



LARISSA ESTEFANE CRUZ DAS GRAÇAS

**ESPECTROSCOPIA DE REFLECTÂNCIA NO
INFRAVERMELHO PRÓXIMO DE FEZES PARA PREDIZER
VARIÁVEIS NUTRICIONAIS DE VACAS LEITEIRAS
CONFINADAS**

LAVRAS – MG

2021

Larissa Estefane Cruz das Graças

**ESPECTROSCOPIA DE REFLECTÂNCIA NO INFRAVERMELHO PRÓXIMO DE
FEZES PARA PREDIZER VARIÁVEIS NUTRICIONAIS DE VACAS LEITEIRAS
CONFINADAS**

Dissertação de Mestrado apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia, área de concentração em Produção e Nutrição de Ruminantes, para obtenção do título de Mestre.

Orientador: Prof. Marina de Arruda Camargo Danés, PhD.

LAVRAS – MG

2021

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Graças, Larissa Estefane Cruz das.

Espectroscopia de reflectância no infravermelho próximo de fezes para prever variáveis nutricionais de vacas leiteiras confinadas / Larissa Estefane Cruz das Graças. - 2021.

57 p. : il.

Orientador(a): Marina de Arruda Camargo Danes.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. NIRS. 2. Nutrição de precisão. 3. Machine learning. I. Danes, Marina de Arruda Camargo. II. Título.

Larissa Estefane Cruz das Graças

**ESPECTROSCOPIA DE REFLECTÂNCIA NO INFRAVERMELHO PRÓXIMO DE
FEZES PARA PREDIZER VARIÁVEIS NUTRICIONAIS DE VACAS LEITEIRAS
CONFINADAS**

Dissertação de Mestrado apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia, área de concentração em Produção e Nutrição de Ruminantes, para obtenção do título de Mestre.

Aprovada em 30 de Março de 2021

Marina de Arruda Camargo Danés, PhD. UFLA

Daniela Rume Casagrande, D. Sc. UFLA

João Ricardo Rebouças Dórea, PhD. UW - Madison

Orientador: Prof. Marina de Arruda Camargo Danés, PhD.

LAVRAS – MG

2021

*Aos meus pais Airton e Cláudia e meus irmãos Júlia e Lucas, por todo incentivo e apoio. E
por sempre me fazer acreditar que tudo é possível.*

Dedico

Agradecimentos

Agradeço primeiramente a Deus autor de toda vida, que me permitiu alcançar todos os objetivos até aqui. Por ser tão onipresente em minha vida e pôr ser tão bom o tempo todo.

A minha família, fonte inesgotável de inspiração e orgulho. Aos meu Pais e Irmãos, grandes incentivadores da minha, caminhada por toda admiração e zelo. Aos meus Tios e Tias pelo grande exemplo que são em minha jornada acadêmica. Aos meus avós por todo carinho e por ser o alicerce da nossa família.

Agradeço ao departamento de Zootecnia da UFLA, por todos esses anos de parceria e troca. Por todos os ensinamentos acadêmicos e pessoais, através de seus colabores, técnicos, professores e orientadores.

Agradeço ao Minions pelo acolhimento, mesmo com o relacionamento um pouco distante. Me deram apoio, trocas de experiências e amizades que irão durar para toda vida. Em especial agradeço ao Daniel que foi essencial na realização deste trabalho e por se tornar um grande irmão.

Agradeço ao centro de pesquisa Better Nature e ao professor Marcos Neves, que através de seus alunos disponibilizaram o material e seus respectivos dados para realização desde trabalho.

Ao professor Paulo Hein por disponibilizar seu laboratório e seus conhecimentos me auxiliando no processo de análise de dados. Ao professor Erick Maziero pelo grande auxílio na análise de dados deste experimento e pela parceria com nosso grupo de pesquisa.

Em especial agradeço a minha orientadora Marina Danés, por ter embarcado comigo nessa aventura, por ter entendido as minhas limitações e pela parceria e cuidado. E principalmente pelo exemplo que ela representa de força e dedicação, que inspira a cada um de nós. Não poderia ter tido melhor orientação.

Aos meus amigos tão presentes em cada momentos da minha vida. Agradeço a Deus por ter amigos tão inspiradores, incentivadores e tão presentes nos bons e nos maus momentos.

Por fim, a equipe 3rlab por permitir e incentivar a conclusão de mais essa etapa em minha carreira.

“Change will not come if we wait for some other person, or if we wait for some other time. We are the ones we've been waiting for. We are the change that we seek.”

Barack Obama.

RESUMO INTERPRETATIVO

MENSURAR CONSUMO INDIVIDUAL DE NUTRIENTES PODE SER UMA TAREFA DIFÍCIL EM SISTEMAS DE PRODUÇÃO DE LEITE COM VACAS CONFINADAS.

MÉTODOS EXISTENTES SÃO TRABALHOSOS E DEMORADOS.

NIRS FECAL É UM MÉTODO NÃO INVASIVO, NÃO DESTRUTIVO E NÃO POLUENTE.

CERCA DE 70% DAS VARIÁÇÕES EM CONSUMO PODEM SER EXPLICADAS ATRAVÉS DAS FEZES.

**234
AMOSTRAS**
**64
VACAS**
**5
EXPERIMENTOS**

**TRATAMENTOS
ESPECTRAIS:**

NENHUM
MSC
SNV

ALGORITMOS:

PLSR
GBR
KNNR
SVM

O NIRS FECAL SE MOSTROU PROMISSOR PARA PREDIZER VARIÁVEIS COMPLEXAS

VÁRIÁVEIS:

- CONSUMO E COMPOSIÇÃO DA DIETA CONSUMIDA
- DIGESTIBILIDADE DE NUTRIENTES
- EFICIÊNCIAS
- SELEÇÃO DE PARTÍCULAS

É POSSÍVEL PREDIZER VARIÁVEIS COMPLEXAS PELO NIRS FECAL.

E NECESSÁRIO QUE SEJA INVESTIGADO O MELHOR ALGORITMO.

É PRECISO VALIDAR A APLICABILIDADE DE MODELOS NIRS EM DADOS EXTERNOS.

RESUMO

A ingestão individual de nutrientes, bem como variáveis relacionadas a esta característica são essenciais dentro de um sistema de nutrição de precisão e ainda assim difíceis de medir individualmente quando se tem animais criados em grupo. Foi proposto que as informações químicas contidas nas fezes são representativas da dieta consumida e estão relacionadas à ingestão e digestibilidade. A espectroscopia no infravermelho próximo (NIRS) para analisar 234 amostras fecais de 64 vacas leiteiras em lactação alimentadas com TMR, de 5 experimentos realizados em condições semelhantes, nos quais foram analisados o CMS individual medido em um tie-stall, digestibilidade e composição da dieta consumida (PB, FDN e amido) e seleção de partículas de alimentos. Foram utilizados dois softwares para as análises estatísticas dos dados, os software The Unscrambler e Python. Os espectros foram analisados com regressor de mínimos quadrados parciais (PLSR), *support vector machine* (SVR), *k-nearest neighbors regressor* (KNNR) e *gradient boosting regressor* (GBR) com ou sem pré-tratamento do conjunto de dados (MSC ou SNV). Duas estratégias de validação foram testadas: remover do conjunto de treinamento todos os pontos de dados de um animal (deixar um animal de fora - LOAO) ou de um experimento (LOEO) e usar o conjunto de dados excluído para testar o algoritmo. Não foram encontradas grandes diferenças entre os algoritmos, o melhor algoritmo variou de acordo com a variável estudada. A melhor estratégia de validação foi a LOAO. Para validação dos modelos utilizamos o RMSE. O CMS apresentou erro de 2,98 Kg/d, acompanhado e valores inferiores a 0,5% para composição da dieta consumida. Para digestibilidades foram obtidos valores de RMSE inferiores a 8,71%. A melhor acurácia para seleção de partículas foi da peneira de 8mm, com erro médio de 8,85%. Para EUN a média foi de 2,65% e para EA igual a 0,23 Kg leite/ Kg MS. Concluímos que a análise de dados espectrais recebe influência de diversos fatores externos que irão determinar qual melhor método a ser utilizado para se obter melhores resultados. É importante notar que, embora os resultados sejam promissores e as estratégias de validação visem minimizar o emaranhamento biológico dentro de uma vaca ou ensaio, o conjunto de dados veio inteiramente do mesmo rebanho. Assim, é possível que o modelo seja menos preciso quando usado em um rebanho e ambientes completamente diferentes. É necessário que estes algoritmos sejam explorados, buscando pelos melhores hiperparâmetros, utilizando conjunto de dados maiores e com validação externa para verificar a aplicabilidade desses modelos.

Palavras-Chaves: Machine learning. Estratégias de validação. F.NIRS. Nutrição de precisão.

ABSTRACT

The individual intake of nutrients, as well as variables related to this characteristic are essential within a system of precision nutrition and yet difficult to measure individually when you have group-raised animals. It has been proposed that the chemical information contained in the faeces is representative of the diet consumed and is related to ingestion and digestibility. Near infrared spectroscopy (NIRS) was used to analyze 234 faecal samples from 64 lactating dairy cows fed with TMR, from 5 experiments carried out under similar conditions, in which the individual CMS measured in a tie-stall, digestibility and composition of the consumed diet (CP, NDF and starch) and food selection (Sorting starch and NDF). Two softwares were used for the statistical analysis of the data, the software The Unscrambler and Python. The spectra were analyzed with partial least squares regressor (PLSR), support vector machine regression (SVR), K- nearest neighbor regression (KNNR) and gradient boosting regression (GBR) with or without pretreatment of the set of data (MSC or SNV). Two validation strategies were tested: remove from the training set all data points of an animal (leaving an animal out - LOAO) or an experiment (LOEO) and use the excluded data set to test the algorithm. The results demonstrate that the most used analysis for NIR spectra (PLSR with pretreatments) was not adequate to deal with the complex interaction existing between the analyzed parameters and the faecal spectra. No major differences were found between the algorithms, the best algorithm varied according to the variable studied. The best validation strategy was LOAO. To evaluate the models, we used the RMSE. The CMS showed an error of 2.98 kg / d, accompanied by values below 0.5% for the composition of the consumed diet. For digestibility, RMSE values below 8.71% were obtained. The best accuracy for particle selection was the 8mm sieve, with an average error of 8.85%. For EUN the average was 2.65% and for EA equal to 0.23 Kg milk / Kg DM. As conclusion the analysis of spectral data is influenced by several external factors that will determine which method is best to be used to obtain better results. It is important to note that, although the results are promising and the validation strategies are aimed at minimizing biological entanglement within a cow or assay, the data set came entirely from the same herd. Thus, it is possible that the model is less accurate when used in a completely different herd and environment. It is necessary that these algorithms be explored, looking for the best hyperparameters, using a larger data set and with external validation to verify the applicability of these models.

Keywords: Machine learning. Validation strategies. F.NIRS. Precision nutrition.

LISTA DE TABELAS

Tabela 1. Composição das dietas dos experimentos ¹ em ingredientes (% da matéria seca)	30
Tabela 2. Descrição das variáveis de interesse dos experimentos utilizados no banco de dados para desenvolvimento das curvas de calibração a partir dos espectros de NIRS fecal.....	32
Tabela 3 - Avaliação dos modelos desenvolvidos por regressão dos quadrados mínimos parciais para prever consumo de matéria seca, composição da dieta consumida, digestibilidade de nutrientes, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal.....	36
Tabela 4 - Avaliação do desempenho (RMSEVC) dos quatro algoritmos para predição de consumo de matéria seca, composição da dieta consumida, digestibilidade, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal com estratégia de validação <i>leave-one-experiment-out</i>	37
Tabela 5 - Avaliação do desempenho (RMSEVC) dos quatro algoritmos para predição de consumo de matéria seca, composição da dieta consumida, digestibilidade, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal com estratégia de validação <i>leave-one-animal-out</i>	38
Tabela 6 - Descrição dos melhores modelos preditores para consumo de matéria seca e composição da dieta consumida a partir de espectros de NIRS fecal nas estratégias de validação <i>leave-one-experiment-out</i> e <i>leave-one-animal-out</i>	39
Tabela 7 - Descrição dos modelos selecionados para variáveis de digestibilidade dos nutrientes nas estratégias de validação <i>leave-one-experiment-out</i> e <i>leave-one-animal-out</i>	40
Tabela 8 - Descrição dos modelos selecionados para variáveis de seleção de partículas nas estratégias de validação <i>leave-one-experiment-out</i> e <i>leave-one-animal-out</i>	40
Tabela 9 - Descrição dos modelos selecionados para variáveis de eficiência nas estratégias de validação <i>leave-one-experiment-out</i> e <i>leave-one-animal-out</i>	41

LISTA DE ABREVIACOES

EA Eficincia alimentar

EUN Eficincia de utilizao de Nitrognio

GBR *Gradient boosting regressor*

KNNR *k-nearest neighbors regressor*

LOAO *Leave-one-animal-out*

LOEO *Leave-one-experiment-out;*

MSC Correo multiplicativa de sinal

NIRS Espectroscopia de refletncia no infravermelho prximo

PLS Mnimos quadrados parciais

RMSE Raiz quadrada do erro mdio

RMSEC Raiz quadrada do erro mdio obtido na calibrao

RMSECV Raiz quadrada do erro mdio obtido na validao cruzada

SNV Normalizao padronizada de sinal

SVR *Support vector machine*

SUMÁRIO

1	Introdução	13
2	Objetivos	14
3	Revisão de literatura	15
3.1	Nutrição de Precisão	15
3.2	Consumo de matéria seca.....	17
3.3	Digestibilidade	19
3.4	Seleção de alimentos.....	21
3.5	Espectroscopia no infravermelho próximo	21
3.5.1	Tratamentos espectrais.....	23
3.5.2	Calibração	24
3.5.3	Validação	25
3.5.4	Utilização do NIRS fecal na produção animal.....	26
3.6	Métodos alternativos de calibração.....	28
4	Materiais e Métodos.....	30
5	Resultados.....	35
6	Discussão	41
7	Conclusão.....	48
8	Referências Bibliográficas	49

1 Introdução

O consumo de alimentos em sistemas de produção animal é uma variável importante no gerenciamento, pois, permite uma formulação nutricional mais precisa e econômica. Quando associado à produção de leite, o consumo de alimentos permite estimar o valor econômico de cada animal em qualquer estágio da lactação. O desempenho dos ruminantes está intimamente relacionado à ingestão, sendo que cerca de 60 a 90% do desempenho dos animais pode ser explicado por variações no consumo (MERTENS, 1994). Medidas de consumo individual são facilmente obtidas em confinamentos do tipo *tie stall* ou em centros de pesquisa. Já em sistemas de produção em que os animais são criados em grupos, o consumo do lote pode ser calculado pela diferença entre quantidade ofertada e sobra, mas não é possível obter o consumo individual.

A maioria das fazendas que buscam maior eficiência na produção, com redução de desperdícios e otimização de desempenho individual, utiliza instalações de confinamento, com animais mantidos em lotes e alimentados com dieta completa (TMR). As TMR são planejadas como uma mistura homogênea para ajudar a minimizar o consumo seletivo de componentes individuais do alimento por vacas leiteiras (COPPOCK et al., 1981). No entanto, foi demonstrado recentemente que o gado leiteiro em lactação ingere uma dieta que pode ser diferente da dieta oferecida. DeVries et al. (2005), reportaram que o conteúdo de FDN da TMR no cocho aumentou ao longo do dia, indicando que as vacas estavam selecionando contra os componentes de fibra da dieta. De igual modo, outros pesquisadores mostraram que a maioria das vacas tende a selecionar partículas menores de concentrado em vez de partículas longas de forragem (LEONARDI & ARMENTANO, 2003). Nesses casos, a dieta realmente consumida pelas vacas é maior em carboidratos fermentáveis e menor em fibra efetiva do que a formulada, aumentando assim o risco de acidose ruminal.

As fontes de variação na TMR incluem a composição nutricional dos ingredientes, a pesagem dos alimentos e a uniformidade da mistura, influenciada também pelas características físicas dos alimentos, de mistura do equipamento e do próprio animal. Estimar essa variação de forma rápida e precisa, propiciaria tomadas de decisão que permitiriam melhorar o manejo nutricional do rebanho. Ao longo dos anos, foram desenvolvidas diversas metodologias para mensurar o consumo individual e a composição do que foi consumido, assim como a digestibilidade e seleção de alimentos. Os principais avanços se deram no uso de marcadores (VELÁSQUEZ et al., 2018) e modelos matemáticos (NRC, 2001; CNCPS; FOX et al., 2004) que ainda carregam as limitações de serem trabalhosos e passíveis de erros de aplicação. Além do fato de que os modelos não são utilizados para avaliação individual. Atualmente, pesquisas com sensores (CHIZZOTTI et al., 2015) e espectroscopia (JOHNSON et al., 2017) mostram

um novo horizonte para avanços em pesquisas com foco na avaliação do consumo real de nutrientes.

A crescente demanda por tecnologias rápidas e precisas para avaliação da composição nutricional da dieta e determinação do consumo em ruminantes fez com que a espectroscopia de reflectância no infravermelho próximo (NIRS) ganhasse espaço nestas avaliações. O NIRS apresenta vantagens em relação às análises convencionais, pois estas são geralmente trabalhosas, apresentam custos altos e demandam mais tempo, além de gerar resíduos químicos de impacto ambiental (SKOOG et al, 2006). A técnica NIRS consiste no desenvolvimento de modelos de calibração através do uso de parâmetros estatísticos e matemáticos, que correlacionam as características dos espectros das amostras (variáveis independentes) com os constituintes da dieta (variáveis dependentes) consumida pelos animais. O NIRS oferece vantagens sobre os métodos mais tradicionais usados para prever as características e a ingestão da dieta, pois fornece uma análise não destrutiva, não invasiva e de baixo custo.

A tecnologia NIRS também pode ser usada para estimar a variação entre animais na composição da dieta, digestibilidade e consumo. Holloway et al. (1981) propuseram o uso de componentes fecais para prever a ingestão em novilhos e determinaram que 70% da variação na ingestão entre animais poderia ser explicada através das fezes. Segundo esses autores, as informações químicas contidas nas fezes são representativas da dieta consumida e estão relacionadas ao consumo e digestibilidade. Embora pesquisas anteriores tenham demonstrado a capacidade dessa técnica de prever o consumo voluntário, qualidade da dieta e digestibilidade (AGNEW et al., 2004, DECRUYENAERE et al., 2009, JOHNSON, 2017), a utilização do NIRS fecal para avaliação de consumo de TMR, aprofundando em aspectos correlacionados com essa variável como seleção de alimentos, digestibilidade de amido, PB e FDN ainda não foram estudados, bem como a composição da dieta consumida. Estas são variáveis zootécnicas importantes dentro de um sistema de produção de vacas leiteiras e que demandam agilidade, facilidade e precisão.

2 Objetivos

O objetivo principal deste trabalho foi desenvolver curvas de calibração de NIRS para estimar o consumo e composição da dieta consumida, eficiência alimentar, digestibilidade de nutrientes e PB e seleção de alimentos de dietas completas, utilizando amostras fecais de vacas leiteiras. Adicionalmente, buscou-se avaliar e comparar diferentes técnicas de regressão e estratégias de validação dos modelos preditivos, levando em conta a interdependência biológica dos dados disponíveis para evitar a inflação de valores de acurácia.

3 Revisão de literatura

3.1 Nutrição de Precisão

A nutrição de precisão se baseia no uso de tecnologia da informação para melhorar o desempenho econômico, social e ambiental na fazenda (SPILKE & FAHR, 2003). Essa abordagem tem potencial de melhorar a produtividade leiteira, atendendo às necessidades de nutrientes de cada animal ou lote de forma mais precisa (GEHMAN, 2011). A implementação de sistemas de alimentação baseados na nutrição de precisão podem melhorar a produção de leite e a lucratividade (ANDRÉ et al., 2007). A implementação de programas nutricionais que permitem agrupar os animais de acordo com suas necessidades nutricionais e visam fornecer uma dieta mais homogênea podem aumentar a produção, a eficiência reprodutiva e a lucratividade (ALLEN, 2009). Uma formulação de ração mais precisa significa que as diferenças entre oferta e demanda de nutrientes são menores dentro de um grupo (ST-PIERRE & THRAEN, 1999). No entanto, um grande desafio na nutrição de vacas leiteiras de alta produção tem sido encontrar o equilíbrio nutricional correto para promover a saúde ruminal e maximizar o consumo de energia e o fluxo de nutrientes para a glândula mamária (VANDEHAAR et al., 2012).

As dietas completas (TMR) surgiram como meio de fornecer suprimento consistente de nutrientes aos microrganismos do rúmen para potencializar as funções do rúmen e melhorar a eficiência da utilização de nutrientes (COPPOCK et al., 1981). O agrupamento dos animais foi o método mais eficaz que permitiu que as vacas fossem alimentadas de forma mais homogênea para atender suas necessidades de nutrientes (MCGILLIARD et al., 1983). Contudo, variações na composição da dieta são inevitáveis. A consistência da TMR pode ser influenciada pela composição nutricional dos alimentos, equipamentos de mistura, condição e ordem de mistura dos ingredientes (MIKUS, 2012). Além disso, alimentar o rebanho com uma TMR, em muitos casos, significa formular uma dieta para o grupo levando em consideração a média do rebanho. Nesse cenário, podem ocorrer os casos de subalimentação ou superalimentação em grande parte do rebanho. A superalimentação de vacas menos produtivas provavelmente aumentará a excreção de N e outros nutrientes (AGUERRE et al., 2011).

Segundo Stone (2008), essa variabilidade pode afetar a saúde e a produção incluindo incidência de deslocamento de abomaso e flutuações na produção média diária de leite. Fornecimento de uma ração consistente é parte essencial de maximizar o desempenho da vaca e obter o melhor retorno sobre o custo alimentar. Inconsistências na ingestão de nutrientes têm muitas causas, dentre elas a seleção pelo animal, que resulta em diferenças entre a dieta

formulada e a consumida (LEONARDI & ARMENTANO, 2003) e afeta a produção à nível de rebanho (SOVA et al., 2013). A seleção da TMR pode reduzir o valor nutritivo da sobra do cocho de alimentação, particularmente nas últimas horas após o horário da entrega do alimento (DEVRIES et al., 2005). Medidas para reduzir a variabilidade diária da TMR, como análises regulares de alimentação, preparação e educação de funcionários, devem ser tomadas para garantir a máxima saúde do rebanho e lucratividade. No entanto, o primeiro passo para enfrentar esse desafio é conseguir mensurar de forma eficiente a seleção da dieta pelos animais.

Recentes avanços na indústria leiteira, relacionados a melhorias na produtividade, saúde animal e redução do impacto ambiental, exigem conhecimento profundo e preciso do valor nutritivo dos alimentos para manter a sustentabilidade econômica das fazendas leiteiras. A eficiência de utilização reduzida de certos nutrientes, como nitrogênio, levanta preocupações ambientais devido ao aumento da excreção de compostos nitrogenados (DIJKSTRA et al., 2011). A digestibilidade da dieta está diretamente relacionada ao suprimento de energia e ao desempenho animal e tem sido usada para expressar o conteúdo de energia de alimentos para ruminantes (VAN SOEST, 1991).

Diferenças individuais no consumo de matéria seca e conversão alimentar têm sido avaliadas também por alguns grupos de pesquisa, com o intuito de selecionar animais geneticamente mais eficientes (VALLIMONT et al., 2010). Avaliações genéticas de consumo de matéria seca (CMS) de vacas em rebanhos de pesquisa fornecem evidências de que existe variação genética para CMS, mas os dados de rebanhos comerciais são escassos, principalmente devido à dificuldade de aplicar as metodologias existentes em nível de rebanho comercial. A seleção genética afeta indiretamente o CMS porque as seleções que alteram a produção de leite e o peso corporal das vacas também afetam o CMS (VEERKAMP, 2012).

Atualmente, existem várias formas de medir essas variáveis dentro de fazendas comerciais e em ambientes de pesquisa. A nutrição de precisão que tem por objetivo oferecer diariamente uma dieta mais próxima o possível das exigências do animal e a redução de desperdício de ingredientes utilizados na alimentação permite mais informações, controle e menor variabilidade da dieta. Métodos validados para mensurar variáveis como consumo de matéria seca, composição da dieta consumida, digestibilidade e composição nutricional dos alimentos são amplamente utilizados como ferramentas para auxiliar nutricionistas nas tomadas de decisão em relação à nutrição do rebanho, com intuito de se ter maior eficiência produtiva. Os métodos mais utilizados no cenário atual para mensurar essas variáveis serão descritos neste trabalho.

3.2 Consumo de matéria seca

O consumo de matéria seca (CMS) é o parâmetro que mais influencia o desempenho animal. Por isso, é essencial que se tenha métodos para estimar esse parâmetro de forma confiável. Diversas metodologias são utilizadas em cenários de pesquisa para se estimar o CMS, como o uso de marcadores indigestíveis, modelagem e desaparecimento da dieta no cocho (UNDI et al., 2008). No entanto, esses métodos carregam problemas que podem causar variações nas estimativas, como, recuperação fecal incompleta de marcadores ou a incapacidade das equações do modelo de contabilizar os efeitos de manejo, dieta, animal e ambiente nos dados utilizados para o desenvolvimento do modelo (MCMENIMAM et al., 2009). De fato, a grande variação entre e dentro dos métodos foi identificada como a principal limitação dos métodos disponíveis atualmente para estimar o CMS (DECRUYENAERE et al., 2009).

O uso de marcadores é atualmente a técnica mais utilizada para estimar o CMS em ruminantes (VELÁSQUEZ et al., 2018). Os marcadores são substâncias indigeríveis que não são secretadas ou absorvidas pelo animal, têm taxas de passagem semelhantes às dos alimentos, podem ser recuperados completamente após a ingestão e permitem análises químicas práticas e precisas (FAHEY & JUNG, 1983). A técnica de marcador para estimar o consumo utiliza um marcador externo para estimar a produção fecal e um marcador interno, que faz parte do alimento, para estimar a digestibilidade da matéria seca. A técnica considera que a mudança na composição da dieta representa o resultado da digestão e a mudança na concentração do marcador representa a produção fecal (GORDON, 1995). O consumo é calculado pela razão entre produção fecal e a porção indigestível da dieta, inversamente proporcional a digestibilidade. As principais críticas a essa técnica são a dificuldade e o trabalho de fornecer os marcadores externos duas vezes ao dia, durante vários dias, a necessidade de interação excessiva entre animais e pessoas, o que pode afetar o consumo, e o trabalho intensivo decorrente da coleta de amostras de fezes.

Outro método tradicional é o desaparecimento de cocho, onde o CMS é calculado como a diferença entre a massa de dieta pré e pós-ingestão. A metodologia se dá pela pesagem, análise e comparação da dieta ofertada e a sobra de cocho. Essa técnica que permite a predição das características da dieta e cálculos de necessidades de energia para o desempenho animal são estimativas adequadas para grupos de animais ou individualmente em centros de pesquisas com animais em instalações tipo *tie-stall* (MOORE, 1996). Segundo Macoon (2003), o método de desaparecimento da dieta fornece estimativas próximas às obtidas com base nas necessidades de energia. No entanto, na rotina de fazendas comerciais este método ainda carrega a limitação

de ser utilizada para mensurar consumo em grupo e não individualmente para cada animal, como se espera em um sistema de nutrição de precisão.

A modelagem matemática utiliza equações para descrever ou simular processos em um sistema e pode ser uma ferramenta para extrair mais informações sobre o consumo dos alimentos. Os métodos de modelagem vêm sendo atualizados regularmente para incorporar conhecimentos recentes, entre eles estão os modelos americanos Nutrient Requirements of Dairy Cattle (NRC, 2001) e Cornell Net Carbohydrate and Protein System (CNCPS; FOX et al., 2004; VAN AMBURGH et al., 2015) e o modelo europeu do Institut National de la Recherche Agronomique (INRA, 2018). Esses sistemas oferecem a possibilidade de formular rações combinando o aporte de nutrientes com as necessidades dos animais. Os principais modelos de predição do CMS de vacas leiteiras foram desenvolvidos utilizando variáveis relacionadas ao animal, à dieta e/ou ao ambiente. No entanto, os modelos de predição são mais eficazes quando as condições do rebanho onde será aplicado se assemelham as condições da população em que foram criados, ou seja, quando em populações diferentes a estimativa pode conter alguma variação (MERTENS, 1987).

Pesquisas recentes vêm se concentrando na possibilidade de medir variáveis que são influenciadas por variações no CMS e que são mais fáceis de serem estimadas. Dórea et al. (2017) avaliou a relação entre CMS e a excreção de derivados de purina na urina para desenvolver equações para prever CMS, partindo do princípio de que a síntese de proteína microbiana é uma variável ruminal que é afetada por diferenças no CMS. Os autores concluíram que os derivados de purina na urina estão fortemente correlacionados com o CMS e juntamente com a produção de leite e o peso corporal dos animais podem ser usados como uma abordagem alternativa para estimar o consumo de ração em rebanhos leiteiros, em pesquisa ou sistemas de produção comercial.

Da mesma maneira, o uso do espectro do leite a partir da espectroscopia de infravermelho médio que tem sido utilizado para estimar componentes do leite, como ácidos graxos do leite e frações de proteína do leite (SOYEURT et al., 2012). A predição do CMS usando espectros de leite foi estudada pela primeira vez por Shetty et al. (2017), mas não obtiveram resultados positivos, o que pode ser devido ao tipo de análise estatística utilizada para análise dos espectros e calibração. No entanto, mais estudos com espectros de leite devem ser realizados com foco em diferentes tipos de validação. Uma vez que, recentes avanços nos estudos de espectros de leite com novas metodologias de análises de dados mostraram que a técnica é promissora para predição de CMS (DÓREA et al., 2018).

3.3 Digestibilidade

A digestibilidade de um alimento é consequência da ação de enzimas microbianas e do animal durante sua passagem pelo trato digestivo do ruminante. A utilização da energia dietética por ruminantes além de ser dependente do perfil nutricional do próprio alimento, sofre influência também do perfil nutricional dos outros alimentos que compõe a dieta. Como exemplo, temos a influência da concentração de amido na digestão de fibras no rúmen. O objetivo principal das técnicas de avaliação de alimentos é prever a disponibilidade de nutrientes em sistemas de produção animal. Os métodos disponíveis incluem análises químicas, digestibilidade *in vivo*, *in vitro* ou *in situ*, marcadores de digesta e espectroscopia de refletância no infravermelho próximo (NIRS).

A mensuração da digestibilidade aparente *in vivo* em fazendas comerciais para melhorar as estimativas da digestibilidade de nutrientes para formulação de dieta ou para solução de problemas não foi bem sucedida. A digestibilidade também pode ser mensurada através do uso de marcadores dietéticos. Esse método, baseia-se no fornecimento de uma quantidade conhecida de marcador na dieta e posterior quantificação desse marcador nas fezes dos animais (KAVANAGH et al., 2001). O custo e o tempo dispendido no manejo de marcadores externos, como óxido de cromo e óxido de lantânio, tornam o método um pouco limitado. A lignina tem sido usada há muito tempo como um marcador interno, no entanto, a recuperação e quantificação completas são difíceis (FAHEY & JUNG, 1983). Lopes et al. (2015) usou com sucesso a fibra em detergente neutro indigestível *in vitro* (iFDN) como um marcador interno para estimar a digestibilidade de nutrientes *in vivo*.

O NDT (nutrientes digestíveis totais) é um método de expressão energética, onde corresponde a soma das frações digestíveis dos alimentos de acordo com análises do sistema proximal de Weende. A modelagem do NDT dietético primeiro requer uma medição precisa dos nutrientes da ração para cada ingrediente da dieta e, em seguida, requer a soma das concentrações de nutrientes e a estimativa dos coeficientes de digestibilidade aparente dos nutrientes (WEISS et al., 1992). Determinar a concentração de nutrientes na ração é uma prática comum para nutricionistas de laticínios comerciais. No entanto, medir a digestibilidade dos nutrientes tem sido um desafio, tanto para ingredientes individuais quanto para a TMR. Os laboratórios de teste de forragem comercial utilizam várias técnicas de digestão ruminal e pós-ruminal *in vitro* e *in situ* para prever digestibilidades de nutrientes individuais como entrada para modelos de NDT (NRC, 2001). No entanto, as metodologias disponíveis comercialmente variam amplamente e poucas foram relacionadas ao desempenho do gado leiteiro.

Um método de avaliação de qualidade da dieta muito usado atualmente é a estimativa da digestibilidade do amido por meio da concentração de amido fecal. O aumento na digestão de amido do trato total pode aumentar a produção de leite, proteína e eficiência alimentar (FIRKINS et al., 2001). A concentração fecal de amido provou ser um indicador confiável de digestibilidade de amido em bovinos confinados (CORONA et al., 2005). Fredin et al. (2014) estudou a relação entre amido e digestibilidade no trato total, com intuito de desenvolver uma equação para digestibilidade do amido e verificar o potencial do NIRS para avaliar o amido fecal. A concentração de amido fecal provou ser precisa como preditora de digestibilidade do amido em vacas leiteiras. O amido fecal demonstrou ser predito com precisão pelo NIRS, permitindo que estimativas de digestibilidade do amido possa ser determinada de forma rápida e econômica. O amido fecal parece útil para permitir alterações na formulação de dietas. A alteração da dieta, tem o objetivo de melhorar a digestibilidade e a eficiência da produção de leite. Atualmente, laboratórios comerciais oferecem análise química úmida e NIRS para medida da concentração de amido nas fezes.

A fibra é também um componente importante na avaliação de alimentos para ruminantes, que afeta diretamente o consumo de alimentos pelos animais. No entanto, apesar de ser um componente crítico na análise do alimento, ele depende também da sua digestibilidade para de fato representar a qualidade do alimento. A digestibilidade da fibra é mais variável do que a de outros componentes da dieta e é responsável por até 40% da energia digestível das forragens (GOESER & COMBS, 2009). A digestibilidade aparente no trato total da FDN é uma ferramenta útil para avaliar a contribuição da fibra na energia digestível.

Um método que vem sendo utilizado e que se compara a medida real de digestibilidade da fibra é o TTFDND (digestibilidade da fibra em detergente neutro no trato total). O método foi proposto por pesquisadores da Universidade de Wisconsin, com o intuito de ser um método rápido, preciso e representativo para tomadas de decisão a campo. Este valor pode ser utilizado na predição dos valores energéticos da forragem. O TTFDND também pode ser utilizado como um classificador de forragens auxiliando nas decisões nutricionais (COMBS, 2013). No entanto, esse valor não é obtido de forma individual como a digestibilidade do amido descrita nesta revisão. Métodos capazes de prever a digestibilidade da FDN e da PB individualmente não são conhecidos, porém são necessários devido à importância desses componentes na dieta e no metabolismo animal.

3.4 Seleção de alimentos

Muitos estudos têm mostrado que os bovinos são capazes de selecionar entre diferentes porções anatômicas presentes nos alimentos (OSAFO et al., 1997; METHU et al., 2001). Vários fatores químicos e físicos da dieta como concentração de FDN e tamanho de partícula, podem afetar a fermentação ruminal. A redução no tamanho das partículas dietéticas também diminui a concentração de gordura do leite, a atividade de mastigação e pH ruminal (LE LIBOUX & PEYRAUD, 1999). Portanto, é importante considerar o FDN e o tamanho das partículas na formulação da dieta. Dietas com base em partículas longas tendem a ter uma diferença maior entre a TMR oferecida e a consumida (LEONARD et al., 2003).

Para determinação do tamanho de partícula de forragem, o Penn State Particle Separator (PSPS) é um método de análise rápido e econômico. A distribuição de tamanho de partícula de dietas e sobras geralmente são determinadas usando o separador de tamanho de partícula, de acordo com o protocolo ASAE S424.1 (ANSI, 1998). A classificação ou seleção é calculada como a ingestão real de cada peneira, expressa como uma porcentagem da ingestão estimada. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção.

Com sua construção e tamanho simples, o método de peneiramento PSPS pode ser implementado na fazenda e usado no momento da colheita ou alimentação para determinar o tamanho de partícula de forragens ou TMR (LAMMERS et al., 1996). No entanto, embora o aparelho tenha sido amplamente aceito e as medições de tamanho de partícula usando o PSPS sejam comumente relatadas na literatura, a TMR normalmente contém 40 a 60% de concentrado, a maior parte do qual passa pela peneira de 8,0 mm. A medição de partículas menores pode ser útil na compreensão dos efeitos do tamanho de partícula da ração na função ruminal e consumo de alimentos.

3.5 Espectroscopia no infravermelho próximo

O método NIRS é considerado uma tecnologia padrão, aceito como método oficial da AOAC para a predição da FDA, PB (AOAC 989.03) e umidade (AOAC 991.01; BATTEN, 1998), além de apresentar resultados positivos na predição da ingestão e da digestibilidade (REEVES III e DELWICHE, 2003; DECRUYENAERE et al., 2009). O NIRS baseia-se no fato de que cada um dos componentes químicos apresenta propriedades de absorção na região do infravermelho próximo que podem ser usadas para quantificar o componente.

A espectroscopia NIRS compreende a região do espectro eletromagnético que compreende a região de 780 a 2500 nm, o qual analisa substâncias orgânicas a partir da absorção de energia eletromagnética emitida, com comprimentos de ondas situadas na região do infravermelho próximo, que em contato com a amostra podem ser absorvidos por meio de ligações covalentes existentes entre os elementos presentes nos compostos orgânicos, que vibram em determinados comprimentos de onda (SKOOG et al., 2006). Com o estudo da quantidade e tipo de ligação covalente presente no material por meio das ondas eletromagnéticas é possível relacionar os espectros com os componentes químicos.

A técnica NIRS se respalda no desenvolvimento de modelos de regressão multivariada, onde se estuda correlações entre os espectros e a dieta selecionada (STUTH et al., 2003). Os espectros são utilizados para criar os dados de calibração e desenvolver os modelos de predição (SHENK & WESTERHAUS, 1996). A metodologia NIRS apresenta as vantagens de ser rápida e não destrutiva, permitindo que o material seja utilizado para outras análises. Além de permitir a análise de maiores quantidades de amostras em menor tempo.

A principal desvantagem ou dificuldade da espectroscopia NIR é a complexidade de seus espectros, que são de difícil interpretação, devido à natureza dos seus sinais (SKOOG et al., 2006). Os espectros NIR apresentam muitas informações químicas irrelevantes, e, além disso, se correlacionam com as propriedades físicas da amostra, dificultando mais ainda a interpretação dos dados obtidos. Por isso se torna necessário à utilização da quimiometria para extrair as informações importantes dos espectros de amostras obtidos na espectroscopia NIR (PASQUINI, 2003; LUYPAERT et al., 2007). A quimiometria é a disciplina da química que utiliza métodos matemáticos e estatísticos para planejar ou selecionar condições ideais de medição e experimentos para extrair o máximo de informações possível de dados químicos. A quimiometria está dividida em algumas áreas, entre elas a calibração multivariada, que é uma das combinações de métodos estatísticos com dados químicos, mais bem-sucedidos (NETO et al., 2006). A calibração pode ser definida como uma série de operações que estabelecem uma relação entre medidas instrumentais e valores para uma propriedade de interesse correspondente (MARTENS & NAES, 2002).

Os principais componentes de um aparelho NIRS são fonte de luz, sistema divisor de feixe (para seleccione o comprimento de onda), detector óptico, suporte para amostras e sistema analisador de dados (MANLEY et al., 2008). Existem vários dispositivos disponíveis para análise NIRS. O mais comumente usado para selecionar comprimentos de onda são monocromadores que varrem toda a faixa de comprimento de onda usando uma grade ou prisma que forneça versatilidade máxima. Baseados em propriedades ópticas, existem três maneiras

para medir o produto da interação da radiação NIR com a amostra: transmitância, reflectância difusa e transflectância, sendo a transmitância empregada para materiais transparentes, a transflectância utilizada para líquidos, sólidos ou semi-sólidos, dependendo das características de absorção e espalhamento das amostras e a reflectância difusa, empregada a amostras sólidas e pó (PASQUINI, 2003).

Os espectrofotômetros NIRS de espectro contínuo, como de Fourier, apresentam melhor precisão e exatidão na recuperação dos comprimentos de onda. Apresentam alta relação sinal/ruído e rapidez em suas varreduras, o que permite a obtenção de espectros em um curto intervalo de tempo (PASQUINI, 2003). Instrumentos com transformação de Fourier são apropriados tanto para medidas qualitativas como quantitativas. O espectrofotômetro com transformação de Fourier utiliza um interferômetro ao invés de redes de difração e apresenta vantagens como: eficiência no transporte da radiação até o detector, melhorando a relação sinal/ruído e alto poder de resolução e reprodutibilidade do comprimento de onda (SKOOG et al, 2006).

A reflectância difusa é usada como uma das várias maneiras para realização de medidas analíticas na região NIR. Este modo de medida é apropriado quando a superfície do material reflectante é pouco absorvente no comprimento de onda incidente. Na reflectância difusa um feixe de luz penetra na camada superficial das partículas, os modos vibracionais das moléculas são excitados e há o espalhamento da radiação em todas as direções. Parte da radiação é absorvida pela amostra, e a outra parte é refletida difusamente. Um espectro de reflectância é obtido como resultado, e neste espectro estão às informações qualitativas e quantitativas sobre a natureza química do analito de interesse (PASQUINI, 2003; WETZEL et al, 1983).

Apesar das diferenças em absorção com diferentes composições químicas, muitas vezes os picos de absorção se sobrepõem em vários pontos da região espectral. Devido a isso, é necessário processar matematicamente os dados espectrais para obter informações valiosas das propriedades químicas das amostras.

3.5.1 Tratamentos espectrais

Os dados espectrais são compostos por sinais relacionados à composição biológica do alimento, assim como sinais sem informação provenientes de ruídos, mudança de linha de base e bandas sobrepostas (RINNAN et al., 2009a). Portanto, o pré-processamento de dados espectrais é uma estratégia comum e crucial que ajuda a eliminar sinais indesejáveis presentes nos dados brutos, maximizando a relação entre o espectro infravermelho e o fenótipo de interesse (RINNAN et al., 2009a; DE MARCHI et al., 2014; MCPARLAND & BERRY, 2016).

O pré-processamento é comumente executado usando técnicas de pré-tratamento matemático ou abordagem de seleção de variável. As técnicas de pré-processamento mais amplamente utilizadas em espectroscopia NIR podem ser divididas em duas categorias: métodos de correção de dispersão e derivadas espectrais.

O método adequado irá depender do tipo de variação presente nos dados espectrais. Os principais métodos utilizados em tratamentos espectrais são método de normalização padronizada de sinal (SNV) ou o método de correção multiplicativa de sinal (MSC). A correção multiplicativa de sinal (MSC) é usada para remover efeitos físicos, como tamanho de partícula e brilho de superfície dos espectros, que não colaboram com nenhuma informação química ou física, corrigindo diferenças na linha de base e na tendência (MARTENS et al., 1983). A normalização padronizada de sinal (SNV) visa remover os efeitos multiplicativos de dispersão e tamanho de partícula, dando à amostra um desvio padrão unitário (BARNES et al., 1989). O SNV tem sua correção baseada na normalização de cada espectro com seu respectivo desvio padrão e o MSC corrige o efeito de espalhamento pela linearização de cada espectro em função de um espectro de referência.

Além desses tratamentos temos também derivadas de primeira e segunda ordem. O grupo de derivação espectral é representado por duas técnicas: derivadas de Norris-Williams (NW) e filtros de derivadas polinomiais de Savitzky-Golay (SG). Ambos os métodos usam uma suavização dos espectros antes de calcular a derivada, a fim de diminuir o efeito prejudicial na razão sinal-ruído que as derivadas podem apresentar. Quando aplicamos as operações de derivação aos espectros brutos, as informações contidas ao longo dos diferentes comprimentos de onda são acentuadas e se tem uma aparente melhora na resolução espectral. Isso pode contribuir para resolver problemas com bandas sobrepostas, destacando pequenas variações espectrais não evidentes nos dados brutos. Em dados multivariados, comumente utilizam-se derivadas de primeira e segunda ordem. A primeira derivada de Savitzky-Golay é usada para melhorar a resolução do espectro eliminando a linha de base constante, enquanto a segunda derivada elimina a linha de base e a tendência linear (SAVITZKY & GOLAY, 1964).

3.5.2 Calibração

Calibração é a construção de um modelo de predição, sendo a parte mais importante no desenvolvimento de uma equação no NIRS. Sendo este, um modelo de análise multivariada que permite a predição da composição química da amostra com base em dados espectrais. Vários passos estão envolvidos no processo de calibração: Aquisição de dados espectrais, utilização de métodos de referência apropriados para determinar a concentração do que foi analisado em

várias amostras, pré-tratamento de dados do espectro para reduzir os efeitos da dispersão do tamanho de partículas e melhorar o espectro, uso da quimiometria para relacionar os espectros com as amostras analisadas e finalmente validar os modelos usando outro conjunto de amostras não usadas no conjunto de calibração (CEN & HE, 2007).

A calibração multivariada é utilizada para a análise sem a presença de seletividade e confiabilidade das análises realizadas em laboratório. Diversos modelos de calibração multivariada podem ser construídos. Entre os modelos de regressão, os mais aplicados são: modelos de regressão por componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLSR).

De todas as técnicas multivariadas o método PLSR é o mais aplicado no processo de calibração multivariada. Esse método desenvolve uma relação linear entre os dados espectrais (matriz \mathbf{X}) e as propriedades químicas de interesse (matriz \mathbf{Y}). O PLS é uma técnica de redução de dimensão utilizada quando o número de variáveis preditoras é maior do que o número de observações, assim como os dados espectrais, bem como quando existe forte colinearidade entre as variáveis. Este método maximiza a covariância entre as variáveis preditoras e a variável resposta selecionando um pequeno conjunto de componentes (variáveis latentes), que são usados para prever fenótipos alvo em um novo conjunto de dados (MARTENS & JENSEN, 1983). Para avaliar uma calibração é preciso observar parâmetros de medidas de tendência central, dispersão e qualidade dos modelos. É importante observar a identificação de amostras anômalas, através de medidas que indicam o quanto à amostra está distante do centro do modelo ou dos dados mais próximos. Entre os recursos utilizados para melhorar a compreensão desses dados está o erro padrão de predição que é definido como um desvio padrão dos resíduos, de acordo com a equação. Outro medidor de erro é a raiz quadrada do erro quadrático médio de predição (MARTENS & NAES, 2002).

3.5.3 Validação

Dados de espectrometria de infravermelho foram indicados como uma importante fonte de informação para gerar muitos novos fenótipos complexos em gado leiteiro e de corte (CHAPMAN et al., 2019). Para prever os fenótipos de interesse, modelos robustos devem ser desenvolvidos usando um conjunto de dados que representa a heterogeneidade da população. A qualidade de predição dos modelos desenvolvidos se dá através do seu desempenho quando implementados em um conjunto de dados de validação, sendo o ideal que esse conjunto de dados seja externo e não utilizado na calibração. No entanto, definir o conjunto de dados de validação externa é uma tarefa difícil em alguns casos, pois a relação entre os conjuntos deve

ser mínima. A necessidade de se ter uma relação mínima é visando à qualidade de predição desses modelos quando implementados na população real, onde essa relação é mínima. Em experimentos de pesquisa é possível utilizar essa validação quando se utiliza vários rebanhos em que um rebanho ou ensaio é retirado do conjunto de dados para validar o modelo. No entanto, em algumas situações, é difícil definir um conjunto de validação externa principalmente quando os conjuntos de dados são pequenos.

Dado a este fato, geralmente é realizada a validação interna. Para criar o conjunto de dados de calibração e validação interna, as principais técnicas de divisão de dados são a validação cruzada *holdout*, validação cruzada *leave-one-out* e validação cruzada *k-fold*. Na estratégia de validação cruzada *holdout*, como exemplo, podemos selecionar aleatoriamente uma porcentagem do conjunto de dados para calibração e utilizar o restante na validação (STONE, 1974). Na validação cruzada *leave-one-out*, outra abordagem de validação, considera-se o conjunto de dados total é feita a retirada de uma amostra, que pode ser animal, rebanho ou unidade experimental, para validação. Essa divisão é realizada N vezes até que todas as observações tenham sido usadas na validação. O problema deste tipo de validação é que o conjunto de dados de calibração contém quase o mesmo número de observações que o conjunto de dados inteiro (HASTIE et al., 2009).

A validação cruzada *k-fold* é realizada dividindo aleatoriamente o banco de dados inteiro em conjuntos diferentes de tamanhos iguais. Onde um grupo é retirado para validação. Esse processo é repetido até que todos os conjuntos tenham sido usados na validação (STONE, 1974). A desvantagem dos três procedimentos adotados para dividir o conjunto de dados, exceto para a estratégia de *leave-one-out*, é que animais do mesmo rebanho ou vários registros do mesmo animal estarão presentes no conjunto de calibração e validação, criando dependência entre eles.

3.5.4 Utilização do NIRS fecal na produção animal

Cerca de 70% da variação na digestibilidade e ingestão entre animais são explicadas através de uma quantidade diversificada de componentes químicos encontrados nas fezes (HOLLOWAY et al., 1981). A aplicação foi avaliada como um meio de prever a digestibilidade (GARNERSWORTHY & UNAL, 2004; TRAN et al., 2010), qualidade da dieta (JANCEWICZ et al., 2017; JOHNSON et al., 2017) e consumo voluntário (TRAN et al., 2010; JOHNSON et al., 2017) em bovinos de corte e leite.

O NIRS fecal oferece muitas vantagens sobre os métodos mais tradicionais usados para prever as características e a ingestão da dieta, pois fornece uma análise não destrutiva e de baixo

custo. A tecnologia do NIRS também pode ser usada para estimar a variação entre animais na composição da dieta, digestibilidade e consumo. Existem muitas relações possíveis de serem utilizadas entre o espectro fecal e a dieta do animal, desde determinações precisas dos componentes da dieta a uma análise indireta das fezes para determinar as características dietéticas (DIXON & COATES, 2009). Um exemplo de utilização do NIRS de grande interesse dentro da área de nutrição de ruminantes é o monitoramento nutricional de animais em pastejo através da análise dos espectros das fezes (LANDAU et al., 2006).

Lyons & Stuth (1992) foram os pioneiros na utilização do NIRS para avaliar a qualidade da forragem ingerida pelos animais a partir dos espectros das fezes. Leite et al. (2002) utilizaram NIRS fecal para determinar a variação sazonal na qualidade da dieta e o balanço nutricional de ovelhas ao longo do ano em pastagem nativa com diferentes níveis de manejo. Alguns estudos apresentaram resultados satisfatórios com o uso de NIRS fecal para predição da ingestão de dieta. Boval et al. (2004) relataram que o consumo voluntário de matéria orgânica de bovinos poderia ser predito por NIRS fecal apresentando baixo valor de erro padrão na validação. Fanchone et al. (2007) e Decruyenaere et al. (2009) mostraram que o consumo de matéria seca (CVMS) poderia ser predito com sucesso por NIRS fecal com baixo valor de erro na calibração (8 a 12%). Os autores ainda concluíram que o CVMS foi mais acurado quando obtido pelo NIRS fecal em relação a amostras de forragem da dieta.

De igual modo, Garnsworthy & Unal (2004) relataram que a predição do consumo por NIRS fecal mostrou maior precisão do que o método indireto por alcanos. Tran et al. (2010) concluíram em seu estudo que o NIRS fecal forneceu boa predição do CMS total, com $R^2 = 0,77$ e RMSEC abaixo de 8%, tanto nas etapas de calibração e validação. Johnson et al. (2017), encontrou correlações significativas entre as características da dieta e os espectros fecais. A precisão preditiva neste estudo foi maior na predição de PB, FDN e DMS do que na predição de CMS. Entretanto, a capacidade do NIRS fecal de predizer o CMS individual de grupo foi semelhante ao encontrado pela técnica de n-alcano.

Estudos com animais confinados recebendo TMR, só foram investigados utilizando a amostra da dieta e não amostras das fezes dos animais que a consumiram. Mentik et al. (2006) realizou um importante trabalho com NIRS de dietas completas, que teve como objetivo explorar a utilidade do NIRS para predizer o teor de nutrientes em combinação com medidas de desempenho animal, como digestibilidade in vitro da matéria orgânica e da FDN e determinação das frações de proteína bruta. A capacidade do NIRS de predizer PB solúvel, frações A, B, e C da proteína, digestibilidade da FDN e cinzas teve pouco sucesso, apresentando coeficientes de determinação inferiores a 0,77. O NIRS teve boa precisão na predição de

nutrientes básicos, como PB, FDN, amido, CNF e gordura. Foi discutido pelos autores que os resultados foram influenciados pela complexidade dos métodos de referência que podem atenuar a ocorrência de erro, ou até mesmo falta de padronização metodológica, relacionado também a heterogeneidade da dieta e também ao banco de dados reduzido que foi utilizado.

3.6 Métodos alternativos de calibração

O método de mínimos quadrados parciais (PLSR) é atualmente o procedimento mais utilizado para desenvolvimento de equações NIRS em relação a outros métodos (BRESOLINI & DÓREA et al., 2020). A forma simples de execução e o fato de considerar as variabilidades existentes na matriz X e Y, permite que ele seja bem estabelecido para análise de dados NIRS (CENTNER et al., 2000). No entanto, apesar do Método PLSR apresentar boas predições de algumas variáveis complexas, em outras apresentou baixa qualidade de predição. Isso se deve ao fato da existência de relações não lineares em alguns espectros e o uso de regressões lineares simples podem ocasionar problemas de colinearidade (SOYEURT, 2020). O PLSR tem um fraco potencial para avaliação de relações não lineares, na maioria das vezes leva a um sobre ajuste do modelo desenvolvido (THISSEN et al., 2004). Dado a isso, surge à necessidade de estudos avaliando outros métodos de análise de dados capazes de solucionar os problemas descritos e melhorar a capacidade preditiva dos modelos.

Alguns algoritmos de *machine learning* como *Support Vector Machine* (SVR) e a rede neural artificial (RNA) têm capacidade de considerar relações não lineares em seus modelos (THISSEN et al., 2004). Esses algoritmos foram testados como alternativa ao PLS, devido a sua capacidade de identificar variáveis preditoras de maior influência e a capacidade de autoaprendizagem dessas técnicas. Alguns autores que usaram essas abordagens obtiveram uma melhor qualidade de predição quando comparado ao PLS (NIE et al., 2018; PRALLE et al., 2018; DÓREA et al., 2018). Os métodos estatísticos dessas técnicas são diferentes, o que pode conferir diferença em suas capacidades preditivas.

O método SVR, aplica um limite de classificação entre os dados e reduz a distância entre eles, definindo um limite de detecção. O SVR permite a seleção de um conjunto reduzido de amostras, chamados vetores de suporte, que determina a melhor correlação entre os dados espectrais e dados de referência (THISSEN et al., 2004). O SVR é superior em evitar sobre ajustes, pois apresentam melhor desempenho no tratamento de dados não lineares e problemas multidimensionais, principalmente quando se tem um número de amostras reduzido (NIE et al., 2008). Pralle et al. (2018) utilizou espectros de leite para avaliar o potencial de mensuração da

concentração de beta-hidroxibutirato no sangue de vacas leiteiras, onde constatou superioridade na qualidade de predição quando utilizando SVR em relação ao PLS.

A rede neural artificial (RNA) é uma técnica de aprendizagem que se espelha na atividade cerebral neural, aprendendo assim por experiência. A rede é composta por várias camadas, sendo uma de entrada e outra de saída e número determinado de camadas ocultas, que são definidas pelo analista (SOYEURT et al., 2020). A RNA determina os pesos de cada relação entre as unidades pelo método de retro propagação. Nos últimos 20 anos, várias pesquisas foram realizadas com RNA em sistemas de produção de leite. E mais recentemente utilizando este algoritmo principalmente em dados MIR (infravermelho médio, usado comumente em amostra de leite) e em menor proporção NIR, como exemplo, para predição do consumo de alimentos por vacas leiteiras (DÓREA et al., 2018) e a eficiência de utilização de nitrogênio (EUN) em vacas através de espectros de amostras de leite (GRELET et al., 2020).

O *gradient boosting regressor* (GBR) é um algoritmo eficiente da técnica de árvore de decisão, originalmente proposto pelo Dr. Chen da Universidade de Washington. Este método tem atraído muita atenção devido à sua eficiência superior e alta precisão de predição. O *Boosting* abrange uma série de procedimentos de modelagem que tentam combinar a saída de vários modelos "fracos" em um único modelo forte. Em geral, um modelo fraco é sequencialmente ajustado a versões ligeiramente modificadas dos dados. Em modelos com *boosting*, cada árvore é ajustada aos resíduos do modelo anterior por regressão. Liu et al. (2016), usaram espectros NIRS combinado com o *gradient boosting* para avaliação quantitativa do solo.

No entanto, a qualidade da predição é influenciada por muitos fatores, incluindo a característica a ser predita (por exemplo, qualitativa ou quantitativa), a qualidade do conjunto de dados de referência (dados observados), a qualidade dos espectros, o pré-processamento dos espectros, o tamanho da amostra usada para desenvolver as equações de predição e as estratégias de validação usadas para o desenvolvimento e validação do modelo (BONFATTI et al., 2017; DÓREA et al., 2018) Além disso, as características preditas diretamente (por exemplo, gordura do leite) geralmente têm um sinal significativo nos dados espectrais, enquanto as características preditas indiretamente (por exemplo, eficiência alimentar e emissão de metano), o sinal no espectro está associado a características complexas (FERRAGINA et al., 2015). O número de estudos de pesquisa implementando métodos alternativos para prever características complexas em sistemas pecuários é pequeno. Portanto, investigações usando uma abordagem analítica diferente, tamanho de amostra, pré-tratamento de dados e seleção de variáveis são necessárias para lançar luz sobre a análise preditiva de fenótipos complexos.

4 Materiais e Métodos

As amostras e dados utilizados no desenvolvimento das curvas de calibração são provenientes de experimentos previamente conduzidos no Centro de Pesquisa Better Nature, localizado em Ijaci, Minas Gerais, Brasil. Ao todo, foram utilizadas 234 amostras de fezes, de 64 vacas leiteiras diferentes, provenientes de 5 experimentos (Tabela 2). Todos os protocolos de realização desses experimentos foram aprovados pelo Comitê de Ética em Pesquisa com Animais da Universidade Federal de Lavras. Todos os estudos foram com vacas em lactação, alojadas em instalação tipo *tie stall* alimentadas com TMR, com protocolos experimentais padronizados entre experimentos. A composição da TMR era semelhante entre os experimentos (Tabela 1). A TMR foi preparada 2 vezes por dia em um misturador vertical estacionário vertical (Unimix 1200, Casale, São Carlos, Brasil) e as vacas tiveram acesso a nova alimentação às 07:00 e 13:00 h. As vacas foram ordenhadas 3 vezes por dia em ordenha tipo espinha de peixe a partir de 05:00, 13:00 e 20:00 h com produção de leite registrada diariamente.

Tabela 1. Composição das dietas dos experimentos em ingredientes (% da matéria seca).

	Junqueira et al.	Carneiro et al.	Barbosa et al.	Ribeiro et al.
Silagem de Milho	31,2	30,2	29,2	45,0
Silagem de Sorgo	15,6	14,7	20,7	-
Feno de Aveia	-	3,7	2,0	4,0
Feno de Tifton	2,2	-	-	-
Grão de milho reidratado	19,1	18,7	15,0	14,0
Caroço de algodão	11,3	8,9	10,2	8,3
Polpa citrica	-	-	4,0	8,1
Farelo de Soja	10,7	20,4	14,5	13,5
Soypass	-	-	-	4,0
DDG**	6,5	-	-	-
Premix	3,4	2,9	3,6	3,1

* Na tabela estão apresentados os dados dos experimentos que continham as informações de composição da dieta

** Grãos secos de destilaria

As amostras de alimentos foram coletadas diariamente e amostras compostas foram feitas por semana. Da mesma forma, as amostras de sobras foram coletadas diariamente no cocho e também compostas por vaca por semana. As fezes foram coletadas em baldes (coleta total) e pesadas concomitantemente à defecação durante três períodos de amostragem de 8 horas, representando as 24 horas do dia. Alíquotas fecais foram congeladas imediatamente ao longo do período de coleta e foi formada uma amostra composta por vaca. Os consumos individuais foram avaliados registrando a quantidade de alimento oferecido e sobras.

A concentração de MS foi determinada por secagem a 105° C por 24 h (AOAC, 1990; method 934.01). As amostras moídas foram analisadas para proteína bruta (AOAC, 2012) e amido foi determinado enzimaticamente de acordo com Hall (2009). A fibra em detergente neutro (FDN) foi analisada por filtração em cadinhos porosos com alfa-amilase estável ao calor e sulfito de sódio (VAN SOEST et al., 1991). A composição da dieta consumida foi calculada de acordo com as diferenças em quantidades e composição química da dieta ofertada e das sobras individuais. As digestibilidades aparente do trato total da MS, FDN, PB e amido foram determinadas pela relação entre o consumido e o excretado nas fezes.

O comportamento de seleção de alimentos foi avaliado de acordo com Leonardi & Armentano (2003). A proporção de partículas acima da peneira de 19 mm de diâmetro e acima e abaixo da peneira de 8 mm do Penn State Particle Separator foi avaliada para a TMR oferecida. O índice de seleção foi: $100 \times (\text{consumo observado} / \text{consumo estimado})$. Valores de classificação abaixo de 100% representam recusa seletiva, acima de 100% representam ingestão preferencial e igual a 100% representam não seleção.

Amostras fecais individuais de cada animal foram submetidas à leitura via NIRS. Antes da leitura, as amostras ficaram mantidas em um ambiente climatizado (20 °C e umidade relativa de aproximadamente 65%) durante 24 h para atingirem um equilíbrio de umidade de aproximadamente 6%. Os espectros foram gravados no modo de refletância difusa usando um espectrômetro de transformação NIR de Fourier (MPA, Bruker Optik GmbH, Ettlingen, Alemanha) e armazenados usando o software OPUS Spectroscopy versão 7.5. As análises espectrais foram realizadas dentro da região espectral de 4000 – 12500 cm^{-1} , com resolução de 8 cm^{-1} (700 - 2500 nm, 2 nm). Os espectros de cada amostra foram obtidos pela média de 16 análises realizadas duas vezes em recipientes de amostra de lente de quartzo. A primeira leitura foi realizada, a amostra foi submetida a uma nova mistura e assim feita uma nova leitura, ou seja, 2×16 digitalizações por amostra.

Tabela 2. Descrição das variáveis de interesse dos experimentos utilizados no banco de dados para desenvolvimento das curvas de calibração a partir dos espectros de NIRS fecal.

Variáveis	Junqueira et al.	Carneiro et al.	Barbosa et al.	Dias et al.	Ribeiro et al.	Estatística Descritiva					
	n	25	30	59	62	58	N	MIN	MAX	MÉDIA	DP
Leite, Kg/d	25,83	27,5	32,24	29,25	30,57		233	16,14	48,6	29,74	5,57
CMS, kg/d ¹											
<i>Composição da dieta consumida</i>	19,47	20,5	25,03	20,77	23,33		234	10,08	36,31	22,29	4,06
PB, % ²	0,17	0,18	0,17	0,17	0,15		230	1,82	6,34	3,71	0,66
FDN, % ³	0,34	0,3	-	0,29	0,37		175	2,9	12,54	7	1,73
Amido, %	0,27	0,21	-	0,25	0,34		175	2,15	11,15	5,94	1,68
<i>Digestibilidade dos nutrientes</i>											
MS, %	68,13	69,1	69,74	66,51	68,95		231	34,91	84,14	68,47	6,5
PB, %	-	-	77	-	-		59	65,13	86,81	77	4,15
FDN, %	-	53,3	-	49,0	-		92	12,72	77,01	48,49	10,91
Amido	-	93,6	-	90,1	-		92	80,96	98,32	91,25	3,28
<i>Eficiências</i>											
EUN, % ⁴	-	-	23,55	24,6	-		88	17,5	32,45	23,93	3,41
EA, Kg Leite/Kg MS ⁵	1,3	1,35	1,31	1,4	1,34		233	0,66	2,32	1,35	0,26
<i>Seleção de Partículas⁶</i>											
19 mm, %	106,33	-	63,48	82,575	-		141	11,49	187,99	79,23	31,53
8 mm, %	97,08	-	108,23	97,825	-		141	19,63	146,93	101,7	12,22
Fundo, %	101,03	-	118,5	110,06	-		141	31,59	214,23	111,83	18,8

¹Consumo de matéria seca ²Proteína bruta ³Fibra em detergente neutro ⁴Eficiência de utilização de nitrogênio ⁵Eficiência alimentar ⁶Classificação ou seleção calculada como a ingestão real de cada peneira expressa como uma porcentagem da ingestão estimada. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção.

Inicialmente, o desenvolvimento das curvas de calibração a partir dos dados espectrais foi realizado com o software The Unscrambler® (CAMO AS, Oslo, Noruega, v. 9.7), utilizando o método de regressão dos quadrados mínimos parciais (PLSR). Antes da calibração, dados foram tratados matematicamente, visando eliminar ruídos nos espectros que não carregavam informações úteis. Foram feitas as transformações matemáticas SNV que é método de normalização padronizada de sinal que visa remover efeitos multiplicativos de dispersão e tamanho de partícula e o MSC método de correção multiplicativa de sinal usada para remover efeitos físicos, como tamanho de partícula e brilho de superfície dos espectros, ambos também foram testados acrescidos de derivadas de primeira e segunda ordem (1.4.4.1, 2.4.4.2, 2.5.5.2), a primeira é usada para melhorar a resolução do espectro eliminando a linha de base constante, enquanto a segunda derivada elimina a linha de base e a tendência linear. Realizamos o teste de incerteza de Mertens do software The Unscrambler para identificação de amostras outlier antes do processo de calibração. Os modelos foram desenvolvidos com o espectro NIRS como variáveis independentes (matriz X) e os valores observados de CMS, composição da dieta consumida (%PB, %FDN, %Amido), digestibilidade dos nutrientes (PB, FDN, Amido) e seleção de partículas (19 mm, 8 mm e fundo), como variáveis dependentes (matriz Y).

A validação dos modelos foi feita pela técnica de validação cruzada, removendo aleatoriamente parte dos dados (15%) do conjunto de treinamento e utilizando para validar. A segmentação do conjunto de dados (entre conjunto de treinamento e conjunto de validação) era repetida até que todas as amostras tenham servido como amostras de validação. O número ótimo de variáveis latentes em cada modelo foi determinado através da maximização dos coeficientes de determinação e minimizando erros padrões de calibração e de validação (SHENK & WESTERHAUS, 1991a).

A seleção dos melhores modelos gerados pelo software The Unscrambler foi baseada nos seguintes critérios: coeficiente de determinação para calibração (R^2C) e validação cruzada (R^2VC); raiz quadrada do erro médio obtido na calibração (RMSEC) e validação cruzada (RMSEVC). O RMSEC representa a variabilidade na diferença entre valores estimados e valores de referência quando a equação for desenvolvida a partir do conjunto de dados de calibração. Além disso, outros dois parâmetros foram calculados para avaliar os modelos. O primeiro é a relação entre desempenho do modelo e variabilidade no banco de dados (ratio of performance to deviation, RPD), calculado como divisão entre o desvio padrão no banco de dados original e o RMSEVC. Um modelo é considerado excelente quando valor de RPD é $> 2,5$ (WILLIAMS, 2004).

A análise de PLSR com pré-tratamentos matemáticos dos dados é a metodologia mais tradicional de processamento de dados de espectros NIRS (BRESOLIN AND DÓREA et al, 2020). No entanto, consideramos provável que métodos lineares não fossem capazes de incorporar adequadamente a complexidade de interações existentes nas variáveis respostas avaliadas neste estudo. Por isso, optamos por analisar os dados com outras ferramentas de *machine learning*, com e sem os pré-tratamentos matemáticos (apenas SNV e MSC), e testar duas estratégias de validação dos modelos.

Para essa segunda avaliação, 3 bancos de dados foram gerados: 1) os dados originais, sem tratamentos matemáticos; 2) os dados tratados pelo método SNV; e 3) os dados tratados pelo método MSC. Além disso, a produção de leite das vacas foi adicionada no banco de dados de variáveis preditoras e foi incluída nos modelos para CMS (sem nenhuma transformação matemática). Os três bancos de dados foram utilizados para as implementações de 4 algoritmos no software Python, da biblioteca Scikit-learn (<https://scikit-learn.org/>): PLSR, *Support Vector Machine Regressor* (SVR), *Gradient Boosting Regressor* (GBR), e *K-Nearest Neighbors Regressor* (KNNR). O PLSR foi novamente utilizado para que a comparação com os outros fosse feita na mesma base (Python), uma vez que no Unscrambler nem todos os parâmetros do modelo eram ajustáveis. O *Support Vector Machine* e o *K-Nearest Neighbors* são ferramentas mais robustas para trabalhar com dados em que as relações entre matriz X e Y são não lineares. Embora amplamente utilizados para modelos de classificação, ambos podem ser utilizados para regressão (SVR e KNNR). Já o *gradient boosting* (GB) é uma técnica de árvore de decisão que tem atraído atenção recente devido à sua eficiência superior e alta precisão. O GB abrange uma série de procedimentos de modelagem que tentam combinar a saída de vários modelos "fracos" em um único modelo forte. Em geral, um modelo fraco é sequencialmente ajustado a versões ligeiramente modificadas dos dados, ou seja, cada árvore é ajustada aos resíduos do modelo anterior por regressão. Essa técnica tem mostrado bons resultados e ainda não foi testada para previsões a partir de dados de NIRS fecal.

As técnicas de *machine learning* em sua maioria dependem do processo de hiperparametrização para um melhor desempenho do modelo. Os hiperparâmetros são variáveis que controlam o processo de treinamento. Por exemplo, faz parte da configuração de uma rede neural profunda decidir quantas camadas ocultas de nós precisam ser usadas entre a camada de entrada e a camada de saída, número de vetores de suporte no algoritmo SVR ou a distância delimitada entre os pontos no algoritmo KNNR. Essas variáveis não estão diretamente relacionadas aos dados de treinamento, elas são variáveis de configuração. Cada algoritmo

apresenta suas configurações de hiperparâmetros, neste estudo foram utilizadas as configurações fixas da biblioteca Scikit-learn (<https://scikit-learn.org/>) para cada algoritmo.

Duas estratégias de validação foram utilizadas no desenvolvimento dos modelos: *leave-one-animal-out* (LOAO) and *leave-one-experiment-out* (LOEO). Para LOAO, todos os dados de uma mesma vaca foram removidos por vez do banco de dados de treinamento e utilizados para validação. O procedimento foi repetido até que todas as vacas tenham sido utilizadas para validação. Esta estratégia foi utilizada para avaliar a performance preditiva dos algoritmos quando aplicados em dados de novos animais. Já para o LOEO, todos os dados de um mesmo experimento foram removidos por vez do banco de dados de treinamento e utilizados para validação. O procedimento foi repetido até que todos os experimentos tenham sido utilizados para validação. O objetivo desta estratégia foi avaliar o desempenho preditivo dos algoritmos quando dados gerados em um novo cenário (i.e. outra época do ano, outra dieta, outro manejo). O RMSEVC foi utilizado para avaliação da acurácia desses modelos e os valores reportados para cada modelo em cada estratégia é o valor médio de todas as rodadas de validação.

5 Resultados

O melhor modelo para cada variável dependente da análise de PLSR do software Unscrambler estão apresentados na Tabela 3. As variáveis dependentes que não estão apresentadas na tabela (CMS, digestibilidade da MS, FDN e amido e seleção de partículas 19mm, 8mm e fundo) apresentaram coeficientes de determinação tão baixos ($R^2 < 0,5$) que não foram considerados importantes na análise. O tratamento matemático que gerou os melhores modelos foi o SNV com derivada 2.5.5.2 para as três variáveis de composição da dieta consumida e digestibilidade do amido e SNV com derivada 1.4.4.1 para digestibilidade de PB, EUN e EA. Apesar dos coeficientes de determinação para calibração para essas variáveis terem se apresentado satisfatórios, os valores para a validação cruzada foram menos promissores. Apenas para digestibilidade do FDN o modelo apresentou R^2VC maior do que 0,8. No entanto, nem mesmo para essa variável o valor de RPD foi maior do que 2,5, mostrando que essa ferramenta de análise de dados não foi eficaz para predição das variáveis desejadas a partir de espectros de NIRS fecal.

Tabela 3. Avaliação dos modelos desenvolvidos por regressão dos quadrados mínimos parciais para prever consumo de matéria seca, composição da dieta consumida, digestibilidade de nutrientes, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal.

Variáveis	Calibração			Validação Cruzada		
	Tratamento	RMSEC	R ²	RMSCV	R ²	RPD
CMS, kg/d ¹	-	-	-	-	-	-
<i>Composição da dieta consumida</i>						
PB, % ²	SNV _{2.5.5.2} [*]	1,4	0,7	1,8	0,6	0,1
FDN, % ³	SNV _{2.5.5.2}	6,7	0,7	7,2	0,5	0,2
Amido, %	SNV _{2.5.5.2}	5,9	0,8	6,3	0,6	0,3
<i>Digestibilidade dos nutrientes</i>						
MS, %	-	-	-	-	-	-
PB, %	-	-	-	-	-	-
FDN, %	SNV _{1.4.4.1} ^{**}	1,2	0,9	1,7	0,8	2,4
Amido	SNV _{2.5.5.2}	1,8	0,7	2,2	0,6	1,5
<i>Eficiências</i>						
EUN, % ⁴	SNV _{1.4.4.1}	1,6	0,8	2,7	0,3	1,3
EA, Kg Leite/Kg MS ⁵	SNV _{1.4.4.1}	1,8	0,6	2,4	0,3	0,1
<i>Seleção de Partículas⁶</i>						
19 mm, %	-	-	-	-	-	-
8 mm, %	-	-	-	-	-	-
Fundo, %	-	-	-	-	-	-

¹Consumo de matéria seca ²Proteína bruta ³Fibra em detergente neutro ⁴Eficiência de utilização de Nitrogênio ⁵Eficiência alimentar ⁶Classificação ou seleção calculada como a ingestão real de cada peneira expressa como uma porcentagem da ingestão prevista. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção. * Normalização padronizada de sinal com derivada de segunda ordem com cinco pontos de suavização. **Normalização padronizada de sinal com derivada de primeira ordem com 4 pontos de suavização.

As tabelas 4 e 5 apresentam a acurácia (RMSE) dos modelos gerados pelos quatro algoritmos avaliados para todas as variáveis dependentes, na estratégia de validação LOEO (Tabela 4) e LOAO (Tabela 5). O objetivo dessas duas tabelas é a avaliação qualitativa do desempenho global das 4 ferramentas de análise e, por isso, os valores de RMSE reportados são a média aritmética entre os valores obtidos nas 3 formas de pré-tratamento dos dados (nenhum, SNV e MSC). A variável digestibilidade de PB foi avaliada apenas no experimento 3, dado a este fato o parâmetro só poderá ser avaliado na estratégia de validação LOAO.

Tabela 4. Avaliação do desempenho (RMSEVC) dos quatro algoritmos para predição de consumo de matéria seca, composição da dieta consumida, digestibilidade, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal com estratégia de validação leave-one-experiment-out.

Variáveis	Algoritmos ⁷			
	GBR	KNNR	PLSR	SVR
CMS, kg/d ¹	3,64	5,12	3,88	4,54
<i>Composição da dieta consumida</i>				
PB, % ²	0,02	0,01	0,01	0,01
FDN, % ³	0,05	0,05	0,06	0,04
Amido, %MS	0,07	0,08	0,08	0,05
<i>Digestibilidade dos nutrientes</i>				
MS, %	8,69	7,10	6,66	6,23
PB, %	-	-	-	-
FDN, %	14,92	13,77	15,90	12,51
Amido, %	5,08	5,46	5,06	5,09
<i>Eficiências</i>				
EUN, % ⁴	4,27	5,12	4,82	4,35
EA, Kg Leite/Kg MS ⁵	0,26	0,32	0,25	0,26
<i>Seleção de Partículas⁶</i>				
19 mm, %	42,07	39,78	39,52	40,62
8 mm, %	16,75	14,94	12,86	13,89
Fundo, %	20,99	22,59	18,52	21,24

¹Consumo de matéria seca ²Proteína bruta ³Fibra em detergente neutro ⁴Eficiência de utilização de nitrogênio ⁵Eficiência alimentar ⁶Classificação ou seleção calculada como a ingestão real de cada peneira expressa como uma porcentagem da ingestão estimada. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção.

⁷GBR=*gradient boosting regressor*; KNNR=*k-nearest neighbors regressor*; SVR=*Support vector machine*; PLSR=método dos mínimos quadrados parciais.

As duas tabelas apresentam a pequena superioridade dos algoritmos GBR e do PLSR em relação aos outros algoritmos, já que apresentou RMSEVC levemente menor do que os outros em todas as variáveis de interesse e em ambas as estratégias de validação. Em geral, a diferença entre os algoritmos não foi tão evidente. O RMSE é uma medida de erro, ou seja, mede a diferença direta entre os valores preditos e observados. Em técnicas de *Machine Learning* em que queremos prever a variável resposta diretamente, é comum optar por métricas de erro, como RMSE, por exemplo, e sempre levando em conta a escala da variável resposta para avaliar se o erro é considerado baixo ou alto. Outra observação interessante é que a acurácia dos modelos foi maior para a estratégia de validação LOAO, indicando que a predição seria melhor quando aplicada em um novo animal do que quando aplicada em um novo cenário (condições diferentes). Vale ressaltar que os valores de RMSEVC apresentados

nesta tabela não devem ser interpretados de forma quantitativa pois eles foram gerados com a média dos três tratamentos matemáticos dos dados. Valores muito altos de RMSECV podem ser gerados quando por modelos que não funcionaram com algum dos tratamentos, e isso inflaciona o valor médio.

Tabela 5. Avaliação do desempenho (RMSEVC) dos quatro algoritmos para predição de consumo de matéria seca, composição da dieta consumida, digestibilidade, eficiências de uso de nutrientes e seleção de partículas a partir de espectros de NIRS fecal com estratégia de validação leave-one-animal-out.

Variáveis	Algoritmos ⁷			
	GBR	KNNR	PLSR	SVR
CMS, kg/d ¹	2,98	3,11	3,35	3,37
<i>Composição da dieta consumida</i>				
PB, %MS ²	0,01	0,00	0,01	0,01
FDN, %MS ³	0,01	0,01	0,03	0,04
Amido, %MS	0,01	0,01	0,04	0,04
<i>Digestibilidade dos nutrientes</i>				
MS, %	6,49	5,84	5,89	5,82
PB, %	4,00	3,98	3,79	3,52
FDN, %	10,90	9,99	9,49	8,98
Amido, %	2,49	2,44	2,53	2,43
<i>Eficiências</i>				
EUN, % ⁴	2,88	2,90	2,66	2,85
EA, Kg Leite/Kg MS ⁵	0,26	0,24	0,23	0,23
<i>Seleção de Partículas⁶</i>				
19 mm, %	26,95	27,39	27,01	27,38
8 mm, %	10,94	9,79	8,94	9,52
Fundo, %	16,56	15,78	15,60	16,04

¹Consumo de matéria seca ²Proteína bruta ³Fibra em detergente neutro ⁴Eficiência de utilização de nitrogênio ⁵Eficiência alimentar ⁶Classificação ou seleção calculada como a ingestão real de cada peneira expressa como uma porcentagem da ingestão estimada. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção. ⁷GBR=*gradient boosting regressor*; KNNR=*k-nearest neighbors regressor*; SVR=*Support vector machine*; PLSR=método dos mínimos quadrados parciais.

Depois da avaliação geral, os melhores algoritmos para cada modelo preditor preditor dentre todos avaliados (4 ferramentas de análise e 3 tratamentos de dados) para cada variável dependente foi então selecionado de acordo com os valores de acurácia (RMSEVC) para cada estratégia de validação. As Tabelas 6, 7, 8 e 9 descrevem os modelos selecionados para consumo e composição da dieta consumida (Tabela 6), digestibilidade dos nutrientes (Tabela 7), seleção de partículas (Tabela 8) e eficiências de uso de nutrientes (Tabela 9).

Tabela 6. Descrição dos melhores modelos preditores para consumo de matéria seca e composição da dieta consumida a partir de espectros de NIRS fecal nas estratégias de validação leave-one-experiment-out e leave-one-animal-out.

Variáveis	Algoritimos	Tratamento	RMSEVC
LOEO			
Composição da dieta consumida			
CMS, kg/d ¹	GBR	nenhum	3,64
PB, %MS ²	PLSR	nenhum	0,01
FDN, %MS ³	SVR	nenhum	0,04
Amido, %MS	SVR	nenhum	0,05
LOAO			
Composição da dieta consumida			
CMS, kg/d	GBR	nenhum	2,98
PB, %MS	GBR	nenhum	0,01
FDN, %MS	GBR	nenhum	0,01
Amido, %MS	GBR	nenhum	0,01

¹Consumo de matéria seca ²Proteína bruta ³Fibra em detergente neutro LOEO=deixar um experimento de fora; LOAO=deixar um animal de fora; GBR=*gradiente boosting regressor*; SVR=*Support vector machine regressor*; PLSR=método dos mínimos quadrados parciais; SNV=normalização padronizada de sinal.

Os modelos para consumo de matéria seca e a composição da dieta consumida apresentaram maior acurácia quando foi utilizado o algoritimo GBR e a estratégia de validação LOAO. Para essas variáveis, o tratamento matemático do banco de dados melhorou muito pouco a predição. Vale ressaltar que as diferenças encontradas entre os algoritimos foram minimas, variando em casas decimais de uma variável para outra.

Tabela 7. Descrição dos melhores modelos preditores para digestibilidade dos nutrientes a partir de espectros de NIRS fecal nas estratégias de validação leave-one-experiment-out e leave-one-animal-out.

Variáveis	Algoritmos	Tratamento	RMSEVC
LOEO			
Digestibilidade dos nutrientes			
MS, % ¹	SVR	SNV	6,23
PB, % ²	-	-	-
FDN, % ³	SVR	nenhum	12,26
Amido	GBR	nenhum	4,52
LOAO			
Digestibilidade dos nutrientes			
MS, %	SVR	nenhum	5,82
PB, %	SVR	nenhum	3,52
FDN, %	SVR	nenhum	8,71
Amido	SVR	nenhum	2,43

¹Matéria seca ²Proteína bruta ³Fibra em detergente neutro LOEO=deixar um experimento de fora; LOAO=deixar um animal de fora; GBR=*gradiente boosting regressor*; SVR=*Support vector machine regressor*; SNV=normalização padronizada de sinal.

As variáveis de digestibilidade dos nutrientes foram melhor preditas com o algoritmo SVR. Além disso, os modelos com LOAO foram mais precisos do que com LOEO. A predição da digestibilidade da matéria seca na estratégia de validação LOEO foi levemente melhorada com o tratamento SNV em relação à utilização dos dados sem tratamento. Para digestibilidade da matéria seca, o RMSEVC dos dados sem tratamento foi 6,24% para LOEO.

Tabela 8. Descrição dos melhores modelos preditores para eficiências a partir de espectros de NIRS fecal nas estratégias de validação leave-one-experiment-out e leave-one-animal-out.

Variáveis	Algoritmos	Tratamento	RMSEVC
LOEO			
Seleção de Partículas ¹			
19 mm, %	PLSR	nenhum	39,10
8 mm, %	PLSR	SNV	12,59
Fundo, %	PLSR	SNV	18,52
LOAO			
Seleção de Partículas			
19 mm, %	GBR	SNV	26,34
8 mm, %	PLSR	SNV	8,85
Fundo, %	PLSR	SNV	15,60

¹Classificação ou seleção calculada como a ingestão real de cada peneira expressa como uma porcentagem da ingestão prevista. Valores abaixo de 100% indicam recusas seletivas, acima de 100% é consumo preferencial e igual a 100% não houve seleção. LOEO=deixar um experimento de fora; LOAO=deixar um animal de fora; GBR=*gradiente boosting regressor*; PLSR=método dos mínimos quadrados parciais; SNV=normalização padronizada de sinal.

Os algoritmo GBR e PLSR foram os que tiveram melhor desempenho para a predição da seleção de partículas estimada pelo método das peneiras. Esta variável foi mais dependente

de algum tratamento matemático, sendo a melhor estratégia de validação LOAO, e as três variáveis tiveram melhor desempenho com o tratamento SNV. A diferença entre tratado e não tratado foi em torno de 0,20 a 0,40 % para todas as variáveis. Levando em consideração a escala dessa variável resposta, os valores de acurácia para esses modelos foram encorajadores nas duas estratégias de validação.

Tabela 9. Descrição dos melhores modelos preditores para seleção de partículas espectros de NIRS fecal nas estratégias de validação leave-one-experiment-out e leave-one-animal-out.

Variáveis	Algoritimos	Tratamento	RMSEVC
LOEO			
Eficiências			
EUN, % ¹	GBR	SNV	3,66
EA, Kg Leite/Kg MS ²	PLSR	nenhum	0,24
LOAO			
Eficiências			
EUN, %	PLSR	nenhum	2,65
EA, Kg Leite/Kg MS	PLSR	nenhum	0,23

¹Eficiência de utilização de Nitrogênio ²Eficiência alimentar LOEO=deixar um experimento de fora; LOAO=deixar um animal de fora; GBR=*gradiente boosting regressor*; PLSR=método dos mínimos quadrados parciais; MSC=correção mutiplicativa de sinal; SNV=normalização padronizada de sinal.

Os parâmetros de eficiência foram melhor preditos com o algoritimo PLSR sem nenhum tratamento. A estratégia de validação LOAO também obteve melhor desempenho para estas variáveis. A diferença encontrada entre os algoritimos para esta variável foi de 0,20% para EUN e 0,02 kg leite/kg MS para EA.

6 Discussão

O objetivo do nosso estudo foi encontrar métodos menos trabalhosos e invasivos para prever variáveis complexas e de extrema importância dentro do sistema de produção leiteiro como consumo, digestibilidade, seleção de partículas e eficiências. O NIRS é um método que apresenta vantagens em relação às análises convencionais, pois fornece uma análise não destrutiva, não invasiva e de baixo custo (SKOOG et al., 2006). Nossa hipótese de que os espectros NIRS seriam capazes de prever variáveis nutricionais parte do princípio de que as informações químicas contidas nas fezes são representativas da dieta consumida e estão relacionadas ao consumo e digestibilidade (HOLLOWAY et al. 1981). Os estudos com NIRS fecal tem demonstrado resultados promissores, no entanto, a maioria destes estudos analisaram composição química das fezes. Johnson et al. (2017) encontrou melhores resultados quando

analisou PB e FDN de fezes. Outro exemplo de utilização do NIRS de grande interesse dentro da área de nutrição de ruminantes é o monitoramento do consumo e composição de nutrientes de animais em pastejo através da análise dos espectros das fezes (LANDAU et al., 2006).

Um trabalho recente realizado em nosso grupo de pesquisa mostrou resultados promissores na predição da composição da dieta consumida e digestibilidade de nutrientes a partir de NIRS fecal com animais consumindo exclusivamente gramíneas tropicais frescas (Oliveira, 2019). No entanto, essa abordagem nunca foi conduzida para animais consumindo TMR. Em uma TMR, as complexidades das ligações orgânicas relacionadas com um processo digestivo são mais complexas do que ligações orgânicas associadas a moléculas mais simples, como proteína ou amido (MENTIK et al., 2006). Neste caso, por ser um alimento bastante heterogêneo, as relações entre a absorvância e o nutriente de interesse se tornam mais desafiadoras, configurando a existência de relações não lineares em alguns espectros. Além disso, as variáveis que nos propusemos a estudar não dependem somente da composição química do alimento ou das fezes, mas são fenótipos que sofrem interferência do comportamento e do metabolismo animal.

Em situações como essas, modelos lineares podem apresentar problemas de colinearidade (SOYEURT, 2020). O PLSR, método mais utilizado em análises de espectros NIRS, tem um fraco potencial para avaliação de relações não lineares, na maioria das vezes leva a um sobre ajuste do modelo desenvolvido (THISSEN et al., 2004). Algoritmos de aprendizagem de máquina como SVR, KNNR e GBR têm capacidade de considerar relações não lineares em seus modelos. Estudos com os algoritmos SVR e com redes neurais artificiais obtiveram uma melhor qualidade de predição quando comparado ao PLSR (PRALLE et al., 2018; DÓREA et al., 2018).

Em nosso estudo o PLSR realizado pelo software The unscrambler não demonstrou bons resultados, nem mesmo quando realizamos tratamentos dos espectros e/ou retiramos amostras outlier. Embora outros estudos utilizando espectros de NIRS fecal tenham reportado bons resultados com PLSR para predizer composição nutricional e consumo (TRAN et al., 2010; JOHNSON et al., 2017), nenhum destes avaliaram variáveis tão complexas como a desse estudo ou mesmo quando avaliaram, estas tiveram pior desempenho quando comparadas a variáveis mais simples como de composição química.

No entanto, o PLSR realizado no software Python apresentou melhores resultados. O intuito de utilizar este algoritmo no Python foi de permitir a comparação entre essas técnicas de regressão já que o software The unscrambler não nos permite tanto controle de parâmetros em análise como o Python. Os algoritmos aqui utilizados, foram selecionados pois reportaram bons

resultados em estudos anteriores, quando comparados ao PLSR (SEXTON et al., 2018). No entanto, em nosso estudo, não foram encontradas grandes diferenças entre os algoritmos, assim como em alguns trabalhos anteriores. Embora os resultados não sejam diretamente comparáveis, eles indicam que os problemas de regressão são específicos de cada aplicação. Isso explica o comportamento diferente de cada variável resposta a um determinado tipo de algoritmo. A influência da relação entre o componente e espectro, parece determinar maior ou menor diferença entre os algoritmos, principalmente devido a linearidade dessas relações. No entanto, as técnicas de *machine learning* muitas vezes necessitam de uma busca pelos melhores hiperparâmetros, antes de treinar os algoritmos para realizar a predição final, o que pode interferir no desempenho dos modelos e conferir maior ou menor diferença entre eles (BENGIO, 2012; DÓREA, 2020).

Técnicas de machine learning, são empregadas devido à sua capacidade de pesquisar em um espaço de alta dimensão, características que melhor descrevem a variável resposta, com capacidade de autoaprendizagem (GIANOLA et al., 2011). Diferente do esperado neste estudo, não identificamos grandes diferenças entre os diferentes algoritmos. Alguns autores avaliando a composição de direta da cana de açúcar também não encontraram diferenças entre os algoritmos, mostrando que o melhor algoritmos também é dependente da natureza da variável estudada. No presente estudo, apesar dos bons resultados, não foi realizada uma investigação profunda sobre os melhores hiperparâmetros para os algoritmos utilizados. Os hiperparâmetros são variáveis que controlam o processo de treinamento. A busca pelos melhores hiperparâmetros confere melhor acurácia aos modelos além de evidenciar as diferenças de desempenho de cada um.

O CMS é o parâmetro que mais influencia o desempenho animal. Por isso, é essencial que se tenha métodos para estimar esse parâmetro de forma confiável. Estudos anteriores com NIRS apresentaram resultados promissores para predição de CMS, no entanto estes trabalhos foram feitos com animais a pasto (TRAN et al., 2010). Nossos modelos apresentaram bom desempenho avaliando esta variável, sendo um dado singular e promissor em análises de NIRS fecal com animais consumido TMR. Dados de produção de leite foram adicionados ao modelo, como na maioria dos modelos para predição de consumo. Produção de leite é o principal fator determinante da necessidade de nutrientes da vaca (NRC, 2001). Outros modelos preditivos reportados na literatura apresentaram melhora significativa quando este dado foi adicionado na equação, o que é esperado devido a relação entre consumo e produção (DÓREA et al., 2017).

O dado de consumo de matéria seca e composição da dieta consumo é de extrema importância dentro de um sistema de nutrição de precisão, tendo em vista que desempenho dos

ruminantes está intimamente relacionado à ingestão, sendo que cerca de 60 a 90% do desempenho dos animais pode ser explicado por variações no consumo de nutrientes (MERTENS, 1994). O modelo do NRC 2001 para CMS apresenta valores de RMSE entre 1,75 e 1,78 Kg/d, em nosso estudo, obtivemos média de RMSE de 2,98 Kg/d na estratégia de validação LOAO, sem considerar outros fatores como peso corporal e DEL que são utilizados no modelos NRC2001 e sem uma investigação profunda sobre hiperparâmetros. Assim como no NRC 2001, a produção de leite foi incluída no modelo, a inclusão de produção de leite em modelos de consumo confere melhoria na predição, uma vez que estas variáveis estão relacionadas (DÓREA et al., 2020). Modelos de predição são mais eficazes quando as condições do rebanho onde será aplicado se assemelham as condições da população em que foram criados, ou seja, quando em populações diferentes a estimativa pode conter alguma variação (MARTENS, 1987). Portanto, os modelos gerados nesse estudo precisam ser validados em outras condições de validações externas para verificar sua acurácia e aplicabilidade em outras populações diferentes.

O objetivo principal das técnicas de avaliação de alimentos é prever a disponibilidade de nutrientes em sistemas de produção animal. A estimativa de digestibilidade de nutrientes a partir do espectro NIRS das amostras de alimentos é bastante utilizada em laboratórios comerciais, principalmente para FDN e amido. A digestibilidade e consumo estão intimamente relacionados e determinam a utilização da energia dietética pelos animais. Decruyenaere et al. (2009) demonstrou que o NIRS fecal mostrou ser uma ferramenta promissora para análise de digestibilidade de outros componentes como FDN e PB, porém estes trabalhos foram com animais a pasto. Há um interesse crescente nas calibrações com NIRS fecal para valores de digestibilidade in vivo, uma vez que permite a substituição de técnicas de laboratório para prever o valor da energia digestível da dieta (STUTH et al., 2003). Neste trabalho, fizemos calibrações para digestibilidade in vivo para estes parâmetros e alcançamos resultados bons resultados que poderão auxiliar na avaliação de dietas em fazendas leiteiras.

A eficiência alimentar tem papel importante em aspectos econômicos nos sistemas de produção de ruminantes, este dado permite tomadas de decisão em relação ao manejo e também auxiliam em programas de melhoramento genético (BERRY & CROWLEY, 2013). O uso de espectrometria NIR e MIR vem sendo explorado como uma ferramenta potencial para prever características relacionadas à eficiência alimentar em bovinos de leite, como por exemplo o consumo individual. No entanto, a avaliação direta dessas características, principalmente avaliando NIRS fecal não foram investigadas. Melhorar a eficiência do uso de nitrogênio (EUN) e a eficiência alimentar (EA) de forma individual ou em nível de rebanho são essenciais para

sistemas de produção de leite, por razões ambientais e econômicas. Grelet et al., (2020) obteve resultados promissores avaliando EUN a partir de MIR de leite com o algoritmo SVR, obteve R^2 na validação acima de 0,7 com erros abaixo de 5%. Muitos estudos com espectros de leite vêm sendo realizados, principalmente devido a facilidade de obtenção da amostra. No entanto, assim como neste trabalho existe a necessidade de inserir mais dados nos modelos para cobrir um máximo da variabilidade em relação às raças, produção e estágios de lactação.

Os animais ruminantes têm a capacidade de selecionar entre diferentes partículas dos alimentos, em alguns casos, a dieta realmente consumida pelas vacas é maior em carboidratos fermentáveis e menor em fibra efetiva do que a formulada, aumentando assim o risco de acidose ruminal subaguda (DEVRIES et al., 2005). No entanto, este é um dado importante dentro do sistema de produção para avaliar problemas nutricionais e de manejo. O método mais utilizado para mensurar a seleção de partículas é o uso da peneira *Penn State*, de acordo com a porcentagem de cada peneira chegamos à conclusão se está havendo seleção de partículas ou não dentro do rebanho. Avaliar este tipo de parâmetro de forma individual pelo NIRS fecal, permitiria avaliar animais selecionadores e identificar problemas relacionados a nutrição. Até então, não existiam relatos de estudos que abordaram calibração NIRS para prever essas variáveis. Os resultados obtidos no nosso estudo se mostraram promissores, com a peneira de 19mm conseguimos avaliar seleção de partículas longas, com isso conseguimos perceber a proporção de fibra fisicamente efetiva ingerida. Em contrapartida, na peneira do fundo permanecem as menores partículas da dieta, que devem ter consumo monitorado para avaliar possíveis efeitos na função ruminal e consumo. A TMR normalmente contém 40 a 60% de concentrado, a maior parte do qual passa pela peneira de 8,0 mm, fazendo com que seja uma porção bastante variável e heterogênea, no entanto em nosso estudo obtivemos uma boa acurácia na predição dessa porção da dieta, assim como as outras. Levando em consideração a escala de valores dessas variáveis reposta.

A estratégia de validação *leave-one-out* é um dos métodos de validação cruzada comumente aplicada quando o tamanho da amostra é pequeno e há preocupação sobre o tamanho limitado do conjunto de calibração (GIANOLA & SCHÖN, 2016). A ideia da técnica de validação cruzada de deixar um de fora também pode ser adaptado para deixar um rebanho de fora (por exemplo, grupo, fazenda, ensaio, ano, entre outros), o que pode ser uma boa estratégia para reduzir a interdependência biológica entre os dados de treinamento e validação quando esses conjuntos são definidos aleatoriamente (BRESOLIN & DÓREA, 2020). No entanto, o ideal é utilizar a validação externa, quando se tem disponível um grande tamanho de

amostra. Como exemplo, pode ser definido como conjunto de validação externa, dados de outra fazenda, rebanho, região ou lote.

As estratégias de validação utilizadas, tiveram como base experimentos anteriores do nosso grupo de pesquisa, onde foram comparadas estratégias como LOEO e LOAO com estratégias onde são selecionados dados aleatórios para conjunto de validação. Neste tipo de estratégia, pode ocorrer interdependência entre dados do mesmo animal ou do mesmo experimento, estratégia esta que utilizamos no software The unscrambler. Podemos notar que a estratégia LOAO obteve menores erros, ou seja, o modelo teria melhor desempenho para prever as variáveis em estudo, em animais externos ao conjunto de dados de calibração. Quando avaliado em condições novas, ou seja, em outro experimento como LOEO, observamos uma queda na acurácia desses modelos. Pois, apesar de um experimento completo ter sido retirado, todos foram feitos sob condições semelhantes, fato este que foi necessário para mantermos a padronização dos métodos de referência.

Embora nossos modelos precisem ser validados para uso em condições de rotina, esses resultados preliminares mostraram que é possível obter informações fenotípicas complexas por meio de espectros NIR de fezes. Isso poderia permitir predições em grande escala para auxiliar estudos genéticos e o desenvolvimento de ferramentas de gestão agrícola. Portanto, conjuntos de dados maiores e abordagens modernas de mineração de dados devem ser investigadas para melhorar a capacidade preditiva e para confirmar a qualidade dos modelos de calibração. Apesar de um número aceitável e representativo de amostras utilizados nesse experimento, algumas variáveis estudadas não foram analisadas em todos os experimentos, reduzindo substancialmente o número de amostras dessas variáveis.

Deve-se levar em consideração que apesar de unidades experimentais diferentes, o ambiente, as condições experimentais e alguns animais eram semelhantes e/ou iguais entre eles, o que pode ser vantajoso pela redução de erros relacionados aos métodos de referência. Em contrapartida, é preciso cautela quanto a acurácia obtida na validação cruzada neste estudo, uma vez que a ocorrência de *overfitting* é comum nesse tipo de validação (HASTIE et al., 2009). Embora essas técnicas de aprendizado de máquina sejam ideais para implementação em espaços multidimensionais, esta abordagem tende a se ajustar facilmente, principalmente devido aos conjuntos de dados reduzidos e dados ruidosos (HEMPSTALK et al., 2015). O sobre ajuste (*overfitting*) é um problema recorrente nesses métodos e costuma ser identificado pela alta precisão de predição no conjunto de dados de treinamento, mas muito pobre no conjunto de dados de validação. Dado a isso, o método de validação se torna uma estratégia importante

no processo, devendo ser mais o imparcial possível. Para tal, é necessária maior independência possível dos dados entre o treinamento e a validação (ROBERTS et al., 2017).

7 Conclusão

A análise de dados espectrais recebe influência de diversos fatores externos e internos que irão determinar qual o melhor método a ser utilizado para se obter resultados mais acurados em análise NIRS. Embora os resultados obtidos sejam promissores e as estratégias de validação visem minimizar o emaranhamento biológico dentro de uma vaca ou ensaio, o conjunto de dados veio inteiramente do mesmo rebanho. Assim, é possível que o modelo seja menos preciso quando usado em um rebanho e ambientes completamente diferentes.

Espectros de amostras fecais de vacas leiteiras alimentadas com TMR, podem ser utilizados como preditores de variáveis complexas de rotina na fazenda, desde que seja avaliado e utilizado o algoritmo adequado. A perspectiva é que em novas pesquisas, estes algoritmos sejam explorados, buscando melhor hiperparametrização dos algoritmos, em conjunto de dados maiores e com validação externa para verificar a aplicabilidade desses modelos.

8 Referências Bibliográficas

- AGNEW, R. E. et al. Potential of near infrared spectroscopy to predict the voluntary intake of grazed grass. **Animal Feed Science and Technology**, v. 115, n. 1-2, p. 169-178, 2004.
- AGRICULTURAL AND FOOD RESEARCH COUNCIL (AFRC). Energy and protein requirement of ruminant. Wallingford, UK. CAB international, p.159, 1993.
- AGUERRE, Matias J. et al. Effect of forage-to-concentrate ratio in dairy cow diets on emission of methane, carbon dioxide, and ammonia, lactation performance, and manure excretion. **Journal of dairy science**, v. 94, n. 6, p. 3081-3093, 2011.
- ALLEN, Michael S. et al. Grouping to increase milk yield and decrease feed costs. In: **20th Annual Tri-State Dairy Nutrition Conference**. Ohio State University Press Columbus, 2009.
- ANDRÉ, Geert et al. Increasing economic profit of dairy production utilizing individual real time process data. **Precision livestock farming**, v. 7, p. 179-186, 2007.
- AOAC. 1990. Official methods of analysis. 15th ed. Association of Official Analysis Chemists, Arlington, VA.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. **Applied Spectroscopy**, v. 43, p. 772-777, 1989.
- BATTEN, G. D. Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. **Australian Journal of Experimental Agriculture**, v. 38, n. 7, p. 697-706, 1998.
- BERRY, D. P.; CROWLEY, J. J. Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. **Journal of animal science**, v. 91, n. 4, p. 1594-1613, 2013.
- BENGIO, Y. "Practical recommendations for gradient-based training of deep architectures," in **Neural Networks: Tricks of the Trade**, eds G. Montavon, G. B. Orr, and K.-R. Müller (Heidelberg: Springer), 437-478. 2012.
- BONFATTI, V. et al. Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations. **Journal of dairy science**, v. 100, n. 9, p. 7306-7319, 2017.
- BOVAL et al. Faecal near infrared reflectance spectroscopy (NIRS) to assess chemical composition, in vivo digestibility and intake of tropical grass by Creole cattle. **Animal Feed Science and Technology**, v. 114, n. 1-4, p. 19-29, 2004.
- BRESOLIN, T.; DÓREA, J. R. R. Infrared Spectrometry as a High-Throughput Phenotyping Technology to Predict Complex Traits in Livestock Systems. **Frontiers in Genetics**, v. 11, 2020.
- CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72-83, 2007.

CENTNER, Vitezslav et al. Comparison of multivariate calibration techniques applied to experimental NIR data sets. **Applied spectroscopy**, v. 54, n. 4, p. 608-623, 2000.

CHAPMAN, James et al. Shining light into meat—a review on the recent advances in in vivo and carcass applications of near infrared spectroscopy. **International Journal of Food Science & Technology**, v. 55, n. 3, p. 935-941, 2020.

CHIZZOTTI, M. L. et al. Validation of a system for monitoring individual feeding behavior and individual feed intake in dairy cattle. **Journal of dairy science**, v. 98, n. 5, p. 3438-3442, 2015.

COATES, David B.; DIXON, Rob M. Development of near infrared analysis of faeces to estimate non-grass proportions in diets selected by cattle grazing tropical pastures. **Journal of Near Infrared Spectroscopy**, v. 16, n. 5, p. 471-480, 2007.

COMBS, Dan. TTNDFD: A new approach to evaluate forages. In: **Cornell Nutrition Conference** (<https://hdl.handle.net/1813/36476>), 2013.

COPPOCK, C. E.; BATH, D. L.; HARRIS JR, B. From feeding to feeding systems. **Journal of dairy science**, v. 64, n. 6, p. 1230-1249, 1981.

CORONA, L. et al. Comparative effects of whole, ground, dry-rolled, and steam-flaked corn on digestion and growth performance in feedlot cattle. **The Professional Animal Scientist**, v. 21, n. 3, p. 200-206, 2005.

DECRUYENAERE, V. et al. Evaluation of green forage intake and digestibility in ruminants using near infrared reflectance spectroscopy (NIRS): Developing a global calibration. **Animal Feed Science and Technology**, v. 148, n. 2-4, p. 138-156, 2009.

DE MARCHI, M. et al. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. **Journal of Dairy Science**, v. 97, n. 3, p. 1171-1186, 2014.

DEVRIES, T. J.; VON KEYSERLINGK, M. A. G. Time of feed delivery affects the feeding and lying patterns of dairy cows. **Journal of dairy science**, v. 88, n. 2, p. 625-631, 2005.

DIXON, R.; COATES, D. Near infrared spectroscopy of faeces to evaluate the nutrition and physiology of herbivores. **Journal of Near Infrared Spectroscopy**, v. 17, n. 1, p. 1-31, 2009.

DIJKSTRA, J. et al. Relationships between methane production and milk fatty acid profiles in dairy cattle. **Animal Feed Science and Technology**, v. 166, p. 590-595, 2011.

DÓREA, J. R. R.; FRENCH, E. A.; ARMENTANO, L. E. Use of milk fatty acids to estimate plasma nonesterified fatty acid concentrations as an indicator of animal energy balance. **Journal of dairy science**, v. 100, n. 8, p. 6164-6176, 2017.

DÓREA, J. R. R. et al. Mining data from milk infrared spectroscopy to improve feed intake predictions in lactating dairy cows. **Journal of dairy science**, v. 101, n. 7, p. 5878-5889, 2018.

FAHEY JR, G. C.; JUNG, H. G. Lignin as a marker in digestion studies: a review. **Journal of Animal Science**, v. 57, n. 1, p. 220-225, 1983.

FANCHONE, Audrey et al. Faecal indices based on near infrared spectroscopy to assess intake, in vivo digestibility and chemical composition of the herbage ingested by sheep (crude

protein, fibres and lignin content). **Journal of Near Infrared Spectroscopy**, v. 15, n. 2, p. 107-113, 2007.

FERRAGINA, A. et al. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. **Journal of dairy science**, v. 98, n. 11, p. 8133-8151, 2015.

FIRKINS, J. L. et al. Effects of grain variability and processing on starch utilization by lactating dairy cattle. **Journal of Animal Science**, v. 79, n. suppl_E, p. E218-E238, 2001.

FOX, D. G. et al. The Cornell Net Carbohydrate and Protein System model for evaluating herd nutrition and nutrient excretion. **Animal Feed Science and Technology**, v. 112, n. 1-4, p. 29-78, 2004.

FREDIN, S. M. et al. Fecal starch as an indicator of total-tract starch digestibility by lactating dairy cows. **Journal of dairy science**, v. 97, n. 3, p. 1862-1871, 2014.

GARNSWORTHY, P. C.; UNAL, Y. Estimation of dry-matter intake and digestibility in group-fed dairy cows using near infrared reflectance spectroscopy. **ANIMAL SCIENCE-GLASGOW THEN PENICUIK-**, v. 79, n. 2, p. 327-334, 2004.

GEHMAN, A. M. Enhanced nitrogen utilisation in dairy cattle with precision protein nutrition. **Recent Advances Animal Nutrition**, v. 18, p. 187-195, 2011.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. **Applied spectroscopy**, v. 39, n. 3, p. 491-500, 1985.

GIANOLA, Daniel et al. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC genetics**, v. 12, n. 1, p. 1-14, 2011.

GIANOLA, D., AND SCHÖN, C.-C. Cross-validation without doing cross-validation in genome-enabled prediction. *G3* 6, 3107–3128. 2016.

GRELET, C. et al. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. **Journal of dairy science**, v. 103, n. 5, p. 4435-4445, 2020.

GOESER, J. P.; COMBS, D. K. An alternative method to assess 24-h ruminal in vitro neutral detergent fiber digestibility. **Journal of dairy science**, v. 92, n. 8, p. 3833-3841, 2009.

GORDON, I. J. Animal-based techniques for grazing ecology research. **Small Ruminant Research**, v. 16, n. 3, p. 203-214, 1995.

HALL, M. B. Determination of starch, including maltooligosaccharides, in animal feeds: Comparison of methods and a method recommended for AOAC collaborative study. **Journal of AOAC International**, v. 92, n. 1, p. 42-49, 2009.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning, 2nd edn New York. **NY: Springer**, p. 587-603, 2009.

HEMPSTALK, K.; MCPARLAND, Sinead; BERRY, D. P. Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. **Journal of dairy science**, v. 98, n. 8, p. 5262-5273, 2015.

- HOLLOWAY, J. W.; ESTELL, R. E.; BUTTS JR, W. T. Relationship between fecal components and forage consumption and digestibility. **Journal of Animal Science**, v. 52, n. 4, p. 836-848, 1981.
- JANCEWICZ, L. J. et al. Predictability of growth performance in feedlot cattle using fecal near-infrared spectroscopy. **Canadian Journal of Animal Science**, v. 97, n. 4, p. 701-720, 2017.
- JOHNSON, J. R. et al. Application of fecal near-infrared reflectance spectroscopy profiling for the prediction of diet nutritional characteristics and voluntary intake in beef cattle. **Journal of animal science**, v. 95, n. 1, p. 447-454, 2017.
- KAVANAGH, S. et al. A comparison of total collection and marker technique for the measurement of apparent digestibility of diets for growing pigs. **Animal feed science and technology**, v. 89, n. 1-2, p. 49-58, 2001.
- LAMMERS, B. P.; BUCKMASTER, D. R.; HEINRICHS, A. J. A simple method for the analysis of particle sizes of forage and total mixed rations. **Journal of dairy science**, v. 79, n. 5, p. 922-928, 1996.
- LANDAU, S.; GLASSER, T.; DVASH, L. Monitoring nutrition in small ruminants with the aid of near infrared reflectance spectroscopy (NIRS) technology: a review. **Small Ruminant Research**, v. 61, n. 1, p. 1-11, 2006.
- LEITE, E. R.; CESAR, M. F.; DE ARAUJO FILHO, J. A. Efeitos do melhoramento da caatinga sobre os balanços protéico e energético na dieta de ovinos. **Embrapa Caprinos e Ovinos-Artigo em periódico indexado (ALICE)**, 2002.
- LE LIBOUX, S.; PEYRAUD, Jean-Louis. Effect of forage particle size and feeding frequency on fermentation patterns and sites and extent of digestion in dairy cows fed mixed diets. **Animal feed science and technology**, v. 76, n. 3-4, p. 297-319, 1999.
- LEONARDI, C.; ARMENTANO, L. E. Effect of quantity, quality, and length of alfalfa hay on selective consumption by dairy cows. **Journal of Dairy Science**, v. 86, n. 2, p. 557-564, 2003.
- LIU et al. Quantitative retrieval of organic soil properties from visible near-infrared shortwave infrared (Vis-NIR-SWIR) spectroscopy using fractal-based feature extraction. **Remote Sensing**, v. 8, n. 12, p. 1035, 2016.
- LOPES, F.; RUH, K.; COMBS, D. K. Validation of an approach to predict total-tract fiber digestibility using a standardized in vitro technique for different diets fed to high-producing dairy cows. **Journal of dairy science**, v. 98, n. 4, p. 2596-2602, 2015.
- LYONS, R. K.; STUTH, J. W. Fecal NIRS equations for predicting diet quality of free-ranging cattle. **45:238-244**. 1992.
- LUYPAERT, J.; MASSART, D. L.; VANDER HEYDEN, Y. Near-infrared spectroscopy applications in pharmaceutical analysis. **Talanta**, v. 72, n. 3, p. 865-883, 2007.
- MACOON, B. et al. Comparison of three techniques for estimating the forage intake of lactating dairy cows on pasture. **Journal of Animal Science**, v. 81, n. 9, p. 2357-2366, 2003.

MANLEY, M., DOWNEY, G., BAETEN, V. Spectroscopic technique: Near-Infrared (NIR) spectroscopy. In: **Modern techniques for food authentication**. Da-Wen Sun 1st ed., Amsterdam; Boston: Elsevier/Academic Press. Pp 65-115. 2008.

MARTENS, Harald; JENSEN, S. A.; GELADI, P. Multivariate linearity transformation for near-infrared reflectance spectrometry. In: **Proceedings of the Nordic symposium on applied statistics**. Stokkand Forlag Publishers Stavanger, Norway, p. 205-234, 1983.

MCGILLIARD, M. L.; SWISHER, J. M.; JAMES, R. E. Grouping lactating cows by nutritional requirements for feeding. **Journal of Dairy Science**, v. 66, n. 5, p. 1084-1093, 1983.

MCMENIMAN, N. P. Methods of estimating intake of grazing animals. **Reunião Anual da Sociedade Brasileira de Zootecnia**, v. 34, p. 131-168, 1997.

MCPARLAND, S.; BERRY, D. P. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. **Journal of Dairy Science**, v. 99, n. 5, p. 4056-4070, 2016.

MENTINK, R. L.; HOFFMAN, P. C.; BAUMAN, L. M. Utility of near-infrared reflectance spectroscopy to predict nutrient composition and in vitro digestibility of total mixed rations. **Journal of Dairy Science**, v. 89, n. 6, p. 2320-2326, 2006

MERTENS D.R. Regulation of forage intake. In: FAHEY, G.C. Jr. et al. (Eds.) **Forage quality, evaluation and utilization**. Madison: American Society of Agronomy, p.450-492. 1994.

MARTENS, D. R. Predicting intake and digestibility using mathematical models of ruminal function. **Journal of animal science**, v. 64, n. 5, p. 1548-1558, 1987.

MARTENS, Harald. Multivariate calibration by data compression. Near-infrared technology in the agricultural and food industries, p. 57-87, 1987.

METHU, J. N. et al. Botanical and nutritional composition of maize stover, intakes and feed selection by dairy cattle. **Livestock Production Science**, v. 71, n. 2-3, p. 87-96, 2001.

MIKUS, J. H. Diet consistency: Using TMR auditsTM to deliver more from your feed, equipment, and people to the bottom line. in **High Plains Dairy Conference Proceedings**, Amarillo, TX. Texas Animal Nutrition Council, Dallas, TX. Pages 27–36, 2012.

MOORE, J. E. Practical approaches to estimating pasture intake. **Nutrient cycling in forage systems**, p. 193, 1996.

NATIONAL RESEARCH COUNCIL – NRC. **Nutrient requirements of dairy cattle**. 7 ed. Washington: Academic Press.v.381. 2001.

NETO, B. B; SCARMINIO, Ieda S.; BRUNS, Roy E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, n. 6, p. 1401-1406, 2006.

NIE, Zhihong et al. An EMD-SVR model for short-term prediction of ship motion using mirror symmetry and SVR algorithms to eliminate EMD boundary effect. **Ocean Engineering**, v. 217, p. 107927, 2020.

- OLIVEIRA, D. M. Predictability of dry matter intake, diet composition and digestibility in beef cattle using fecal near-infrared reflectance spectroscopy. **Dissertação** (Mestrado em Zootecnia) – Universidade Federal de Lavras, Lavras, 2019.
- OSAFO, E. L. K. et al. Effects of amount offered and chopping on intake and selection of sorghum stover by Ethiopian sheep and cattle. **Animal Science**, v. 65, n. 1, p. 55-62, 1997.
- PASQUINI, Celio. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian chemical society**, v. 14, n. 2, p. 198-219, 2003.
- PRALLE, R. S.; WEIGEL, K. W.; WHITE, H. M. Predicting blood β -hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. **Journal of Dairy Science**, v. 101, n. 5, p. 4378-4387, 2018.
- QIAO, T. et al. Quantitative prediction of beef quality using visible and NIR spectroscopy with large data samples under industry conditions. **Journal of Applied Spectroscopy**, v. 82, n. 1, p. 137-144, 2015.
- REEVES III, J.B.; DELWICHE, S.R. SAS®. Partial least squares regression for analysis of spectroscopic data. **Journal Near Infrared Spectroscopy**, v.11, p.415-431. 2003.
- RINNAN, Å.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in **Analytical Chemistry**, v. 28, n. 10, p. 1201-1222, 2009.
- ROBERTS, David R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. **Ecography**, v. 40, n. 8, p. 913-929, 2017.
- SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical chemistry**, v. 36, n. 8, p. 1627-1639, 1964.
- SEXTON et al. A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a sugarcane quality measure. **Journal of Near Infrared Spectroscopy**, v. 26, n. 5, p. 297-310, 2018.
- SHENK, J. S.; WESTERHAUS, M. O. Calibration the ISI way. In: DAVIES, A. M. C.; WILLIAMS, P. (Eds.), **Near Infrared Spectroscopy: The Future Waves**. NIR Publications, West Sussex, UK. P.198202, 1996.
- SHETTY, N. et al. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. **Journal of dairy science**, v. 100, n. 1, p. 253-264, 2017.
- SKOOG.; WEST.; HOLLER.; CROUCH. **Fundamentos de Química Analítica**, Tradução da 8ª Edição norte-americana, Editora Thomson, São Paulo-SP, 2006.
- ST-PIERRE, N. R.; THRAEN, C. S. Animal grouping strategies, sources of variation, and economic factors affecting nutrient balance on dairy farms. **Journal of Animal Science**, v. 77, n. suppl_2, p. 72-83, 1999.
- SPIPKE, Joachim et al. Precision Dairy Farming—integrativer Ansatz für eine nachhaltige Milcherzeugung. **Zeitschrift für Agrarinformatik**, v. 2, n. 03, p. 19-25, 2003.

- SOYEURT et al. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. **Journal of Dairy Science**, v. 103, n. 12, p. 11585-11596, 2020.
- SOVA, A. D. et al. Associations between herd-level feeding management practices, feed sorting, and milk production in freestall dairy farms. **Journal of dairy science**, v. 96, n. 7, p. 4759-4770, 2013.
- STONE, B. Reducing the variation between formulated and consumed rations. In: **Advances in dairy technology: proceedings of the... Western Canadian Dairy Seminar**. 20:145–162, 2008.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 36, n. 2, p. 111-133, 1974.
- STUTH, J.; JAMA, A.; TOLLESON, D. Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. **Field Crops Research**, v. 84, n. 1-2, p. 45-56, 2003.
- TRAN, H. et al. “Global” and “local” predictions of dairy diet nutritional quality using near infrared reflectance spectroscopy. **Journal of dairy science**, v. 93, n. 10, p. 4961-4975, 2010.
- UNDI, M. et al. Comparison of techniques for estimation of forage dry matter intake by grazing beef cattle. **Canadian Journal of Animal Science**, v. 88, n. 4, p. 693-701, 2008.
- VAN AMBURGH, M. E. et al. The Cornell Net Carbohydrate and Protein System: Updates to the model and evaluation of version 6.5. **Journal of Dairy Science**, v. 98, n. 9, p. 6361-6380, 2015.
- VALLIMONT, J. E. et al. Heritability of gross feed efficiency and associations with yield, intake, residual intake, body weight, and body condition score in 11 commercial Pennsylvania tie stalls. **Journal of Dairy Science**, v. 94, n. 4, p. 2108-2113, 2011.
- VAN SOEST, P. V.; ROBERTSON, J.; LEWIS, B. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. **Journal of Dairy Science**, v. 74, n. 10, p. 3583-3597, 1991
- VANDEHAAR, M. J. et al. Considerations in using residual feed intake to define feed efficiency in dairy cattle. 2012.
- VEERKAMP et al. Genome-wide associations for feed utilisation complex in primiparous Holstein–Friesian dairy cows from experimental research herds in four European countries. **Animal**, v. 6, n. 11, p. 1738-1749, 2012.
- VELÁSQUEZ, A. V. et al. Evaluating internal and external markers versus fecal sampling procedure interactions when estimating intake in dairy cows consuming a corn silage-based diet. **Journal of dairy science**, v. 101, n. 7, p. 5890-5901, 2018.
- WEISS, W. P.; CONRAD, H. R.; PIERRE, NR St. A theoretically-based model for predicting total digestible nutrient values of forages and concentrates. **Animal Feed Science and Technology**, v. 39, n. 1-2, p. 95-110, 1992.

WETZEL, D. L.; REFFNER, J. A. Using spatially resolved Fourier transform infrared microbeam spectroscopy to examine the microstructure of wheat kernels. **Cereal Foods World**, v. 38, n. 1, p. 9-20, 1993.

WILLIAMS, P. **Near-infrared technology, getting the best out of light**. A short course in the practical implementation of Near-infrared spectroscopy for the user. PDK Grain, Nanaimo, British Columbia, and Winnipeg, Manitoba, Canada, pp. 12–13, 2004.