). A Pāmu-funded study by Kruger et al. (2023) reported improvements in nutritional status among older women at risk of malnutrition, with increases in muscle mass observed in participants with higher body mass index. As this evidence comes from a single company-funded study, further independent research is required to confirm these findings. Currently, Pāmu offers a rich, creamy milk powder that's suitable as a drink or mixed into other foods at \$143.75 per 420g (Pāmu n.d.). It has been claimed that demand for this powder exceeds supply (Piddock 2023). Pāmu also markets two additional deer milk powder products. Doe Nutrition Beauty is promoted as supporting a youthful appearance, while Doe Nutrition Revive is marketed for brain development (Pāmu n.d.). These are product marketing claims rather than conclusions established through independent scientific evaluation. Pāmu also supplies powdered deer milk for a probiotic range specifically formulated for children in Hong Kong, as well as for cosmetic products sold in Korea (Pāmu n.d.).

Pāmu Deer Milk is currently being used by chefs in fine dining restaurants in New Zealand and Australia. It has also gained international exposure by appearing on the menu at prestigious events such as the Asia's Best Restaurant Awards in Macau in 2019 (Pāmu 2019). Separately, DMNZ has developed the world's first deer milk cheese, in Gouda and Havarti varieties, with the aim of exporting (NZIPIM 2020). These developments demonstrate the commercial potential of deer milk as a premium, value-added product. However, the industry's growth is currently constrained by the limited number of commercial deer dairies, the short seasonal lactation period, and the relatively low milk yield per hind. These factors restrict production volume and create challenges in maintaining a consistent year-round supply. Consequently, while New Zealand's large farmed deer population presents opportunities for future industry

expansion and diversification of farm income beyond venison and velvet production, the current scale of the industry also limits the availability of representative milk samples for research. This presents a challenge for developing robust and broadly applicable MIR prediction equations, highlighting the importance of studies that evaluate the performance of existing prediction equations and improve understanding of the spectral regions associated with deer milk composition.

2.4 Composition of deer milk

2.4.1 Milk proteins

Published studies have reported total protein concentrations in red deer milk ranging from 6.8% to 10.6%, with several studies observing increases in protein concentration as lactation progresses (Li et al. 2023; Wang et al. 2017). Comparing these studies with common dairy species, red deer milk is reported to have a higher protein concentration than in bovine milk, although the exact difference depends on the stage of lactation and the population studied. Wang et al. (2017) reported that the average protein concentration is approximately 7.5% in red deer milk compared with around 4.1% in bovine milk. It is also generally higher in protein than milks from caprine or ovine, though comparable to or slightly higher than buffalo and other cervids like reindeer, whose milk averages 9.9% protein (Berruga et al. 2021; Gjølstein et al. 2004).

Caseins have consistently been reported as the predominant protein fraction in red deer milk. Li et al. (2023) reported that caseins accounted for approximately 81.7% of total milk protein. β -casein was the most abundant casein type and remained relatively stable throughout lactation, at a higher concentration (32.9 g/L) than in bovine (15.1 g/L) and caprine (26.3 g/L) milk (Li et al. 2023; Wang et al. 2017). Whereas κ -casein tends to increase over lactation, whereas α s2-casein decreases. Studies have reported relatively low concentrations of total whey protein in red deer milk, primarily β -lactoglobulin and α -lactalbumin, accounting for less than 1.0% (~11-15 g/L) of the milk, which is comparable to bovine milk despite the overall higher total protein in deer milk (Li et al. 2023; Roy et al. 2020; Wang et al. 2017). Wang et al. (2017) reported lactoferrin concentrations of approximately 2.6 g/L in red deer milk. Reported casein-to-whey protein ratios in red deer milk range from approximately 80:20 to 85:15, which is comparable to that of bovine and buffalo milk, both 82:18. Caprine (78:22) and ovine (76:24) milks also exhibit different ratios, lower in casein (Roy et al. 2020).

2.4.2 Milk fats

Published studies have reported fat concentrations in red deer milk ranging from 6.0–19.7 g/100 mL, generally exceeding those reported for bovine, caprine and ovine milk (Roy et al. 2020). Gathercole and Reis (2021) reported that fat accounted for approximately 47% (42–52%) of total milk solids in red deer milk, making it the primary component, in contrast to bovine and caprine milk, where lactose is dominant. The cream-like consistency resulting from these elevated fat concentrations is a biological adaptation designed to meet the intensive energy demands associated with the rapid growth and early development of fawns in often challenging environments (Holand et al. 2017).

According to Gathercole and Reis (2021), approximately 98% of deer milk fat consists of triacylglycerols (TAGs), which serve as an important energy source and are composed of fatty acids (FAs), including saturated, monounsaturated, polyunsaturated, and branched-chain FAs. Reported FA profiles indicate that saturated FAs (SFAs) comprise approximately 53% of total FAs, with palmitic acid (C16:0) being the most abundant at about 24%. Other significant SFAs include myristic acid (C14:0) at roughly 10% and stearic acid (C18:0) at around 12%. Monounsaturated FAs (MUFAs) account for approximately 32% of total FAs, predominantly oleic acid (C18:1), which represents around 20% of total FAs (Gathercole and Reis 2021; Holand et al. 2017; Li et al. 2023). Polyunsaturated FAs (PUFAs) comprise approximately 6% of total FAs and include linoleic acid, conjugated linoleic acid (CLA), and α -linolenic acid (Gathercole and Reis 2021; Li et al. 2023). Notably, α -linolenic acid (C18:3 n-3) is more abundant in red deer milk fat than in bovine, ovine, and caprine milk. Li et al. (2023) also identified six branched-chain FAs (BCFAs) in red deer milk: iso C14, iso C15, anteiso C15, iso C16, iso C17, and anteiso C17, which collectively accounted for approximately 2.85% of total FAs.

2.4.3 Lactose

Lactose is the major carbohydrate and milk sugar found in the milk of most mammalian species, where it primarily functions as an osmotic regulator, influencing the milk's water content (Wang et al. 2017). Published studies report considerable variation in lactose concentration among deer species and across lactation, with lactose concentrations in red deer milk ranging from 2.6 to 6.2%, with reported mean values between approximately 4.3 and 5.9% (Berruga et al. 2021; Li et al. 2023). In reindeer, reported lactose concentrations are generally lower,

averaging approximately 3–3.5% during early lactation before declining throughout lactation (Gjøstein et al. 2004; Holand et al. 2017; Li et al. 2023). When compared to milk from common dairy animals, deer milk has generally demonstrated lower lactose content than bovine milk (4.4–5.6%), and sheep milk (4.1–5.9%) (Roy et al. 2020).

2.4.4 Vitamins and minerals

Mineral composition contributes to the nutritional value of red deer milk and may influence physicochemical properties associated with milk processing. Malacarne et al. (2015) reported that red deer milk contains, on average, 273 ± 33 mg/100 g calcium, 191 ± 15 mg/100 g phosphorus and 18 ± 2 mg/100 g magnesium, with concentrations exceeding those commonly reported in caprine and bovine milk. The high calcium and phosphorus concentrations are associated with the elevated casein content characteristic of deer milk. Sodium concentrations averaged 41 ± 7 mg/100 g, while potassium and chloride concentrations averaged 126 ± 6 mg/100 g and 68.91 ± 14.16 mg/100 g, respectively, with variation compared with other ruminant milks (Malacarne et al. 2015).

Mineral concentrations can also change throughout lactation. In red deer milk, calcium, phosphorus and magnesium concentrations remain relatively stable, whereas sodium and chloride increase and potassium decreases as lactation progresses. Trace minerals such as iron and zinc increase during later lactation, potentially supporting continued nutrient transfer to the developing fawn despite declining milk yield. Selenium concentrations also increase during late lactation, while copper decreases after early lactation (Gallego et al. 2006; Li et al. 2023).

Information regarding the vitamin profile of red deer milk remains limited. Studies in other cervid species have reported the presence of fat-soluble vitamins, including vitamins A, D and E, with reindeer milk containing higher concentrations than bovine milk during early lactation (Holand et al. 2017). However, due to the limited availability of vitamin data in red deer milk and the limited ability of MIR spectroscopy to directly predict vitamin and mineral concentrations, these components are not a primary focus of the current study.

2.4.5 Known factors affecting milk composition

The composition of deer milk varies both among species and between studies of the same species (Table 2.1). However, direct comparison of reported values should be interpreted

cautiously because studies differ substantially in sample size, sampling design, stage of lactation, management system, diet and analytical methodology. Consequently, the reported variation likely reflects both biological differences and methodological variation rather than definitive species-level composition.

Despite the variability, several consistent patterns emerge across the literature. Reindeer milk generally contains higher fat and protein concentrations than other cervid species, reflecting an adaptation to the energetic demands of calves developing in cold northern environments (Gjøstein et al. 2004; Holand et al. 2017). In contrast, red deer and fallow deer exhibit broader variation in milk composition among studies, largely reflecting differences in lactation stage, nutritional management and study design rather than clear species-specific differences (Berruga et al. 2021; Landete-Castillejos et al. 2000; Li et al. 2023; Malacarne et al. 2015). Across all species, lactose concentrations are consistently lower than fat and protein concentrations and generally exhibit less variation than the other major milk components.

One of the major factors influencing deer milk composition is stage of lactation. Across red deer, fallow deer and reindeer, fat and protein concentrations generally increase as lactation progresses, whereas lactose either decreases gradually or remains relatively stable (Bovolenta et al. 2013; Gjøstein et al. 2004; Landete-Castillejos et al. 2000; Li et al. 2023; Wang et al. 2017). For example, Li et al. (2023) reported that fat concentration in New Zealand red deer increased from 7.3% during week 3 of lactation to 13.0% by week 16, while Landete-Castillejos et al. (2000) observed a similar increase from 8.3% to 14.4% over 34 weeks of lactation. Comparable trends have also been reported in reindeer, where fat concentration increased from 11.4% to 21.5% between weeks 3 and 24 postpartum (Gjøstein et al. 2004). Protein concentration followed a similar pattern, increasing from approximately 7.0% to 8.6% across lactation in New Zealand red deer (Li et al. 2023), while the proportion of casein remained relatively constant. These compositional changes occur as milk yield declines, increasing the energy density of milk and helping maintain nutrient transfer to the growing fawn during late lactation (Gjøstein et al. 2004; Holand et al. 2017; Wang et al. 2017).

Table 2.1. Reported ranges of deer milk composition from different deer species.

Species	Study	Animals (n)	Lactation stage	Study conditions	Protein (%)	Fat (%)	Lactose (%)
Red deer (<i>Cervus elaphus</i>)	Berruga et al. (2021)	5 hinds	Weeks 4, 6, 10, 14 and 18	Captive Iberian red deer (Spain); irrigated pasture with concentrate and total mixed ration	6.77–7.78	9.00–12.70	4.17–4.72
	Landete-Castillejos et al. (2000)	14 hinds	Weeks 2–34	Captive Iberian red deer (Spain); barley straw, hay and concentrate; oxytocin-assisted milking	6.3–8.8	8.3–14.4	5.5–6.2
	Li et al. (2023)	120 lactating does	Weeks 3–16	Commercial NZ deer dairy farm; pasture-based with grain supplementation	7.0–8.6	7.3–13.0	5.0–5.8
	Malacarne et al. (2015)	23 hinds	Months 1–4	Two commercial captive herds (Italy); each hind sampled once	7.32 ± 1.00	10.85 ± 1.68	4.17 ± 0.48

Table 2.1. Continued. Reported ranges of deer milk composition from different deer species.

Species	Study	Animals (n)	Lactation stage	Study conditions	Protein (%)	Fat (%)	Lactose (%)
Fallow deer (<i>Dama dama</i>)	Bovolenta et al. (2013)	13 lactating does (40 females enrolled over 2 years)	45, 90 and 135 days postpartum	Pasture-based system (Italy); herbage-only vs concentrate- supplemented does	5.66–9.29	10.0–20.4	4.27–4.91
	Malacarne et al. (2015)	31 does	Approximately month 3	Captive experimental herd (Italy); each doe sampled once	7.45 ± 0.80	14.09 ± 3.12	3.62 ± 0.47
Reindeer (<i>Rangifer tarandus</i>)	Gjøstein et al. (2004)	11 does (11 doe- lactations)	Weeks 3–24 postpartum	Semi-domesticated herd (Norway); hay and concentrate offered ad libitum	7.7–12.2	11.4–21.5	1.2–3.6
White-tailed deer (<i>Odocoileus virginianus</i>)	Silver (1961)	1 doe	24–48 h, 1 week, 1 month and 5 months postpartum	Captive doe; late- lactation sample from wild deer during hunting season	8.83–11.50	7.50–18.00	2.22–3.82
	Youatt et al. (1965)	10 does (34 milk samples)	1, 3, 7 and 21 days postpartum	Captive herd (Michigan, USA); oxytocin-assisted milking	Not reported	8.1–10.3	Not reported

In contrast, lactose generally shows the opposite pattern. In New Zealand red deer, lactose declined modestly throughout lactation (Li et al. 2023), whereas Landete-Castillejos et al. (2000) reported relatively stable lactose concentrations in Iberian red deer. Reindeer exhibited a much greater decline, with lactose decreasing from 3.6% during early lactation to 1.2% by week 24 (Gjøstein et al. 2004), while a similar decline has also been reported in white-tailed deer (Youatt et al. 1965). Overall, these lactational changes resemble those reported in bovine, ovine and caprine milk, suggesting conserved physiological adaptations across ruminant species (Li et al. 2023).

Maternal nutrition also influences deer milk composition. Bovolenta et al. (2013) demonstrated that supplementary feeding increased lactose concentration in fallow deer milk compared with herbage-fed does (4.5 vs. 4.2 g/100 mL). The study also found that the relationship between offspring sex and milk composition depended on maternal nutritional status. Among herbage-fed does, milk produced for female fawns contained higher fat (16.1 vs. 14.2 g/100 mL) and protein (8.76 vs. 8.04 g/100 mL) concentrations than milk produced for males, whereas the opposite pattern was observed in supplementary-fed does. These findings indicate a diet × offspring sex interaction rather than a consistent sex-specific pattern of maternal investment.

Mineral composition also changes throughout lactation. In red deer milk, calcium, phosphorus and magnesium remain relatively stable, whereas sodium and chloride concentrations increase and potassium decreases as lactation progresses (Gallego et al. 2006; Li et al. 2023). Iron, zinc and selenium concentrations also increase during late lactation, while copper concentrations decline after early lactation. These changes are thought to compensate for declining milk yield by maintaining the daily transfer of essential micronutrients required for skeletal development, immune function and growth of the fawn (Gallego et al. 2006; Li et al. 2023).

2.4.6 Nutritional value

Deer milk's nutritional value distinguishes it from other dairy species' milks. Its high protein content contains greater amounts of essential amino acids, reindeer milk itself contains at least double the essential amino acids as bovine milk. β -casein is the predominant protein in deer milk. In an in vitro simulated duodenal digestion model, deer milk casein was digested more rapidly than bovine casein, reaching complete digestion within 10 minutes. While these findings suggest differences in digestive behaviour between species, further in vivo research is

required before implications for human nutrition, including infant nutrition, can be established (Gathercole and Reis 2021; Li et al. 2023; Wang et al. 2017). Lactoferrin is an iron-binding protein that has been reported to possess antimicrobial and antiviral properties in experimental studies. During simulated digestion, deer milk lactoferrin generated a different peptide profile from bovine lactoferrin, suggesting potential functional differences that require further investigation. While β -lactoglobulin resists digestion, the difference in whey protein amino acid sequences suggests unique bioactive peptides when compared to other species (Gathercole and Reis 2021; Wang et al. 2017). Deer milk contains relatively high concentrations of branched-chain fatty acids (BCFAs), compounds that have been investigated for potential anti-inflammatory and anti-cancer activities in experimental studies. However, evidence for these effects in humans remains limited (Li et al. 2023). The high mineral content of deer milk may contribute to a relatively high buffering capacity, which has been proposed as one possible explanation for its historical use in digestive health. However, direct clinical evidence supporting benefits for conditions such as gastric ulcers is limited (Wang et al. 2017). Deer milk's rich nutrition, favourable cheese-making properties, and high vitamin and lipid content make it advantageous for high-value food and cosmetic products.

Although the present study focuses on the major milk components of fat, protein and lactose, understanding the broader composition of red deer milk provides important biological and commercial context. The premium value of deer milk is due to its distinctive nutritional composition, including its high fat and protein concentrations, relatively low lactose content, favourable fatty acid profile, and elevated concentrations of selected minerals and vitamins reported in the literature. These characteristics distinguish deer milk from traditional dairy species and underpin its marketing as a premium nutritional product. Therefore, a comprehensive understanding of deer milk composition provides an important foundation for interpreting the application of MIR spectroscopy in this species.

2.5 Mid-infrared spectroscopy

Mid-infrared (MIR) spectroscopy is an analytical technique used to identify and quantify chemical compounds by measuring their interaction with infrared radiation. The MIR region of the electromagnetic spectrum spans approximately $4000\text{--}400\text{ cm}^{-1}$ ($2.5\text{--}25\text{ }\mu\text{m}$) and contains the fundamental vibrational frequencies of most organic molecules (Hsu 1997; Stuart 2004). As atoms within molecules are in continuous motion, chemical bonds vibrate through

stretching (changes in bond length) and bending (changes in bond angle). When the frequency of incident infrared radiation matches the natural vibrational frequency of a chemical bond, energy is absorbed, producing characteristic absorption bands that are unique to the molecular structure (Gendrin et al. 2008; Ní Fhuaráin et al. 2024). These absorption bands collectively form an infrared spectrum that can be used to identify chemical compounds and quantify their concentrations.

The MIR spectrum is commonly divided into two regions: the functional group region (4000–1300 cm^{-1}) and the fingerprint region (1500–600 cm^{-1}). The functional group region contains absorption bands associated with specific chemical bonds, allowing individual molecular groups to be identified (Hsu 1997). Within this region, the X–H stretching region (4000–2500 cm^{-1}) contains characteristic absorptions arising from O–H, N–H and C–H stretching vibrations. Hydrogen bonding can broaden and shift O–H and N–H absorption bands towards lower frequencies. The triple-bond region (2500–2000 cm^{-1}) contains stretching vibrations from nitrile ($\text{C}\equiv\text{N}$) and alkyne ($\text{C}\equiv\text{C}$) bonds, while the double-bond region (2000–1500 cm^{-1}) is dominated by strong carbonyl ($\text{C}=\text{O}$) stretching vibrations, typically occurring between 1830 and 1650 cm^{-1} depending on the chemical environment (Hsu 1997; Stuart 2004).

The functional group region also contains the principal absorption bands used for routine milk analysis because the major milk constituents possess characteristic chemical bonds that absorb at predictable frequencies (Etzion et al. 2004; Kaylegian et al. 2009). The principal MIR absorption bands associated with milk fat include ester carbonyl ($\text{C}=\text{O}$) stretching vibrations near 1743 cm^{-1} , CH_2 and CH_3 bending vibrations between 1390 and 1454 cm^{-1} , and C–H stretching vibrations between 2862 and 2927 cm^{-1} (Kaylegian et al. 2009; Mohamed et al. 2021). Milk proteins are characterised by the amide I (approximately 1600–1700 cm^{-1}) and amide II (approximately 1540–1570 cm^{-1}) absorption bands, which arise predominantly from peptide $\text{C}=\text{O}$ stretching together with C–N stretching and N–H bending vibrations (Etzion et al. 2004; López-Lorente and Mizaikoff 2016). Lactose and other milk carbohydrates exhibit their strongest absorptions between 1045 and 1250 cm^{-1} , primarily due to C–O, C–C and C–O–C stretching vibrations associated with glycosidic bonds (Kaylegian et al. 2009; Mohamed et al. 2021; Pinto et al. 2021). Consequently, MIR spectroscopy does not directly measure milk fat, protein or lactose; rather, it quantifies the characteristic molecular vibrations associated with these components and relates spectral absorbance to composition using calibration models developed from reference analytical methods (Mohamed et al. 2021). Chemometric methods,

such as partial least squares regression (PLSR), exploit absorbance information across multiple spectral regions simultaneously to predict milk composition with high accuracy (Etzion et al. 2004; Pinto et al. 2021). Studies across cattle, sheep, goats and camel milk have consistently demonstrated that these conserved absorption regions underpin the successful prediction of milk composition using MIR spectroscopy, despite differences in milk composition between species (Grelet et al. 2021; Mohamed et al. 2021).

Although several important milk-component absorption bands occur within the functional group region, additional discriminatory information is provided by the fingerprint region (1500–600 cm^{-1}) (Hsu 1997). Unlike the functional group region, where absorption bands can be assigned to individual chemical bonds, the fingerprint region contains complex combinations of molecular vibrations that are unique to the complete molecular structure. Minor steric and electronic differences between molecules produce distinctive spectral patterns, allowing compounds with similar functional groups but different molecular arrangements to be differentiated (Waynant et al. 2001). Consequently, the fingerprint region acts as a unique molecular "fingerprint" and is particularly valuable for confirming compound identity and distinguishing structural isomers (Hsu 1997; Stuart 2004).

2.5.1 Comparative studies completed on bovine, ovine, and caprine

Model performance in MIR spectroscopy studies is typically evaluated using a combination of calibration, cross-validation and external validation, together with several complementary performance metrics. Calibration statistics, such as the coefficient of determination (R^2) and root mean square error of calibration (RMSE_C), describe how well a model fits the dataset used for model development but generally provide optimistic estimates of predictive performance because they are calculated using the same data on which the model was trained. Cross-validation, commonly reported as the coefficient of determination (R^2_{CV}) and the root mean square error of cross-validation (RMSE_{CV} or SE_{CV}), provides a more realistic estimate by repeatedly withholding subsets of the calibration data during model development. However, because all samples originate from the same calibration population, predictive performance may still be overestimated when related animals, herds or sampling conditions are represented in both the calibration and validation subsets. External validation, reported using an independent dataset that was not involved in model development (e.g., R^2_v , R^2_{ExV} , R^2_{Ext} or r_v , together with RMSE_P or SE_P), is therefore considered the most rigorous assessment of model

generalisability and is preferred when evaluating the robustness of MIR prediction equations (Ho et al. 2021; Ezenarro and Schorn-García 2025; Ramspek et al. 2021)

In addition to R^2 and prediction error, many studies report the residual predictive deviation (RPD) or ratio of error range (RER), which relate prediction error to the natural variation present within the reference dataset, with higher values indicating greater predictive performance. Regression slope and prediction bias are also frequently reported to assess systematic over- or under-prediction and agreement between predicted and reference values, although terminology varies slightly among studies (e.g., $RMSEP$ versus SEP , $RMSECV$ versus $SECV$, and R^2_v versus R^2_{Ext}).

Several bovine studies have evaluated MIR spectroscopy for predicting milk protein composition using PLSR-based calibration models, although the validation strategies differed between studies. De Marchi et al. (2009) developed calibration equations using random cross-validation, whereas McDermott et al. (2016) additionally assessed model performance using an independent external validation dataset. Etzion et al. (2004) validated prediction equations using fresh milk samples collected from individual cows after calibration on standardised milk samples. Across these studies, prediction performance was generally strongest for total protein and major protein fractions, with cross-validation or external validation R^2 (or correlation coefficients) typically ranging from approximately 0.50 to 0.75. Prediction errors were relatively low ($RMSE \approx 0.08\text{--}0.22\%$ protein), although RPD values generally remained below 2 for individual protein fractions, indicating that the models were better suited to screening or discrimination than precise analytical quantification. Together, these studies demonstrate that MIR spectroscopy can reliably predict major protein traits, although prediction accuracy declines for lower-abundance protein fractions and free amino acids.

Numerous studies have investigated MIR prediction of milk fatty acid composition using PLSR models combined with cross-validation and independent external validation (Rutten et al. 2009; Soyeurt et al. 2011; Maurice-Van Eijndhoven et al. 2013; Ferrand et al. 2011; De Marchi et al. 2011; Caredda et al. 2016). Most studies used either random calibration-validation splits or repeated cross-validation, while several additionally assessed model robustness using completely independent validation datasets. Across these studies, major fatty acids and fatty acid classes consistently achieved high predictive performance, with cross-validation and external validation R^2 commonly exceeding 0.90 and RPD values generally between 3 and 10. Conversely, minor or low-abundance fatty acids typically produced R^2 values below 0.70 and

RPD values below 2, indicating limited quantitative predictive ability. Several authors further demonstrated that prediction accuracy could be improved through spectral pre-processing, variable selection techniques such as genetic algorithms, or the inclusion of more diverse calibration populations, all of which enhanced model robustness during external validation.

Studies predicting milk coagulation properties generally used modified PLSR models with either cross-validation alone or a combination of cross-validation and independent external validation (Dal Zotto et al. 2008; De Marchi et al. 2009a; Manuelian et al. 2019; Cellesi et al. 2019; De Marchi et al. 2013; Goi et al. 2025). Regardless of species, predictive performance was consistently lower than that reported for the major milk components. Rennet coagulation time (RCT) generally produced the most accurate models, with validation R^2 values typically between 0.60 and 0.77 and RPD values around 2, whereas curd-firming time and curd firmness frequently produced R^2 values below 0.50 and RPD values below 1.5. These findings indicate that MIR spectroscopy is suitable for screening coagulation characteristics but is not yet sufficiently accurate to replace conventional laboratory measurements.

To this author's knowledge, there have been no published studies using MIR spectroscopy to predict any components or technological properties of deer milk. The studies reviewed in this section show that the ability of MIR spectroscopy to predict milk traits depends on both the properties of the trait being measured and the quality of the reference dataset used to develop the prediction equation. Across bovine, ovine and caprine milk, MIR spectroscopy has consistently produced high prediction accuracies for the major milk components, including fat, protein and lactose. Researchers have also investigated its ability to predict traits such as fatty acid composition, milk coagulation properties, cheese yield, minerals and milk metabolites, with varying levels of success. In general, prediction accuracy is highest for the major milk components, while traits present in lower concentrations or with more complex chemical structures require larger and more diverse reference datasets before reliable prediction equations can be developed.

These findings are particularly relevant to red deer milk. Compared with the milk of other dairy species, red deer milk contains higher concentrations of fat and protein, lower lactose, and a distinctive fatty acid profile that contributes to its nutritional value and premium market position. Based on the success of MIR spectroscopy in other dairy species, it is reasonable to expect that the MilkoScan equations used in this study will provide strong predictions for the major milk components of fat, protein and lactose. There is an opportunity for future MIR

research for the distinctive fatty acid profile, vitamin and mineral composition of red deer milk. As larger and more representative deer milk datasets become available, future studies could investigate whether these nutritionally important compounds can also be predicted using MIR prediction equations.

2.5.2 Advantages of mid-infrared spectroscopy

MIR spectroscopy is a valuable tool for predicting milk composition due to being cost-effective, rapid, non-destructive, and suitable for large-scale application, making it an essential analytical method in the dairy industry (Gengler et al. 2016). Unlike reference methods such as gas chromatography or high-performance liquid chromatography, which are expensive and time-consuming, MIR allows the analysis of hundreds of samples per hour (Dal Zotto et al. 2008; Maurice-Van Eijndhoven et al. 2013). Machefert et al. (2024) reported that between 100 and 600 samples can be processed in a single hour, such a level of output is unfeasible with traditional methods.

A key advantage of MIR is its low cost, as once the equipment is installed, spectra can be collected and stored electronically without ongoing chemical or consumable costs (Cellesi et al. 2019). These stored spectra can be re-analysed at any point to develop new predictive models with no extra costs, adding long-term value to routine milk testing (Franzoi et al. 2023; Ní Fhuaráin et al. 2024). This cost-efficiency has led to the adoption of MIR worldwide, particularly in milk analysis laboratories, where it is routinely used to predict fat, protein, lactose, and urea (Gengler et al. 2016). MIR is also non-destructive, meaning the milk sample remains intact and additional testing is able to be completed. MIR can also be operated without highly trained technicians (Baba et al. 2021; Caredda et al. 2016).

Beyond routine traits, MIR is capable of predicting a wide range of complex milk components. These include fatty acid profiles, mineral content, protein fractions, lactoferrin, as well as milk coagulation properties. MIR can also predict metabolic indications present in milk, such as beta-hydroxybutyrate and acetone, which are linked to energy balance and ketosis. MIR not only assesses milk quality and composition but can also provide indirect indicators of animal health, nutrition, and environmental impact that can be used to monitor herd health and inform farm management decisions (Baba et al. 2021; Gengler et al. 2016; Grelet et al. 2021).

MIR spectroscopy also plays an increasingly important role in genetic improvement and breeding programmes. Beyond predicting milk composition, individual MIR wavelengths can be treated as quantitative phenotypes, enabling their use in phenomic selection and genomic analyses. Machefert et al. (2024) demonstrated moderate to high heritability across the MIR spectrum (0.13–0.42) and strong genomic correlations between highly heritable spectral regions and economically important traits such as milk fat and protein content, supporting the use of MIR spectra as indirect selection criteria. Furthermore, MIR phenotypic data can be integrated into genome-wide association studies (GWAS) and PLSR analyses to identify genomic regions associated with milk composition. For example, Marie-Etancelin and Rupp (2025) demonstrated that analysing the complete MIR spectrum in Lacaune dairy ewes identified quantitative trait loci (QTL) associated with both spectral wavelengths and predicted milk composition traits, while also revealing wavelength-specific QTL that were not detected using predicted milk traits alone. These findings indicate that the complete MIR spectrum contains additional genetically informative variation beyond conventional milk composition traits, improving the ability to detect and characterise loci influencing milk composition (Benedet et al. 2019; Marie-Etancelin and Rupp 2025). Integrating MIR data with complementary information, including genomic, pedigree and on-farm management data, has also been shown to improve the accuracy of predictive models and animal evaluation (Baba et al. 2021).

The use of MIR-derived phenotypes in genetic evaluation has the potential to improve breeding programmes by increasing the accuracy of breeding value estimation while reducing the cost and time associated with direct laboratory phenotyping. Because numerous milk traits can be predicted from a single spectral measurement, MIR data provide a high-throughput source of phenotypic information that can be incorporated into genetic evaluations without requiring separate laboratory analyses for each trait (Dagnachew et al. 2013). From an industry perspective, accurate MIR predictions also have practical value for milk processors, enabling payment systems based on specific milk quality characteristics and technological properties, thereby allowing milk to be rewarded according to its suitability for different dairy products and processing requirements (Careda et al. 2016; Gengler et al. 2016; Goi et al. 2025).

2.5.4 Limitations and considerations of mid-infrared spectroscopy

Different prediction equations vary in accuracy and reliability due to differences in calibration sets and an incomplete understanding of spectral variability, leading to equations being

developed from different spectrometer bands (Soyeurt et al. 2019). Validation strategies like random or k-fold cross-validation can inflate accuracies and predictions, as herd data may overlap in training and testing data. Herd-independent validation reveals lower accuracy, suggesting models developed under specific conditions may not perform well in diverse real-world conditions (Baba et al. 2021; Grelet et al. 2021; Tiplady 2022). Limited datasets often lead to complex and overfitted models. Even reliable models will face issues when they encounter samples outside calibration variability, and potential sources of error include inconsistencies between laboratory technicians, a lack of replicate measurements, and methods that are subjective (Grelet et al. 2021; Ní Fhuaráin et al. 2024).

Prediction equations developed for one spectrometer may not be trusted on another, even of the same model, due to differences in spectral responses. This makes transferring models between research and routine use difficult (Gengler et al. 2016; Grelet et al. 2021; Tiplady 2022). While standardisation methods can improve reproducibility, they require identical reference samples to be analysed across all instruments, which creates logistical and biosecurity barriers for international collaborations. Individual spectrometers also display instability over time due to factors like component deterioration, environmental conditions, maintenance, and electronic drifts, leading to incorrect predictions (Grelet et al. 2021; Tiplady 2022). Raw milk samples display species, breed, dietary and seasonal variations in composition and functionality. So to build reliable predictive models, MIR spectra and reference methods must be collected over multiple years and across all seasons. Which is expensive and time-consuming, as there is an ongoing need for adjustment. (Ní Fhuaráin et al. 2024).

While MIR has high prediction accuracies for major components like fat, protein, and lactose, MIR models for minor components (e.g. specific fatty acids, minerals) or animal-related traits (e.g. energy balance, metabolic status, methane emissions, pregnancy status) are less accurate or indirect (Gengler et al. 2016; Grelet et al. 2021; Tiplady 2022). MIR often struggles to detect molecules present in small quantities and may not discriminate well between chemically similar compounds. These models with relatively low precision may still be valuable for comparing groups, discriminating between high and low values, or detecting extreme cases (Grelet et al. 2021). Strong water absorption bands at 1,600-1700 cm⁻¹ and 3,000-3,500 cm⁻¹, mask important information and contribute to the high noise levels in the MIR spectrum. Wavenumbers in these noisy regions often show poor reproducibility across instruments (Grelet et al. 2021; Ní Fhuaráin et al. 2024). Noise in the spectra can result in outliers caused

by sample or instrument anomalies, requiring identification and removal. While pre-processing can improve accuracy by correcting signal noise, there is no universal agreement on the best treatment, and effectiveness varies by dataset (Tiplady 2022).

Even with high genetic correlations, the QTL for directly measured and MIR-predicted traits may not always be the same. Implying that selection based on MIR predictions may not target the same genetic pathways as direct selection (Gengler et al. 2016; Tiplady 2022). The use of MIR-derived traits in breeding objectives requires more research on the socio-economic weights and genetic correlations with other traits (Gengler et al. 2016). Individual MIR wavenumbers can capture genetic signals, but they do not directly represent breeding values. This would require routine storage of spectral data, which has not been a historical practice (Soyeurt et al. 2019). Integrating genomic data with MIR spectra can enhance prediction for some traits, but standard Partial Least Squares regression (PLSR) does not effectively weight heterogeneous sources, and may underperform when genomic data is added without appropriate methods (Baba et al. 2021).

While the chemical information contained within MIR spectra is determined by molecular vibrations, transforming spectral data into accurate predictions of milk composition requires calibration equations developed from reference datasets with known milk component concentrations. The predictive performance and generalisability of these equations depend on the diversity and representativeness of the calibration data, including variation across herds, lactation stages, and management systems. As a result, models developed from a limited calibration population may perform well during model development or cross-validation, but exhibit reduced predictive accuracy when applied to independent populations if they fail to capture the biological and environmental variability present in new datasets (Ezenarro and Schorn-García, 2025; Ramspek et al., 2021). Consequently, applying MIR prediction equations to new populations, such as red deer milk, requires independent external validation to determine whether the equations maintain acceptable predictive performance and generalisability under different biological and production conditions (Ho et al., 2021).

2.6 Conclusions

The reviewed literature highlights both the distinctive composition of red deer milk and the potential of MIR spectroscopy as a rapid, non-destructive tool for milk analysis. Compared

with conventional dairy species, red deer milk contains higher concentrations of fat and protein, lower lactose, and a distinctive nutritional profile that underpins its value in premium food, nutraceutical and cosmetic products. However, milk composition is influenced by factors such as lactation stage, nutrition and management, creating biological variation that must be considered when interpreting MIR predictions.

Comparative studies in bovine, ovine and caprine milk demonstrate that MIR spectroscopy consistently provides accurate prediction of the major milk components, particularly fat, protein and lactose, while prediction of lower-abundance or more complex traits, including individual fatty acids, minerals and technological properties, generally requires larger and more diverse reference datasets. Despite these advances, no published studies have evaluated the predictive performance of proprietary MIR prediction equations in red deer milk or investigated the MIR wavelengths associated with the major milk components. This represents an important knowledge gap given the increasing commercial interest in deer milk and the need for rapid, high-throughput methods of milk composition analysis.

The literature also demonstrates that robust MIR prediction equations are typically developed and evaluated using large, diverse datasets collected across multiple herds, seasons and years. In contrast, the New Zealand deer dairy industry is small, highly seasonal and currently limited to a small number of commercial operations, restricting the availability of representative milk samples. Consequently, this study was designed around the largest practical dataset currently available, comprising milk samples collected from commercial New Zealand red deer hinds representing four source herds during the 2024–2025 lactation season. Following data quality screening, 202 milk samples were available for analysis. While this dataset is smaller than those commonly used in bovine MIR studies, it reflects the current structure of the New Zealand deer dairy industry and provides an appropriate proof-of-concept dataset for evaluating the applicability of MIR prediction equations to red deer milk.

This thesis aims to evaluate the predictive performance of the FOSS MilkoScan equations for estimating fat, protein and lactose concentrations in New Zealand red deer milk. It also aims to identify the MIR wavelengths most strongly associated with each of these milk components, providing greater insight into the spectral regions contributing to prediction accuracy. Together, these objectives address an important gap in the literature and provide a foundation for future refinement and evaluation of MIR prediction equations as larger and more representative deer milk datasets become available.

Chapter 3: Materials and methods

3.1 Milk sample collection

A total of 296 milk samples were initially collected from red deer (*Cervus elaphus*) hinds aged 4-11 years originating from Aratiatia, Raft Creek, Rangitaiki, and Raurimu farms. However, only 202 samples were successfully processed for analysis due to insufficient milk yield or incomplete milking, often caused by cups being dislodged during milking. All milking was conducted at the Aratiatia Deer Milk Farm, Wairakei, Taupō, New Zealand, which operated both as a commercial farm for production and research purposes.

Milking occurred once daily in the morning between 07:30 and 08:00 am from January to March 2025. Before each milking, hinds received an intramuscular injection of oxytocin (10 IU; 1 ml) into the rump approximately one minute before milking to facilitate milk let-down. Milk yields and samples were collected three times per week: Monday, Wednesday, and Friday. Milk yield was recorded using in-line Waikato flock meters, while milk samples were collected via Tru-Test meters. Farm staff manually recorded milk volume. Samples were immediately analysed

Following the collections, samples were maintained under ambient air temperature conditions during transportation to the laboratory. Upon arrival at the Livestock Improvement Corporation (LIC) laboratory in Hamilton, New Zealand, the samples were analysed using a MilkoScan 4000 mid-infrared spectrometer (Foss, Hillerød, Denmark). Fat, protein and lactose concentrations were predicted using a pre-existing MIR calibration equation developed specifically for red deer milk used by LIC. Each sample was analysed once using the existing calibration, and no additional wet chemistry or calibration development was undertaken as part of this study.

3.2 Farming conditions and environment

All deer were maintained under a pasture-grazing system. Hinds grazed on a mixture of ryegrass, white and red clover, and chicory, with an additional 1 kg/day of grain supplementation, Deer Nut Elite, available to lactating hinds. The average pasture intake was estimated at 4kg DM per hind per day. The total effective grazing area was 192.3 hectares, including a 7-aside milking herringbone. November to March is the lactation season for red deer, aligning with late spring to early autumn in New Zealand. The farm maintains a mother-

fawn suckling system for a minimum of eight weeks before weaning, to meet the criteria of 32kg body weight. In the 2024-2025 season, the average suckling period was 56 days. During the milking season, the environmental conditions supported pasture growth with an average minimum temperature ranging from 2.4 °C to 9.9 °C, while the average maximum temperature ranged from 24.7 °C to 28.6 °C. Monthly rainfall varied between 22.8 mm and 64.2 mm, with February being the wettest month.

3.3 Data editing

The original dataset consisted of MIR spectral data and corresponding milk composition measurements (fat, protein and lactose) from 202 red deer milk samples. Additional animal information, including date of birth and herd of origin, was also available. The MIR spectra comprised pins 240–1299, representing 1,060 spectral variables. Wavelength assignments provided by the instrument manufacturer were used to identify the corresponding wavelengths for each spectral pin, as the conversion method is not publicly available. Spectral regions known to be strongly influenced by water absorption were removed to reduce spectral noise. Specifically, pins 398-447, 601-624 and 770-1296 were excluded, leaving 459 spectral variables for analysis. The dataset was randomly divided into a calibration (training) dataset containing 40% of the samples ($n = 81$) and a hold-out external validation dataset containing the remaining 60% ($n = 121$). The calibration dataset was used for model development and cross-validation, while the external validation dataset was withheld until final model evaluation.

3.4 Statistical analysis

3.4.1 Partial least square regression

All statistical analyses were performed in RStudio, including the descriptive statistics. PLSR was selected for the analysis of MIR spectral data rather than traditional Multiple Linear Regression (MLR) because MIR spectral datasets are characterised by a large number of highly correlated predictor variables. In this study, the dataset comprised 1,060 spectral variables but only 202 milk samples, meaning the number of predictors substantially exceeded the number of observations. Under these conditions, conventional MLR is unsuitable for predictive modelling because coefficient estimates become unstable, predictor collinearity inflates variance, and the risk of chance correlations is increased. In contrast, PLSR is specifically

designed to address these issues by transforming the original spectral variables into a smaller set of orthogonal latent variables that maximise the covariance between the predictor ($\mathbf{X}$) and response ($\mathbf{Y}$) matrices. This approach produces more stable and robust predictive models, even when predictor variables are highly correlated and outnumber observations, and can be integrated into local calibration strategies to account for the common non-linear relationships observed between spectral data and milk component concentrations (Benoudjit et al. 2009; Cramer III 1993). PLSR establishes a linear relationship between the predictor matrix ($\mathbf{X}$), representing the MIR spectral variables, and the response matrix ($\mathbf{Y}$), representing the milk component concentrations, through latent variables that maximise the covariance between the two matrices. The decomposition of the predictor ($\mathbf{X}$) and response ($\mathbf{Y}$) matrices in PLSR is expressed as (Rosipal and Krämer 2005):

$$\begin{aligned}\mathbf{X} &= \mathbf{TP}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ}^T + \mathbf{F}\end{aligned}$$

where $\mathbf{T}$ and $\mathbf{U}$ are the matrices of latent scores derived from the predictor ($\mathbf{X}$) and response ($\mathbf{Y}$) matrices, respectively. $\mathbf{P}$ and $\mathbf{Q}$ are the corresponding loading matrices, while $\mathbf{E}$ and $\mathbf{F}$ represent the residual variation not explained by the model.

The regression relationship between the predictor matrix ($\mathbf{X}$) and response matrix ($\mathbf{Y}$) can then be written as (Rosipal and Krämer 2005):

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F}^*$$

where $\mathbf{B}$ is the matrix of regression coefficients and $\mathbf{F}^*$ represents the residual error. The regression coefficient matrix is calculated as (Rosipal and Krämer 2005):

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T$$

where $\mathbf{W}$ is the weight matrix defining the linear combinations of spectral variables contributing to each latent variable, $\mathbf{P}$ is the loading matrix for the predictor variables, and $\mathbf{C}$ describes the relationship between the latent variables and the response variable. Through this

approach, PLSR reduces the dimensionality of highly collinear spectral data while retaining the information most relevant for predicting the response variable.

3.4.2. Model performance evaluation

Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and the residual predict deviation (RPD). The coefficient of determination describes the proportion of variation in the reference milk component values explained by the model and was calculated as (Cramer III 1993):

$$R^2 = \frac{SS_{\text{original}} - SS_{\text{residual}}}{SS_{\text{original}}}$$

where SS_{original} represents the total sum of squared deviations of the observed values from their mean, and SS_{residual} is the sum of squared differences between the observed and predicted values.

The RMSE measures the average magnitude of prediction error and was calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where y_i represents the observed value, $\hat{y}_i$ represents the predicted value, and n is the number of observations. RMSE was calculated separately for calibration ($RMSE_C$), cross-validation ($RMSE_{CV}$), and external validation ($RMSE_P$). This approach to evaluating calibration, cross-validation, and prediction performance is consistent with previous MIR spectroscopy studies (Franzoi et al. 2023; Goi et al. 2025; Manuelian et al. 2019).

The ratio of performance to deviation (RPD) was used to assess predictive performance relative to the natural variability of the reference data and was calculated as (Manuelian et al. 2019):

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

where SD is the standard deviation of the reference values and RMSE is the corresponding root mean square error. Higher RPD values indicate greater predictive performance.

3.4.3. Cross- validation

For each milk component, the optimal number of latent variables was determined using 10-fold cross-validation implemented within the pls package in R (Mevik and Wehrens 2007). Ten-fold cross-validation was selected because it provides a more reliable estimate of prediction accuracy while making efficient use of the limited calibration dataset (Hastie et al. 2009; Kohavi 1995). The calibration dataset was randomly divided into ten approximately equal subsets. During each repetition, nine subsets were used to fit the model, and the remaining subset was used for validation. This process was repeated until every subset had served once as the validation set. The $RMSE_{CV}$ was calculated for models containing increasing numbers of latent variables, allowing model complexity to be assessed while reducing the risk of overfitting. The optimal number of latent variables was then selected using the one-standard-error (1σ) rule, where the simplest model with an $RMSE_{CV}$ within one standard error of the minimum $RMSE_{CV}$ was selected. This approach recognises that cross-validation error estimates are subject to sampling variability and therefore favours a more parsimonious model when several models exhibit similar prediction performance, reducing the likelihood of overfitting while maintaining predictive accuracy (Hastie et al. 2009).

3.4.4. Identification of influential spectral regions

Following PLSR model development, the absolute PLSR regression coefficients were extracted from each model to quantify the contribution of individual spectral wavelengths to the prediction of fat, protein, and lactose concentrations. Manhattan plots were constructed by plotting the absolute regression coefficients against wavelength, enabling visual identification of the spectral regions with the greatest influence on each PLSR model. To facilitate the subsequent exploratory backward stepwise regression, the analysis was initially restricted to the 80 wavelengths with the largest absolute PLSR regression coefficients. This restriction was necessary because the stepwise regression package used in RStudio could not efficiently accommodate the full set of 459 spectral variables. Consequently, the exploratory analysis was confined to the spectral variables identified by the PLSR models as contributing most strongly

to prediction, while the PLSR models themselves were fitted using all available spectral variables.

An exploratory backward stepwise multiple regression was then performed using the 30 most influential wavelengths selected from the initial subset of 80 wavelengths. Reducing the number of predictors to 30 ensured that the number of predictor variables remained lower than the number of observations while retaining the spectral regions identified by the PLSR models as contributing most strongly to prediction. The exploratory analysis was undertaken solely to identify which influential spectral regions remained associated with variation in fat, protein, or lactose concentration and was not used for model development or validation. Because MIR spectral variables remain highly collinear, the retained wavelengths should not be interpreted as uniquely independent predictors and may vary between datasets. Consequently, the PLSR models remained the primary predictive models, while the exploratory stepwise regression was used only to assist in the interpretation of influential spectral regions.

Chapter 4: Results

4.1 Descriptive statistics

Descriptive statistics for fat, protein, and lactose concentrations in red deer milk are presented in Table 4.1. Fat had the highest mean concentration ($9.75 \pm 1.55\%$), followed by protein ($7.58 \pm 0.54\%$) and lactose ($4.19 \pm 0.39\%$). Fat also showed the greatest variation among samples ($SD = 1.55\%$), whereas lactose exhibited the lowest variation ($SD = 0.39\%$).

Table 4.1. Mean, standard deviation (SD) and number (N) for red deer milk composition traits (fat, protein, and lactose).

Trait	N	Mean	SD
Fat (%)	202	9.75	1.55
Protein (%)	202	7.58	0.54
Lactose (%)	202	4.19	0.39

4.2 Evaluation of the existing FOSS calibration by partial least squares regression

Calibration, 10-fold cross-validation, and hold-out validation statistics for the PLSR models reproducing the predictions generated by the existing FOSS calibration are summarised in Table 4.2.

Table 4.2. Calibration, 10-fold cross-validation, and validation statistics for the PLSR models predicting fat, protein, and lactose concentrations in red deer milk.

Component	Dataset	Components	R ²	RMSE (%)	RPD
Fat	Calibration	13	0.9998	0.0201	—
	10-fold CV	13	0.9995	0.0319	—
	Hold-out	13	0.9994	0.0405	40.43
Protein	Calibration	16	0.9994	0.0122	—
	10-fold CV	16	0.9979	0.0229	—
	Hold-out	16	0.9975	0.0282	19.93
Lactose	Calibration	21	0.9993	0.0094	—
	10-fold CV	21	0.9962	0.0224	—
	Hold-out	21	0.9953	0.0281	14.66

4.2.1 Fat

A PLSR model was fitted using the calibration dataset ($n = 81$). Ten-fold cross-validation was used to determine the optimal number of latent components. The minimum $RMSE_{CV}$ (0.0298%) occurred at 30 latent components; however, application of the one-standard-error rule selected a simpler model containing 13 latent components ($RMSE_{CV} = 0.0319\%$). The 13-component model was retained for subsequent analyses. The calibration model achieved an R^2_C of 0.9998 with an $RMSE_C$ of 0.0201%, while 10-fold cross-validation produced an R^2_{CV} of 0.9995 with an $RMSE_{CV}$ of 0.0319% (Table 4.2).

The hold-out validation dataset ($n = 121$) produced an R^2_V of 0.9994, an $RMSE_P$ of 0.0405%, and an RPD of 40.43 (Table 4.2). Figures 4.1 and 4.2 show the observed and predicted values for the cross-validation and hold-out validation datasets.

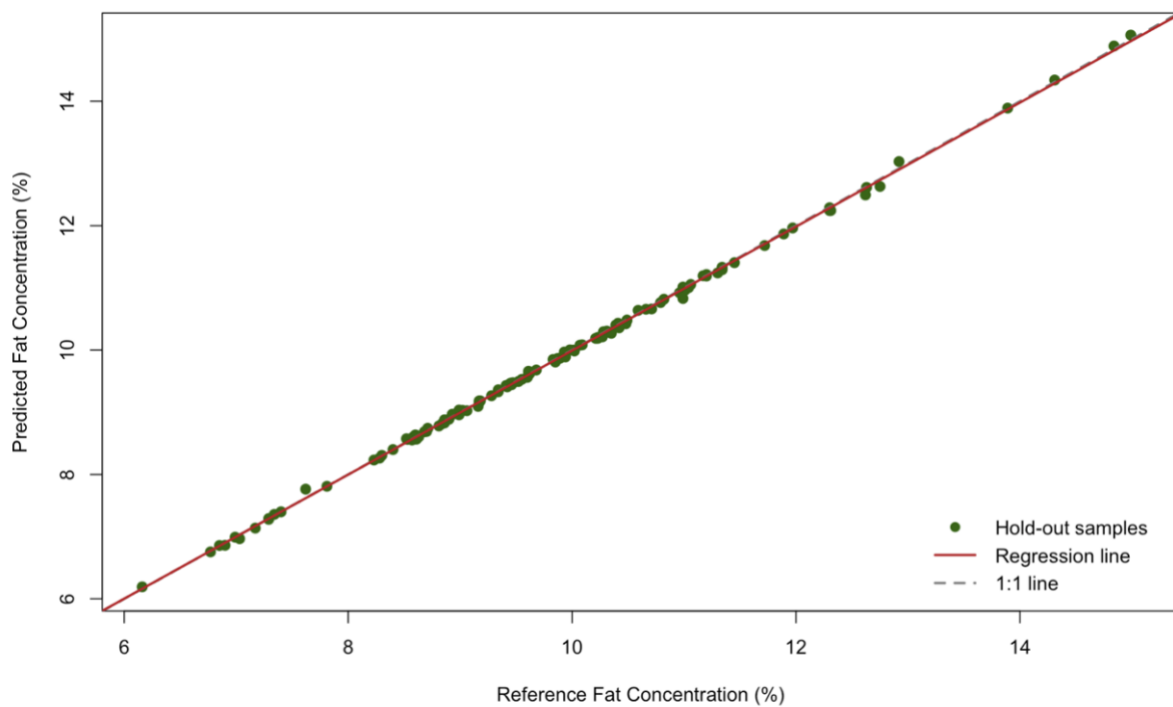


Figure 4.1. Observed versus cross-validated predicted red deer milk fat concentration (%) for the 13-component PLSR model.

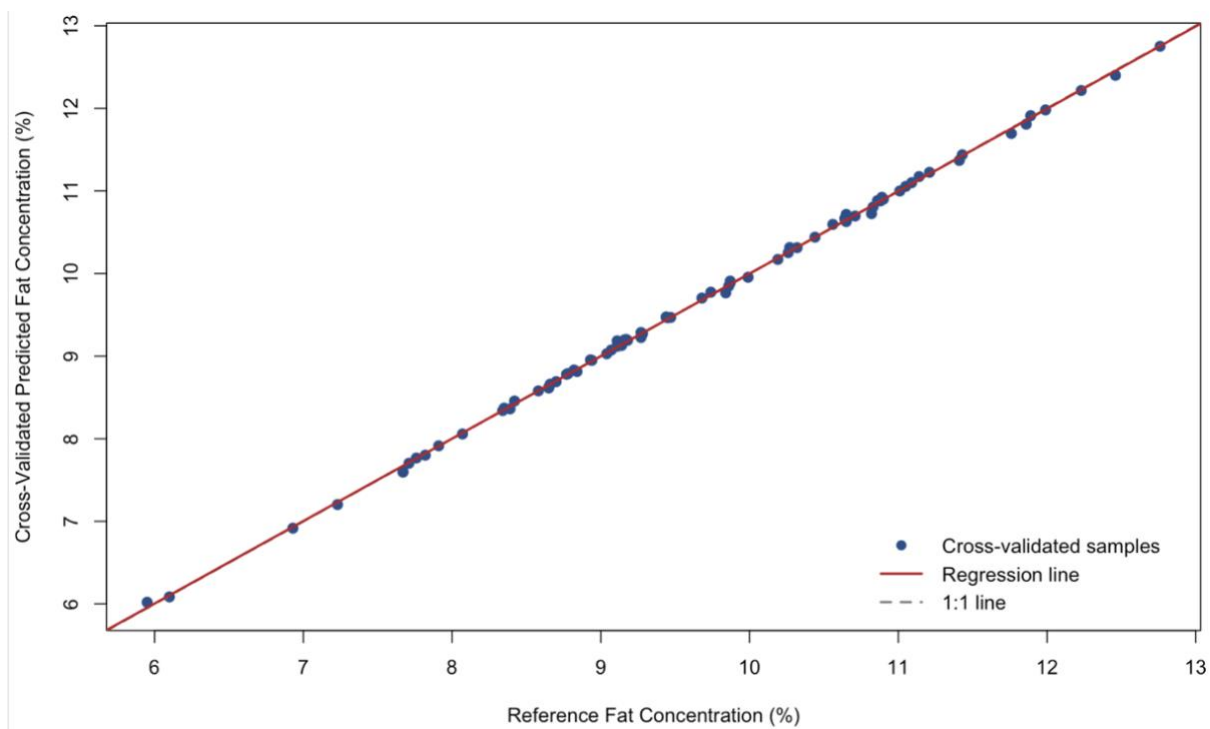


Figure 4.2. Observed versus predicted red deer milk fat concentration (%) for the validation dataset using the 13-component PLSR model.

4.2.2 Protein

A PLSR model was fitted using the calibration dataset ($n = 81$). Ten-fold cross-validation was used to determine the optimal number of latent components. The minimum $RMSE_{CV}$ (0.0222%) occurred at 19 latent components; however, application of the one-standard-error rule selected a simpler model containing 16 latent components ($RMSE_{CV} = 0.0229\%$). The 16-component model was retained for subsequent analyses. The calibration model achieved an R^2_C of 0.9994 with an $RMSE_C$ of 0.0122%, while 10-fold cross-validation produced an R^2_{CV} of 0.9979 with an $RMSE_{CV}$ of 0.0229% (Table 4.2).

The hold-out validation dataset ($n = 121$) produced an R^2_V of 0.9975, an $RMSE_P$ of 0.0282%, and an RPD of 19.93 (Table 4.2). Figures 4.3 and 4.4 show the observed and predicted values for the cross-validation and hold-out validation datasets.

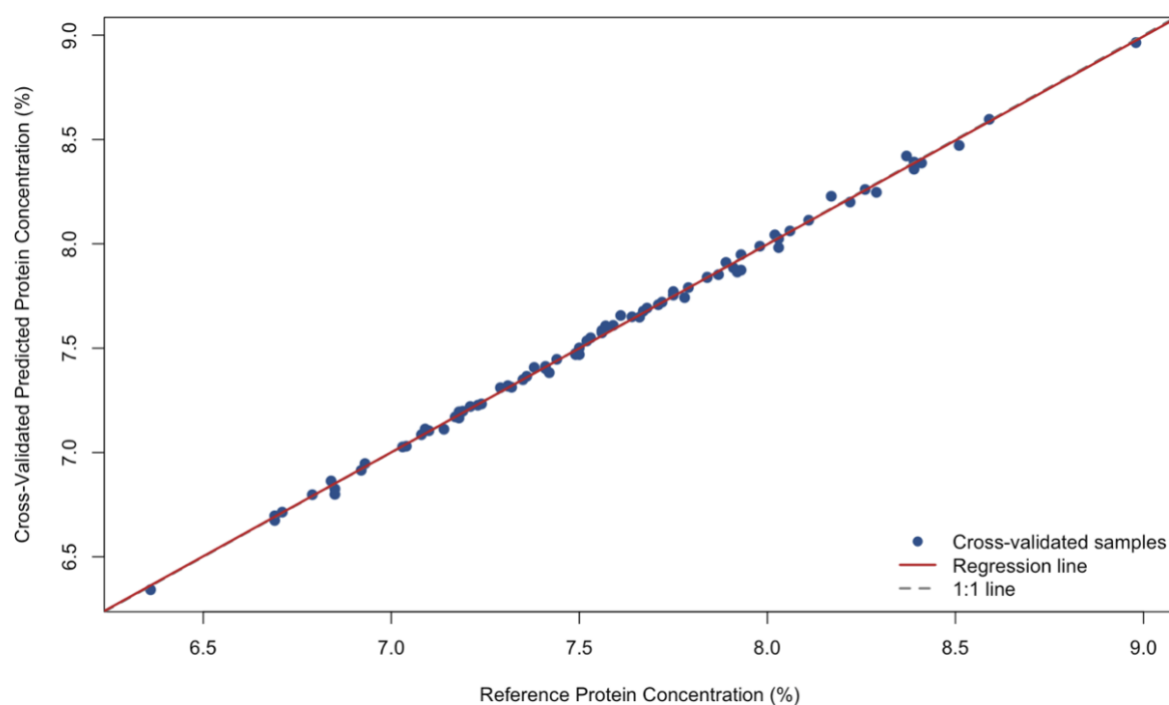


Figure 4.3. Observed versus cross-validated predicted red deer milk protein concentration (%) for the 16-component PLSR model.

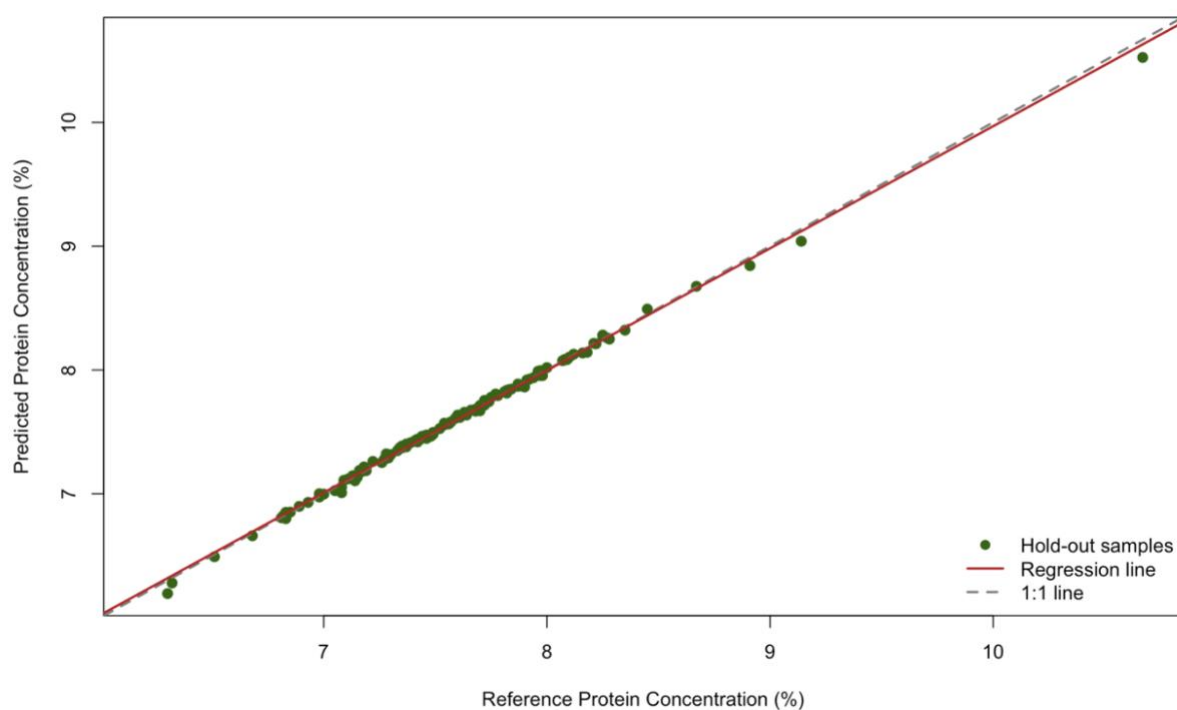


Figure 4.4. Observed versus predicted red deer milk protein concentration (%) for the validation dataset using the 16-component PLSR model.

4.2.3 Lactose

A PLSR model was fitted using the calibration dataset ($n = 81$). Ten-fold cross-validation was used to determine the optimal number of latent components. The minimum RMSE_{CV} (0.0224%) occurred at 21 latent components, and application of the one-standard-error rule also selected a model containing 21 latent components. The 21-component model was retained for subsequent analyses. The calibration model achieved an R^2_{C} of 0.9993 with an RMSE_{C} of 0.0094%, while 10-fold cross-validation produced an R^2_{CV} of 0.9962 with an RMSE_{CV} of 0.0224% (Table 4.2).

The hold-out validation dataset ($n = 121$) produced an R^2_{V} of 0.9953, an RMSE_{P} of 0.0281%, and an RPD of 14.66 (Table 4.2). Figures 4.5 and 4.6 show the observed and predicted values for the cross-validation and hold-out validation datasets.

Overall, the PLSR models closely reproduced the predictions generated by the existing FOSS calibration, with fat exhibiting the highest predictive performance, followed by protein and lactose.

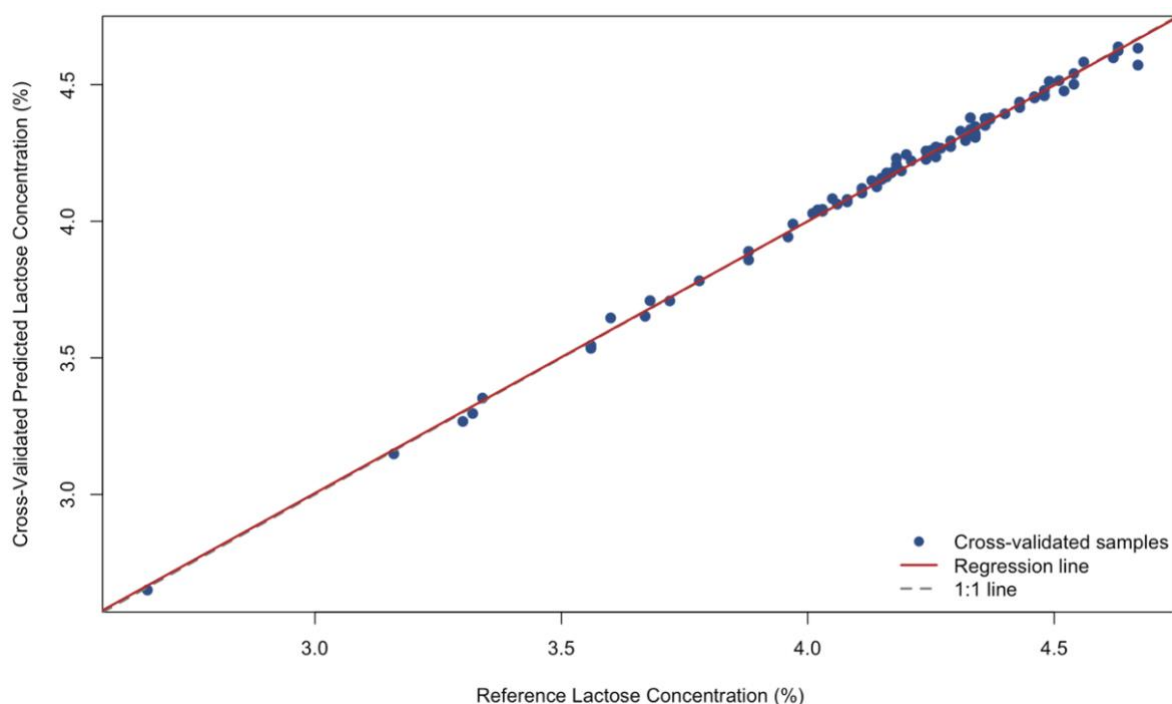


Figure 4.5. Observed versus cross-validated predicted lactose concentration (%) for the 21-component PLSR model.

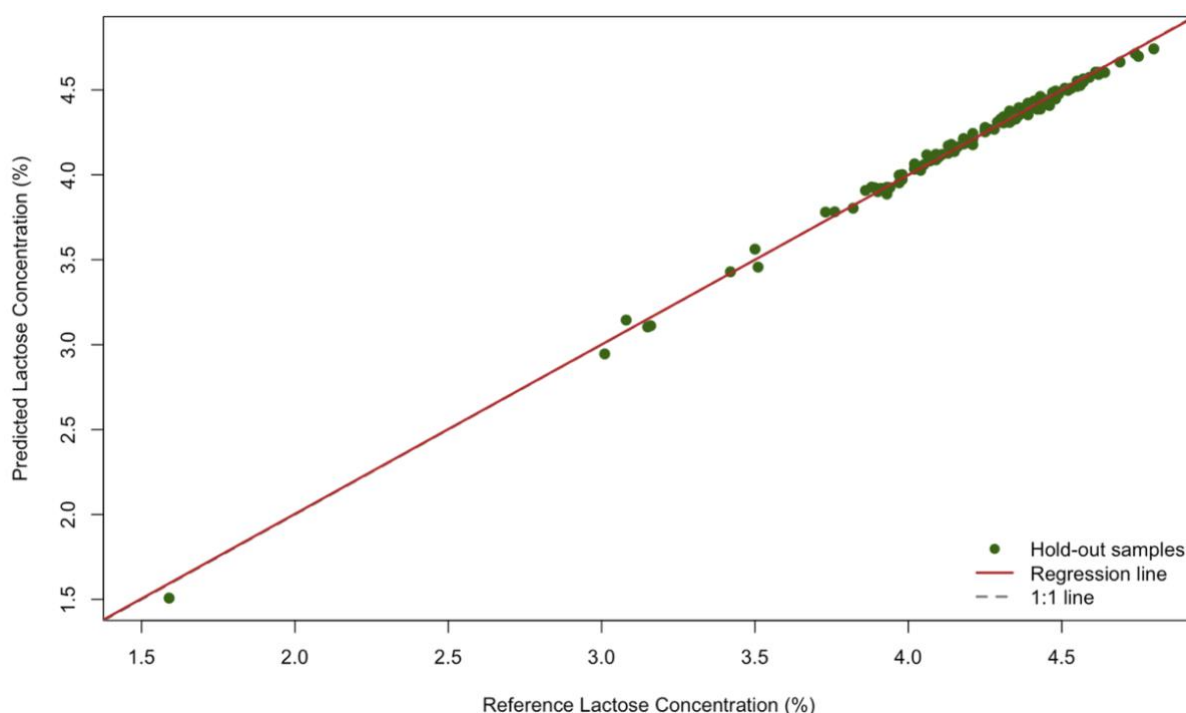


Figure 4.6. Observed versus predicted red deer milk lactose concentration (%) for the validation dataset using the 21-component PLSR model.

4.3 Identification of influential MIR spectral pins using PLSR regression coefficients

4.3.1 Fat

Figure 4.7 shows the absolute PLSR regression coefficients for the 13-component fat model (Appendix 2). The largest coefficients were concentrated around pins 380–390, 448–461, and 734–769, with the highest absolute coefficients occurring at pins 457, 456, 458, 448, 759, and 455. Exploratory backward stepwise regression of the 30 most influential spectral pins retained 22 pins, primarily within the regions encompassing pins 380–388, 448–460, and 735–761 (Appendix 1).

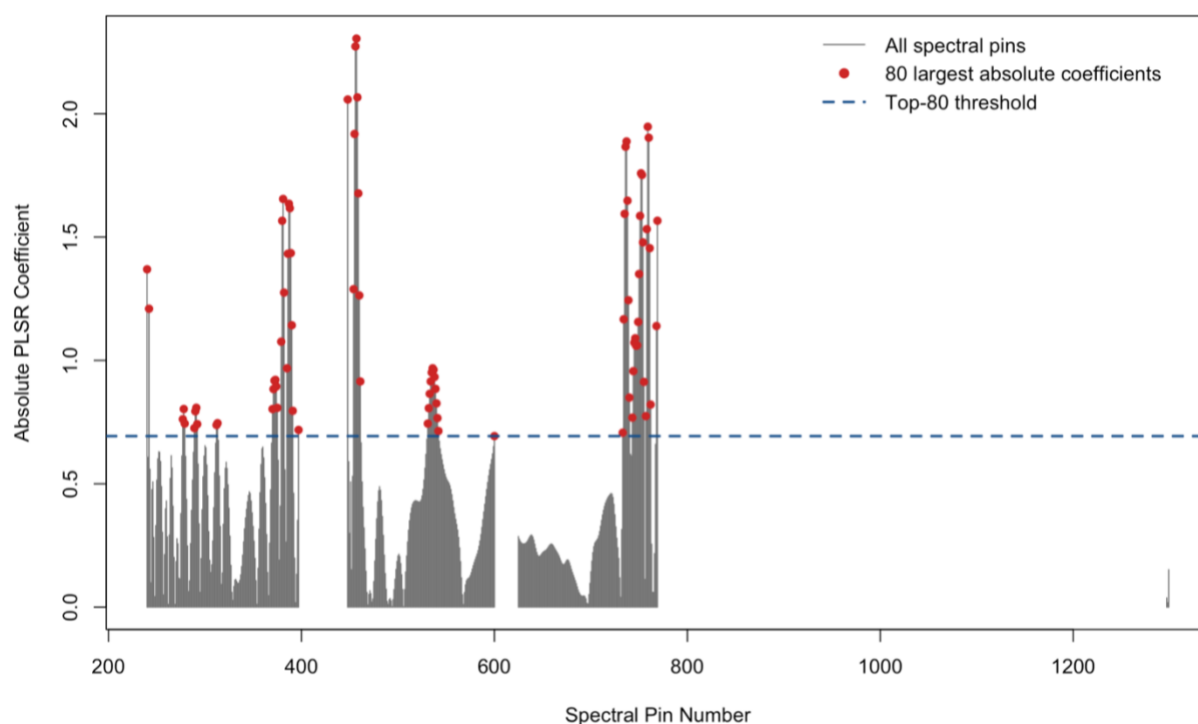


Figure 4.7. Manhattan plot of the absolute PLSR regression coefficients for the 13-component red deer milk fat model, highlighting the 80 most influential spectral pins.

4.3.2. Protein

Figure 4.8 shows the absolute PLSR regression coefficients for the 16-component protein model (Appendix 3). The largest coefficients were concentrated around pins 376–397, 448–457, and 734–769, with the highest absolute coefficients occurring at pins 397, 396, 395, 381, 380, and 379. Exploratory backward stepwise regression of the 30 most influential spectral pins retained 24 pins, primarily within the regions encompassing pins 359–383, 448–451, and 734–741 (Appendix 1).

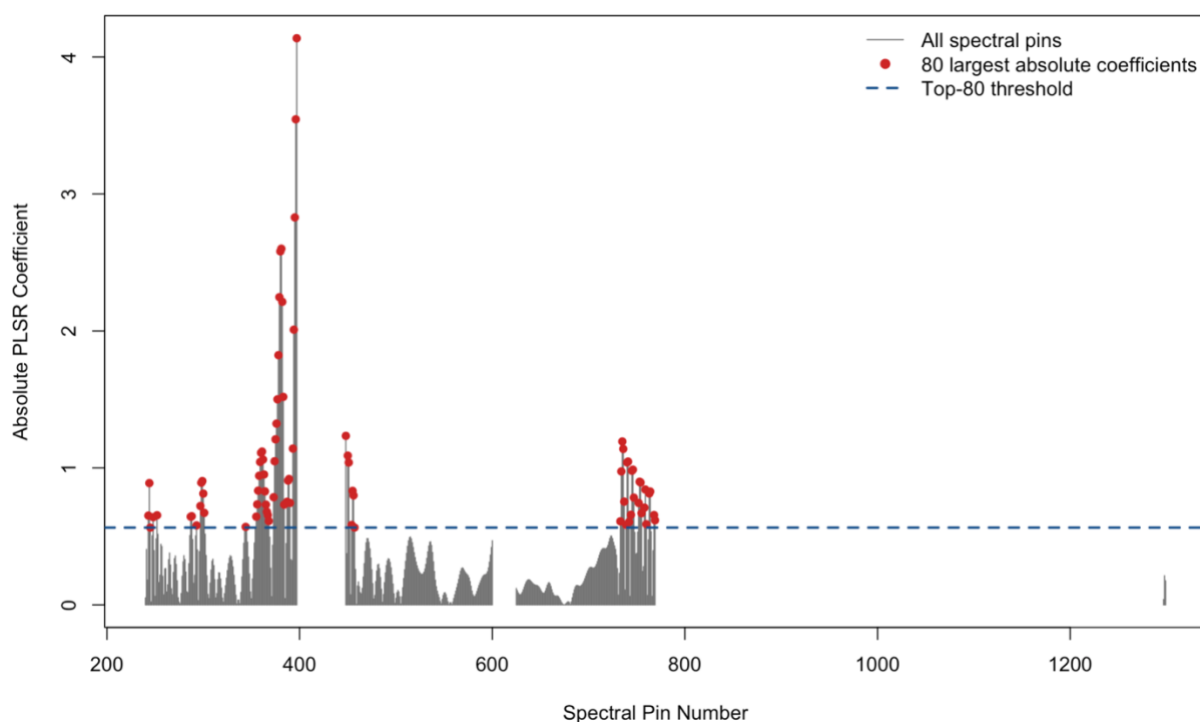


Figure 4.8. Manhattan plot of the absolute PLSR regression coefficients for the 16-component protein model, highlighting the 80 most influential spectral pins.

4.3.3 Lactose

Figure 4.9 shows the absolute PLSR regression coefficients for the 21-component lactose model (Appendix 4). The largest coefficients were concentrated around pins 274–303, 349–390, 448–452, and 739–765, with the highest absolute coefficients occurring at pins 448, 763, 764, 388, 762, and 387. Exploratory backward stepwise regression of the 30 most influential spectral pins retained 23 pins, primarily within the regions encompassing pins 274–302, 351–389, 451–452, 740–741, and 763 (Appendix 1).

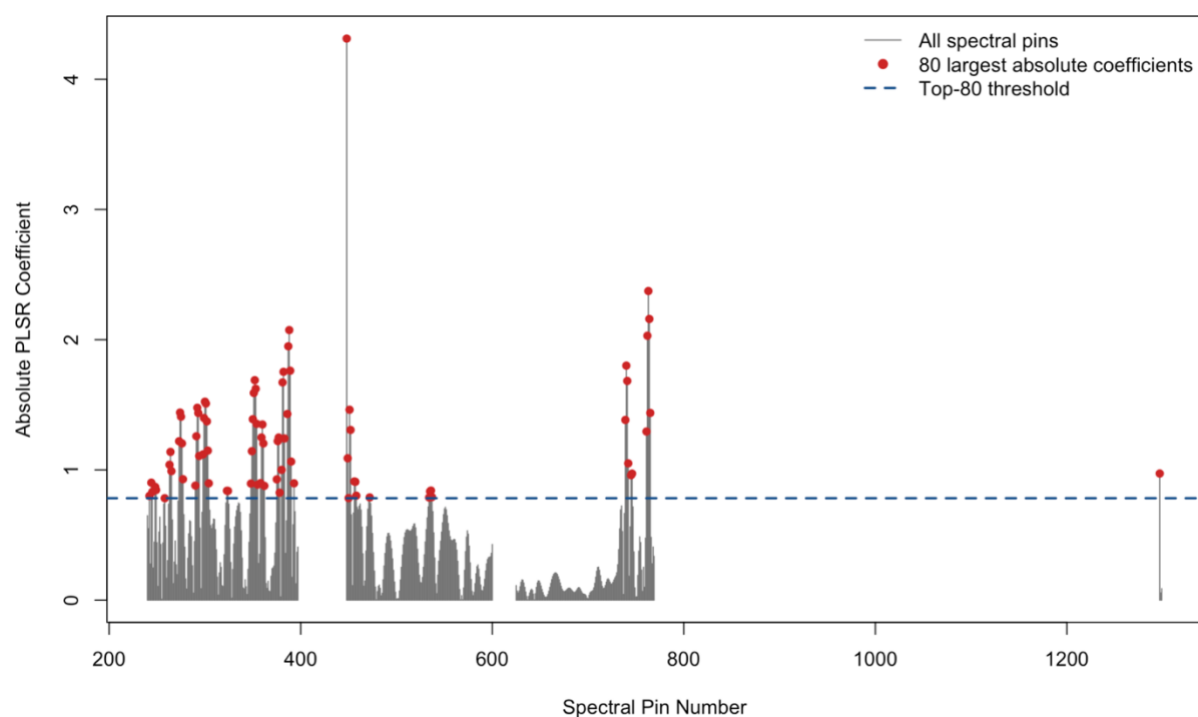


Figure 4.9. Manhattan plot of the absolute PLSR regression coefficients for the 21-component lactose model, highlighting the 80 most influential spectral pins.

Chapter 5: Discussion and conclusions

5.1 Model accuracy and predictive performance

The PLSR models developed in this study reproduced the fat, protein and lactose estimates generated by the existing FOSS calibration with a high degree of agreement. Across all three milk components, calibration, 10-fold cross-validation and hold-out validation produced coefficients of determination exceeding 0.995, together with low RMSE values and high hold-out validation RPD values. Fat exhibited the closest agreement with the FOSS-derived estimates (validation $R^2 = 0.9994$; $RMSEP = 0.0405\%$; $RPD = 40.43$), followed by protein (validation $R^2 = 0.9975$; $RMSEP = 0.0282\%$; $RPD = 19.93$) and lactose (validation $R^2 = 0.9953$; $RMSEP = 0.0281\%$; $RPD = 14.66$). These findings demonstrate that the spectral information contained within red deer milk is strongly associated with the outputs of the existing FOSS calibration.

The hold-out validation RPD values were exceptionally high, reaching 40.43 for fat, 19.93 for protein and 14.66 for lactose. RPD is commonly reported alongside R^2 and $RMSEP$ when evaluating the predictive performance of MIR calibration models, with larger values indicating lower prediction error relative to the natural variation within the reference dataset (Manuelian et al., 2019; Franzoi et al., 2023; Goi et al., 2025). The exceptionally high RPD values observed in this study are expected given the combination of a relatively large standard deviation within the external validation dataset and the very low $RMSEP$ values. Because the PLSR models were developed to reproduce the existing FOSS calibration outputs rather than predict independent wet chemistry reference values, prediction errors were minimal, resulting in high RPD values. Consequently, the combination of high coefficients of determination, low $RMSEP$ values and exceptionally high RPD values indicates that the retained spectral information closely reproduced the variation present within the FOSS-derived estimates. However, because the reference values used for model development were themselves generated by the existing FOSS calibration rather than independent wet chemistry analyses, these performance statistics should be interpreted as evidence of successful reproduction of the existing calibration rather than independent validation of true milk composition.

The close correspondence between the influential spectral regions identified in this study and the established absorption bands for fat, protein and lactose further suggests that the models captured chemically meaningful information rather than random statistical relationships, consistent with previous MIR studies in cattle, sheep and goats (Careda et al., 2016;

Dagnachew et al., 2013; Ferrand et al., 2011). Although the present models do not independently validate the accuracy of the existing FOSS calibration, they demonstrate that the identified MIR wavelengths contain sufficient spectral information to reliably reproduce its outputs. These findings provide a strong foundation for the future development and independent validation of red deer-specific MIR calibration equations using reference wet chemistry.

5.2 Identified top influential wavelengths

5.2.1 Fat

The prediction of milk fat using MIR spectroscopy is fundamentally based on the absorption of infrared radiation by specific covalent bonds within triacylglycerol molecules, which comprise the majority of milk lipids. These absorptions arise from characteristic molecular vibrations associated with ester carbonyl (C=O), aliphatic carbon–hydrogen (C–H), and carbon–carbon (C–C) bonds, and are commonly grouped into the Fat A ($\sim 1700\text{ cm}^{-1}$), Fat B ($\sim 2800\text{ cm}^{-1}$), and Fat C ($\sim 1470\text{ cm}^{-1}$) spectral regions (Kaylegian et al. 2009; Mohamed et al. 2021).

In the present study, the largest absolute PLSR regression coefficients were concentrated around pins 380–390, 448–461 and 734–769, while the subsequent backward stepwise regression retained 22 influential pins primarily within the regions of pins 380–388, 448–460 and 735–761. Following conversion to wavenumber, these retained pin regions correspond closely to the established Fat B ($\sim 2800\text{--}3000\text{ cm}^{-1}$), Fat A ($\sim 1700\text{--}1800\text{ cm}^{-1}$) and Fat C ($\sim 1400\text{--}1500\text{ cm}^{-1}$) absorption bands. Previous studies using FTIR and MilkoScan instruments have shown that the Fat B region is dominated by C–H stretching vibrations of CH_2 and CH_3 groups within fatty acid chains, making it one of the most informative regions for estimating milk fat concentration. Likewise, the Fat A region corresponds to the ester carbonyl (C=O) stretching vibration that is characteristic of triglycerides, while the Fat C region contains CH_2 and CH_3 bending vibrations that provide additional information relating to fatty acid structure (Kaylegian et al. 2009; Mohamed et al. 2021).

The close agreement between the influential pin regions identified in this study and these well-established MIR fat absorption bands indicates that the PLSR coefficients and subsequent stepwise selection identified chemically meaningful spectral regions rather than random

statistical associations. This finding demonstrates that the spectral regions contributing most strongly to the existing FOSS calibration correspond closely with established lipid absorption bands, providing evidence that the calibration is based on chemically meaningful spectral information rather than random statistical relationships. Furthermore, the retention of multiple spectral regions, rather than a single absorption band, is consistent with previous work showing that milk fat prediction benefits from information distributed across several complementary lipid-associated vibrations, reflecting variation in fatty acid composition and improving the robustness of MIR-based fat estimation (Kaylegian et al. 2009).

5.2.2 Protein

The prediction of milk protein using MIR spectroscopy is primarily governed by the vibrational behaviour of the polypeptide backbone, particularly the amide bonds present within proteins. Protein absorption in the MIR region is dominated by the amide I (1600–1700 cm^{-1}), amide II (~1550–1570 cm^{-1}), and amide III (~1200–1300 cm^{-1}) bands, which arise primarily from C=O stretching, N–H bending, and C–N stretching vibrations of peptide bonds (Etzion et al. 2004; López-Lorente and Mizaikoff 2016; Mohamed et al. 2021). MIR spectroscopy is therefore highly sensitive to protein concentration because these vibrational modes are characteristic of the protein backbone rather than individual amino acids.

In the present study, the largest absolute PLSR regression coefficients were concentrated around pins 376–397, 448–457 and 734–769, while backward stepwise regression retained 24 influential pins primarily within the regions of pins 359–383, 448–451 and 734–741. Following conversion to wavenumber, these regions correspond closely to the established amide II (~1550–1570 cm^{-1}), amide I (~1600–1700 cm^{-1}), and amide III (~1200–1300 cm^{-1}) absorption bands reported for proteins. Previous studies have consistently shown that the amide I band is the dominant spectral feature for protein analysis because it is primarily associated with C=O stretching vibrations of peptide bonds, whereas the amide II band arises from N–H bending coupled with C–N stretching vibrations. The amide III region provides complementary information through additional C–N stretching and N–H bending vibrations and contributes to characterising protein conformation (Etzion et al. 2004; López-Lorente and Mizaikoff 2016; Mohamed et al. 2021).

The close agreement between the influential spectral regions identified in this study and these well-established protein absorption bands indicates that the PLSR coefficients and subsequent

stepwise selection identified chemically meaningful wavelengths rather than random statistical associations. Although the influential wavelengths were distributed across multiple amide regions, this is consistent with previous studies demonstrating that MIR-based protein prediction relies on complementary information distributed across several protein-associated absorption bands rather than a single spectral feature (Etzion et al. 2004; López-Lorente and Mizaikoff 2016). Collectively, these findings support one of the primary objectives of this study by identifying MIR spectral regions strongly associated with the FOSS-derived protein estimates in red deer milk.

5.2.3 Lactose

The prediction of milk lactose using MIR spectroscopy is governed by the vibrational behaviour of carbohydrate functional groups, primarily hydroxyl (–OH) and ether (C–O–C) bonds within the lactose molecule. Lactose absorption is concentrated within the carbohydrate fingerprint region (approximately 1000–1250 cm^{-1}), where overlapping C–O, C–C and C–H stretching vibrations generate characteristic absorption features (Kaylegian et al. 2009; Mohamed et al. 2021; Pinto et al. 2021).

In the present study, the largest absolute PLSR regression coefficients were concentrated around pins 274–303, 349–390, 448–452 and 739–765, while backward stepwise regression retained 23 influential pins primarily within the regions of pins 274–302, 351–389, 451–452, 740–741 and 763. Following conversion to wavenumber, the retained pins around 740–763 correspond closely to the established carbohydrate fingerprint region ($\sim 1000\text{--}1250\text{ cm}^{-1}$), whereas the regions around pins 351–389 and 451–452 correspond to higher wavenumber regions that include additional absorptions associated with carbohydrate functional groups and neighbouring spectral features. Previous studies have identified the strongest lactose absorption near 1048 cm^{-1} within the broader fingerprint region, where C–OH, C–O and C–C stretching vibrations dominate the MIR spectrum of lactose (Kaylegian et al. 2009; Mohamed et al. 2021; Pinto et al. 2021).

The agreement between the influential spectral regions identified in this study and these established carbohydrate absorption bands indicates that the PLSR coefficients and subsequent stepwise selection identified chemically meaningful wavelengths rather than random statistical associations. The retention of influential pins across several spectral regions, rather than solely within the principal carbohydrate fingerprint region, suggests that lactose prediction is

informed by complementary spectral information distributed across multiple wavelengths. This is consistent with previous studies demonstrating that MIR-based lactose prediction relies on multivariate relationships across the spectrum rather than a single dominant absorption peak (Pinto et al. 2021; Mohamed et al. 2021). Collectively, these findings support one of the primary objectives of this study by identifying MIR spectral regions strongly associated with the FOSS-derived lactose estimates in red deer milk.

5.3 Spectral overlap

The influential spectral pins identified for fat, protein and lactose showed substantial overlap across several regions of the MIR spectrum. In particular, all three components retained influential wavelengths around pins 351–389, 448–452 and 734–763, although each component also exhibited unique spectral regions. This overlap is expected because the characteristic absorption bands associated with lipids, proteins and carbohydrates occur within adjacent or partially overlapping regions of the MIR spectrum rather than as isolated peaks (Andrade et al. 2019; Pinto et al. 2021; Soyeurt et al. 2010). Furthermore, the concentrations of the major milk constituents often co-vary biologically across animals and throughout lactation, meaning that absorbance measured at a single wavelength may contain information from more than one milk component (De Marchi et al. 2014; Soyeurt et al. 2010).

The retention of influential pins within the same spectral regions across all three models indicates that multiple milk constituents contribute to the measured spectral signal, despite arising from different molecular vibrations. Consequently, individual wavelengths are generally insufficient for accurately characterising specific milk components. Instead, predictive information is distributed across many correlated wavelengths, reinforcing the need for multivariate chemometric approaches such as PLSR. By modelling covariance across the entire spectrum rather than relying on isolated wavelengths, PLSR is better able to utilise overlapping spectral information and capture the complex relationships between MIR spectra and the MIR-derived estimates of milk composition (De Marchi et al. 2014; Andrade et al. 2019; Soyeurt et al. 2010).

Although substantial overlap was observed among the influential spectral regions for fat, protein and lactose, each model also retained unique spectral pins that aligned with the characteristic absorption bands discussed in Sections 5.2.1–5.2.3. This demonstrates that, despite considerable spectral overlap, chemically meaningful MIR regions associated with

each milk component could still be identified. These findings therefore fulfil one of the primary objectives of this study by identifying the spectral regions most strongly associated with the FOSS-derived estimates of fat, protein and lactose in red deer milk, providing a foundation for future refinement and validation of deer-specific MIR calibration equations.

5.4 Coefficients

The regression coefficient profiles differed between milk components (Appendix 2). Fat exhibited a relatively simple profile characterised by a small number of dominant peaks concentrated within well-defined spectral regions. Protein displayed a greater number of influential coefficients distributed across the main protein-associated absorption regions, whereas lactose was characterised by one highly dominant coefficient together with numerous secondary coefficients distributed across several regions of the MIR spectrum. These differences are consistent with the underlying chemistry of the major milk constituents, where lipid absorption is dominated by a relatively small number of strong absorption bands, whereas proteins and carbohydrates exhibit broader and more overlapping absorption features (De Marchi et al. 2014; Mohamed et al. 2021).

The coefficient profiles were also reflected in the complexity of the PLSR models. Fat achieved optimal performance using 13 latent variables, whereas protein and lactose required 16 and 21 latent variables, respectively. Regression coefficients identify the wavelengths contributing most strongly to each model, whereas latent variables represent combinations of correlated spectral variables that collectively explain covariance between the MIR spectra and the reference values (Mevik and Wehrens, 2007). The larger number of latent variables required for protein and particularly lactose therefore suggests that additional spectral information was required to reproduce the FOSS-derived reference values, despite the influential wavelengths remaining concentrated within distinct spectral regions.

Together, the regression coefficient profiles and latent variable structure demonstrate that the PLSR models extracted information from chemically meaningful regions of the MIR spectrum while effectively accounting for the substantial collinearity characteristic of MIR spectral data. Rather than relying on individual wavelengths, each model utilised information distributed across multiple correlated spectral regions, further supporting the suitability of PLSR for modelling complex MIR spectra.

5.5 Implications of mid-infrared technology in the deer dairy industry

Although deer-specific MIR calibration equations have yet to be independently validated against reference wet chemistry, the present findings indicate that MIR spectroscopy has considerable potential for application within the deer dairy industry. MIR spectroscopy is a rapid, non-destructive and cost-effective analytical technique routinely used in the dairy industry to quantify major milk components, including fat, protein and lactose, through large-scale herd testing programmes (Dal Zotto et al., 2008; Gengler et al., 2016; Maurice-Van Eijndhoven et al., 2013). In the New Zealand dairy industry, routine MIR analysis underpins genetic evaluation by providing the phenotypic records used to estimate Breeding Worth (BW) and Production Worth (PW) (LIC; CRV). While also supporting nutritional management, milk quality monitoring and animal health surveillance through the identification of high- and low-performing cows and those affected by conditions such as mastitis (Baba et al., 2021; Gengler et al., 2016; Grelet et al., 2021). If independently validated deer-specific calibration equations become available, routine MIR analysis could provide similar benefits to those currently realised within the bovine dairy industry. Beyond routine milk composition analysis, the establishment of a standardised MIR recording system could enable the collection of large-scale phenotypic data required for genetic evaluation, facilitating the selection of animals with favourable milk composition traits and improving the consistency of milk supplied for processing (Dagnachew et al., 2013; Machefert et al., 2024). As the deer dairy industry expands, the integration of MIR spectroscopy into routine milk recording could therefore support both genetic improvement and product standardisation, strengthening the long-term productivity and competitiveness of the industry.

5.6 Study limitations

The principal limitation of this study is that the reference milk composition values used for model development were generated by the existing FOSS MIR calibration rather than independent wet chemistry analyses. Consequently, the PLSR models developed in this study were trained to reproduce the outputs of the existing calibration, meaning the exceptionally high agreement observed between predicted and reference values reflects successful reproduction of the FOSS-derived estimates rather than independent validation of true fat, protein and lactose concentrations. Although this approach reflects current industry practice and enabled evaluation of the spectral regions contributing to the existing calibration, it does not establish the analytical accuracy of the FOSS calibration for red deer milk. Independent

reference analyses, such as Gerber or Röse-Gottlieb methods for fat, Kjeldahl or Dumas methods for protein, and enzymatic or HPLC methods for lactose, are therefore required before the predictive performance of MIR spectroscopy for true red deer milk composition can be fully established (Crocker et al., 1955; Thompson et al., 2002; Trani et al., 2017).

A second limitation relates to the size and representativeness of the dataset. Although 296 milk samples were initially collected, only 202 samples were available for analysis because incomplete milking or insufficient milk yield prevented successful sample collection from all animals. Furthermore, all deer were managed and milked under the same environmental conditions despite originating from multiple herds. Meaning the standardisation has the potential to reduce variation arising from factors such as nutrition, management and climate, it also limits the range of biological variation represented within the dataset. Consequently, the developed models may not fully capture the diversity of milk composition likely to occur across different farms, production systems and environmental conditions.

The study was also limited to a single sampling period during lactation, meaning each hind contributed only one milk sample. Milk composition and MIR spectra are known to vary throughout lactation in response to physiological and nutritional changes. Consequently, the present models represent only the stage of lactation sampled and may not fully reflect spectral variation occurring during early or late lactation or across multiple production seasons. Expanding future datasets to include repeated sampling throughout lactation and across successive years would improve the robustness and general applicability of deer-specific MIR calibrations.

Although the PLSR models demonstrated strong agreement with the FOSS-derived reference values, the relatively small sample size also limited calibration development. Robust MIR calibrations typically benefit from large datasets encompassing substantial biological variation, including multiple herds, management systems, and production years. Given the relatively small size of the commercial deer dairy industry in New Zealand, assembling datasets of this scale remains challenging but will be essential for developing broadly applicable calibration equations.

Finally, interpretation of the influential wavelengths identified through backward stepwise regression should be undertaken with caution. To maintain statistical feasibility, the stepwise analysis was restricted to the most influential wavelengths identified by the PLSR models.

Although this approach improved interpretability by reducing the number of predictors, stepwise regression remains sensitive to multicollinearity and small changes in the calibration dataset. Consequently, the retained wavelengths should be regarded as exploratory indicators of informative spectral regions rather than definitive markers of individual milk components. The consistency between these retained regions and established MIR absorption bands for fat, protein and lactose nevertheless provides confidence that the identified wavelengths are chemically plausible.

5.7 Further research

The highest priority for future research is the development of a red deer-specific MIR calibration validated against independent reference chemistry. Because the present study reproduced milk composition estimates generated by the existing FOSS calibration, rather than independently measured fat, protein and lactose concentrations, the analytical accuracy of the calibration for red deer milk remains unknown. Future studies should therefore establish deer-specific calibration equations using recognised reference methods, including Gerber or Röse-Gottlieb analyses for fat, Kjeldahl or Dumas analyses for protein, and enzymatic or high-performance liquid chromatography (HPLC) methods for lactose. Validation against these independent reference methods is essential before the predictive accuracy of MIR spectroscopy for true red deer milk composition can be established (Crocker et al., 1955; Thompson et al., 2002; Trani et al., 2017).

Following the development of an independently validated calibration, future studies should focus on improving model robustness by expanding the calibration population. This should include larger numbers of animals sampled across multiple commercial deer dairies, management systems, lactation stages, and production years. Repeated sampling of individual hinds throughout lactation would allow the effects of physiological changes on MIR spectra to be quantified and would improve understanding of lactational variation in milk composition. Although the limited number of commercial deer dairies currently operating in New Zealand presents a practical challenge, continued growth of the industry should improve opportunities to develop larger and more representative calibration datasets.

Once robust calibration equations have been established for the major milk components, future research could investigate the application of MIR spectroscopy to additional deer milk traits. These may include minor milk constituents, such as individual fatty acids, protein fractions,

minerals and metabolites, before extending to more complex functional traits, including cheese-making properties and indicators of animal health. Similar applications have been successfully investigated in bovine, ovine and caprine milk, but reliable prediction of these more complex traits depends on first establishing accurate calibration equations for the major milk components.

5.8 Conclusion

The objectives of this thesis were to identify the MIR wavelengths most strongly associated with fat, protein and lactose estimates in New Zealand red deer milk, and to evaluate the performance of the existing FOSS MilkoScan machine. Both objectives were achieved.

The present study identified the MIR spectral regions most strongly associated with the existing FOSS MilkoScan calibration for fat, protein and lactose in New Zealand red deer milk. The influential wavelengths identified for each milk component corresponded closely with established absorption regions associated with lipid, protein and carbohydrate functional groups, demonstrating that the spectral information captured by the existing calibration is chemically meaningful and consistent with the known molecular composition of milk. Although substantial overlap was observed among the informative spectral regions for fat, protein and lactose, each component also exhibited distinct wavelength regions, supporting the use of multivariate chemometric approaches such as PLSR to resolve the complex spectral information contained within MIR data.

The PLSR models reproduced the outputs of the existing FOSS calibration with exceptionally high agreement across calibration, cross-validation and hold-out validation datasets. However, because the reference values were themselves generated by the FOSS calibration rather than independent wet chemistry analyses, these findings should be interpreted as successful reproduction of the existing calibration rather than independent validation of true milk composition. Consequently, the principal contribution of this research is not the development of a new prediction equation, but the identification of the MIR wavelength regions most strongly associated with the current FOSS calibration for red deer milk.

Overall, this study addresses an important knowledge gap by providing the first detailed evaluation of the MIR spectral regions associated with fat, protein and lactose estimates in New Zealand red deer milk. These findings provide a scientifically informed foundation for the

future development of deer-specific MIR calibration equations using independently measured wet chemistry reference data. As larger and more representative datasets become available, these wavelength regions may contribute to the development of robust calibration equations capable of supporting genetic improvement, routine milk composition analysis and the continued growth of the New Zealand deer dairy industry.

**Appendix 1: Fat, protein, lactose spectral pins retained
following exploratory backward stepwise
regression**

The table lists the spectral pins retained following backward stepwise regression applied to the 30 highest-ranked PLSR regression coefficients for fat (22 pins retained), protein (24 pins retained), and lactose (23 pins retained).

Fat	Protein	Lactose
pin457	pin396	pin763
pin456	pin381	pin388
pin458	pin380	pin387
pin448	pin379	pin740
pin455	pin382	pin389
pin760	pin378	pin382
pin737	pin383	pin352
pin752	pin377	pin741
pin753	pin376	pin381
pin459	pin448	pin353
pin381	pin375	pin351
pin738	pin735	pin300
pin387	pin736	pin301
pin388	pin361	pin292
pin735	pin360	pin451
pin751	pin450	pin274
pin380	pin362	pin293
pin758	pin374	pin386
pin761	pin741	pin275
pin386	pin359	pin299
pin382	pin740	pin302
pin460	pin451	pin354
	pin734	pin452
	pin363	

**Appendix 2: Top 80 influential spectral pins identified by
PLSR coefficients for predicting milk fat
concentration**

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk fat concentration

Pin	Coefficient	Absolute coefficient
240	-1.36928	1.369277172
242	1.209506	1.209506384
277	0.761767	0.761767165
278	0.803048	0.803048439
279	0.744018	0.744017966
289	-0.72561	0.72560864
290	-0.79376	0.793758089
291	-0.80811	0.808112762
292	-0.74173	0.741726508
312	0.738126	0.738126298
313	0.745661	0.745660621
370	0.802967	0.802966728
371	0.883915	0.883915142
372	0.918116	0.91811623
373	0.921152	0.921152458
374	0.894538	0.894538398
375	0.807729	0.807729377
379	-1.07605	1.076051854
380	-1.56613	1.566131909
381	-1.65446	1.654464918
382	-1.275	1.274998344
385	0.968076	0.968075735
386	1.43233	1.43232966
387	1.635149	1.635148996
388	1.617	1.617000388
389	1.435009	1.435008694
390	1.142505	1.142504506
391	0.795764	0.795763756
397	-0.71821	0.7182145
448	-2.05772	2.05772026

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk fat concentration continued

Pin	Coefficient	Absolute coefficient
454	-1.28891	1.288906224
455	-1.9179	1.917895673
456	-2.27289	2.272885687
457	-2.30459	2.304585262
458	-2.06636	2.066360451
459	-1.67708	1.677081114
460	-1.26348	1.26348427
461	-0.91471	0.91471499
531	0.743974	0.743973971
532	0.806504	0.806504485
533	0.864848	0.864847999
534	0.914766	0.914765666
535	0.951012	0.951011805
536	0.968134	0.968134215
537	0.962233	0.962232529
538	0.93298	0.932980418
539	0.88475	0.884749672
540	0.825909	0.825909151
541	0.766296	0.766295816
542	0.713962	0.713961607
600	-0.69325	0.69325233
733	-0.70699	0.706988122
734	-1.16677	1.166767189
735	-1.59416	1.594157794
736	-1.86614	1.866141908
737	-1.88768	1.887676732
738	-1.6475	1.647496004
739	-1.24405	1.244052892
740	-0.84938	0.849381612
743	-0.76814	0.768144415

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk fat concentration continued

Pin	Coefficient	Absolute coefficient
744	-0.95646	0.956464061
745	-1.07216	1.072162417
746	-1.08996	1.08995859
747	-1.05952	1.059521213
748	-1.06123	1.061228996
749	-1.15602	1.156016475
750	-1.35023	1.350228943
751	-1.58599	1.585989397
752	-1.75851	1.758508038
753	-1.75179	1.751794478
754	-1.47851	1.478510994
755	-0.9129	0.912898169
757	0.774234	0.774233539
758	1.532157	1.532156732
759	1.947224	1.947224375
760	1.902647	1.902646874
761	1.454959	1.454958867
762	0.820998	0.820998346
768	1.139306	1.139306263
769	1.566263	1.566262995

**Appendix 3: Top 80 influential spectral pins identified by
PLSR coefficients for predicting milk
protein concentration**

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk protein concentration

Pin	Coefficient	Absolute coefficient
243	-0.6516	0.651598
244	-0.88897	0.888968
245	-0.56406	0.564062
248	0.640099	0.640099
252	-0.65416	0.654157
287	0.643885	0.643885
288	0.647233	0.647233
293	-0.57996	0.579965
297	0.721988	0.721988
298	0.891575	0.891575
299	0.904004	0.904004
300	0.81237	0.81237
301	0.671953	0.671953
344	-0.56929	0.56929
355	0.644109	0.644109
356	0.733656	0.733656
357	0.833785	0.833785
358	0.942499	0.942499
359	1.043227	1.043227
360	1.110187	1.110187
361	1.119186	1.119186
362	1.061128	1.061128
363	0.951834	0.951834
364	0.828738	0.828738
365	0.731909	0.731909
366	0.679877	0.679877
367	0.655715	0.655715
368	0.612181	0.612181
373	-0.78591	0.78591
374	-1.04849	1.048494

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk protein concentration continued

Pin	Coefficient	Absolute coefficient
375	-1.20869	1.208688
376	-1.32338	1.32338
377	-1.50084	1.50084
378	-1.8231	1.823098
379	-2.24686	2.246864
380	-2.58087	2.580871
381	-2.60043	2.600433
382	-2.2127	2.212702
383	-1.51874	1.518739
384	-0.73207	0.73207
387	0.752653	0.752653
388	0.907452	0.907452
389	0.919011	0.919011
390	0.74436	0.74436
393	-1.14117	1.141167
394	-2.00912	2.009116
395	-2.82886	2.828863
396	-3.54526	3.545265
397	-4.13698	4.13698
448	-1.23381	1.233814
450	1.08975	1.08975
451	1.039173	1.039173
454	-0.583	0.582997
455	-0.83229	0.832294
456	-0.79927	0.799269
457	-0.56448	0.564476
733	-0.6099	0.609904
734	-0.97431	0.974312
735	-1.19284	1.192836
736	-1.13871	1.138715

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk protein concentration continued

Pin	Coefficient	Absolute coefficient
737	-0.75382	0.753824
739	0.58487	0.58487
740	1.040165	1.040165
741	1.046067	1.046067
742	0.604868	0.604868
744	-0.65772	0.657722
745	-0.97912	0.979123
746	-0.98724	0.98724
747	-0.78291	0.782914
752	-0.74393	0.743933
753	-0.89903	0.89903
754	-0.89308	0.893085
755	-0.66952	0.669516
758	0.709747	0.709747
759	0.841804	0.841804
760	0.589061	0.589061
763	-0.81321	0.81321
764	-0.82756	0.827559
768	0.655646	0.655646
769	0.617744	0.617744

**Appendix 4: Top 80 influential spectral pins identified by
PLSR coefficients for predicting milk lactose
concentration**

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk lactose concentration

Pin	Coefficient	Absolute coefficient
448	-4.31256	4.312559799
763	-2.37385	2.373847203
764	-2.15929	2.159293842
388	2.074067	2.074066991
762	-2.03048	2.030484499
387	1.949343	1.949342824
740	1.800835	1.800835339
389	1.761975	1.761975477
382	-1.75313	1.753131935
352	-1.68911	1.68910823
741	1.683364	1.683364319
381	-1.67207	1.672070885
353	-1.62312	1.623119548
351	-1.59084	1.590838303
300	1.523899	1.523898616
301	1.509921	1.509920505
292	-1.47733	1.477333915
451	1.462072	1.46207182
274	-1.44046	1.440455748
765	-1.43763	1.437627546
293	-1.43728	1.437280953
386	1.429575	1.429574869
275	-1.40891	1.408906124
299	1.398606	1.398605725
350	-1.38936	1.38936296
739	1.383534	1.38353399
302	1.372842	1.372841541
354	-1.35377	1.353771356
360	1.348857	1.348856834

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk
lactose concentration continued

Pin	Coefficient	Absolute coefficient
360	1.348857	1.348856834
452	1.307041	1.307040552
761	-1.29487	1.294874954
291	-1.25861	1.258612578
359	1.249099	1.249099112
377	1.248804	1.248803671
383	-1.24183	1.241827667
376	1.221113	1.221112647
273	-1.22017	1.220171723
276	-1.20343	1.20342835
361	1.203147	1.203147464
303	1.147871	1.147871308
349	-1.14369	1.143687586
264	-1.13824	1.138241042
298	1.119167	1.119166612
294	-1.10728	1.107275546
449	-1.08935	1.089353048
390	1.063816	1.063815698
742	1.050335	1.050335125
263	-1.0395	1.039495449
380	-1.00041	1.000411141
265	-0.99069	0.990693402
1297	0.97244	0.972440405
746	-0.97179	0.971788797
745	-0.95954	0.959537068
277	-0.92866	0.928656616
375	0.927537	0.927536687
456	-0.91133	0.911327932
457	-0.9086	0.908595268

Top 80 influential MIR spectral pins identified by PLSR coefficients for predicting milk
lactose concentration continued

Pin	Coefficient	Absolute coefficient
358	0.898405	0.898405488
304	0.897084	0.897084262
393	-0.897	0.896998616
348	-0.89506	0.895058685
355	-0.88674	0.886735762
290	-0.87954	0.879540864
362	0.87768	0.877679914
248	0.869874	0.869874055
249	0.846047	0.846046752
536	0.841349	0.841348706
323	0.839831	0.839830925
324	0.838743	0.838742573
535	0.835292	0.835292331
245	-0.83057	0.830573754
378	0.825042	0.825042079
458	-0.80187	0.801870066
242	0.800464	0.800464116
537	0.791441	0.791441217
472	0.788871	0.788870871
534	0.787849	0.787849366
450	0.783562	0.783562195
258	0.782607	0.782607173

References

- Andrade, J., Pereira, C.G., de Almeida Junior, J.C., Viana, C.C.R., de Oliveira Neves, L.N., da Silva, P.H.F., Bell, M.J.V. and dos Anjos, V.d.C. 2019. FTIR-ATR determination of protein content to evaluate whey protein concentrate adulteration. *LWT - Food Science and Technology* 99: 166–172.
- Baba, T., Pegolo, S., Mota, L.F., Peñagaricano, F., Bittante, G., Cecchinato, A. and Morota, G. 2021. Integrating genomic and infrared spectral data improves the prediction of milk protein composition in dairy cattle. *Genetics Selection Evolution* 53: 29.
- Benedet, A., Ho, P., Xiang, R., Bolormaa, S., De Marchi, M., Goddard, M. and Pryce, J. (2019). The use of mid-infrared spectra to map genes affecting milk composition. *Journal of Dairy Science* 102: 7189–7203.
- Benoudjit, N., Melgani, F. and Bouzgou, H. 2009. Multiple regression systems for spectrophotometric data analysis. *Chemometrics and Intelligent Laboratory Systems* 95:144–149.
- Berruga, M.I., de la Vara, J.Á., Licón, C.C., Garzón, A.I., García, A.J., Carmona, M., Chonco, L. and Molina, A. 2021. Physicochemical, microbiological and technological properties of red deer (*Cervus elaphus*) milk during lactation. *Animals* 1: 906.
- Bovolenta, S., Corazzin, M., Messina, M., Focardi, S. and Piasentier, E. 2013. Supplementary feeding of farmed fallow deer: effect on milk composition and fawn performance. *Italian Journal of Animal Science* 12(4): e98.
- Caredda, M., Addis, M., Ibba, I., Leardi, R., Scintu, M.F., Piredda, G. and Sanna, G. 2016. Prediction of fatty acid content in sheep milk by mid-infrared spectrometry with a selection of wavelengths by genetic algorithms. *LWT* 65: 503–510.
- Cellesi, M., Correddu, F., Manca, M.G., Serdino, J., Gaspa, G., Dimauro, C. and Macciotta, N.P.P. 2019. Prediction of milk coagulation properties and individual cheese yield in sheep using partial least squares regression. *Animals* 9: 663.
- Cramer III, R.D. 1993. Partial least squares (PLS): its strengths and limitations. *Perspectives in Drug Discovery and Design* 1: 269–278.

Crocker, W. P., Jenkins, D. I., Provan, A. L., Macdonald, F. J., Rowland, S. J., and White, J. C. D. 1955. A comparison of the Gerber and Röse Gottlieb methods for the determination of fat in milk. *Journal of Dairy Research* 22: 336-339.

CRV. n.d. Herd testing. Retrieved 24 April from: <https://crv4all.co.nz/nz/service/herd-testing>

Dagnachew, B., Meuwissen, T. and Ådnøy, T. 2013. Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. *Journal of Dairy Science* 96: 5933–5942.

Dal Zotto, R., De Marchi, M., Cecchinato, A., Penasa, M., Cassandro, M., Carnier, P., Gallo, L. and Bittante, G. 2008. Reproducibility and repeatability of measures of milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy. *Journal of Dairy Science* 91: 4103–4112.

De Marchi, M., Bonfatti, V., Cecchinato, A., Di Martino, G. and Carnier, P. 2009a. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. *Italian Journal of Animal Science* 8: 399–401.

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

De Marchi, M., Toffanin, V., Cassandro, M. and Penasa, M. 2013. Prediction of coagulating and noncoagulating milk samples using mid-infrared spectroscopy. *Journal of Dairy Science* 96: 4707–4715.

De Marchi, M., Toffanin, V., Cassandro, M. and Penasa, M. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. *Journal of Dairy Science* 97: 1171–1186.

DINZ. 2025. Thrive 2035 – New Zealand Deer Industry Strategy. Retrieved 3 August from: <https://www.deernz.org/assets/DINZ/Strategy-Workbook.pdf>

DINZ. n.d. Statistics. Retrieved 3 August from: <https://www.deernz.org/home/deer-industry-new-zealand/statistics/>

Etzion, Y., Linker, R., Cogan, U. and Shmulevich, I. 2004. Determination of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy. *Journal of Dairy Science* 87: 2779–2788.

Ferrand, M., Huquet, B., Barbey, S., Barillet, F., Faucon, F., Larroque, H., Leray, O., Trommenschlager, J. and Brochard, M. 2011. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: interest of applying a variable selection by genetic algorithms before a PLS regression. *Chemometrics and Intelligent Laboratory Systems* 106: 183–189.

Fennessy, P. 1995. Deer and deer farming – the New Zealand experience.

Figgins, G. and Holland, P. 2012. Red deer in New Zealand: game animal, economic resource or environmental pest? *New Zealand Geographer* 68: 36–48.

Fondahl, G. 1989. Reindeer dairying in the Soviet Union. *Polar Record* 25: 285–294.

Franzoi, M., Niero, G., Meoni, G., Tenori, L., Luchinat, C., Penasa, M., Cassandro, M. and De Marchi, M. 2023. Effectiveness of mid-infrared spectroscopy for the prediction of cow milk metabolites. *Journal of Dairy Science* 106: 5288–5297.

Gallego, L., Landete-Castillejos, T., Garcia, A. and Sánchez, P. 2006. Seasonal and lactational changes in mineral composition of milk from Iberian red deer (*Cervus elaphus hispanicus*). *Journal of Dairy Science* 89: 589–595.

Gathercole, J. and Reis, M. 2021. Deer milk: production and composition. *Encyclopedia*: 514–521.

Gendrin, C., Roggo, Y. and Collet, C. 2008. Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review. *Journal of Pharmaceutical and Biomedical Analysis* 48: 533–553.

Gengler, N., Soyeurt, H., Dehareng, F., Bastin, C., Colinet, F., Hammami, H., Vanrobays, M.-L., Laine, A., Vanderick, S. and Grelet, C. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. *Journal of Dairy Science* 99: 4071–4079.

- Gjøstein, H., Holand, Ø. and Weladji, R.B. 2004. Milk production and composition in reindeer (*Rangifer tarandus*): effect of lactational stage. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology* 137: 649–656.
- Goi, A., Magro, S., Lanni, L., Boselli, C. and Marchi, M.D. 2025. Mid-infrared spectroscopy for predicting goat milk coagulation properties. *Foods* 14: 2403.
- Grelet, C., Dardenne, P., Soyeurt, H., Fernandez, J., Vanlierde, A., Stevens, F., Gengler, N. and Dehareng, F. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: key factors affecting the quality of predictions. *Methods* 186: 97–111.
- Hall, R. 2023. New Zealand Deer Production and Trends. United States Department of Agriculture, Foreign Agricultural Service, Report No. NZ2022-0019.
- Hastie, T., Tibshirani, R., Friedman, T. 2009. The elements of statistical learning: data mining, inference, and prediction. Pages 219-257
- Ho, P., Luke, T. and Pryce, J. 2021. Validation of milk mid-infrared spectroscopy for predicting the metabolic status of lactating dairy cows in Australia. *Journal of Dairy Science* 104: 4467–4477.
- Holand, Ø., Gjøstein, H., Nieminen, M. and Haenlein, G.F. 2017. Reindeer milk. In: *Handbook of Milk of Non-Bovine Mammals*, pp. 686–718.
- Hsu, C.-P.S. 1997. Infrared spectroscopy. In: *Handbook of Instrumental Techniques for Analytical Chemistry*, p. 249.
- Kaylegian, K., Lynch, J., Fleming, J. and Barbano, D. 2009. Influence of fatty acid chain length and unsaturation on mid-infrared milk analysis. *Journal of Dairy Science* 92: 2485–2501.
- Kruger, M.C., Mazahery, H., Mugridge, O., Turner, S. and von Hurst, P. 2023. A comparative intervention trial of deer milk and an oral nutritional supplement efficacy for improving older adults' nutritional status, muscle mass and physical performance. *Clinical Nutrition ESPEN* 57: 346–357.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of International Joint Conference of Artificial Intelligence 2*: 1137 – 1145.

Landete-Castillejos, T., Garcia, A., Molina, P., Vergara, H., Garde, J. and Gallego, L. 2000. Milk production and composition in captive Iberian red deer (*Cervus elaphus hispanicus*): effect of birth date. *Journal of Animal Science* 78: 2771–2777.

Li, S., Saharawat, A., Ye, A., Dave, A. and Singh, H. 2023. Characteristics of red deer (*Cervus elaphus*) milk: lactational changes in composition and processing impacts on structural and gelation properties. *Foods* 12: 1517.

LIC. n.d. Herd testing overview. Retrieved 24 April from: <https://www.lic.co.nz/products-and-services/herd-testing/herd-testing-overview/>

López-Lorente, Á.I. and Mizaikoff, B. 2016. Mid-infrared spectroscopy for protein analysis: potential and challenges. *Analytical and Bioanalytical Chemistry* 408: 2875–2889.

Machefert, C., Robert-Granié, C., Astruc, J. and Larroque, H. 2024. Genetic parameters of milk mid-infrared spectra and their genetic relationships with milk production and feed efficiency traits in French Lacaune dairy sheep. *Journal of Dairy Science* 107: 11239–11253.

Malacarne, M., Franceschi, P., Formaggioni, P., Pisani, G.M., Petrera, F., Abeni, F., Soffiantini, C.S. and Summer, A. 2015. Minerals content and distribution in milk from red deer (*Cervus elaphus*), fallow deer (*Dama dama*) and roe deer (*Capreolus capreolus*). *Small Ruminant Research* 130: 208–215.

Manuelian, C.L., Penasa, M., Giangolini, G., Boselli, C., Currò, S. and De Marchi, M. 2019. Fourier-transform mid-infrared spectroscopy to predict coagulation and acidity traits of sheep bulk milk. *Journal of Dairy Science* 102: 1927–1932.

Marie-Etancelin, C. and Rupp, R. 2025. Genetic parameters and genome-wide association studies of milk infrared spectra from Lacaune dairy ewes. *Journal of Dairy Science* 108: 8648–8665.

Maurice-Van Eijndhoven, M., Soyeurt, H., Dehareng, F. and Calus, M. 2013. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. *Animal* 7: 348–354.

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

Mevik, B. H., and Wehrens, R. 2007. The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* 18: 1-23.

Mohamed, H., Nagy, P., Agbaba, J. and Kamal-Eldin, A. 2021. Use of near and mid infra-red spectroscopy for analysis of protein, fat, lactose and total solids in raw cow and camel milk. *Food Chemistry* 334: 127436.

Ní Fhuaráin, A.M., O'Donnell, C.P., Luo, J. and Gowen, A.A. 2024. A review on MIR, NIR, fluorescence and Raman spectroscopy combined with chemometric modeling to predict the functional properties of raw bovine milk. *ACS Food Science and Technology* 4: 2258–2271.

Nixon, C. 2004. Innovation and trade liberalisation: a case study of the New Zealand deer industry.

NZIPIM. 2020. The Journal 24: 16–19. Retrieved 1 August from: https://www.irpnz.co.nz/Folder?Action=View%20File&Folder_id=120&File=Journal%20September%202020.pdf

Pāmu. 2019. Media statement – immediate release. Retrieved 5 August from: <https://www.pamunewzealand.com/pamunewzealand-prod/corporate/images/Deer-milk-on-the-menu-at-prestigious-Asian-Restaurants-Awards-night.pdf>

Pāmu. n.d. Pāmu deer milk. Retrieved 5 August from: <https://www.pamumilk.com/deer-milk>

Pearse, T. 2022. Farming red deer in New Zealand: industry history, structure and administration. *The Management of Enclosed and Domesticated Deer: International Husbandry Systems and Diseases*. Springer 3-33.

Piddock, G. (2023). Deer milk proves to be a high-value venture for a Taupō farm. Retrieved 5 August from: <https://www.farmersweekly.co.nz/people/deer-milk-proves-to-be-a-high-value-venture-for-a-taupo-farm/>

Pinto, P.A., Anconi, A.C.S.A., de Abreu, L.R., Magalhães, E.J. and Nunes, C.A. 2021. Strategies to determine lactose in cow milk by mid infrared spectroscopy. *Journal of Food Composition and Analysis* 104: 104176.

Rosipal, R. and Krämer, N. 2005. Overview and recent advances in partial least squares. *International Statistical and Optimization Perspectives Workshop: Subspace, Latent Structure and Feature Selection*.

Roy, D., Ye, A., Moughan, P.J. and Singh, H. 2020. Composition, structure, and digestive dynamics of milk from different species – a review. *Frontiers in Nutrition* 7: 577759.

Rutten, M., Bovenhuis, H., Hettinga, K., Van Valenberg, H. and Van Arendonk, J. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. *Journal of Dairy Science* 92: 6202–6209.

Silver, H. 1961. Deer milk compared with substitute milk for fawns. *The Journal of Wildlife Management* 25: 66–70.

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

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

Soyeurt, H., Gengler, N. and Dehareng, F. 2019. Interest, limitations and perspectives of using milk mid-infrared spectra for management and breeding of dairy cattle. *Proceedings of the 6th Symposium on Dairy Cow Nutrition and Milk Quality*.

Soyeurt, H., Misztal, I. and Gengler, N. 2010. Genetic variability of milk components based on mid-infrared spectral data. *Journal of Dairy Science* 93: 1722–1728.

Stats NZ. 2024. Agricultural production statistics: year to June 2024 (final). Statistics New Zealand. Retrieved 3 August from: <https://www.stats.govt.nz/information-releases/agricultural-production-statistics-year-to-june-2024-final/>

Stuart, B.H. 2004. Infrared Spectroscopy: Fundamentals and Applications. John Wiley and Sons.

Thompson, M., Owen, L., Wilkinson, K., Wood, R., and Damant, A. 2002. A comparison of the Kjeldahl and Dumas methods for the determination of protein in foods, using data from a proficiency testing scheme. *Analyst* 127:1666-1668.

Tiplady, K. 2022. The role of Fourier-transform mid-infrared spectroscopy in improving the prediction of new and existing traits in New Zealand dairy cattle. PhD thesis, Massey University, Hamilton, New Zealand.

Trani, A., Gambacorta, G., Loizzo, P., Cassone, A., Fasciano, C., Zambrini, A. V., and Faccia, M. 2017. Comparison of HPLC-RI, LC/MS-MS and enzymatic assays for the analysis of residual lactose in lactose-free milk. *Food Chemistry* 233: 385-390.

Wang, Y., Bekhit, A.E.-D.A., Morton, J.D. and Mason, S. 2017. Nutritional value of deer milk. In: *Nutrients in Dairy and Their Implications on Health and Disease*, pp. 363–375. Elsevier.

Waynant, R.W., Ilev, I.K. and Gannot, I. 2001. Mid-infrared laser applications in medicine and biology. *Philosophical Transactions of the Royal Society A* 359: 635–644.

Youatt, W.G., Verme, L.J. and Ullrey, D.E. 1965. Composition of milk and blood in nursing white-tailed does and the blood composition of their fawns. *The Journal of Wildlife Management* 29: 79–84.