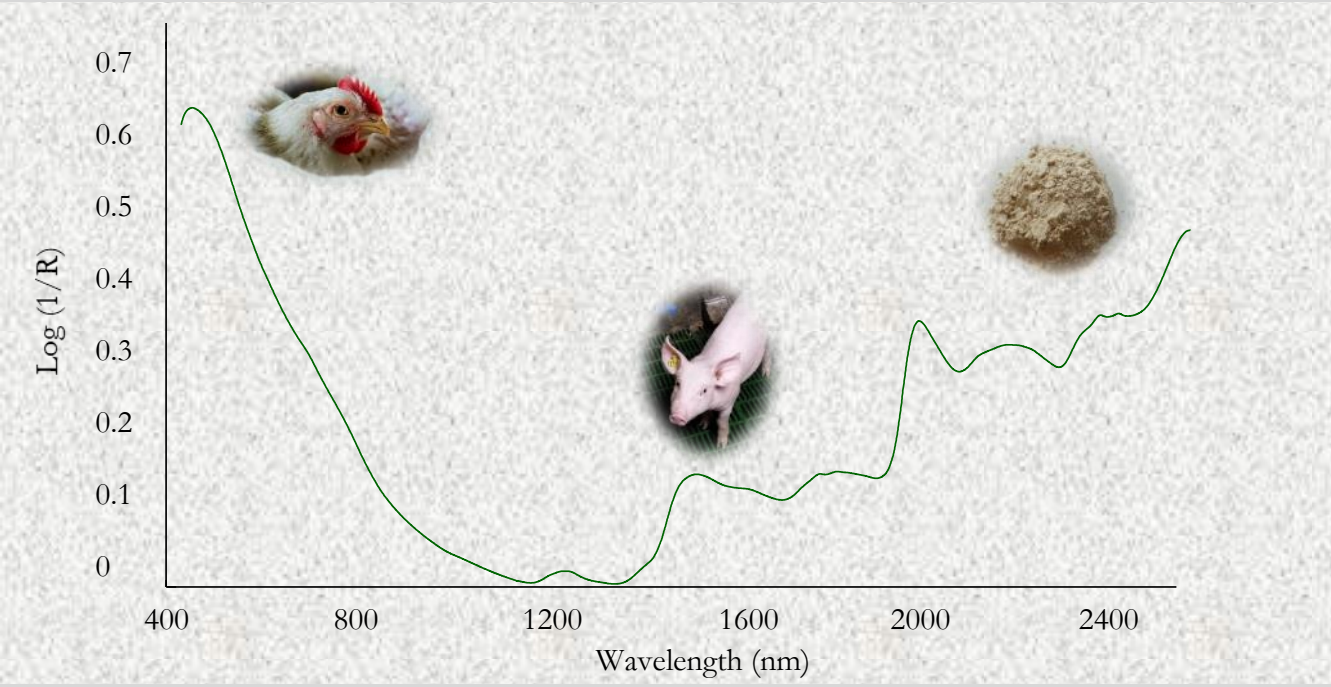
NOVEL APPLICATIONS OF NIR SPECTROSCOPY FOR THE OPTIMIZATION OF MONOGASTRIC ANIMAL DIETS

Andrés Cruz Conesa

UNIVERSITAT ROVIRA i VIRGILI

Novel applications of NIR spectroscopy for the optimization of monogastric animal diets

ANDRÉS CRUZ CONESA



DOCTORAL THESIS
2023

Presentation

This thesis has been the result of a colaboration between the University Rovira i Virgili (URV) and the Institute of Agrifood Research and Technology (IRTA). The research center in Mas Bové (Constantí, Tarragona) is one of the several IRTA centers in Catalonia and includes facilities such as monogastric experimental farms, a feed factory and an analytical laboratory. A key research line at IRTA Mas Bové is carried out by the IRTA-Animal Nutrition (IRTA-AN) program on monogastric nutrition. An important part of the research consists of studying how diets are digested by monogastric animals and how they affect animal productivity, health and waste generation. In order to simplify these studies as much as possible and reduce the laboratory load, IRTA acquired a benchtop NIR instrument (FOSS NIR DS2500) in 2020. The expertise of the Chemometrics, Qualimetrics and Nanosensors group at the URV was applied to develop, apply and validate the NIR multivariate calibration models for the determination of the nutrient content of different types of samples generated during in vivo assays. This thesis involves the use of stored feed, faeces, excreta and ileal digesta contents from research projects that have been carried out at IRTA in the recent past or during the development of this thesis. The analytical reference data from these samples was provided by the analytical laboratory at IRTA. I also did a research stay of three months in Viborg (Denmark) in the Department of Animal and Veterinary Sciences, AU Viborg research centre Foulum. This agri-food research centre belongs to the Aarhus University and has research lines like those of IRTA.

Andrés Cruz Conesa

Novel applications of NIR spectroscopy for the optimization of monogastric animal diets

Doctoral Thesis

Supervised by

Dr. Joan Ferré Baldrich

Dr. Itziar Ruisánchez Capelastegui

Dr. Anna M. Pérez Vendrell

Departament de Química Analítica i Química Orgànica (URV)

Institute of Agrifood Research and Technology (IRTA)

Tarragona

2023

UNIVERSITAT ROVIRA i VIRGILI
Dept. de Química Analítica
i Química Orgànica

IRTA
Institute
of Agrifood Research
and Technology

WE STATE that the present study, entitled "Novel applications of NIR spectroscopy for the optimization of monogastric animal diets", presented by Andrés Cruz Conesa for the award of the degree of Doctor, has been carried out under our supervision at the Department of Analytical Chemistry and Organic Chemistry of the Universitat Rovira i Virgili (URV) and at the Institute of Agrifood Research and Technology (IRTA) in Mas Bové (Constantí, Spain) and that it has the requirements to opt to the International Mention.

Tarragona, June 23rd, 2023

Doctoral Thesis Supervisors

JOAN FERRE BALDRICH - DNI 39876285N
Digitally signed by JOAN FERRE BALDRICH - DNI 39876285N
Date: 2023.06.24 19:38:02 +02'00'

María Iciar Ruisanchez Capelastegui - DNI 35079545Z (TCAT)
Firmado digitalmente por María Iciar Ruisanchez Capelastegui - DNI 35079545Z (TCAT)
Fecha: 2023.06.26 10:21:05 +02'00'

Anna Maria Pérez Vendrell
Firmado digitalmente por Anna Maria Pérez Vendrell
Fecha: 2023.06.24 13:54:30 +02'00'

Dr. Joan Ferré Baldrich Prof. Dr. Itziar Ruisánchez Capelastegui Dr. Anna M. Pérez Vendrell

UNIVERSITAT
ROVIRA i VIRGILI

Aknowledgments

En primer lugar, quiero agradecer a mis directores de tesis Anna Pérez, Joan Ferré, e Itziar Ruisánchez por darme la oportunidad de realizar esta tesis y estar siempre pendientes de mi trabajo, enseñándome y ayudándome en todo lo posible. Durante este tiempo, gracias a ellos, no sólo he aprendido sobre quimiometría, NIR o nutrición animal, también creo que he adquirido una forma de trabajar y de pensar con respecto a la ciencia y a la investigación que espero me ayude mucho en mi porvenir.

Gracias también a Pilar Callao por su disponibilidad, colaboración y apoyo sobre todo en las primeras etapas de esta tesis.

También me gustaría agradecer a Jesús Anzano, profesor de la Universidad de Zaragoza, con quién di mis primeros pasos en investigación y quién me animó a realizar una tesis doctoral.

Todo este trabajo no habría sido posible sin la colaboración de los investigadores del grupo de Nutrición Animal de IRTA Mas Bove (Anna, Rosil, María F, David, Borja, Nuria, Joan, Enric, Joaquim, Maria D) de cuyos estudios he obtenido gran parte de las muestras utilizadas en este trabajo. Agradezco además todos sus consejos, su ayuda y su compañía a lo largo de este periodo. Tampoco habría sido posible sin Josep y el personal del Laboratorio de Nutrición Animal que siempre han estado ahí proporcionándome todo lo que he necesitado.

Agradecer a Lluis y al resto de las personas de la fábrica de piensos por su acogida y su ayuda durante los primeros meses en los que estuve con ellos y a todo el personal de IRTA que siempre me ha hecho sentir a gusto y me han facilitado las cosas en todo lo posible.

Gracias con especial cariño a mi compañero Eudald que ha estado a mi lado en IRTA durante todo el tiempo realizando su tesis y al resto de personas que me habéis hecho mucho más agradable mi día a día: Zacharia, Muzahir, Bea, Helena, Tom, Gerard, Luismi, Boris …

Gracias a Samantha J. Noel y a Knud Eric Bach Knudsen por hacer posible mi estancia en Dinamarca de la que siempre guardaré un recuerdo maravilloso y que me sirvió muchísimo para crecer tanto a nivel de conocimientos como a nivel personal. Gracias con mucho cariño a mis compañeros de despacho y de departamento en Foulum y especialmente a la gente de Norreso.

Gracias a mis amigos de Alcañiz de toda la vida: Samuel, Rixi, Mitor y Helio. A mis amigos del instituto, de Zaragoza, de Tarragona y a todos los de Pitarque que hacéis que Agosto sea el mejor mes del año.

Gracias Sofía por tu cariño y tu apoyo en todo momento.

Por último, mi más profundo agradecimiento a mi familia. A mis padres Amadeo y Elena y a mi hermano Martín. También quiero agradecer especialmente a mis tíos Pili y Fran que siempre han tenido su casa de Tarragona abierta para mí y a mis primos David y Alejandro.

Abbreviations

AA, amino acid

ADF, acid detergent fibre

AID, apparent ileal digestibility

AME, apparent metabolizable energy

ANN, artificial neural networks

ATTD, apparent total tract digestibility

CF, crude fibre

CHO, carbohydrates

CP, crude protein

CV, coefficient of variation

DE, digestible energy

DF, dietary fibre

DM, dry matter

GE, gross energy

GLSW, generalized least squares weighting

HPLC, high performance liquid chromatography

IF, insoluble fibre

LV, latent variables

LOOCV, leave-one-out cross-validation

MSC, multiplicative scatter correction

NCP, non-cellulosic polysaccharides

ME, metabolizable energy

MLR, multiple linear regression

NDF, neutral detergent fibre

NE, net energy

NIRS, near-infrared spectroscopy

NSP, non-starch polysaccharides

OM, organic matter

OS, oligosaccharides

OSC, orthogonal signal correction

P, phosphorus

PCA, principal component analysis

PCR, principal component regression

PLSR, partial least squares regression

R, correlation coefficient

R_c^2, coefficient of determination of calibration

R_{cv}^2, coefficient of determination of cross-validation

R_p^2, coefficient of determination of prediction

RMSEC, root mean square error of calibration

RMSECV, root mean square error of cross-validation

RMSEP, root mean square error of prediction

RSD, relative standard deviation

RPD, ratio of performance of deviation

RS, resistant starch

SD, standard deviation

SEL, standard error of the laboratory

SNV, standard normal variate

TME, true metabolizable energy

Summary

To maximize the productivity of monogastric animals (e.g. poultry and swine) while avoiding nutritional excesses or deficiencies, it is essential to adopt a formulation approach that incorporates not only accurate values of nutrient content and the energy provided by the feed but also the digestibility of the nutrients and the metabolizable or digestible energy. In-vivo assays are the preferred method to obtain information about digestibility, although they are expensive, laborious, and generate a large number of samples (faeces, excreta and/or intestinal contents) for analysis. These samples are traditionally analysed by wet chemistry analytical methods. Near-infrared spectroscopy (NIRS) coupled with chemometrics presents several advantages against wet chemistry, such as rapidness, cleanliness, and the fact that it is a non-destructive, multi-parametric technique. The proven ability of NIRS to determine the nutrient content (protein, fat, fibre, and others) of feedstuffs and animal diets has promoted the routine use of this technique in the agri-food sector. We have confirmed the applicability of NIRS to determine the nutrient content of compound feeds and demonstrated that NIRS is also a reliable tool to characterize the carbohydrate and lignin fraction (gaining importance in today's formulations) of a variety of feedstuffs included in monogastric diets. We have shown that global Vis-NIR calibration models (involving a wide range of experimental variability) predict the nutrient composition of poultry excreta and pig faeces accurately. We have also developed accurate calibration models for protein and most amino acids in poultry ileal digesta. A fast and inexpensive access to the nutrient content of all these samples will help reduce the cost of digestibility studies and improve the search for optimal feeds.

Animal nutrition research involves testing different diets (containing new ingredients, new proportions of ingredients, additives, etc.) on animals that may vary their digestive capacity from one in-vivo assay to another (e.g. different breed or different age). Therefore, new samples from compound feeds, faeces, excreta, or ileal digesta can introduce sources of variability that were not modelled during the calibration development. We have presented two strategies to incorporate these new sources of variability into the models. The first strategy was applied to compound feeds and involved identifying the most interesting samples from their spectra using the model diagnostic measures Hotelling's T^2 and Q residuals. It has been proven that models periodically updated with these samples improve their robustness and predictive ability over time. The second strategy was applied to pig faeces. We have developed a sample selection

algorithm based on D-optimality, which efficiently selects a reduced number of new samples from a large batch that is used to update the model. This way, the remaining samples from the batch can be predicted accurately, and the model incorporates new information that can be useful to predic samples from other batches in the future.

An additional drawback regarding in-vivo assays is the ethical problems related to animal experimentation. Therefore, there is a tendency to simplify or minimize them as much as possible. In the last part of this thesis, we explored different strategies involving NIRS to predict the digestibility of nutrients and the metabolizable or digestible energy with the ultimate aim of replacing in-vivo assays. We achieved varying degrees of success for the studied parameters and strategies, laying the groundwork for future studies in this field.

Table of contents

List of figures

Chapter 2. General introduction

Section 2.1. Monogastric nutrition

Section 2.2. Near-infrared spectroscopy and multivariate analysis

Section 2.3. Near-infrared spectroscopy in monogastric nutrition

Chapter 3. Replacing wet chemistry by near-infrared spectroscopy to improve formulation and reduce the costs of in-vivo assays

Section 3.2. Determination of carbohydrate and lignin content in feedstuffs for monogastric animals using near-infrared spectroscopy

Section 3.3. Development of near-infrared calibration models for the nutrient content of poultry and swine compound feeds

List of tables

Chapter 2. General introduction

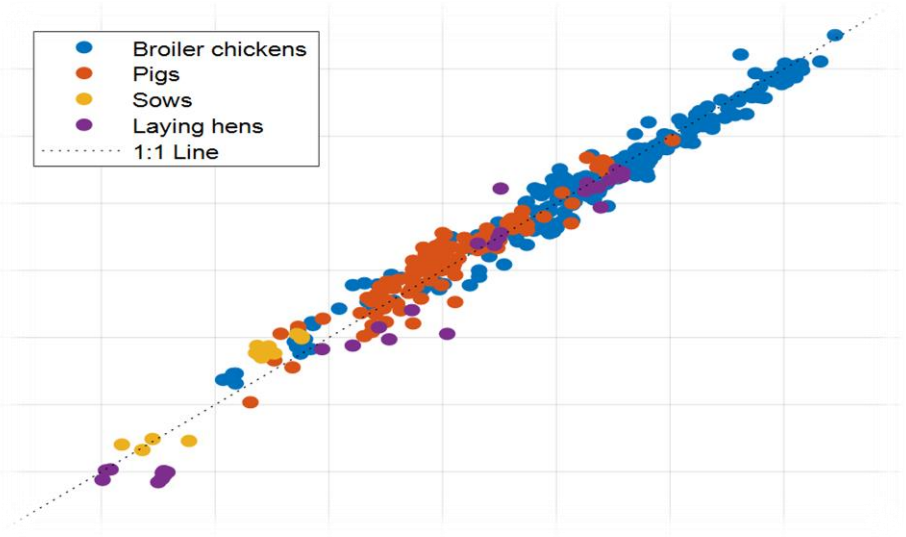
Section 2.3. Near-infrared spectroscopy in monogastric nutrition

Chapter 3. Replacing wet chemistry by near-infrared spectroscopy to improve formulation and reduce the costs of in-vivo assays

Section 3.2. Determination of carbohydrate and lignin content in feedstuffs for monogastric animals using near-infrared spectroscopy

Chapter 5. Strategies based on near-infrared spectroscopy to predict digestibility of monogastric animal diets

Chapter 1.
Objectives and structure



1.1. Objectives

The main objective of this thesis is to develop, validate and implement multivariate calibration models based on near-infrared spectroscopy (NIRS) or visible-near infrared spectroscopy (Vis-NIRS) to obtain the nutrient content of samples related with in-vivo assays with farm animals, concretely poultry and swine. The fast and inexpensive access to the nutrient content of these samples, that ultimately inform about the feed and how has been digested by the animal, may have applications in the agri-food sector among them the reduction of the cost of nutritional in-vivo assays and the easier identification of optimal diets for monogastric animals. This main objective gave rise to three specific objectives.

Specific objective 1. The development of methods based on NIRS or Vis-NIRS as an alternative to the routine wet chemistry analytical methods used to determine the nutrient content of feedstuffs, compound feeds, excreta, faeces and intestinal contents involved in in-vivo assays of poultry and swine. The purpose is to reduce the cost and increase the speed of the analytical determinations and avoid the drawbacks of wet chemistry analysis.

Specific objective 2. The development of strategies to maintain and update some of the calibration models developed in Specific objective 1. Calibration models used in routine determinations can fail with samples having unmodelled sources of variability. The purpose of these strategies is to detect those situations and update the models to maintain their predictive ability over time.

Specific objective 3. The study of different approaches based on NIRS for the determination of energy and digestibility. The purpose is to investigate whether NIRS can help simplify in-vivo assays even to the point of reducing the use of animals in nutritional studies focused on the determination of these nutritional parameters.

1.2. Structure

This thesis has six chapters. Chapter 1 contains the objectives and the structure. Chapter 2 introduces the analytical technique and the field of application. Chapters 3 to 5 contain the main experimental results and Chapter 6 has the conclusions and indicates future perspectives.

Chapter 1: *Objectives and structure*. This chapter.

Chapter 2: *General introduction*. It has three sections. Section 2.1 introduces the context of this thesis that is monogastric animal nutrition. Section 2.2 explains some theoretical aspects of near-infrared spectroscopy (NIRS), that is the main analytical technique used in this thesis, and multivariate calibration. Section 2.3 is a literature review of applications of NIRS in monogastric animal nutrition.

Chapter 3: *Replacing wet chemistry by NIRS to improve feed formulation and reduce the costs of in-vivo assays*. This chapter is related to the first objective and focuses on the development and validation of NIR calibration models to determine relevant parameters and nutrients in various types of samples related to nutritional in-vivo assays. Section 3.1 introduces the topic. The following five sections contain the main developments written in publication format (*). Section 3.2 studies the capability of NIRS to characterize carbohydrate and lignin fractions of a large variety of ingredients often included in monogastric diets. Section 3.3 deals with the development and validation of NIR calibration models for the nutrient content (crude protein, fat, crude fibre, gross energy, ash and phosphorus) of monogastric compound feeds and compares the performance of these models (global) with commercial and specific calibrations. Section 3.4 evaluates the usefulness and accuracy of Vis-NIR calibration models in the prediction of the nutrient content (organic matter, protein, fat, gross energy, uric acid and phosphorus) of poultry excreta (broilers, laying hens and turkeys) comparing global and specific calibrations. Section 3.5 deals with the development of Vis-NIR calibration models for the nutrient content (organic matter, protein, fat, gross energy, CF, NDF, ADF, lignin and phosphorus) of pig faeces. Finally, section 3.6 deals with the evaluation of the potential of NIRS to determine total protein content and 17 amino acids in poultry ileal digesta.

Chapter 4: *Maintenance of NIR calibration models*. This chapter is related to the second objective of this thesis and deals with the actions required to maintain and update some of the NIR and vis-NIR calibration models developed in Chapter 3. Section 4.1 introduces the topic. Section 4.2 describes a strategy to maintain calibration models for monogastric compound feeds (crude protein, fat, crude fibre and ash). Section 4.3 describes a strategy to update calibration models for pig faeces (phosphorus). These last two sections are written in publication format.

Chapter 5: *Strategies based on near-infrared spectroscopy to predict digestibility of monogastric animal diets*. This chapter is related to the third objective of this thesis and compares different strategies based on NIRS to obtain apparent metabolizable energy for broilers, apparent digestible energy for pigs, and apparent ileal digestibility of phosphorus for broilers. This chapter contains a single section written in publication format.

Chapter 6: *Conclusions and perspectives*. This chapter contains the conclusions of the thesis. It also comments on some future perspectives that can be explored based on the findings and implications presented throughout the thesis.

Appendix A lists the scientific contributions published by the author of this thesis and the presentations in national and international congresses. Appendix B lists the research stay and training courses attended.

(*) Note that the main experimental sections have been formatted as a publication to give this part of the thesis a uniform format, but not all correspond to actual publications.

Chapter 2.
General introduction

2.1. Monogastric nutrition

Introduction

Precision farming is a modern farming management concept that takes advantage of digital technology to monitor and optimize farming production processes. Different from traditional farming, precision farming uses a data driven approach to choose the most adequate settings for each production stage [1]. This concept covers many aspects from crops (sowing, growth control and harvesting) to the production of animals for meat, eggs, and milk (feeding, sanitary control and performance). The most advanced production methods in these fields constitute some type of precision farming even though they are not always explicitly called as such.

The need of evolving from traditional farming to precision farming has been recognized in foresight studies from the European Parliamentary Research Service [2,3] which stress the future challenges of farming. These are, basically, the demand of sustainably feeding an expanding world population, with less soil and water available, while combating soil pollution and climate change and ensuring food safety and economic growth. So far, the increasing demand for food has been met by increasing supply and improving production efficiency sometimes with not sustainable practices. The next step in the necessary evolution is precision farming. The production of monogastric animals is not an exception.

This thesis focuses on monogastric animals, specifically swine and poultry which are the most important production animals in the European Union [4]. Production of swine was 23.4 million tons in 2021 in the EU. The production has increased by 6% in the last 10 years. In Spain, the increase has been much more pronounced in this period (53%) until reaching in 2021 the first position as European producer surpassing Germany [5]. Poultry production rose to 13.5 million tons in 2021 in the EU. It increased by approximately 33% in the last 10 years both in the EU and in Spain that it is the second European producer behind Poland. Broiler chicken is the most produced poultry meat by far (more than 80% of poultry meat) followed by hen and turkey [6].

A key research subject in precision farming of monogastric animals is optimal feedstuff formulation. Feedstuff represents the major production cost, estimated to be between 60% and 70% of the total costs of swine [7] and poultry [8]. Feed is the key factor influencing the growth,

health and waste production and, consequently, the profit and environmental impact of this industry.

Feedstuff formulation criteria have evolved over the years. The classical approach is to formulate for total content of nutrients, such as total protein content. Compositional tables of the main ingredients (wheat, maize …) and their recommended levels for each animal species and age are used to decide the percentage of each ingredient in the formulation [9]. A disadvantage is that the listed values of nutrients are average values that may not correspond to the actual ingredient being used, whose properties vary depending on the origin (production conditions, climate …), transport and storage conditions. Additionally, raw ingredients are mainly purchased based on price and not based on the required proportion of nutrients for the feedstuff being formulated. This obliges the feed industry to determine the properties of the ingredients that are available at each moment with costly chemical analyses and continuously reformulate the combination of ingredients in the feedstuff seeking a compromise between the product's cost and nutritional properties.

Advances on optimal feedstuff formulation arrived by replacing wet-chemistry analytical methods (e.g., Kjeldahl method) by others based on faster techniques such as near-infrared spectroscopy (NIRS). NIRS is a fast, reliable and non-destructive analytical technique that is well suited for the analysis of organic constituents from plant and animal origin. The applications of NIRS to the agri-food industry have extended remarkably, in hand with the development of chemometric methods that are required to analyse the NIR spectra. NIRS has proved to be a better tool to formulate than table values and allows maintaining the desired amount of nutrients for each batch ensuring the quality of each manufactured diet [10,11].

Other important advances arrived by redefining the performance criteria that feedstuff must meet. In this sense, industry and animal research centres moved from total protein content to formulations based on amino acid content since the relative proportion of amino acids is a major factor affecting the animal development [12]. Another advance is the growing use of enzymes like carbohydrase and phytase in the diets, that has made necessary the determination of constituents traditionally ignored like non-starch polysaccharides (NSP) or phytic phosphorus [13]. Moving a step forward into precision farming, the next trend is formulation based on available content, which considers how the nutrients are digested and metabolized by the animal. When the same content of protein or amino acids can be obtained with different combinations

of raw ingredients, one is interested in the most cost-effective formulation that results in the highest digestibility. The combination of ingredients, the animal species and their age affect how the ingredients are effectively digested, so a variety of indicators such as the digestibility coefficients of the main nutrients (energy, fat, protein and amino acids) are used to guide the feedstuff formulation.

The main deterrent for the routine use of digestibility values is that they are not readily available. They must be obtained from extensive (and expensive) carefully planned in-vivo assays, where animals are fed different diets of perfectly known composition and their faeces or excreta (in poultry, the combination of faeces and urine) and/or intestinal contents are analysed to determine the effectiveness of the diet. For example, digestible energy is determined in in-vivo assays by subtracting the gross energy of the faeces from the gross energy of the feed. Therefore, in addition to the logistics and infrastructure required to work with animals fulfilling legal and ethical considerations, in-vivo assays require well-equipped analytical laboratories able to accurately determine the chemical composition of feedstuffs, compound feeds, faeces or excreta and intestinal contents [14]. Running such a variety of analytical methods is not only slow and costly but also has an environmental impact that cannot be ignored, because they use solvents and generate chemical waste that must be processed properly.

Different approaches have been studied to replace in-vivo assays. Tabular values [15], empirical models based on the chemical composition of feed [16], in-vitro methods [17] and NIRS [18] have been applied to estimate digestible energy of swine and poultry. The estimation of the digestibility of nutrients has been also studied although it is still generally based on in-vivo assays.

Feedstuffs in monogastric nutrition

In Spain, the total production of animal feed reached almost 40 millions of tonnes in 2020 of which 18 millions were for the swine sector and 7 for the poultry sector. The average proportion of feedstuffs used for feed production was 62% for cereals, 17% for oilseed co-products (protein-rich feedstuffs), 7% for cereal co-products, 3% for minerals such as calcium carbonate and 2% for sources of fat such as animal fat or vegetal oils. The remaining 9% were products or co-products from legumes, tubers and roots, non-oil seeds, forages or silages, other plants

and animal by-products [19]. The cereals are used so extensively because they are widely available, easy to transport and handle, readily consumed by the animal and provide a large amount of energy, mainly through the digestion of the starch. The most employed cereals are corn, barley and wheat. Corn is the main cereal used in monogastric animal feeds. Barley is more often used in swine diets than in poultry diets and wheat is quite used in cold regions [20]. In the context of circular economy, the use of cereal co-products in the diets is increasing, especially flours, middlings and cereal brans from the previously mentioned cereals and by-products from distillery and brewery (DDGS) [21]. Since cereals and cereal co-products contain low amounts of essential nutrients such as protein, fat or minerals, protein sources and typically low amounts of fat and minerals are also included to prepare balanced diets. The quintessential source of protein is soybean meal because it has a high protein content with an adequate amino acid profile and a stable nutrient composition. This crop, however, has undesirable environmental impact such as the emission of nitrous oxide from the fields, the land-use change or the transport of the product to areas far from the producers [22]. As an alternative, sunflower and rapeseed meal are sometimes used in EU as home-grown protein-rich feedstuffs and nowadays more and more research is being conducted about the use of new sources of protein such as insects [23] or algae [24]. The nutritional profiles and the importance of the principal ingredients for poultry and swine nutrition can be found in references [9] and [25], respectively.

Controlled parameters in feedstuffs

Figure 1 shows the parameters that are commonly considered when a diet is formulated. The parameters that are determined more often in the agri-food industry are those belonging to the proximate system of feed analysis by Weende [26]. This system has been used for more than 150 years and consists of the analysis of dry matter (DM), crude protein (CP), fat, crude fibre (CF) and ash. Nowadays, however, it is not possible to formulate an adequate diet based only on these parameters.

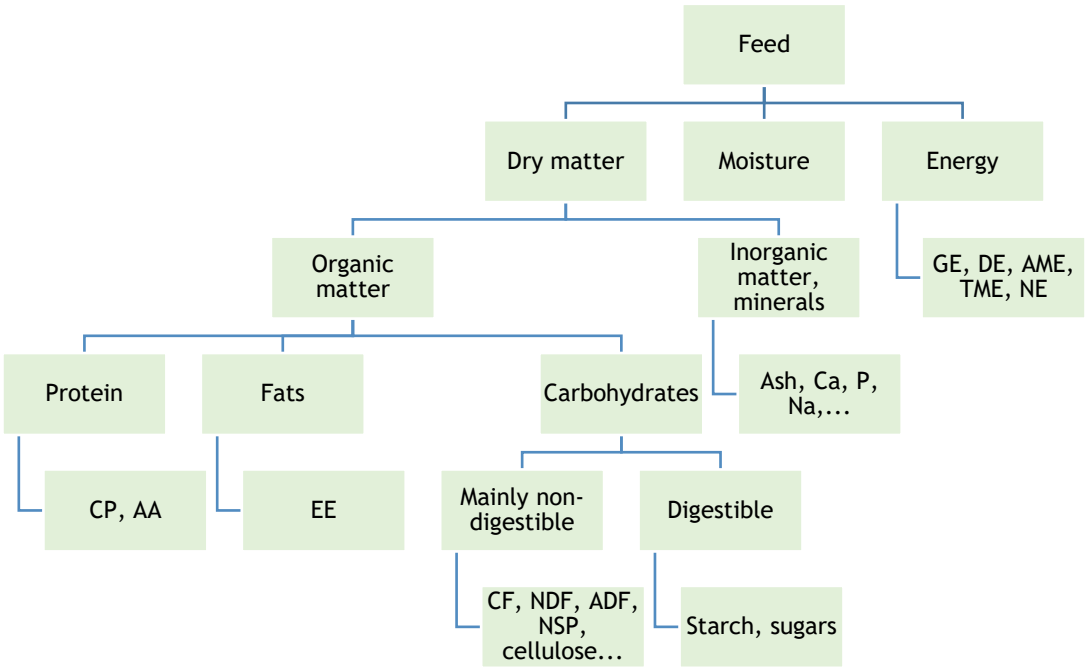


Figure 1. Parameters that are commonly controlled in feedstuffs and diets.

Swine and poultry have specific amino acid (AA) requirements. The AAs that usually lack in conventional ingredients are known as limiting AAs. Lys, followed by Met and Cys (the sulphured AAs) and Thr are the main limiting AAs for poultry and swine diets [27,28]. Depending on the composition of the diet, other AAs could be limiting, e.g. Trp in corn based diets or Val in wheat based diets. Traditionally, feeds have been formulated with high levels of CP to ensure that the animal receives the required amount of each AA. The formulation based on the AAs profile (supplementing some AAs if needed) makes it possible to reduce the CP content of the feed and thus reducing the environmental impact caused by an excess of nitrogen in the diets [22]. Diets can be further improved by considering the available protein and AAs instead of the total amount. The fraction of CP and AAs non-digested by monogastric animals is preferably measured before the microbial degradation of AAs and the synthesis of AAs in the hindgut. In the case of poultry, it is also important to avoid the interfering influence of urine on AA digestibility. Because of that, it is more precise and repeatable to determine the non-digested CP and AAs in the ileal digesta than in faeces or excreta samples [29]. The digestibility calculated using the ileal digesta is called apparent ileal digestibility (AID). However, ileal digesta contains also AAs of endogenous origin that are called ileal endogenous AA losses and can be divided into basal and specific. True ileal digestibility (TID) is the most reliable way to consider the AA digestibility because it includes both types of endogenous losses in the calculations but these

values are rarely available. Standardized ileal digestibility (SID), that takes into account only the basal endogenous losses is often used to formulate because SID values are much easier to obtain and because, unlike TID values, they are additive when a diet is formulated [30].

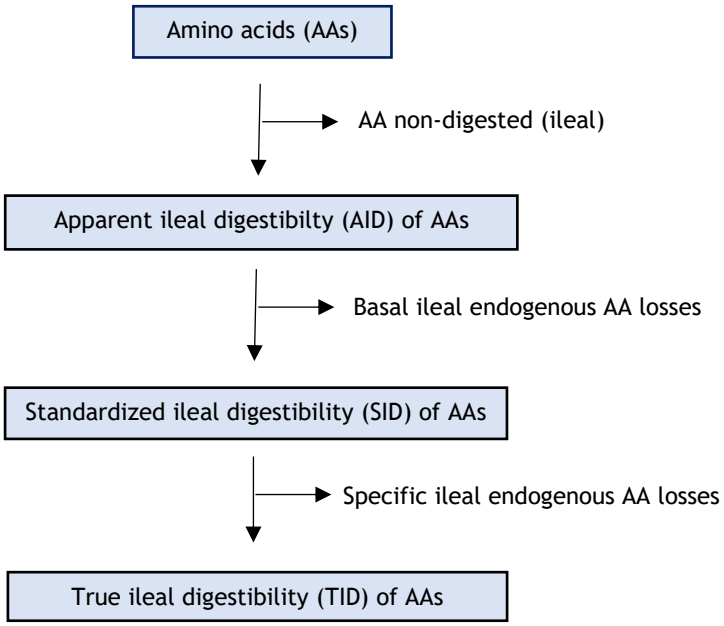


Figure 2. Partitioning of AA digestibility for feedstuffs and diets.

Feed contains large amounts of different carbohydrates that can be grouped in monosaccharides, disaccharides, oligosaccharides (3-12 sugars) and polysaccharides [31]. Monogastric animals can directly absorb the monosaccharides only, although disaccharides and starch are also usable thanks to endogenous enzymes. In fact, starch represents the main source of energy in a typical poultry or swine diet [32]. Nevertheless, these animals do not digest well oligosaccharides, non-starch polysaccharides (NSP) and lignin. The sum of these constituents is called dietary fibre (DF). The soluble fraction of the fibre can be fermented by the microbiota, positively influencing the intestinal health, while the insoluble fraction is often considered an anti-nutritive factor [33]. Nowadays, in order to mitigate this effect, supplementation with carbohydrases such as xylanase, β-glucanase or manannase have gained popularity in poultry [34] and swine industry [35]. According to European Food Safety Authority (EFSA) the companies and research centres that test and use supplementary enzymes must demonstrate their usefulness by analysing their substrates [36]. However, the complexity and cost of the analyses to characterize the carbohydrate fraction described by Bach Knudsen 1997 [37] lead to the fact that in most cases, the analysis of carbohydrates in feedstuffs is limited to the

determination of starch and the use of simplified and approximate methods to determine the fibre fraction. In the poultry industry, crude fibre (CF) is the most widely used fibre-related parameter when formulating a diet and represents variable portions of insoluble NSP, including cellulose and some hemicellulose. Neutral detergent fibre (NDF), which is a measure of hemicellulose, cellulose, and lignin, and acid detergent fibre (ADF), which includes only cellulose and lignin, are the most widely used fibre-related parameters in swine alimentation [38].

Phosphorus (P) is a constituent that is attracting attention from the animal production sector. In conventional feedstuffs, P is mainly in the form of phytic P, which is poorly digestible by poultry and swine. In consequence, unused P accumulates in animal waste impacting the environment and inorganic P must be supplemented to the diet to satisfy the animal's physiological needs. The inclusion of phytase doses in the diets improves the digestibility of phytic P reducing the P in animal waste and the amount of inorganic P supplemented [39]. Hence, in addition to total P, phytic P need to be determined in feedstuffs and diets when phytase is going to be used.

Energy is a key parameter to consider when formulating animal diets. The amount of feed consumed by the animal depends on its energetic needs. Therefore, it is possible to predict feed consumption if the energy concentration of a diet is accurately provided and essential nutrients are not lacking. This information is the basis of diet formulation [40]. Gross energy (GE) is defined as the total chemical energy measured upon complete combustion of the feed in a calorimeter bomb. However, GE alone is not a good energy indicator because the animal does not retain all of it. There are losses in the faeces and urine and as gases and heat. Different energy systems have been proposed depending on which of these losses are considered (Figure 3). Digestible energy (DE) is the difference between the GE in the feed and the energy losses found in the faeces, that is the GE of the faeces. Metabolizable energy (ME) also considers the losses in urine and gases from digestive fermentation. Net energy (NE) considers the increase in heat in addition to all the above. In poultry, since faeces and urine are excreted together, ME is the most determined energy form while DE is the most common for swine. DE and ME are also referred to as apparent DE (ADE) and apparent ME (AME), respectively. In poultry, AME values are often corrected on a nitrogen basis (AME_N) to account for the part of N that remains as tissue protein or egg protein or that is converted to uric acid and excreted [41]. Another energy system, the so-called true ME (TME) was developed for poultry [42]. In this case, the GE in the excreta is corrected by subtracting the endogen energy losses that are calculated as

the GE of the excreta during long periods of fasting. TME was used especially in North America but has been progressively abandoned due to the difficulty in estimating endogenous losses [43]. NE should be a more accurate estimate than DE or ME of the actual energy used by the animal but is more expensive and ethically problematic to obtain since the animals must be kept in respiration chambers to estimate the heat increment. The use of NE is more established in swine than in poultry industry. According to Van der Klis & Jansman 2019 [44] the efficiency of the AME system compared with the NE system depends less on the nutrient contents of the feedstuffs for poultry than for swine. Therefore, the NE system for poultry would not have a clear benefit over the AME system. Noblet et al. 2022 [45] have recently reviewed how the different energy parameters are obtained and which are the benefits or limitations of using one or another to formulate swine and poultry diets.

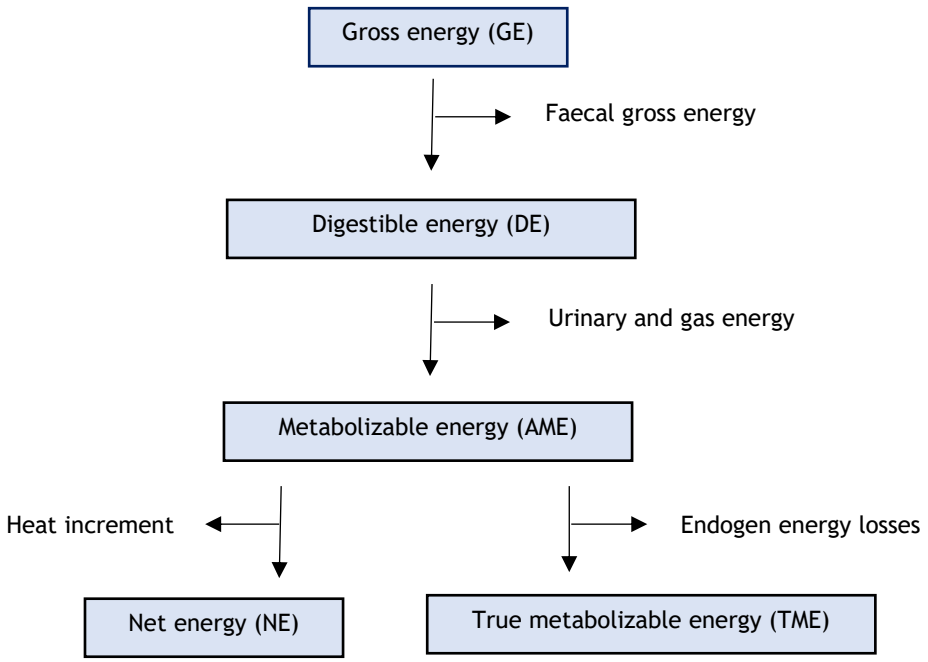


Figure 3. Partitioning of energy for feedstuffs and diets.

In-vivo assays to obtain digestibility

In-vivo assays measure the animal response to variations in the diets. They are the most reliable methods to determine the digestible energy and the digestibility of nutrients [46]. The faecal digestibility coefficient of energy and nutrients ($D_{Nut}(\%)$) is obtained as GE or nutrient in the feed consumed minus GE or nutrient in the faeces and is equal to:

$$D_{Nut}(\%) = 100 \times \left(\frac{[Nut]_{feed} \times Feed_{intake} - [Nut]_{faeces} \times Feces_{excreted}}{[Nut]_{feed} \times Feed_{intake}} \right) \quad (1)$$

where $[Nut]_{feed}$ and $[Nut]_{faeces}$ are the amount of GE or the concentration of the nutrient in feed or faeces, respectively, $Feed_{intake}$ is the weight of feed fed to the animal and $Faeces_{excreted}$ is the weight of faeces collected.

For the ileal digestibility coefficients, the nutrient in the ileal digesta is used instead of the nutrient in faeces.

In-vivo assays are conducted keeping the animals in digestibility pens or cages. Before starting the measure of feed intake and faeces collection the animal needs time to adapt to the location and to the feed. The adaptation time varies depending on the specie, the diet and the weight of the animal. At least 3 days are necessary for poultry and 5 days for swine in the growth phase, but longer periods are recommended for adult animals or when fibre-rich diets are used. After that, pig faeces are typically collected for 2-10 days and poultry excreta for 2-5 days [45].

Confining animals in digestibility pens/cages is an animal welfare issue and measuring feed intake and faecal excretion over long periods of time is laborious. As a solution, indigestible markers are now used in most in-vivo assays. The marker is added to the feed and remains in the faeces. It is assumed that the marker is completely indigestible, distributes evenly in the feed and faeces, and can be measured accurately even at low concentrations [29]. Then:

$$Feed_{intake}[Marker]_{feed} = Faeces_{excreted}[Marker]_{faeces} \quad (2)$$

where $[Marker]$ is the concentration of the indigestible marker in feed or faeces, $Feed_{intake}$ is the weight of feed fed to the animal and $Faeces_{excreted}$ is the weight of faeces collected.

Eq. 1 can now be rewritten as:

$$D_{Nut}\ (\%) = 100 \times \frac{[Nut]_{faeces} \times [Marker]_{feed}}{[Nut]_{feed} \times [Marker]_{faeces}} \quad (3)$$

where [Nut] is the amount of GE or the concentration of the nutrient in feed or faeces and [$Marker$] is the concentration of the indigestible marker in feed or faeces.

Therefore, knowing the concentration of marker in the feed and in the faeces/excreta, it is not necessary to measure the feed intake and collect all the faeces, and the animals can be kept under conventional husbandry conditions.

Some studies have compared the accuracy of the total collection method with the marker method using different markers [47,48]. In brief, the repeatability with the marker method is lower than with the total collection method but there is no systematic difference between them and between markers. The most used markers are acid-insoluble ash (silicon dioxide), titanium dioxide and chromium oxide. The inclusion of titanium dioxide in animal diets has been recently restricted in UE [49]. Animals fed diets with titanium dioxide cannot be used for consumption and must be slaughtered at the end of the in-vivo assay. Some alternatives such as the use of yttrium and ytterbium oxide are being considered. The use of an indigestible marker is even more necessary in in-vivo assays whose objective is the determination of ileal digestibility because the total collection of ileal digesta is more complex than that of faeces. Among the methods for collecting ileal digesta, collection from the ileum after euthanasia of animals is the most common practice in poultry. For swine, a through surgically fitted simple T-cannula is commonly used [50].

There are many factors affecting the digestibility values obtained by in-vivo assays. The particle size of the feed and the way the animal is fed are important mainly for poultry. Pelleted feed is used in the first weeks of life because consumption increases, and mash feed when the animal is older, as it is believed to positively affect the development of the gastrointestinal tract [43]. Although there are ways to control the feed that the animal intakes, ad libitum feeding is the most recommendable if indigestible markers are used. Differences in the digestive capacity of the animals lead to different digestibility values. It cannot be assumed that the digestibility values of an ingredient or diet obtained in an in-vivo assay for an animal from one species or breed will be the same as those obtained for an animal from other species or breed. Adult animals digest feed nutrients better, so the digestibility values obtained depend on the growth phase in which the animals are. Sex can also produce slight differences in digestibility [51]. Possible interactions between nutrients is another factor that may alter the digestibility values obtained. Formulation assumes that the nutrients are additive but the presence of one nutrient can affect

the digestion of the others. The best-known example is the extra-caloric effect of the fats that may produce a higher than expected DE, possibly due to its interaction with the non-starch polysaccharides (NSP) [52]. Other factors such as thermal processing of the feed, anti-nutritional factors, the addition of feed additives such as enzymes, probiotics and prebiotics, or environmental conditions can also make the digestibility values vary [43].

To obtain reliable digestibility values, the key is to control as much as possible the mentioned sources of variation and to use replicate units. This means groups of animals of the same breed, age, and sex receiving exactly the same diet under the same conditions. Digestibility is then calculated for each replicate unit and averaged.

Determining the digestibility of an ingredient is more complicated than that of a diet. The most widely used experimental design to obtain digestibility values of ingredients is called basal substitution. A basal diet is formulated to be balanced in nutrients and energy. Then a portion of the diet is substituted with the test ingredient to produce a test diet. The test animals are separated into two groups, one is fed the basal diet and the other is fed the test diet. Digestibility values are determined for each group and the digestibility of the test ingredient can be calculated considering the substitution ratio of the test ingredient using the following equation:

$$D_{Nut}\ test\ ingredient = D_{Nut}\ basal\ diet - \frac{D_{Nut}\ basal\ diet - D_{Nut}\ test\ diet}{Proportion\ of\ test\ ingredient} \quad (4)$$

where D_{Nut} is the digestibility coefficient of energy or nutrients.

The included amount of the test ingredient in the test diet can vary at graduated levels of inclusion. In this case, the digestibility of the test ingredient is obtained by extrapolation on a regression line formed by the digestibility values at different ingredient inclusion levels. It is also possible to calculate the digestibility of various ingredients by applying multiple linear regression (MLR). It consists of using multiple diets with several test ingredients at different levels and measuring the digestibility values of each of them. The number of diets must be greater than the number of test ingredients. The coefficients of the regression equation will correspond to the digestibility value of each of the tested ingredients [53].

Alternative methods to determine digestibility

In-vivo assays are the reference method to study the animal's response to the diets. Nevertheless, in-vivo assays are time demanding, require dedicated personnel, need facilities to house and keep the animals in conditions suitable for testing, require a large amount of feed that must be administered to all animals participating in the assay and require the analysis of a large number of samples including feeds, faeces/excreta and/or intestinal contents. These analyses are traditionally done by wet chemistry methods that are usually slow, expensive, require reagents and produce waste. An additional concern is society's prejudice towards animal experimentation [54]. These limitations have made in-vivo assays not a highly desirable option for routine evaluation of feed digestibility and institutions, companies and research centres are searching for other methodologies. This section summarizes some alternative methods to determine in a more economical and rapid way the digestibility of nutrients and the digestible energy of feedstuffs and diets for monogastric animals. These include tabulated values, in vitro analyses and prediction models from the chemical composition of the feed. The use of near-infrared spectroscopy (NIRS), the subject of this thesis, is explained in Sections 2.2 and 2.3.

Tabulated values

Under most practical situations, published values are the basis of feed formulation [43]. Organizations worldwide (e.g. NRC, CVB, FEDNA, INRA) provide the nutrient content of diverse ingredients. For the most commonly used parameters (DM, CP, fat, CF, ash, starch) the values derive from large number of samples analysed by laboratories associated with the organization or by trusted laboratories. To formulate a diet the desired concentration of nutrients is introduced in a software package like Brill Formulation ® together with the nutritional profile of the ingredients found in the tables. Based on this information and in the price of the ingredients the software provides the most cost-effective diet that has the desired concentration of nutrients. Values of digestible energy and of some nutrients digestibility also appear in some tables and can be used to formulate the diets but the average values come from a lower number of measurements than for the total nutrient composition and the variability between values of different tables is high. The complexity of in-vivo assays and the lack of standardization is the cause. Mateos et al. 2018 [43] compared table values of several important institutions in animal nutrition such as CVB, Evonik, WPSA, Premier Atlas, INRA or FEDNA.

They found a wide variability in the case of the energy values of many ingredients and concluded that this uncertainty limits the applicability of table values, and that nutritionists should be aware of the procedures used by the research institutions for feed evaluation.

Other important drawback of formulating based on tables is that the ingredient nutrient profile is an average and may not correspond to the actual ingredient being used, whose properties vary depending on the origin (production, conditions, climate …), transport and storage conditions.

In vitro analyses

In vitro analyses are used to estimate the digestibility of constituents such as DM, OM, CP and others. In vitro techniques have been developed for swine [55–57] and have been recently reviewed for poultry [58]. These techniques employ enzymes and other substances or chemical products trying to imitate the reactions of the digestive processes. The obtaining of in-vitro digestibility values is cheaper and faster than in-vivo, but still requires expensive equipment and reagents, as well as long reaction times. Another drawback is that in-vitro values are sometimes higher than in-vivo mainly due to minimal, or lack of, endogenous losses of nutrients in the in vitro digestibility [54]. Because of that, the data provided by in vitro analyses is typically used to develop specific regression equations for prediction of ileal or total tract digestibility of nutrients and energy [59].

Prediction models

Empirical prediction models relate the nutrient content of the feed obtained by chemical analysis or the in-vitro digestibility values with the digestibility of the nutrients or energy determined by in-vivo assays. Mathematical models have been developed to relate the nutrient content of feeds and diets to their ADE value for swine [16,60–62] and to their AME value for poultry [53,63–65]. The predictors of the models for AME in poultry are generally CP, fat, ash, starch and sugars and in a lesser extent CF or NDF. For DE in swine NDF is usually included together with CP, ash and fat. Other authors have tried to predict the digestibility of nutrients such as CP, fat, starch [66] or AA [67] but they did not find a strong relationship between the digestibility and the proximal composition.

Most of the models to predict energy are obtained by multiple linear regression (MLR). MLR is an extension of simple linear regression for multiple predictors in order to predict a single variable [68]. The model has the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_j x_j \ldots + \cdots + \beta_p x_p + \varepsilon \quad (5)$$

where x_j is the j^{th} predictor, β_j is the regression coefficient for x_j , y is the response and ε is error term that represents the variability that cannot be explained by the model. In our case, y is the energy (AME, ADE or other form of energy) and the $x_j's$ are the values of the parameters we want to use, for example, CP, EE or CF.

To obtain the best model, several models are created stepwise. Models are developed adding variables sequentially to the first model, keeping each time the variable that produces the greatest improvement in model quality (decreasing the prediction error). After adding each new variable, the method can also remove variables that do not improve the model fit. These models are well explained in reference [68] for example.

MLR assumes a linear relationship between the predictors and the response. In some studies, non-linear models such as Artificial Neural Networks (ANN) have performed better in predicting digestibility [69–71].

References

[1] H. Auernhammer, Precision farming - The environmental challenge, Comput. Electron. Agric. 30 (2001) 31–43.

[2] R. Schrijver, Precision agriculture and the future of farming in Europe Scientific Foresight Study. Scientific Foresight Unit, European Parliamentary Research Service, 2016.

[3] Mi. Kritikos, Precision agriculture in Europe: Legal and ethical reflections for law-makers, 2017.

[4] FAO, World Food and Agriculture - Statistical Yearbook, Rome, Italy, 2021.

[5] El sector de la carne de cerdo en cifras: Principales indicadores económicos. Subdirección General de Producciones Ganaderas y Cinegéticas, Dirección General de Productores y Mercados Agrarios, 2021.

[6] El sector de la avicultura de carne en cifras: Principales Indicadores Económicos, Subdirección General de Producciones Ganaderas y Cinegéticas, Dirección General de Productores y Mercados Agrarios, 2021.

[7] J.F. Patience, M.C. Rossoni-Serão, N.A. Gutiérrez, A review of feed efficiency in swine: Biology and application, J. Anim. Sci. Biotechnol. 6 (2015) 1–9.

[8] M. Zampiga, F. Calini, F. Sirri, Importance of feed efficiency for sustainable intensification of chicken meat production: implications and role for amino acids, feed enzymes and organic trace minerals, Worlds. Poult. Sci. J. 77 (2021) 639–659.

[9] S. Leeson, J.D. Summers, Commercial Poultry Nutrition, 3ª Ed., Nottingham University Press, Nottingham, England, 2008.

[10] C. Soto, E. Avila, J. Arce, F. Rosas, D. McIntyre, Evaluation of different strategies for broiler feed formulation using near infrared reflectance spectroscopy as a source of information for determination of amino acids and metabolizable energy, J. Appl. Poult. Res. 22 (2013) 730–737.

[11] R. Montanhini Neto, E. N'Guetta, C. Gady, M. Francesch, A. Preynat, Combined effect of using near-infrared spectroscopy for nutritional evaluation of feed ingredients and non-starch polysaccharide carbohydrase complex on performance of broiler chickens, Anim. Sci. J. 88 (2017) 1979–1986.

[12] G. Santomá, G.G. Mateos, Necesidades nutricionales en avicultura. Normas FEDNA, 2ª Ed., Improtalia S.L., Madrid, Spain, 2018.

[13] O. Adeola, A.J. Cowieson, Board-invited review: Opportunities and challenges in using exogenous enzymes to improve nonruminant animal production, J. Anim. Sci. 89 (2011) 3189–3218.

[14] E. da Silva, C.B.V. Rabello, M.B. De Lima, E.M.F. De Arruda, J. V. Ludke, M.C.M.M. Ludke, Determination of the chemical composition, amino acid levels and energy values of different poultry offal meals for broilers, Rev. Bras. Cienc. Avic. 14 (2012) 97–107.

[15] D.Y. Kil, B.G. Kim, H.H. Stein, Invited review - Feed energy evaluation for growing pigs, Asian-Australasian J. Anim. Sci. 26 (2013) 1205–1217.

[16] J. Noblet, J.M. Perez, Prediction of digestibility of nutrients and energy values of pig diets from chemical analysis, J. Anim. Sci. 71 (1993) 3389–3398.

[17] M. Clunies, S. Leeson, J.D. Summers, In Vitro Estimation of Apparent Metabolizable Energy, Poult. Sci. 63 (1984) 1033–1039.

[18] P.C. Garnsworthy, J. Wiseman, K. Fegeros, Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy, J. Agric. Sci. 135 (2000) 409–417.

[19] Ministerio de Agricultura Pesca y Alimentación, Datos De Producción De Piensos 2020, 2020.

[20] P. McDonald, R.A. Edwards, J.F.D. Greenhalgh, C.A. Morgan, L.A. Sinclair, R.G. Wilkinson, Animal nutrition, 7th ed., Prentice Hall/Pearson, Harlow, England, 2011.

[21] A. Seidavi, M. Azizi, A.A. Swelum, M.E.A. El-Hack, Practical application of some common agro-processing wastes in poultry diets, Worlds. Poult. Sci. J. 77 (2021) 913–927.

[22] L. Cappelaere, J. Le Cour Grandmaison, N. Martin, W. Lambert, Amino Acid Supplementation to Reduce Environmental Impacts of Broiler and Pig Production: A Review, Front. Vet. Sci. 8 (2021) 1–14.

[23] T. Veldkamp, A.G. Vernooij, Use of insect products in pig diets, J. Insects as Food Feed. 7 (2021) 781–793.

[24] E. Coudert, E. Baéza, C. Berri, Use of algae in poultry production: a review, Worlds. Poult. Sci. J. 76 (2020) 767–786.

[25] A.J. Lewis, L.L. Southern, Swine Nutrition, 2ª Ed., CRC Press, Boca Raton, Florida, 2001.

[26] W. Henneberg, F. Stohmann, Über das Erhaltungsfutter volljährigen Rindviehs, J. Landwirtschaft. 3 (1859) 485–551.

[27] Y. Wang, J. Zhou, G. Wang, S. Cai, X. Zeng, S. Qiao, Advances in low-protein diets for swine, J. Anim. Sci. Biotechnol. 9 (2018) 1–14.

[28] C.W. Maynard, M.T. Kidd, P. V. Chrystal, L.R. McQuade, B. V. McInerney, P.H. Selle, S.Y. Liu, Assessment of limiting dietary amino acids in broiler chickens offered reduced crude protein diets, Anim. Nutr. 10 (2022) 1–11.

[29] C. Kong, O. Adeola, Evaluation of amino acid and energy utilization in feedstuff for swine and poultry diets, Asian-Australasian J. Anim. Sci. 27 (2014) 917–925.

[30] H.H. Stein, M.F. Fuller, P.J. Moughan, B. Sève, R. Mosenthin, A.J.M. Jansman, J.A. Fernández, C.F.M. de Lange, Definition of apparent, true, and standardized ileal digestibility of amino acids in pigs, Livest. Sci. 109 (2007) 282–285.

[31] M. Choct, Feed non-starch polysaccharides for monogastric animals: Classification and function, Anim. Prod. Sci. 55 (2015) 1360–1366.

[32] K.E. Bach Knudsen, Fiber and nonstarch polysaccharide content and variation in common crops used in broiler diets, Poult. Sci. 93 (2014) 2380–2393.

[33] T.O. Adebowale, K. Yao, A.O. Oso, Major cereal carbohydrates in relation to intestinal health of monogastric animals: A review, Anim. Nutr. 5 (2019) 331–339.

[34] N.E. Ward, Debranching enzymes in corn/soybean meal–based poultry feeds: a review, Poult. Sci. 100 (2021) 765–775.

[35] A.L. Petry, J.F. Patience, Xylanase supplementation in corn-based swine diets: A review with emphasis on potential mechanisms of action, J. Anim. Sci. 98 (2020) 1–12.

[36] European Food Safety Authority (EFSA), Guidance on the assessment of the efficacy of feed additives, EFSA J. 16 (2018) 5274.

[37] K.E. Bach Knudsen, Carbohydrate and lignin contents of plant materials used in animal feeding, Anim. Feed Sci. Technol. 67 (1997) 319–338.

[38] K.E. Bach Knudsen, H.N. Laerke, Carbohydrates and lignin in the feed - from sugars to complex composed fibres, in: P.J. Moughan, W.H. Hendriks (Eds.), Feed evaluation science, Wageningen Academic Publishers, Wageningen, Netherlands, 2018: pp. 113–144.

[39] M.E.A. El-Hack, M. Alagawany, M. Arif, M. Emam, M. Saeed, M.A. Arain, F.A. Siyal, A. Patra, S.S. Elnesr, R.U. Khan, The uses of microbial phytase as a feed additive in poultry nutrition - A review, Ann. Anim. Sci. 18 (2018) 639–658.

[40] D.J. Farrel, In vivo and in vitro techniques for the assessment of the energy content of feed grains for poultry: a review, Aust. J. Agric. Res. 50 (1999) 881–888.

[41] S.B. Wu, M. Choct, G. Pesti, Historical flaws in bioassays used to generate metabolizable energy values for poultry feed formulation: a critical review, Poult. Sci. 99 (2020) 385–406.

[42] I.R. Sirbald, A bioassay for true metabolizable energy in feedingstuffs, Poult. Sci. 55 (1976) 303–308.

[43] G.G. Mateos, L. Cámara, G. Fondevila, R.P. Lázaro, Critical Review of the Procedures Used for Estimation of the Energy Content of Diets and Ingredients in Poultry, J. Appl. Poult. Res. 0 (2018) 1–20.

[44] J.D. Van Der Klis, A.J.M. Jansman, Net energy in poultry: Its merits and limits, J. Appl. Poult. Res. 28 (2019) 499–505.

[45] J. Noblet, S.B. Wu, M. Choct, Methodologies for energy evaluation of pig and poultry feeds: A review, Anim. Nutr. 8 (2022) 185–203.

[46] A. Bourdillon, B. Carre, B. Leclercq, M. Lessire, L. Cannon, M. Francesch, M. Fuentes, J. Wiseman, G. Huyghebaert, W.M.M.A. Janssen, M. Rigoni, European reference method of in vivo determination of metabolisable energy in poultry: Reproducibility, effect of age, comparison with predicted values, Br. Poult. Sci. 31 (1990) 567–576.

[47] N. Smeets, F. Nuyens, L. Van Campenhout, E. Delezie, J. Pannecoucque, T. Niewold, Relationship between wheat characteristics and nutrient digestibility in broilers: Comparison between total collection and marker (titanium dioxide) technique, Poult. Sci. 94 (2015) 1584–1591.

[48] C. Huang, P. Li, X. Ma, N.W. Jaworski, H.H. Stein, C. Lai, J. Zhao, S. Zhang, Methodology effects on determining the energy concentration and the apparent total tract digestibility of components in diets fed to growing pigs, Asian-Australasian J. Anim. Sci. 31 (2018) 1315–1324.

[49] V. Bampidis, G. Azimonti, M. de L. Bastos, H. Christensen, B. Dusemund, M. Fašmon Durjava, M. Kouba, M. López-Alonso, S. López Puente, F. Marcon, B. Mayo, A. Pechová, M. Petkova, F. Ramos, Y. Sanz, R.E. Villa, R. Woutersen, G. Aquilina, G. Bories, J. Gropp, J. Galobart, M.V. Vettori, Safety and efficacy of a feed additive consisting of titanium dioxide for all animal species (Titanium Dioxide Manufacturers Association), EFSA J. 19 (2021).

[50] O. Adeola, P.C. Xue, A.J. Cowieson, K.M. Ajuwon, Basal endogenous losses of amino acids in protein nutrition research for swine and poultry, Anim. Feed Sci. Technol. 221 (2016) 274–283.

[51] R.J. Hughes, M. Choct, Chemical and physical characteristics of grains related to variability in energy and amino acid availability in poultry, Aust. J. Agric. Res. 50 (1999) 689–701.

[52] V. Ravindran, P. Tancharoenrat, F. Zaefarian, G. Ravindran, Fats in poultry nutrition: Digestive physiology and factors influencing their utilisation, Anim. Feed Sci. Technol. 213 (2016) 1–21.

[53] S.B. Wu, R.A. Swick, J. Noblet, N. Rodgers, D. Cadogan, M. Choct, Net energy prediction and energy efficiency of feed for broiler chickens, Poult. Sci. 98 (2019) 1222–1234.

[54] M. Yegani, D.R. Korver, Review: Prediction of variation in energetic value of wheat for poultry, Can. J. Anim. Sci. 92 (2012) 261–273.

[55] P.R. Regmi, W.C. Sauer, R.T. Zijlstra, Prediction of in vivo apparent total tract energy digestibility of barley in grower pigs using an in vitro digestibility technique, J. Anim. Sci. 86 (2008) 2619–2626.

[56] P.R. Regmi, N.S. Ferguson, R.T. Zijlstra, In vitro digestibility techniques to predict apparent total tract energy digestibility of wheat in grower pigs, J. Anim. Sci. 87 (2009) 3620–3629.

[57] T.A. Woyengo, R. Jha, E. Beltranena, R.T. Zijlstra, In vitro digestion and fermentation characteristics of canola co-products simulate their digestion in the pig intestine, Animal. 10 (2016) 911–918.

[58] F. Zaefarian, A.J. Cowieson, K. Pontoppidan, M.R. Abdollahi, V. Ravindran, Trends in feed evaluation for poultry with emphasis on in vitro techniques, Anim. Nutr. 7 (2021) 268–281.

[59] B. Losada, P. García-Rebollar, C. Álvarez, P. Cachaldora, M.A. Ibáñez, J. Méndez, J.C. De Blas, The prediction of apparent metabolisable energy content of oil seeds and oil seed by-products for poultry from its chemical components, in vitro analysis or near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 160 (2010) 62–72.

[60] P. V. Anderson, B.J. Kerr, T.E. Weber, C.J. Ziemer, G.C. Shurson, Determination and prediction of digestible and metabolizable energy from chemical analysis of corn coproducts fed to finishing pigs, J. Anim. Sci. 90 (2012) 1242–1254.

[61] J.Y. Sung, B.G. Kim, Prediction equations for digestible and metabolizable energy concentrations in feed ingredients and diets for pigs based on chemical composition, Anim. Biosci. 34 (2021) 306–311.

[62] P.E. Urriola, M. Li, B.J. Kerr, G.C. Shurson, Evaluation of prediction equations to estimate gross, digestible, and metabolizable energy content of maize dried distillers grains with solubles (DDGS) for swine based on chemical composition, Anim. Feed Sci. Technol. 198 (2014) 196–202.

[63] B. Carré, M. Lessire, H. Juin, Prediction of metabolisable energy value of broiler diets and water excretion from dietary chemical analyses, Animal. 7 (2013) 1246–1258.

[64] B. Losada, C. De Blas, P. García-Rebollar, P. Cachaldora, J. Méndez, M. Ibáñez, Prediction of apparent metabolisable energy content of cereal grains and by-products for poultry from its chemical composition, Spanish J. Agric. Res. 13 (2015) 1–5.

[65] S. Barzegar, S.B. Wu, J. Noblet, M. Choct, R.A. Swick, Energy efficiency and net energy prediction of feed in laying hens, Poult. Sci. 98 (2019) 5746–5758.

[66] S. Cerrate, R. Ekmay, J.A. England, C. Coon, Predicting nutrient digestibility and energy value for broilers, Poult. Sci. 98 (2019) 3994–4007.

[67] B.S. Sheikhhasan, H. Moravej, M. Shivazad, F. Ghaziani, E. Esteve-Garcia, W.K. Kim, Prediction of the total and standardized ileal digestible amino acid contents from the chemical composition of soybean meals of different origin in broilers, Poult. Sci. 99 (2020) 4947–4957.

[68] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, New York, 2013.

[69] A. Faridi, A. Golian, M. Mottaghitalab, S. López, J. France, Predicting the metabolizable energy content of corn for ducks: A comparison of support vector regression with other methods, Spanish J. Agric. Res. 11 (2013) 1036–1043.

[70] H. Ahmadi, A. Golian, M. Mottaghitalab, N. Nariman-Zadeh, Prediction model for true metabolizable energy of feather meal and poultry offal meal using group method of data handling-type neural network, Poult. Sci. 87 (2008) 1909–1912.

[71] H. Ahmadi, M. Rodehutscord, Application of Artificial Neural Network and Support Vector Machines in Predicting Metabolizable Energy in Compound Feeds for Pigs, Front. Nutr. 4 (2017) 1–8.

2.2. Near-infrared spectroscopy and multivariate analysis

NIRS fundamentals

Near-infrared spectroscopy (NIRS) is a type of vibrational spectroscopy in the wavelength range from 700 to 2500 nm. Molecular vibrations produce the fundamental absorption bands in the mid-infrared (MIR) range. In the NIR region, the absorption bands are 10 to 1000 times weaker. They correspond to overtones whose frequencies are multiples of the frequencies of the fundamental absorption bands and to combinations of the fundamental absorption bands [1].

Figure 1 shows the typical location of the NIR absorption bands.

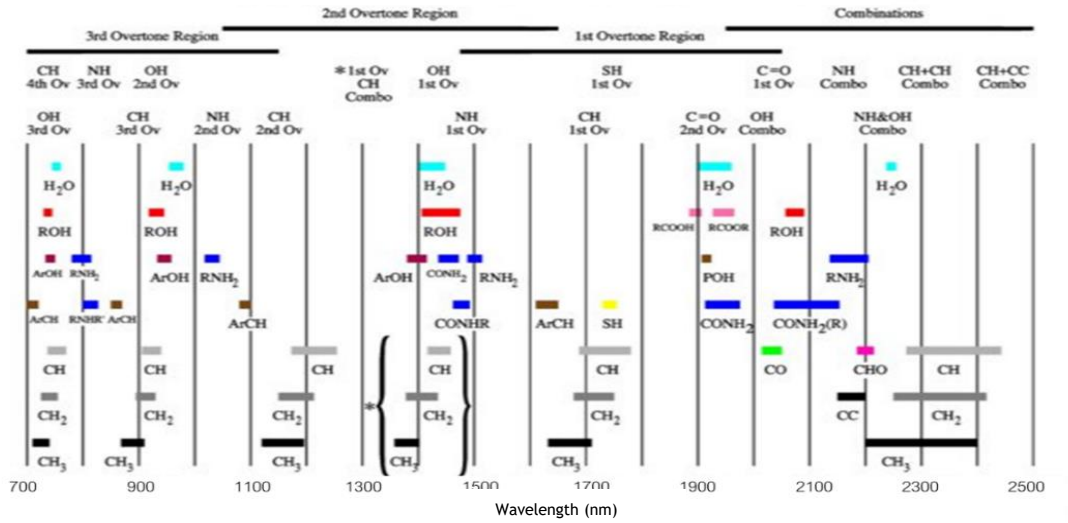


Figure 1. Near-infrared absorption bands (adapted from Harris and Altaner, 2013 [2]).

The NIR spectrum has three main overtone regions with a fourth being very weak and normally ignored. These overtones correspond to the absorptions of groups with elements attached to the hydrogen atom and from the C=O groups. The first case includes the absorptions of C-H (methyl, methylene, methoxy, aromatic and carbonyl associated), N-H (amides, amines and amine salts), O-H (water and alcohols) and S-H. Combination bands are vibrational combinations of these chemical groups formed by interactions between molecular vibrational frequencies, overlapped information from Fermi resonances and inactive MIR bounds among

other phenomena. A comprehensive account of the theory of NIR absorption bands is explained in references [1,3,4].

Measurement modes in NIRS

When NIR radiation hits a sample (Figure 2) the light is scattered (i.e., randomly redirected in all directions when it encounters a rough or particulate surface), transmitted through the material, reflected and/or absorbed. The proportions vary with the wavelength of the light and the composition and physical characteristics of the sample. Light reflection can be specular or diffuse. The specular reflection has only one direction and does not provide relevant chemical information since its contact with the sample is very low. Diffuse reflection occurs when light slightly penetrates the sample, scatters within the sample, and returns to the surface in any direction [5].

The two main measurement modes in NIRS are transmission and diffuse reflection. Absorbance is related to transmittance and diffuse reflectance as follows:

$$Absorbance = \log\left(\frac{P_0}{P}\right) = \log\left(\frac{1}{T}\right) \quad (1)$$

$$Absorbance = \log\left(\frac{1}{R_{relative}}\right) = \log\left(\frac{R_{reference\ material}}{R_{sample}}\right) \quad (2)$$

where P_0 is the initial radiation power, P is the radiation passing per unit area, T is transmittance and R is reflectance.

The NIR instrument used in this thesis works in diffuse reflection mode that is more appropriate than the transmission mode to analyse ingredients, compound feeds or animal faeces due to several reasons. Firstly, unlike in transmission mode the position of the sample and the changes in its geometry do not critically affect the measurements in reflection mode. This flexibility allows for easier handling of solid samples and better repeatability in the analysis of agri-food samples that can be very heterogeneous. Secondly, diffuse reflection induces less heat compared to transmission mode. This is particularly beneficial when analysing thick and dense feed samples. Moreover, the versatility of the reflection mode makes it well-suited for remote sensing

or field applications like the use of NIR sensors for on-site and real time monitoring of dry matter, yield, nitrogen, pest and diseases in cereal crops [6].

Transmission mode is used mainly for the analysis of liquids. In the agri-food sector, for example has been employed to determine nutritive parameters in cow's milk [7].

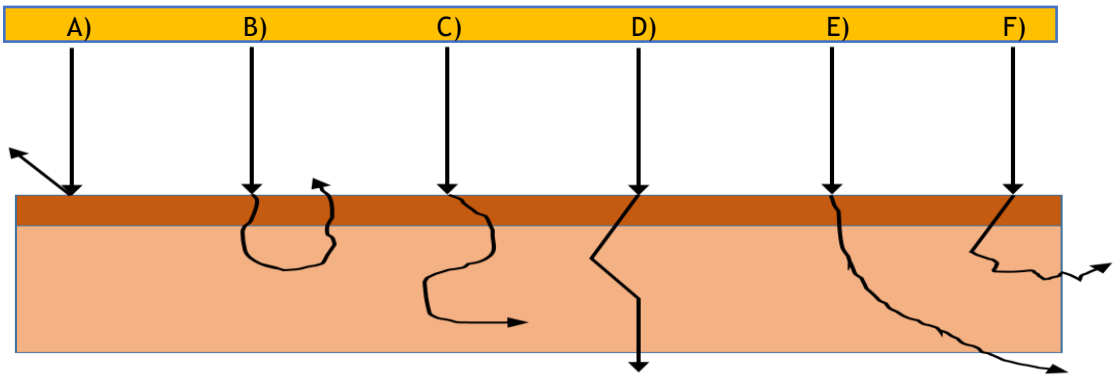


Figure 2. Interaction between solid particles in a sample and near-infrared radiation. A) Specular reflection, B) Diffuse reflection, C) Absorption, D) Transmission, E) Refraction, F) Scattering.

Instrumentation in NIRS

The three most common configurations of NIR instruments are pre-dispersive configuration, post-dispersive configuration and Fourier Transform (FT) [8]. Pre-dispersive and post-dispersive configurations (Figure 3) have a light source (usually a tungsten halogen lamp), a wavelength selector or monochromator (provided by a grating that diffracts the incident light at different degrees converting it in discrete wavelengths) and the detector (usually a silicon detector (Si) to work in the wavelength range between 400 and 1100 nm and a lead sulphide (PbS) detector to work between 1100 and 2500 nm, or a combination of both to cover a wide wavelength range). FT-NIR instruments do not have a grating since the measurements are made in time domain to obtain an interferogram that is processed with the Fourier transform to obtain the spectrum [9].

Pre-dispersive and FT-NIR configurations are commonly used in benchtop instruments because they offer high linearity, accuracy, and sensibility. On the other hand, post-dispersive configurations are preferred for in-line or portable instruments because they have faster

scanning times, thanks to a set of detectors placed in array. Additionally, they are less affected by external factors like vibrations or temperature because they have fewer moving parts compared to pre-dispersive or FT-NIR instruments. In this thesis, due to the type of samples (ingredients and compound feeds, excreta, faeces or ileal digesta) and the accuracy required in the determination of the nutritional parameters it has been used a pre-dispersive benchtop instrument able to work from 400 to 2500 nm.

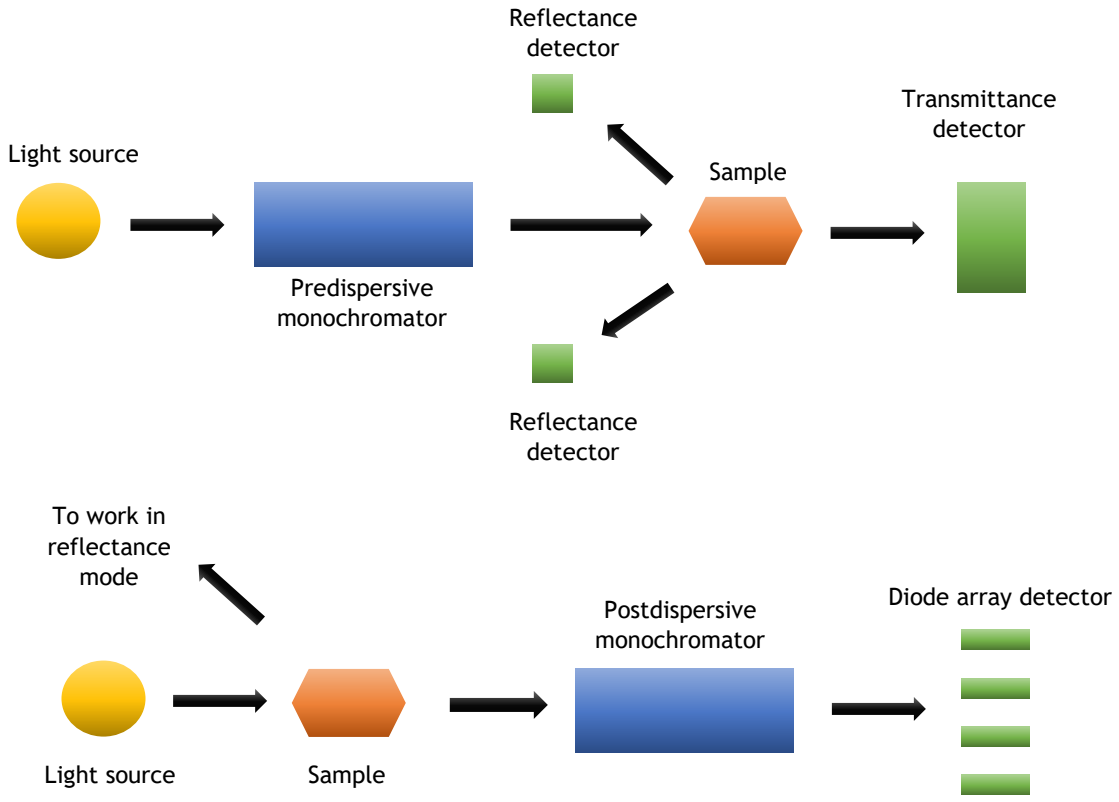


Figure 3. Pre-dispersive (above) and post-dispersive (below) configurations illustrating the position that the detectors would have in transmission or diffuse reflection mode.

Advantages and disadvantages of NIRS

Two of the main advantages of NIRS as analytical technique is that it is fast and non-destructive. Three factors contribute to the rapidness. First, scans are acquired quickly, which results in spectra being recorded sometimes in less than a minute and highly skilled personnel are not needed. Second, NIRS is a multi-parametric technique. This means that, for example, using NIRS we can determine simultaneously humidity, protein, fat and fibre in a feedstuff by recording only one spectrum while the alternative would be to use four different analytical

methods to determine these parameters. And third, sample treatments other than grinding or drying (and not always) are rarely needed since NIRS can measure solids in diffuse reflection mode. This is particularly beneficial when samples are solid, complex and challenging-to-handle, such as faeces or excreta. In addition, the absence of chemical treatments means that chemical reagents are not needed, thus reducing the cost and the environmental impact of the analysis. As NIRS is non-destructive it can also be used to perform on-line analyses during mixing of compound feeds to adjust the ingredient proportions to have the required chemical composition [10].

An important drawback of NIRS is its lower sensitivity compared to other analytical techniques such as HPLC or GC although the most common nutritional constituents (protein, fat, fibre, ash, starch,…) analysed to formulate a compound feed or to check the quality of an ingredient are not at low concentrations (their concentrations are in the order of g/100g). Other limitations are related to the calibration process. Firstly, the development and maintenance of the calibration models require skilled personnel. Secondly, databases containing a large number of spectra and reference values are needed to develop a reliable calibration. The obtaining of the spectra of the samples is fast but to obtain the reference values often slow and costly wet chemistry methods are needed. Furthermore, the maintenance and update of the calibrations might require the periodical analysis of extra samples. Lastly, transferring calibrations between instruments can be challenging.

Spectra and data analysis

The understanding and use of a NIR spectrum to analyse or classify samples based on their chemical features requires overcoming the inherent complexity associated with the interpretation of the spectrum. A NIR spectrum consists of several overlapped bands, with strong collinearity. In addition, several factors can produce baseline fluctuations, low signal-to-noise ratios and poor reproducibility. The most influential one for solid samples is light scattering, which is associated to variations in particle size. Other factors such as variations in temperature or humidity and instrument-related factors (e.g. instrumental drift) also affect [3]. To minimize the effect of these sources of variability, that are not related to the chemical composition of the sample, the spectrum can be mathematically corrected. Multivariate analysis

contribute also to minimize these effects and allow overcoming the lack of selectivity and the high information redundancy contained in the sample spectrum [5].

Multivariate analysis methods

Some multivariate methods that are used to relate a NIR spectrum and the composition of a sample assume a linear relationship between the analyte or property to be predicted and its absorbance following Lambert-Beer's law. In inverse multivariate calibration, the analyte concentration is modelled against the response as:

$$c_k = b_{0,k} + \sum_{j=1}^{J} r_j \, b_{j,k} + e_k \quad (3)$$

where c_k is the analyte concentration, $b_{0,k}$ the constant term, r_j the response at sensor j, $b_{j,k}$ is the model coefficient for the response at sensor j and e_k is the error. In matrix notation, the model for the training set can be written as (Figure 4):

$$\mathbf{c} = \mathbf{R}\mathbf{b} + \mathbf{e} \quad (4)$$

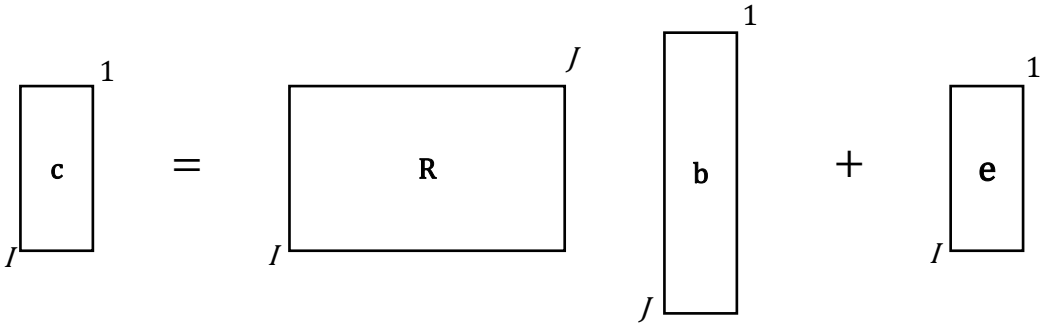


Figure 4. Equation 4 in matrix notation for the training set, where **c** is the vector of analyte concentrations for the *I* calibration samples, **R** is the matrix of spectra in rows, adapted for the form of the model (it has a row of ones at the beginning due to the constant term $b_{0,k}$), **b** is the vector of regression coefficients and **e** the error vector.

The regression coefficients vector can be obtained by:

$$\mathbf{b} = \mathbf{R}^{+}\mathbf{c} \quad (5)$$

where $\mathbf{R}^{+}$ is the pseudoinverse of **R**. In inverse least squares $\mathbf{R}^{+}$ is estimated as:

$$\mathbf{R}^{+} = (\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}} \quad (6)$$

that requires that the number of calibration samples *I* be greater than number of variables *J*, that is, the number of wavelengths used in the analysis. Using many wavelengths would require analysing a large number of samples, and, moreover, the collinearity among the wavelengths would degrade the quality of the estimated coefficients that is affected by the stability of the inverse of $\mathbf{R}^T\mathbf{R}$. It is said that the matrix $\mathbf{R}^T\mathbf{R}$ is ill-conditioned when its determinant is very close to zero, which happens due to the high collinearity among the wavelengths. In this case, small variations due to random error in the measurement of the spectrum can lead to big changes in $(\mathbf{R}^T\mathbf{R})^{-1}$and, in consequence, large variance in the estimated coefficients and the predictions. Furthermore, changes in the correlation structure within **R** would negatively affect the prediction of new samples that do not have the same correlation structure as the training samples [11].

There are several strategies to reduce the problem of collinearity in the spectral data. One of them is to combine inverse least squares with wavelength selection. The objective is to select a reduced number of wavelengths related to the analyte to be predicted. These methods have been recently reviewed by Yun et al. in 2019 [12]. It must be noticed that wavelength selection is also used with other modelling strategies because it may improve the predictive ability and interpretability of the model [13]. Other approaches for finding **b** are principal components regression (PCR) and partial least squares regression (PLSR), that decompose **R** in orthogonal factors. These methods approximate **R** as a product of a scores matrix, **T** and a loadings matrix **P** as follows:

$$\mathbf{R} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{7}$$

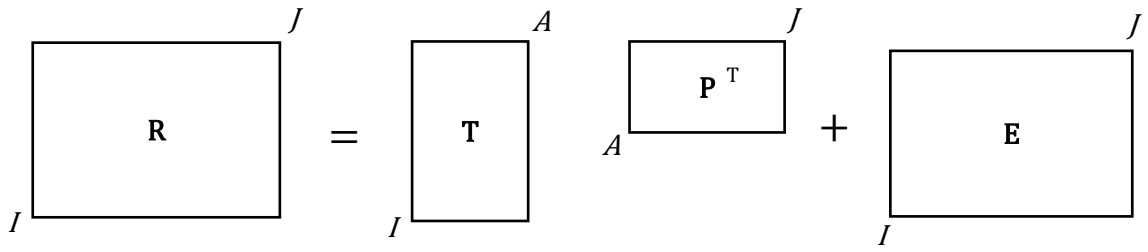


Figure 5. Equation 7 in matrix notation where **R** is the response matrix (the spectra), **T** the scores matrix, **P** the loadings matrix and **E** the error matrix.

The information in **R** is condensed in a smaller set of orthogonal variables called factors or latent variables in general and principal components in PCR. The scores capture the relationship between samples and the loadings the relationship between variables. The loadings can be used, for example to study the influence of the variables used to build the model (the wavelengths) and the scores to detect outliers or clusters within the dataset.

There are two main diagnostics for outlier detection in multivariate models [14]: the leverage, that measures the distance of a sample to the centre of the model, and the spectral residual that measures the orthogonal distance between a sample and the model. In this thesis we use Hotelling's T^2 as a measure of leverage and Q residuals for the spectral residuals [15].

Hotelling's T^2 represents a measure of the variation in each sample within the model and for PCR is calculated as follows:

$$T_i^2 = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{t}_i^{\mathrm{T}} = \mathbf{x}_i \mathbf{P}_{\mathrm{A}} \boldsymbol{\lambda}^{-1} \mathbf{P}_{\mathrm{A}}^{\mathrm{T}} \mathbf{x}_i^{\mathrm{T}} \quad (8)$$

where T_i^2 is the value of the Hotelling's T^2 statistic for the i sample, $\mathbf{t}_i$ is the i^{th} row of the scores matrix $\mathbf{T}$, $\boldsymbol{\lambda}$ is the matrix containing the eigenvalues corresponding to A factors, $\mathbf{x}_i$ is the response vector for the i sample and $\mathbf{P}_{\mathrm{A}}$ is the matrix of loadings for A factors.

Q residuals are a lack-of-fit statistic calculated as follows:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}} = \mathbf{x}_i \left(\mathbf{I} - \mathbf{P}_{\mathrm{A}} \mathbf{P}_{\mathrm{A}}^{\mathrm{T}}\right) \mathbf{x}_i^{\mathrm{T}} \quad (9)$$

where Q_i is the value of the Q residuals statistic for the i sample, $\mathbf{e}_i$ is the i^{th} row of the error matrix $\mathbf{E}$, $\mathbf{P}_{\mathrm{A}}$ is the matrix of A loadings and $\mathbf{x}_i$ is the response vector for the i sample.

A sample whose spectrum has high Q_i may show the presence of features that were not accounted for in the calibration set, whereas a sample whose spectrum has a high Hotelling's T^2 may indicate the presence features that, even if accounted for in the calibration set, are more extreme.

References [16–18] show how PCR and PLSR decompose **R** and calculate $\mathbf{R}^+$. PCR decomposes **R** in an unsupervised way and generates principal components, which are the directions that explain most of the variability in **R**. These directions, however, may not be the best for predicting the analyte concentration. PLSR find directions that better explain both the response and the predictors so that the scores are as closely correlated with the response as possible. Because of this, PLSR generally uses less factors than PCR to achieve a similar predictive ability [19]. Nowadays PLSR is the preferred method for NIR calibration. The factors in PLSR are ranked. The first factors describe mainly systemic variation which is necessary to predict. The subsequent factors describe more and more random variation in the spectrum. One must include enough factors to capture most of the systemic variation while avoiding the inclusion of factors that model too much noise [20]. Cross-validation (CV) is one of the most widely used methods to estimate the optimal number of factors and can be used to guide wavelength selection and verify the performance of a model when a test set is not available. This approach consists of randomly dividing the data set into k groups, or folds, that have the same size. All folds except one are used to create the model, the remaining fold is used to validate it. This process is repeated k times using a different fold each time, and the average prediction error is calculated. When k is equal to the number of samples, the method is called leave-one-out cross-validation (LOOCV) [21].

PCR and PLS can accommodate non-linearities to some degree, but their performance degrades [22]. When non-linearities predominate, calibration methods such as Artificial Neural Networks (ANN) are recommended [23,24]. The application of ANN and other non-linear algorithms to the NIR spectrum in food analysis was reviewed by Zareef et al. in 2020 [25].

Calibration process

There are five stages in the development and use of a NIR-based calibration model once the spectra have been obtained: selection of samples, pretreatment of the spectra, calculation of the multivariate model, validation of the model and maintenance of the calibration. The process is iterative and it is usual to come back to previous stages and modify the initial settings. For example, once the model has been calculated, some samples may be detected as outliers and must be removed and the model recalculated.

The objective of sample selection is to obtain a calibration set that is representative of the chemical and physical characteristics of the population being analysed [5]. Typically, principal components analysis (PCA) or clustering are used to find groups of samples and detect spectral outliers. If samples are clearly separated in a few groups, then one may attempt specific calibrations for each group, and these may perform better than a single calibration with the whole dataset. Samples far away are considered spectral outliers and removed prior to sample selection and model calculation, as they would alter these stages and the predictive ability of the model. When the dataset is large and/or when the samples have similar characteristics (which translated into a low spectral variance) some samples might provide redundant information to the model. In those cases, analysis of the entire dataset by reference methods may be unnecessary and algorithms can be applied to select a small subset of samples that is representative of the dataset [26]. Two widely used algorithms with this purpose are Kennard-Stone's algorithm [27] and the duplex [28] algorithm. They try to uniformly cover the multidimensional spectral space by selecting the samples with the maximum distance (commonly Euclidean distance) between the selected samples. Sample selection is discussed in more detail in Chapter 4 in the context of model updating.

Spectra are usually preprocessed to minimize the effects of the sources of spectral variability that are not related to the property being predicted, and filter noise [29]. Rinnan et al. 2009 [30] reviewed some common pretreatments used in NIRS. They can be divided into two groups: methods that minimize the light scattering such as standard normal variate (SNV) [31] or multiplicative signal correction (MSC) [32] and methods that minimize the baseline variations such as Savitzky-Golay (SG) polynomial derivative filters [33]. Methods that minimize the scattering also can correct baseline shifts. SNV centres and scales each spectrum individually to have a mean equal to 0 and a standard deviation equal to 1. MSC depends on the full dataset. The average spectrum is regressed against each spectrum and the slope and intercept are used to correct that spectrum. The slope represents the multiplicative correction factor and the intercept the additive effect of light scattering. The derivative methods use a smoothing of the spectra before the derivative is calculated in order to decrease the detrimental effect on the signal-to-noise ratio that derivatives would have. The first derivative removes only the baseline while the second removes also the linear trend. SG uses a polynomial function of a specific degree to smooth a window of spectral points. When the parameters of this polynomial are calculated, the derivative of any order of this function is applied to the central point of the window. This operation is applied to each point in the spectrum [30]. The polynomial degree

and the window size need to be carefully selected. If high-degree polynomials and small window sizes are used, the function can fit very well the data, but the noise is barely reduced, leading to poor smoothing. Conversely, a low-degree polynomial and a large window size can smooth the data too much and information-bearing spectral features dissipates. It is not frequent to apply higher-degree derivatives other than the second. Other spectral preprocessing methods less used in NIRS are those that make use of the y-values in addition to the spectra. Orthogonal signal correction (OSC) [34] and generalized least squares weighting (GLSW) [35] are methods that remove the spectral information that is orthogonal to variations of the property values.

The next stage after preprocessing is the calculation of the multivariate model. As explained, PLS is the most used regression model in NIRS and specifically to determine the nutrient content of feedstuffs, compound feeds or animal faeces. Every model has a corresponding calibration error that is a measure of how well the model fits the calibration data. It quantifies the differences between the predicted values obtained from the model and the reference values. Although this statistic is not optimal to assess the performance of the model since calibration data might be overfitted, it can be used to find reference outliers (those samples that have a much larger calibration error than the average of the samples). Cross-validation (CV) allows the model to be tested on data that it has not been trained on and therefore the CV error provides a more realistic assessment than the calibration error. However, the best way to verify the actual performance of a model is to use a test set. This test set can be selected during the sample selection stage or can be created with new samples measured once the model is calculated.

The most commonly used statistics to check the model performance are the coefficient of determination (R^2),

$$R^2 = \frac{\left(\sum_{i=1}^{n} \hat{y}_i y - \sum_{i=1}^{n} \hat{y}_i \sum_{i=1}^{n} y_i / n\right)^2}{\left(\sum_{i=1}^{n} \hat{y}_i{}^2 - \left(\sum_{i=1}^{n} \hat{y}_i\right)^2 / n\right)\left(\sum_{i=1}^{n} y_i{}^2 - \left(\sum_{i=1}^{n} y_i\right)^2 / n\right)} \tag{10}$$

the standard error of prediction (SE),

$$\mathrm{SE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i - bias)^2}{n-1}} \tag{11}$$

and the root mean standard error of prediction (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (12)$$

where $\hat{y}_i$ is i^{th} validation sample predicted value, y_i is i^{th} validation sample reference value and n is the number of samples in validation set.

To a lesser extent, the ratio of performance of deviation or relative predictive determinant (RPD) and bias (b).

$$\mathrm{RPD} = \frac{\mathrm{s_y}}{\mathrm{SE}} \quad (13)$$

$$b = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \quad (14)$$

where s_y is the standard deviation of reference values from the validation set.

NIRS repeatability is usually omitted because actual NIR instruments are very repeatable with homogenous samples. When measuring heterogeneous samples, such as whole grain or wet forage, repeatability may decrease and it is recommended to use more than one subsample [36]. RMSE, SE and bias have the same units as reference values while R^2 and RPD are unitless.

R^2 is an estimate of the percentage of explained variance of the reference values. SE provides information about the precision of the model. Since it is corrected for the bias, if this statistic is reported the bias must be reported as well. RMSE is similar to SE but accounting for bias ($\mathrm{RMSE}^2 = \mathrm{SE}^2 + \mathrm{b}^2$). When bias is low, RMSE≈SE. RPD is related to the capacity of the model to predict future samples in relation to the initial variability of the data [5]. All these statistics can refer to the calibration data, in which case are typically written as R_C^2, SEC, RMSEC, $\mathrm{RPD_C}$; to cross-validation (R_{CV}^2, SECV, RMSECV, $\mathrm{RPD_{CV}}$) or to an external validation set (R_P^2, SEP, RMSEP, $\mathrm{RPD_P}$).

The next section reviews applications of NIRS in monogastric animal nutrition. To compare results of published works we use the coefficient of determination of cross validation (R_{cv}^2) and the coefficient of variation of cross validation ($\mathrm{CV_{cv}}$) reported by those authors or calculated

from their data (the standard error of cross validation (SECV) and the mean of the calibration set). CV_{cv} enables correction for differences in the means of the calibration sets used in the various studies [38]. Because of RPD is fairly close related with R^2 [37] for simplicity we used only R^2. RPD_{CV} is only shown in a few cases when R^2_{cv} was not reported. Even though, criteria to assess the quality of a calibration cannot be generalized to all types of products or all NIR instruments, according to Shenk & Westerhaus, 1995 [39] the fit of a calibration may be considered excellent if R^2>0.90, good if R^2=0.70-0.90 and insufficient if R^2<0.70. For RMSEP, SECV or CV the criteria is that it should not be much larger than the SEC and than the standard error of the laboratory method used to obtain the reference data (SEL). If RMSEP or SECV are much larger than SEL then the model is not modelling well all the sources of variation and if they are much larger than SEC probably the model is overfitted and it cannot predict properly new samples [40].

According to ISO 12099:2017 [41], to accept a NIR calibration model it must pass three statistical tests: a t-test on bias, a t-test on slope and a F-test on SEP. The aim of these tests is to affirm with a specific level of significance (typically α=0.05) that the bias can be considered negligible, that the slope of the predicted vs measured line can be considered one and that the RMSEP can be considered similar to the SEC.

A successfully validated model should ideally be used to predict new samples for an extended period. However, samples and instruments do not necessarily remain stable over time. Unseen variability in samples or measurements can affect the predictions and render calibrations invalid. Model maintenance have to be periodically applied to preserve the predictive capability and improve the models over time with the least amount of cost and effort [42]. This topic will be discussed in Chapter 4.

References

[1] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-infrared spectroscopy. Principles, instruments, applications, Wiley-VCH, Weinheim, Germany, 2002.

[2] P.J. Harris, C.M. Altaner, Workshop on commercial application of IR spectroscopies to solid wood, Wood Technology Research Centre, University of Canterbury, Christchurch, New Zealand, 2013.

[3] Y.O. Ozaki, W.F. McClure, A.A. Christy, Near-infrared spectroscopy in food science and technology, Jhon Wiley & Sons Inc., Hoboken, New Jersey, 2007.

[4] P.C. Williams, K. Norris, Near Infrared Technology in the Agricultural and Food Industries, 2nd Edition, American Association of Cereal Chemists, St Paul, MM, USA, 2001.

[5] L.E. Agelet, C.R. Hurburgh, A tutorial on near infrared spectroscopy and its calibration, Crit. Rev. Anal. Chem. 40 (2010) 246–260.

[6] D. Cozzolino, K. Porker, M. Laws, An Overview on the Use of Infrared Sensors for in Field, Proximal and at Harvest Monitoring of Cereal Crops, Agriculture. 5 (2015) 713–722.

[7] M.F. Laporte, P. Paquin, Near-infrared analysis of fat, protein, and casein in cow's milk, J. Agric. Food Chem. 47 (1999) 2600–2605.

[8] D.I. Givens, J.L. De Boever, E.R. Deaville, The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans, Nutr. Res. Rev. 10 (1997) 83–114.

[9] C. Pasquini, Near infrared spectroscopy: A mature analytical technique with new perspectives – A review, Anal. Chim. Acta. 1026 (2018) 8–36.

[10] E. Fernández-Ahumada, J.E. Guerrero-Ginel, D. Pérez-Marín, A. Garrido-Varo, Near infrared spectroscopy for control of the compound-feed manufacturing process: Mixing stage, J. Near Infrared Spectrosc. 16 (2008) 285–290.

[11] R.H. Myers, Classical and modern regression with applications, 2nd ed., Duxbury Press, Belmont, USA, 1990.

[12] Y.H. Yun, H.D. Li, B.C. Deng, D.S. Cao, An overview of variable selection methods in multivariate analysis of near-infrared spectra, TrAC - Trends Anal. Chem. 113 (2019) 102–115.

[13] C.M. Andersen, R. Bro, Variable selection in regression-a tutorial, J. Chemom. 24 (2010) 728–737.

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

[15] J.E. Jackson, A User's Guide to Principal Components, Jhon Wiley & Sons Inc., Hoboken, USA, 1991.

[16] W.G. Glen, W.J. Dunn, D.R. Scott, Principal components analysis and partial least squares regression, Tetrahedron Comput. Methodol. 2 (1989) 349–376.

[17] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.

[18] R. Kramer, Chemometric techniques for quantitative analysis, CRC Press, Boca Raton, USA, 1998.

[19] E. V. Thomas, D.M. Haaland, Comparison of multivariate calibration methods for quantitative spectral analysis, Anal. Chem. 62 (1990) 1091–1099.

[20] J. Ferré, Calibración multivariante en análisis cuantitativo: el modelo inverso, Universitat Rovira i Virgili, Tarragona, Spain, 2006.

[21] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, New York, 2013.

[22] E. V. Thomas, A Primer on Multivariate Calibration, Anal. Chem. 66 (1994) 795–804.

[23] R. Leardi, Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks, Oxford, UK, 2003.

[24] K. Suzuki, Artificial Neural Netwokrs-Architectures and Applications, InTech, Rijeka, Croatia, 2013.

[25] M. Zareef, Q. Chen, M.M. Hassan, M. Arslan, M.M. Hashim, W. Ahmad, F.Y.H. Kutsanedzie, A.A. Agyekum, An Overview on the Applications of Typical Non-linear Algorithms Coupled With NIR Spectroscopy in Food Analysis, Food Eng. Rev. 12 (2020) 173–190.

[26] J.S. Shenk, M.O. Westerhaus, Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy, Crop Sci. 31 (1991) 469–474.

[27] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics. 11 (1969) 137–148.

[28] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428.

[29] N. Shetty, Å. Rinnan, R. Gislum, Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses, Chemom. Intell. Lab. Syst. 111 (2012) 59–65.

[30] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC - Trends Anal. Chem. 28 (2009) 1201–1222.

[31] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[32] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[33] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[34] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[35] B.M. Zorzetti, J.M. Shaver, J.J. Harynuk, Estimation of the age of a weathered mixture of volatile organic compounds, Anal. Chim. Acta. 694 (2011) 31–37.

[36] P. Dardenne, Some Considerations about NIR Spectroscopy: Closing Speech at NIR-2009, NIR News. 21 (2010) 8–14.

[37] T. Fearn, Assessing Calibrations: SEP,RPD,RER,R2, NIR News. 13 (2002) 12–14.

[38] A. Garrido-Varo, S. Vega, F. Maroto-Molina, M.J. De La Haba, D. Pérez-Marín, On-site quality control of processed land animal proteins using a portable micro-electro-mechanical-systems near infrared spectrometer, J. Near Infrared Spectrosc. 24 (2016) 47–58.

[39] J. Shenk, M. Westerhaus, Routine Operation, Calibration, Develpment and Network System Management Manual. NIRSystems, Silver Spring, MD, 1995.

[40] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[41] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[42] B.M. Wise, R.T. Roginski, A calibration model maintenance roadmap, IFAC-PapersOnLine. 48 (2015) 260–265.

2.3. Near-infrared spectroscopy in monogastric nutrition

History

The first study about the NIR region dates back to 1800 when Herschel dispersed the sunlight using a glass prism and measured the rising of the temperature beyond the red end of the visible spectrum. During the following decades, MIR was the infrared region that received most attention from the scientific community and the NIR region was considered to lack relevant chemical information. In 1960s, Karl Norris of the United States Department of Agriculture began to study the optical properties of dense, light scattering biological materials [1] and he and his team were capable to determine the moisture content from seed extracts using the NIR spectrum and multivariate calibration [2]. The use of multivariate analysis to obtain information of the spectra promoted NIR region to have practical uses. Firstly, in the grain industry where Williams in Canada in 1975 [3] and Hunt et al. in USA in 1977 [4] gave NIR spectroscopy the status of official protein testing method for wheat. Afterwards, its expansion continued with methods to determine protein and oil in ground and whole soybean and protein in small grains. In the agriculture and animal feed sector, NIRS is currently consolidated as alternative to wet chemistry methods since it allows fast determinations with hardly any sample treatment.

Predicting the nutrient content of feedstuffs and diets by NIRS

After the first NIRS practical applications [2–4], in the ninety's, companies and public organizations began to develop NIR calibrations for feedstuffs using large datasets [5]. Nowadays, NIR instrument manufacturers (e.g. FOSS) and specialized companies (e.g. Evonik) sell calibration models for the most frequently determined parameters in the agri-food industry (e.g. dry matter, ash, fat, crude protein and crude fibre). In the market, there are also handheld NIR instruments for quality control of feedstuffs and diets. Although handheld instruments are not as accurate as benchtop instruments, they are cheaper and enable in-situ determinations, eliminating the need of sample collection, conservation and transport to the laboratory [6]. Haughey et al. 2015 [7] and Modroño et al. 2017 [8] showed the benefits of using handheld instruments instead of benchtop instruments for the proximate analysis of compound feeds.

Table 1 summarizes the publications about the use of NIR calibrations for the proximate analysis of feedstuffs and diets from 2000 to the present. In addition to dry matter (DM), ash, fat, crude protein (CP) and crude fibre (CF), the nutritional parameters traditionally determined in feedstuffs and diets, other important parameters for animal nutrition have also been studied such as gross energy (GE), acid detergent fibre (ADF), neutral detergent fibre (NDF) or phosphorus (P).

Table 1. Near-infrared (NIR) calibration models to predict the nutrient content of feedstuffs and diets for swine and poultry described in the literature from 2000. Type of sample, analyte or constituent predicted, coefficient of determination cross validation (R^2_{cv}), coefficient of variatio of cross validation (CV_{cv}) and reference (Ref).

Sample	Analyte	R^2_{cv}	CV_{cv} (%)	Ref
Wet corn	DM, CP, Ash, ADF	0.75, 0.72, 0.81, 0.88	5.3, 6.9, 19.1, 21.4	[14]
Wet corn	DM, CP, ADF, NDF	0.85, 0.91, 0.86, 0.84	7.9, 8.5, 7.5, 10.3	[15]
Corn	DM, CP, Ash, Fat, ADF, NDF	0.98, 0.96, 0.95, 0.93, 0.92, 0.92	0.32, 2.0, 11.4, 9.6, 28.0, 2.6	[16]
Corn	CP, Ash, Fat	0.99, 0.81,0.72	4.4, 11.7, 13.7	[17]
Wheat	DM, CP, Ash, Fat, GE	0.93, 0.88, 0.93, 0.65, 0.47	0.29, 4.1, 3.5, 5.1, 0.59	[11]
Wheat	CP, NDF, GE	0.98, 0.50, 0.76	1.6, 6.4, 0.28	[9]
Cereal and co-products	CP, Ash, Fat, CF, NDF, ADF	1.00, 0.97, 0.95, 0.95, 0.96, 0.91	3.8, 10.3, 23.5, 12.2, 11.0, 16.7	[18]
Oil seeds and by-products	CP, Ash, Fat, CF, NDF, ADF, GE	0.99, 0.79, 0.99, 0.98, 0.98, 0.98, 0.99	1.9, 6.7, 11.3, 8.2, 7.9, 9.7, 0.77	[19]
Soybean	CP, Ash, Fat	0.81, 0.63, 0.71	5.9, 7.8, 6.9	[20]
Rapeseed	DM, CP	0.95, 0.93	0.33, 1.6	[21]
Barley	DM, CP	0.95, 0.98	0.39, 2.0	[22]
Triticale	DM, CP	0.95, 0.98	0.39, 2.0	[22]
Sunflower	DM, CP	0.97, 0.96	0.28, 2.8	[21]
Field peas	DM, CP	0.97, 0.94	0.34, 2.2	[21]
Fish meal	CP, Ash, Fat	0.84, 0.90, 0.90	0.86, 2.6, 4.6	[12]
Meat meal	DM, CP	0.94, 0.96	0.33, 2.3	[21]
Poultry meal	DM, CP	0.87, 0.92	0.39, 2.6	[21]
Fava bean	CP, Fat	0.94, 0.66	1.2, 14.2	[10]
Broiler diets	Moisture, CP, CF, Fat, Ca, P	0.84, 0.95, 0.82, 0.95, 0.75, 0.95	2.0, 2.1, 10.0, 5.6, 7.0, 4.5	[23]
Laying hens	Moisture, CP, CF, Fat, Ca, P	0.89, 0.95, 0.82, 0.92, 0.73, 0.91	2.1, 3.4, 10.0, 6.7, 17.9, 8.0	[23]
Ostrich diets	DM, CP, Ash, Fat, CF, ADF, NDF, GE	0.57, 0.97, 0.67, 0.89, 0.94, 0.89, 0.95, 0.80	0.31, 5.3, 13.6, 18.2, 8.0, 11.7, 7.5, 1.8	[24]
Poultry diets	DM, CP, Fat, CF, Ash	5.5, 4.2, 6.3, 4.5, 5.8, 2.6 [1]	0.35, 3.6, 6.1, 8.3, 15.2	[25]
Poultry diets	Moisture, CP, Fat, CF, Ash	0.88, 0.98, 0.94, 0.94, 0.79		[26]
Pig diets	Moisture, CP, Fat, CF, Ash	0.82, 0.97, 0.86, 0.95, 0.91		[26]
Pig diets	DM, CP, Fat, CF, ash	0.96, 0.90, 0.95, 0.96, 0.81	0.23, 3.2, 7.1, 9.4, 5.9	[27]
Pig and poultry diets	CP, CF	0.95, 0.96	3.9, 11.5	[28]

[1] RPD_{cv} values instead of R^2_{cv}

CP and DM are commonly the best predicted parameters in feedstuffs and compound feeds. The coefficient of variation of cross validation (CV_{cv}) was in most of the cases lower than 4% for CP with several models yielding values lower than 2%. For DM the values were lower than

1%. The coefficients of determination of cross validation (R^2_{cv}) were higher than 0.8 and higher than 0.9 in many cases. It should be noted that the worst predictions were obtained in high moisture corn and wet whole maize. This highlights that a high humidity in the samples worsens the predictions. The values given in the table from the Owens et al. 2009 work [9] were in dry wheat samples. They predicted the same parameters in undried wheat samples, obtaining worse results. The differences in particle size between samples can also affect the predictions, in most cases the samples are ground before the analysis. Owens et al. 2009 [9] and Wang et al. 2014 [10] obtained better predictions in powder than in intact seeds. The quality of the ash predictions varied among authors. Garnsworthy et al. 2000 [11] and Cozzolino et al. 2002 [12] predicted ash very well, with low CV_{cv} and high R^2_{cv} but the rest of the predictions were poor. Ash is difficult to predict because it does not have characteristic NIR absorption bands. Its prediction is possible because the mineral material is correlated with organic components and form salts that affect the hydrogen bonds in the samples [13]. The fat content or ether extract was better predicted in oil seeds, fishmeal and in compound feeds than in cereals due to the low average values and narrow interval that cereals have. CF predictions in general had good determination coefficients (R^2_{cv}>0.90) but high errors (CV_{cv}≥10%), probably due to the reference method to determine CF that involves two successive digestions [29]. Something similar occurs in most cases when ADF or NDF are used instead of or in addition to CF. GE is not a chemical species, but it can be predicted by NIRS because it is highly correlated with some nutrients, especially fat content and starch. GE was very well predicted in terms of CV_{cv} values, but the R^2_{cv} values were low when the dataset contained only one type of feedstuff. This could be due to the small variability in energy that feedstuffs such as wheat have. One major constraint of R^2 is its dependence on the range of reference values in the calibration set and their uncertainty [30]. Losada et al. 2010 [19] used a dataset with different oil seed and by-products and Kays & Barton, 2002 [31] (not shown in the table) included different cereal species. With this larger variability, their calibration models had very good CV_{cv} and R^2_{cv}. Tahir et al. [32] developed specific total P and phytic P calibrations for corn, soybean meal, DDGS or wheat, among others. The results were better for phytic P (R^2_p>0.70) than for total P (R^2_p<0.70). This could be expected because phytate is an organic molecule with C-H and O-H bonds that show up in the NIR region while the determination of total P, like the determination of ash is related to the correlations between the constituent and the organic molecules. Aureli et al. [33] predicted total P and phytic P by merging several feedstuffs (cereals, cereal by-products, legumes seeds and oil meals) into a single dataset. They obtained more robust models with higher determination coefficients (R^2_{cv}>0.90). The errors were high (CV_{cv}>15%) compared to the total P predictions made by Khaleduzzaman

et al. 2017 [23] in broiler and laying hen diets (CV_{cv}<10%). It should be considered that the amount of P and phytic P in some feedstuffs is very small (e.g., total P<0.30% and P-phytic<0.20% in corn and wheat) while the concentration of P in a diet is normally higher than 0.50% and can be higher than 1%. Generally, the lower the concentration, the greater the reference method error and the more difficult it is to obtain accurate NIR predictions. In the same work, Khaleduzzaman et al. predicted also calcium with poor results. Ikoyi & Younge 2020 [34] tried to predict minerals in plant materials but obtained a low accuracy (R^2_{cv}<0.75 and CV_{cv}>10%). This is attributed to the fact that minerals do not absorb in the NIR region. The addition of correctors with vitamins to animal feeds is a common practice. González-Martín et al. 2005, 2006 [35,36] determined vitamin E in the form of α-, (β+γ)- and δ-tocopherol obtaining acceptable results in compound feeds (R^2_{cv}>0.75 and CV_{cv}<25%) and good results for alfalfa (R^2_{cv}>0.90 and CV_{cv}≈10%) considering the low concentration of these species (<1 g/kg).

Prediction of amino acids in feedstuffs and diets by NIRS

The first studys focused on the prediction of the limiting AAs for swine and poultry (See Section 2.1). Rubenthaler and Bruinsma 1978 predicted Lys in a variety of cereals [37] and Gill et al. 1979 in barley of different breeds [38]. Williams et al. 1984 [39] predicted Lys, Met, Thr and Trp in wheat. The prediction errors were relatively low (CV_{cv}<8%) for all except Met (CV_{cv}=13%). In the same work, they predicted 14 AAs in barley with high coefficients of determination for all of them (R^2_{cv}>0.90) except again Met (R^2_{cv}=0.81). The reason could be the narrow range of Met concentration in barley and the higher error of the Met analysis due to oxidative losses during sample preparation. The following year they were more successful in predicting Met (R^2_{cv}>0.90) in peas [40]. Table 2 shows the coefficient of determination (R^2_{cv}) of AA calibrations published in the last 20 years.

The quality of the AAs predictions can vary due to different factors. An important one is the correlation between AAs and protein. This relationship is well known and served to build linear or non-linear regression models that predicted relatively well highly correlated AAs with protein in feedstuffs using the protein content as predictor [41,42]. Fontaine et al. 2001, 2002 [21,22] and Kovalenko et al. 2006 [43] developed linear regression models from the protein content and compared them with NIR models. They showed that NIR models outperformed linear

regression models from protein and that the NIR predictions were better the higher the correlation between the AA and protein.

Table 2. Near-infrared (NIR) calibration models to predict amino acids in feedstuffs and diets for swine and poultry described in the literature from 2000. Type of sample, coefficient of determination of cross validation (R^2_{cv}) and reference (Ref).

Sample	R^2_{cv}<0.70	0.70<R^2_{cv}<0.90	R^2_{cv}>0.90	Ref
Soybean seeds	Asp, Ser, Ala, Met, Cys, Lys, His	Glu, Thr, Gly, Pro, Val, Arg, Ile, Leu, Phe, Tyr		[47]
Soybean meal		Met, Cys	Lys, Thr, Trp, Arg, Ile, Leu, Val	[21]
Rapeseed meal	Thr	Met, Cys, Lys, Arg, Ile, Val, Trp	Leu	[21]
Sunflower		Met, Cys, Lys	Thr, Trp, Arg, Ile, Leu, Val	[21]
Field peas	Met, cys	Lys, Thr, Trp, Ile, Leu, Val	Arg	[21]
Wheat		Lys, Trp, Arg	Met, Cys, Thr, Ile, Leu, Val	[22]
Barley		Met, Cys, Lys, Trp	Thr, Arg, Ile, Leu, Val	[22]
Corn		Met, Cys, Lys, Trp, Arg	Thr, Ile, Leu, Val	[22]
Triticale		Met, Cys, Lys, Trp	Thr, Arg, Ile, Leu, Val	[22]
Wheat bran		Cys, Trp	Met, Lys, Thr, Arg, Ile, Leu, Val	[22]
Rice bran		Cys	Met, Lys, Thr, Trp, Arg, Ile, Leu, Val	[22]
Sorghum		Met, Cys	Lys, Thr, Trp, Arg, Ile, Leu, Val	[22]
Milled rice	Cys, Met, His	Thr, Ser, Ala, Val, Ile, Tyr, Lys, Pro	Asp, Glu, Gly, Leu, Phe, Arg	[48]
Soybeans	Trp, Cys, Met, Ser	Ala, Glu, Ile, Pro, Thr, Val, Arg, Asp, Gly, His, Leu, Lys, Phe, Tyr		[43]
Soybeans	Ala, Cys, Gly, His, Lys, Met, Trp, Tyr	Arg, Asp, Glu, Ile, Leu, Phe, Pro, Ser, Thr, Val		[49]
Animal diets	Lys	Asp, Glu, Gly, Pro, Val	Phe, Tyr	[50]
Poultry diets		Lys, Met+Cys		[51]
Brown rice	Cys, Met	Lys, Pro, Tyr	Asp, Thr, Ser, Glu, Gly, Ala, Val, Ile, Leu, Phe, His, Arg	[52]
Rapeseed	Val	Cys, Met, Pro, Ala, Tyr	Asp, Thr, Ser, Glu, Gly, Ile, Leu, Phe, Lys, His, Arg	[53]
DDGS	Arg, Trp, Val	His, Ile, Lys, Thr, Phe	Leu, Met	[54]
Quality protein maize	Ser, His, Thr, Cys, Met, Lys, Ile	Glu, Gly, Pro, Tyr, Met, Phe	Asp, Arg, Ala, Leu	[55]
Cereals	Lys, Met, Leu, Trp,	Phe, Glu, Gly, Pro, Ser, Tyr, Cys, Thr, Ile, His, Val, Arg, Ala, Asp		[56]
Supplemental feed ingredients		Trp	Lys, Met, Cys, Thr, Ile, Leu, His, Phe, Val, Arg, Ala, Asp, Glu, Gly, Pro, Ser, Tyr	[56]
Pig diets	Lys, Met, Thr, His, Arg, Tyr	Ile, Leu, Phe, Val, Trp, Ala, Asp, Gly, Cys, Glu, Pro, Ser		[56]

The accuracy of the reference method of analysis is another factor that could affect the accuracy of the predictions. The official method of analysis consists of an acid hydrolysis of proteins followed by HPLC analysis [44]. Met and Cys are determined separately, and a peroxidation step is carried out before the hydrolysis. This extra step increases the standard error of the laboratory

(SEL) of Met and Cys as it can be seen in references [21,22,43]. Furthermore, according to Fontaine et al. [21,22] the chromatographic peak of Cys usually suffers from baseline interferences that make the integration of the peak difficult. Hence, the poor predictions of Met and Cys in most of the presented works is not unexpected. To determine Trp, an alkaline hydrolysis instead of acid hydrolysis is carried out [45]. Although this does not increases the SEL significantly, the predictions for Trp are generally worse than for other AAs. The low concentration and range of values in most feedstuffs (it is the AA least abundant) could be the reason. The rest of the AAs could exhibit also large errors due to incomplete hydrolysis in the case of the stable AAs (the aliphatic) and degradation in the case of unstable AA (those with functional groups) [46]. However, as Table 2 shows many AAs have been predicted well (R^2_{cv}>0.90) or good (0.70< R^2_{cv}<0.90) in many feedstuffs.

As stated in Section 2.1 the trend now to avoid unnecessary costs and reduce environmental impacts is to formulate based on available CP and AAs instead of totals. Very few authors have developed NIR calibrations for digestible AAs probably due to the high cost of the in-vivo assays needed to obtain the reference values. Van Kempen & Bodin 1998 [57] predicted the ileal digestibility of Lys, Met and Thr. It was determined in caecectomized cockerels for samples of soybean meal (n=21) and for samples of wheat (n=23). Due to the low number of samples only results of calibration were shown. The coefficients of determination were not very satisfactory in soybean meal (0.60<R^2_c<0.70). They were better in wheat for Met (R^2_c=0.84), similar for Thr (R^2_c=0.69) and worse for Lys (R^2_c=0.55). Even though these calibrations could not be considered accurate, they outperformed nitrogen-based regression equations. Pujol et al. 2007 [58] developed NIR calibrations to predict the pig ileal digestibility of Lys, Met and Cys in barley. The prediction statistics were promising (R^2_P=0.97 for Lys, R^2_P=0.95 for Met and R^2_P=0.85 for Cys), although they only used 15 samples to calibrate and 5 to validate. Recently, Noel et al. 2021 [56] presented calibrations for predicting the pig ileal digestibility of 18 AAs. The dataset (n=38 for Trp, n≈100 for the rest) contained cereals, cereal by-products, protein-rich feedstuffs, fibre-rich feedstuffs and others. The coefficients of determination were good for most of the AAs (R^2_{cv}>0.70) and low errors (CV_{cv}<5%) were found. Trp had worse statistics due to the reduced number of samples and lower concentration of this AA in the samples. No calibrations for the standardized ileal digestibility (SID) of AAs have been presented so far although it is a topic that it is attracting the attention of the most important agri-food research centres in Europe. They are collaborating since 2021 in the framework of a long-term European project about sustainable pig production called PIGWEB [59]. One of the main tasks within this project

is the development of NIR calibrations to predict SID of AAs for pigs. To do this, the participating research centres are providing SID values and spectra of different feedstuffs and diets to create a large dataset.

NIRS to predict carbohydrates and lignin

Starch, the principal source of energy for poultry and swine, has been predicted well in feedstuffs like corn [16], wheat [9,11] or faba bean [10]. Nieto-Ortega et al. 2022 [60] used a set that included different cereal and cereal by-products and Losada et al. 2009 [18] included cereal, cereal by-products and peas. For a single feedstuff, the calibrations were accurate (CV_{cv}<3%) but the coefficient of determination was not very good (R^2_{cv}<0.90) while in the works where different feedstuffs are included in the dataset the calibrations presented higher CV_{cv}>3% but very high R^2_{cv}≈0.99. The increasing in the variability and the wider concentration intervals caused that. Losada et al. 2009 [18] proved that NIRS is capable also to predict sugars (the sum of monosaccharides and disaccharides) with a high coefficient of determination (R^2_{cv}=0.94) but a higher error than starch (CV_{cv}=12%) probably due to the fact that sugars are found in small amounts in feedstuffs and the SEL might be high. Oligosaccharides (OS) have not received much attention. Hollung et al. 2005 [61] predicted α-galactosides (raffinose, stachyose and verbascose) in soybean meals. These are the most important OS in protein-rich feedstuffs. They only used 16 samples and only obtained R^2_{cv}>0.50 for raffinose, not for stachyose or verbascose. Regarding non-starch polysaccharides (NSP), Blakeney et al. 2005 [62] employed a cereal sample set and predicted cellulose, some non-cellulosic polysaccharide (NCP) residues (arabinose, xylose and glucose) and the total NSP content showing also results for the insoluble and soluble fractions. They obtained acceptable results for the total and the insoluble fractions (R^2_{cv}>0.80, 10%<CV_{cv}<25%) considering that the SEL for these constituents is high. The soluble fraction was poorly predicted (R^2_{cv}<0.70, CV_{cv}>30%). The reason is that the soluble fraction is not obtained by chemical analysis, but it is obtained by subtracting the insoluble content from the total content. The prediction error of the soluble fraction is influenced by the accumulation of analytical errors derived from the combination of the two determinations. Blakeney et al. 2005 [62] also predicted the total content of β-glucan. They obtained a good coefficient of determination (R^2_{cv}=0.93) but a high error (CV_{cv}=29%). Bellato et al. 2011 [63] and more recently, Meenu et al. 2022 [64] developed β-glucan calibrations for naked oats. Their dataset variability was considerably low. Because of that, they obtained lower prediction errors

(CV_{cv}<10%) but worse coefficients of determination (R^2_{cv}<0.70). Gomes et al. 2020 [65] divided a large dataset into cereals, protein- and fibre-rich feedstuffs and developed accurate, specific calibrations (0.86<R^2_{cv}<0.97) for most of the total and insoluble NCP residues (xylose, arabinose, glucose, galactose and mannose) and for the total and insoluble NSP content (0.91<R^2_{cv}<0.97). Nieto-Ortega et al. 2022 [60] used a dataset containing cereals and cereal by-products and accurately predicted most of the total and insoluble NCP residues, cellulose, lignin and NSP. The R^2_{cv} values for cellulose, total and insoluble arabinose, xylose and glucose and for total and insoluble NSPs were excellent with R^2_{cv} ranging from 0.97 to 0.98 and CV_{cv} from 10% to 25%. The lignin prediction was also quite good (R^2_{cv}=0.93, CV_{cv}=22%). Worse results were found for total galactose (R^2_{cv}=0.87, CV_{cv}=38%) and mainly total mannose (R^2_{cv}=0.54, CV_{cv}>100%). This is a consequence of the low concentrations of galactose (mean=2.1 g/kg) and mannose (mean=1.7 g/kg) in cereals and cereal by-products.

NIRS to obtain the digestibility of energy and nutrients

Digestible energy and the digestibility of the nutrients (e.g. digestibility of CP) are the parameters that best reflect the quality and adequacy of an ingredient or diet (see Section 2.1). These values are obtained by in-vivo assays that are associated to high costs and ethical problems related to the use of animals. Different strategies involving NIRS have been used to reduce the costs of in-vivo assays, although only the prediction of digestibility from the feed spectra would avoid the use of animals. These strategies are summarized in Figure 1.

As it was shown in Section 2.1, other alternatives in addition to NIRS exist to obtain the digestibility values. In a very recent work, Paternostre et al. 2023 [66] compared the performance of feed tables, empirical models based on chemical composition with or without the inclusion of in-vitro digestibility of nutrients, and NIRS to predict NE of pig diets. They found that the NE values calculated from feed tables approximated the expected value unless diets contained some uncommon ingredients such as beet pulp. The empirical models were more precise than the use of feed tables with the drawback that the samples must be analysed by wet chemistry methods. The inclusion of in-vitro digestibility parameters like organic matter digestibility as predictors in empirical models improved the predictions with the extra costs that in-vitro analysis entail. The use of the NIR feed spectrum resulted in worse predictions than the

empirical models using in-vitro digestibility parameters. To obtain more accurate predictions based on NIRS they had to use both the spectra of feed and faeces.

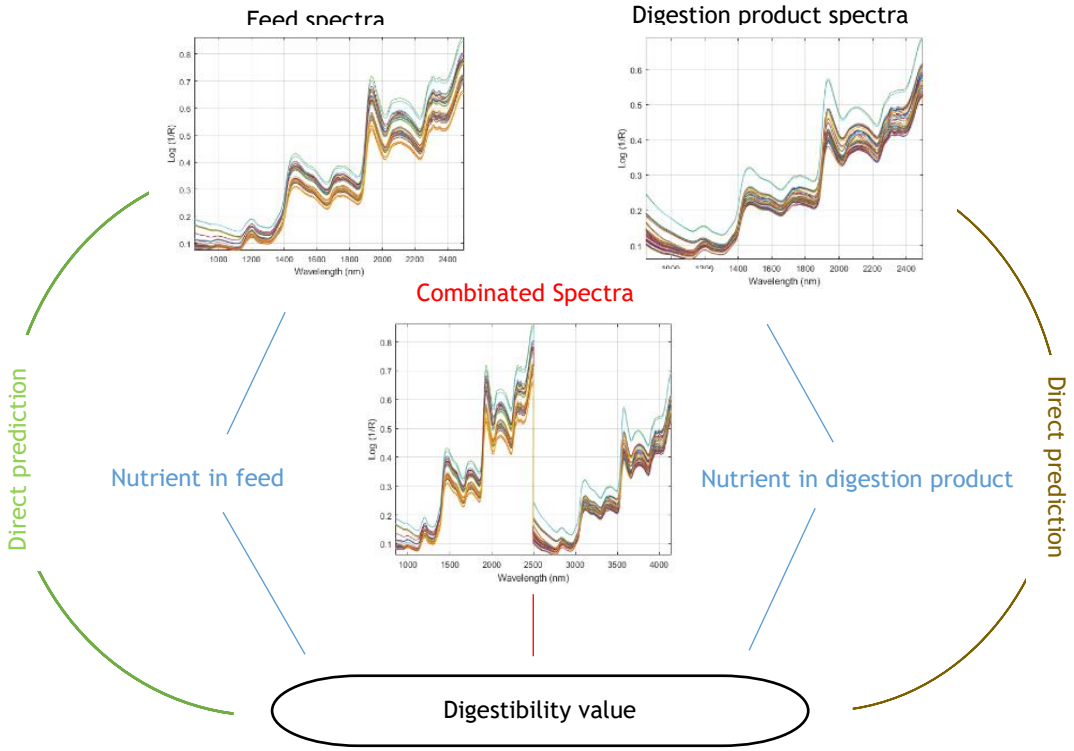


Figure 1. Different ways to obtain digestibility values involving NIRS. Green via: direct prediction from the feed spectra. Brown via: direct prediction from the digestion product (faeces, excreta or ileal digesta) spectra. Blue via: calculation from the nutrient content in feed and in digestion product determined by NIRS. Red via: prediction from the combination of both spectra, feed and digestion product. This combination could be a concatenation, a subtraction, a mean spectra, a combination of PCA or PLS scores and others.

Prediction of digestibility from the NIR feed spectra

Predicting digestibility using the feed spectra has the great advantage of avoiding the use of animals. Table 3 shows the cross-validation statistics of the works where digestible energy and digestibility of nutrients for pig and poultry are predicted from the feed spectra. Same approach was investigated for other animals such as cattle [78] or sheep [79]. The results shown in the table vary but there are several calibrations that could be considered very satisfactory (R^2_{cv}>0.90 and CV_{cv}<5%). The calibrations to predict AME or TME for poultry were generally better than those for DE or ME for pigs. Perhaps, it could be due to chickens are genetically similar and big differences in their digestive capacity are not expected while pigs show more differences between individuals [80]. Since only feed spectra are being considered to predict, the source of

variation associated to the digestive process cannot be modelled when the calibration is developed. It is interesting that the calibrations for wheat samples were the worst and the fact that Valdes & Leeson 1992 [68] considered the samples of wheat as outliers but more studies would be necessary to affirm that there are ingredients in which is more difficult to predict digestible energy than others. Probably the results of the predictions depend primarily on how the in-vivo assays have been carried out and how the laboratory works in each case. Moreover, the variability and range of values in the data set that can be obtained by testing just one ingredient is vastly different compared to a multi-ingredient or compound feed database. The low number of samples used in most studies (n<100) also affects the quality of the calibrations and the predictions.

Table 3. Near-infrared (NIR) calibration models to predict the digestible energy (DE, AME, TME or NE) and the digestibility of nutrients in feedstuffs and diets for poultry and swine described in the literature. Type of sample, analyte or constituent predicted, coefficient of determination (R^2_{cv}) and coefficient of variation of cross validation (CV_{cv}) and reference (Ref).

Sample	Analyte	R^2_{cv}	CV_{cv} (%)	Ref
Poultry digestibility				
Poultry diets	AME_n	0.92	2.0	[67]
Poultry feeds	AME_n	0.93	3.3	[68]
Feed fats	AME_n	0.77	3.6	[69]
Barley	TME	0.95	1.5	[70]
Wheat	AME, TME	0.45, 0.74	9.9, 2.7	[11]
Cereals	AME	0.83	3.7	[71]
Wheat	AME	0.59	2.4	[9]
Cereals	AMEn	0.82	5.8	[18]
Oil seeds	AMEn	0.95	7.1	[19]
Poultry diets	AME, dDM, dCP, dStarch	0.74, 0.76, 0.50, 0.33	3.8, 4.3, 4.5, 2.9	[72]
Pig digestibility				
Cereals	DE	0.87	2.8	[73]
Wheat	DE	0.17	2.5	[11]
Barley	DE	0.79	1.9	[74]
Corn	DE, ME	0.87, 0.86	0.34, 2.2	[75]
Sorghum	DE, ME	0.88, 0.86	1.2, 1.3	[76]
Pig diets	NE, dCP, dFat, dNSP, dCF, dOM	0.79, 0.76, 0.61, 0.61, 0.72, 0.73	3.2, 3.5, 5.5, 8.6, 5.6, 3.2	[27]
Cereals	ME, dDM, dOM, dCP, dFat, dCF	0.75, 0.82, 0.85, 0.64, 0.51, 0.35	2.5, 2.0, 1.8, 4.3, 16, 39	[77]
Supplemental ingredients	ME, dDM, dOM, dCP, dFat, dCF	0.89, 0.89, 0.92, 0.68, 0.33, 0.61	7.4, 7.9, 6.9, 7.0, 26, 32	[77]
Pig diets	ME, dDM, dOM, dCP, dFat, dCF	0.71, 0.80, 0.83, 0.64, 0.63, 0.43	5.1, 3.4, 3.0, 4.5, 13, 35	[77]

Digestibility predictions based on models for feed and digestion product

In-vivo assays usually involve the analysis of many replicate digestion product (faeces, excreta and/or ileal digesta) samples per diet trying to minimize external effects such as temperature

and humidity in the farm or the particular digestion differences between animals of the same species, age and gender. As shown in the previous sections, it is possible to obtain the nutrient content of feedstuffs and diets from the NIR spectra. The use of calibration models to predict the nutrient content of the digestion product samples would be also very useful to reduce the costs of in-vivo assays. As far as we know, there are no published papers where ileal digesta samples are analysed by NIRS. Several authors focused on the use of the faeces/excreta as fertilizer and developed calibrations for OM, N (total, organic and ammoniacal) and P among others [81]. Table 4 shows the prediction statistics of NIR calibrations for the nutrient content of poultry excreta and pig faeces. All the excreta calibrations were developed for broilers. GE was well predicted in most of the works. These results for GE were better than those obtained in feedstuffs maybe due to a higher variability of GE in the digestion product compared to feedstuffs. The fat predictions showed R^2_{cv} values in the order to those found for feedstuffs and diets although presented higher CV_{cv} because, unlike for feedstuffs, the reference method includes an acid extraction [82].

Table 4. Near-infrared (NIR) calibration models to predict the nutrient content of broiler excreta and pig faeces described in the literature. Type of sample, analyte or constituent predicted, coefficient of determination of cross validation (R^2_{cv}), coefficient of variation of cross validation (CV_{cv}) and reference (Ref).

Sample	Analyte	R^2_{cv}	CV_{cv} (%)	Ref
Broiler excreta	Moisture, N, GE, P, Ca	0.96, 0.88, 0.86, 0.91, 0.84	3.1, 3.5, 1.8, 9.3, 8.8	[83]
Broiler excreta	GE, Starch, Fat, N, CP	9.7, 10.2, 10.2, 4.3, 4.4 [1]	0.69, 7.7, 4.0, 5.0, 5.0	[84]
Broiler excreta	N, CP, UA	4.3, 4.4, 5.2 [1]	5.0, 5.0, 5.8	[85]
Broiler excreta	Ash, GE, Starch, Fat, N, UA, CP	4.3, 8.9, 13.7, 9.5, 4.2, 4.1, 5.0 [1]	5.1, 1.0, 5.5, 5.2, 3.2, 7.8, 4.0	[86]
Broiler excreta	GE	0.92	1.0	[87]
Pig faeces	GE	0.97	1.1	[87]
Pig faeces	OM, CP, NDF, ADF, CF	0.73, 0.74, 0.90, 0.94, 0.90	1.6, 8.5, 6.2, 6.8, 7.6	[88]
Pig faeces	OM, CP, GE, Fat, NDF	0.92, 0.88, 0.92, 0.69, 0.94	7.9, 11.1, 8.2, 17.0, 23.0	[89]
Pig faeces	DM, CP, Fat, Ash, CF, OM	0.92, 0.90, 0.91, 0.90, 0.84, 0.90	0.5, 4.2, 6.2, 5.1, 7.2, 1.1	[27]

[1] RPD_{cv} values instead of R^2_{cv}

[2] Values obtained for a test set instead of CV

Uric acid (UA) had high CV_{cv} (>5 %) justified by the fact that SEL for this determination is the highest [90]. CP was worse predicted in terms of CV_{cv} than in feedstuffs and diets probably due to the method of analysis that consists of precipitating the proteins by lead acetate and that introduces some approximations [91]. P was predicted better in excreta that in feedstuffs and compound feeds. The reason could be that P in excreta is mainly in the form of phytate that it is the undigested fraction of P. These results reveal that it is possible to predict properly several

of the most important nutritional parameters in the digestion product. Therefore, it would be possible to calculate the energy and the digestibility of the nutrients using the predicted values obtained from the NIR spectrum of feeds and digestion products.

Prediction of digestibility from the NIR faeces spectra

Another approach investigated is whether it is possible to directly predict the digestibility in the feed from the digestion product spectra. According to Bastianelli et al. 2005 [92] this would avoid the need to use and determine an indigestible marker or precisely measure feed intake and faecal output thereby simplifying the experimental work. GE digestibility for poultry was well predicted in this work from the excreta spectrum (R^2_{cv}=0.98, CV_{cv}=3.0%). The diet was the same, but the animals had different genetics and digestive capacities. They published another work in which DE and digestibility of DM, OM and CP were predicted from pig faeces [93]. Again, the diet was the same for all the animals but in this case, the variation in the animals was moderate. Hence, it is not unexpected that they obtained lower errors (CV_{cv}<2.0%) but worse coefficients of determination (R^2_{cv}<0.70). These works proved that the faeces/excreta spectrum reflects how the animal digests the feed. Other authors used datasets containing pig faeces that came from different animals but also different diets [27,88,89]. Although the variability was very high in these datasets the coefficients of determination were generally poor (R^2_{cv}<0.80) and the errors were higher than those reported in the Bastianelli's works. These unsatisfactory predictions indicate that although the composition of the digestion product is feed-dependent, to obtain accurate calibrations valid for different type of diets the digestion product spectrum is not enough.

Prediction of digestibility from the combination of feed and digestion product NIR spectra

A promising approach to reduce the cost of the in-vivo assays is to combine both the feed and the digestion product spectra. This would avoid the use of any wet chemistry method and the need to use and determine an indigestible marker or precisely measure feed intake and faecal output. Coulibaly et al. 2013 [72] showed that concatenating the spectra of feeds and poultry excreta the SECV values for AME and dDM were two times lower than the errors obtained from only the feed spectrum. In the case of dStarch, SECV was four times lower. These three parameters presented R^2_{cv}>0.95 that can be considered exceptional. dCP prediction statistics

improved significantly but not as much as AME, dDM and dStarch did. Recently, Paternostre et al. 2021 [27] observed also an important improvement when spectra are concatenated in the predictions of NE and nutrient content digestibility in pigs. Predicting only from the feed or the faeces spectra they obtained R^2_{cv}<0.80 and merging them R^2_{cv}>0.80 with parameters such as dOM and NE with R^2_{cv}>0.90. Similar conclusions were drawn from the prediction of digestibility in ruminants [94] and rabbits [95]. The improvement in the predictive ability of models that combine feeds and digestion product spectra in comparison to those that used only the feed spectrum or the digestion product spectrum is expected. By combining both spectra, the original composition of the feed is being considered as well as the variation associated to the digestive process.

References

[1] K.H. Norris, W.L. Butler, Techniques for Obtaining Absorption Spectra on Intact Biological Samples, Ire Trans. Biomed. Electron. 8 (1961) 153–157.

[2] J.R. Hart, K.H. Norris, Direct spectrophotometric determination of moisture content of grain and seeds, in: Proceedings of the 1963 International Symposium on Humidity and Moisture, Principal Methods of Measuring Moisture on Liquids and Solids, Reinhold Publishing Co., New York, 1965: pp. 19–25.

[3] P.C. Williams, Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds, Cereal Chem. 52 (1975) 561–576.

[4] W. Hunt, D. Fulk, B. Elder, K. Norris, Collaborative study on infrared reflectance devices for determination of protein in hard red winter wheat, and for protein and oil in soybeans., Cereal. Foods World. 22 (1977) 534–536.

[5] D.I. Givens, J.L. De Boever, E.R. Deaville, The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans, Nutr. Res. Rev. 10 (1997) 83–114.

[6] K.B. Beć, J. Grabska, C.W. Huck, Miniaturized NIR Spectroscopy in Food Analysis and Quality Control: Promises, Challenges, and Perspectives, Foods. 11 (2022) 1465.

[7] S.A. Haughey, P. Galvin-King, A. Malechaux, C.T. Elliott, The use of handheld near-infrared reflectance spectroscopy (NIRS) for the proximate analysis of poultry feed and to detect melamine adulteration of soya bean meal, Anal. Methods. 7 (2015) 181–186.

[8] S. Modroño, A. Soldado, A. Martínez-Fernández, B. de la Roza-Delgado, Handheld NIRS sensors for routine compound feed quality control: Real time analysis and field monitoring, Talanta. 162 (2017) 597–603.

[9] B. Owens, M.E.E. McCann, K.J. McCracken, R.S. Park, Prediction of wheat chemical and physical characteristics and nutritive value by near-infrared reflectance spectroscopy, Br. Poult. Sci. 50 (2009) 103–122.

[10] J. Wang, H. Liu, G. Ren, Near-infrared spectroscopy (NIRS) evaluation and regional analysis of Chinese faba bean (Vicia faba L.), Crop J. 2 (2014) 28–37.

[11] P.C. Garnsworthy, J. Wiseman, K. Fegeros, Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy, J. Agric. Sci. 135 (2000) 409–417.

[12] D. Cozzolino, A. Chree, I. Murray, J.R. Scaife, The assessment of the chemical composition of fishmeal by near infrared reflectance spectroscopy, Aquac. Nutr. 8 (2002) 149–155.

[13] L. Chen, Z. Yang, L. Han, A review on the use of near-infrared spectroscopy for analyzing feed protein materials, Appl. Spectrosc. Rev. 48 (2013) 509–522.

[14] A. Fassio, E.G. Fernández, E.A. Restaino, A. La Manna, D. Cozzolino, Predicting the nutritive value of high moisture grain corn by near infrared reflectance spectroscopy, Comput. Electron. Agric. 67 (2009) 59–63.

[15] D. Cozzolino, A. Fassio, E. Fernández, E. Restaino, A. La Manna, Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 129 (2006) 329–336.

[16] Y. Fang, X. Chengwei, H. Dan, Analysis and estimate of corn quality by near infrared reflectance (NIR) spectroscopy, 2011 Symposium on Photonics and Optoelectronics (SOPO), Wuhan, China, 2011, pp. 1–4.

[17] C.O. Egesel, F. Kahriman, Determination of Quality Parameters in Maize Grain by NIR Reflectance Spectroscopy, J. Agric. Sci. 18 (2012) 31–42.

[18] B. Losada, P. García Rebollar, P. Cachaldora, C. Álvarez, J. Méndez, C. de Blas, A comparison of the prediction of apparent metabolisable energy content of starchy grains and cereal by-products for poultry from its chemical components, in vitro analysis or near-infrared reflectance spectroscopy, Spanish J. Agric. Res. 7 (2009) 813–823.

[19] B. Losada, P. García-Rebollar, C. Álvarez, P. Cachaldora, M.A. Ibáñez, J. Méndez, J.C. De Blas, The prediction of apparent metabolisable energy content of oil seeds and oil seed by-

products for poultry from its chemical components, in vitro analysis or near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 160 (2010) 62–72.

[20] D.S. Ferreira, O.F. Galão, J.A.L. Pallone, R.J. Poppi, Comparison and application of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for determination of quality parameters in soybean samples, Food Control. 35 (2014) 227–232.

[21] J. Fontaine, J. Hörr, B. Schirmer, Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal, J. Agric. Food Chem. 49 (2001) 57–66.

[22] J. Fontaine, B. Schirmer, J. Horr, Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum, J. Agric. Food Chem. 50 (2002) 3902–3911.

[23] A.B.M. Khaleduzzaman, M.A.A. Mamun, H.M. Salim, Development of local calibrations for the nutritional evaluation of commercial poultry diets by using near infrared reflectance spectroscopy, J. Appl. Anim. Res. 45 (2017) 8–14.

[24] E. Swart, T. Brand, J. Engelbrecht, The use of near infrared spectroscopy (NIRS) to predict the chemical composition of feed samples used in ostrich total mixed rations, S. Afr. J. Anim. Sci. 42 (2012) 550–554.

[25] D. Bastianelli, E. Fermet-Quinet, F. Davrieux, D. Friot, C. Hervouet, Use of near infrared spectroscopy for the prediction of quality of poultry feeds in East Africa, in: G. Burling-Claridge, S. Holroyd, R. Summer (Eds.), Proceedings of the 12th International Conference on Near Infrared Spectroscopy., IM Publications, Auckland, New Zealand: pp. 623–625.

[26] L. Molognoni, L.A. de Sá Ploêncio, C.T.P. Deolindo, L.V.A. de Oliveira, R.B. Hoff, H. Daguer, FT-NIR combined with chemometrics versus classic chemical methods as accredited analytical support for decision-making: Application to chemical compositional compliance of feedingstuffs, Microchem. J. 158 (2020) 105126.

[27] L. Paternostre, V. Baeten, B. Ampe, S. Millet, J. De Boever, The usefulness of NIRS calibrations based on feed and feces spectra to predict nutrient content, digestibility and net energy of pig feeds, Anim. Feed Sci. Technol. 281 (2021) 115091.

[28] E. Fernández-Ahumada, J.E. Guerrero-Ginel, D. Pérez-Marín, A. Garrido-Varo, Near infrared spectroscopy for control of the compound-feed manufacturing process: Mixing stage, J. Near Infrared Spectrosc. 16 (2008) 285–290.

[29] Official Method of Analysis of AOAC International 978.10, Fiber (Crude) in Animal Feed and Pet Food, Gaithersburg, MD, 2005.

[30] T. Fearn, The Overuse of R2 , NIR News. 25 (2014) 32–32.

[31] S.E. Kays, F.E. Barton, Rapid prediction of gross energy, and utilizable energy in cereal food products using near-infrared reflectance spectroscopy, J. Agric. Food Chem. 50 (2002) 1284–1289.

[32] M. Tahir, M.Y. Shim, N.E. Ward, M.O. Westerhaus, G.M. Pesti, Evaluation of near-infrared reflectance spectroscopy (NIRS) techniques for total and phytate phosphorus of common poultry feed ingredients, Poult. Sci. 91 (2012) 2540–2547.

[33] R. Aureli, Q. Ueberschlag, F. Klein, C. Noël, P. Guggenbuhl, Use of near infrared reflectance spectroscopy to predict phytate phosphorus, total phosphorus, and crude protein of common poultry feed ingredients, Poult. Sci. 96 (2017) 160–168.

[34] A.Y. Ikoyi, B.A. Younge, Influence of forage particle size and residual moisture on near infrared reflectance spectroscopy (NIRS) calibration accuracy for macro-mineral determination, Anim. Feed Sci. Technol. 270 (2020) 114674.

[35] I. González-Martín, J.M. González-Cabrera, M. Bustamante-Rangel, M. Milagros Delgado-Zamarreño, Near infrared spectroscopy (NIRS) reflectance technology for determination of tocopherols in animal feeds, Anal. Chim. Acta. 558 (2006) 132–136.

[36] I. González-Martín, J.M. Hernández-Hierro, M. Bustamante-Rangel, N. Barros-Ferreiro, Near-infrared spectroscopy (NIRS) reflectance technology for the determination of tocopherols in alfalfa, Anal. Bioanal. Chem. 386 (2006) 1553–1558.

[37] G. Rubenthaler, B. Bruinsma, Lysine Estimation in Cereals by Near-Infrared Reflectance, Crop Sci. 18 (1978) 1039–1042.

[38] A.A. Gill, C. Starr, D.B. Smith, Lysine and nitrogen measurement by infra-red reflectance analysis as an aid to barley breeding, J. Agric. Sci. 93 (1979) 727–733.

[39] P.C. Williams, K.R. Preston, K.H. Norris, P.M. Starkey, Determination of Amino Acids in Wheat and Barley by Near-Infrared Reflectance Spectroscopy, J. Food Sci. 49 (1984) 17–20.

[40] P.C. Williams, S.L. Mackenzie, P.M. Starkey, Determination of Methionine in Peas by Near-Infrared Reflectance Spectroscopy (NIRS), J. Agric. Food Chem. 33 (1985) 811–815.

[41] D.L. Pazdernik, A.S. Killam, J.H. Orf, Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy, Agron. J. 89 (1997) 679–685.

[42] J.G. Wu, C. Shi, X. Zhang, Estimating the amino acid composition in milled rice by near-infrared reflectance spectroscopy, F. Crop. Res. 75 (2002) 1–7.

[43] I. V. Kovalenko, G.R. Rippke, C.R. Hurburgh, Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy, J. Agric. Food Chem. 54 (2006) 3485–3491.

[44] D. Shi, J. Hang, J. Neufeld, S. Zhao, J.D. House, Estimation of crude protein and amino acid contents in whole, ground and defatted ground soybeans by different types of near-infrared (NIR) reflectance spectroscopy, J. Food Compos. Anal. 111 (2022) 104601.

[45] I. González-Martín, N. Álvarez-García, J.M. González-Cabrera, Near-infrared spectroscopy (NIRS) with a fibre-optic probe for the prediction of the amino acid composition in animal feeds, Talanta. 69 (2006) 706–710.

[46] D. Bastianelli, E. Fermet-Quinet, F. Davrieux, C. Hervouet, L. Bonnal, Calibration strategies for prediction of amino acid content of poultry feeds, in: G. Burling-Claridge, S. Holroyd, R. Summer (Eds.), Proceedings of the 12th International Conference on Near Infrared Spectroscopy., IM Publications, Auckland, New Zealand, 2005: pp. 314–318.

[47] B. Zhang, Z.Q. Rong, Y. Shi, J.G. Wu, C.H. Shi, Prediction of the amino acid composition in brown rice using different sample status by near-infrared reflectance spectroscopy, Food Chem. 127 (2011) 275–281.

[48] G.L. Chen, B. Zhang, J.G. Wu, C.H. Shi, Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 165 (2011) 111–119.

[49] L.J. Zhou, L.Y. Zhang, E.X. Zhang, J.T. Li, W.J. Yang, Z.Y. Wang, Rapid determination of swine available energy and amino acids in corn distillers dried grains with solubles by near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 175 (2012) 198–202.

[50] E.O. Alamu, A. Menkir, M. Adesokan, S. Fawole, B. Maziya-dixon, Near-Infrared Reflectance Spectrophotometry (NIRS) Application in the Amino Acid Profiling of Quality Protein Maize (QPM), Foods. 11 (2022) 2779.

[51] S.J. Noel, H.J.H. Jørgensen, K.E. Bach Knudsen, Prediction of protein and amino acid composition and digestibility in individual feedstuffs and mixed diets for pigs using near-infrared spectroscopy, Anim. Nutr. 7 (2021) 1242–1252.

[52] R.J. Boila, S.C. Stothers, L.D. Campbell, The relationships between the concentrations of individual amino acids and protein in wheat and barley grain grown at selected locations throughout Manitoba, Can. J. Anim. Sci. 76 (1996) 163–169.

[53] W.B. Roush, T.L. Cravener, Artificial Neural Network Prediction of Amino Acid Levels In Feed Ingredients, Poult. Sci. 76 (1996) 721–727.

[54] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[55] Official Methods of Analysis of AOAC International 988.15, Tryptophan in Foods and Food and Feed Ingredients, Gaithersburg, MD, 2016.

[56] A. Lamp, M. Kaltschmitt, O. Lüdtke, Improved HPLC-method for estimation and correction of amino acid losses during hydrolysis of unknown samples, Anal. Biochem. 543 (2018) 140–145.

[57] T. Van Kempen, J.-C. Bodin, Near-infrared reflectance spectroscopy (NIRS) appears to be superior to nitrogen-based regression as a rapid tool in predicting the poultry digestible amino acid content of commonly used feedstuffs, Anim. Feed Sci. Technol. 76 (1998) 139–147.

[58] S. Pujol, A.M. Pérez-Vendrell, D. Torrallardona, Evaluation of prediction of barley digestible nutrient content with near-infrared reflectance spectroscopy (NIRS), Livest. Sci. 109 (2007) 189–192.

[59] J.J. van Milgen, S. Millet, R. Westin, C. Larzul, C. Kaya, J.V. Norgaard, C.C. Metges, G. Bee, PIGWEB: an infrastructure for experimental research for sustainable pig production, in: 72 Annual meeting of the European federation of animal science (EAAP), Davos, Switzerland, 2021: p. 544.

[60] B. Nieto-Ortega, J.J. Arroyo, C. Walk, N. Castañares, E. Canet, A. Smith, Near infrared reflectance spectroscopy as a tool to predict non-starch polysaccharide composition and starch digestibility profiles in common monogastric cereal feed ingredients, Anim. Feed Sci. Technol. 285 (2022) 115214.

[61] K. Hollung, M. Øverland, M. Hrustić, P. Sekulić, J. Miladinović, H. Martens, B. Narum, S. Sahlstrøm, M. Sørensen, T. Storebakken, A. Skrede, Evaluation of nonstarch polysaccharides and oligosaccharide content of different soybean varieties (Glycine max) by near-infrared spectroscopy and proteomics, J. Agric. Food Chem. 53 (2005) 9112–9121.

[62] A.B. Blakeney, P.C. Flinn, Determination of non-starch polysaccharides in cereal grains with near-infared reflectance spectroscopy, Mol. Nutr. Food Res. 49 (2005) 546–550.

[63] S. Bellato, V. Del Frate, R. Redaelli, D. Sgrulletta, R. Bucci, A.D. Magrì, F. Marini, Use of near infrared reflectance and transmittance coupled to robust calibration for the evaluation of nutritional value in naked oats, J. Agric. Food Chem. 59 (2011) 4349–4360.

[64] M. Meenu, Y. Zhang, U. Kamboj, S. Zhao, L. Cao, P. He, B. Xu, Rapid determination of β-glucan content of hulled and naked oats using near infrared spectroscopy combined with chemometrics, Foods. 11 (2022) 43.

[65] G.A. Gomes, T.T. dos Santos, C. Piotrowski, R. García, Development of near infrared calibrations for determination of non-starch polysaccharides content in feedstuff, in: 31st Annual Australian Poultry Science Symposium., Sidney, Australia, 2020: pp. 54–57.

[66] L. Paternostre, S. Millet, J. De Boever, Comparison of feed tables , empirical models and near-infrared spectroscopy to predict chemical composition and net energy of pelleted pig feeds, Anim. Feed Sci. Technol. 297 (2023) 115578.

[67] L.J. Jancewicz, M.L. Swift, G.B. Penner, K.A. Beauchemin, K.M. Koenig, G.E. Chibisa, M.L. He, J.J. McKinnon, W.Z. Yang, T.A. McAllister, Development of near-infrared spectroscopy calibrations to estimate fecal composition and nutrient digestibility in beef cattle, Can. J. Anim. Sci. 97 (2017) 51–64.

[68] E.R. Deaville, D.J. Humphries, D.I. Givens, Whole crop cereals. 2. Prediction of apparent digestibility and energy value from in vitro digestion techniques and near infrared reflectance spectroscopy and of chemical composition by near infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 149 (2009) 114–124.

[69] U. Aftab, Energy and amino acid requirements of broiler chickens: Keeping pace with the genetic progress, Worlds. Poult. Sci. J. 75 (2019) 507–514.

[70] E.V. Valdes, S. Leeson, Research Note: The Use of Near Infrared Reflectance Spectroscopy to Measure Metabolizable Energy in Poultry Feed Ingredients, Poult. Sci. 71 (1992) 1559–1563.

[71] E.V. Valdes, S. Leeson, Near Infrared Reflectance Analysis as a Method to Measure Metabolizable Energy in Complete Poultry Feeds, Poult. Sci. 71 (1992) 1179–1187.

[72] E.V. Valdes, S. Leeson, Measurement of Metabolizable Energy, Gross Energy, and Moisture in Feed Grade Fats by Near Infrared Reflectance Spectroscopy, Poult. Sci. 73 (1994) 163–171.

[73] M.J. Edney, J.E. Morgan, P.C. Williams, L.D. Campbell, Analysis of Feed Barley by near Infrared Reflectance Technology, J. Near Infrared Spectrosc. 2 (1994) 33–41.

[74] J.L. Black, R.J. Hughes, S.G. Nielsen, A.M. Tredrea, P.C. Flinn, Near infrared reflectance analysis of grains to estimate nutritional value for chickens, in: 20th Annual Australian Poultry Science Symposium, The Poultry Research Fundation, Sidney, Australia, 2009: pp. 31–34.

[75] I. Coulibaly, J.P. Métayer, C. Pascal, B. Mahaut, I. Bouvarel, P. Hogrel, D. Bastianelli, La combinaison des infromations issues des aliments et des fientes ameliore la prediction par spir de la digestibilite chez le poulet, in: 10th Poultry Research Days, La Rochelle, France, 2013: pp. 640–644.

[76] R.J. Van Barneveld, J.D. Nuttall, P.C. Flinn, B.G. Osborne, Near infrared reflectance measurement of the digestible energy content of cereals for growing pigs, J. Near Infrared Spectrosc. 7 (1999) 1–7.

[77] R.T. Zijlstra, M.L. Swift, L.F. Wang, T.A. Scott, M.J. Edney, Short communication: Near infrared reflectance spectroscopy accurately predicts the digestible energy content of barley for pigs, Can. J. Anim. Sci. 91 (2011) 301–304.

[78] J. Li, Q. Li, D. Li, Y. Chen, X. Wang, W. Yang, L. Zhang, Use of near-infrared reflectance spectroscopy for the rapid determination of the digestible energy and metabolizable energy content of corn fed to growing pigs, J. Anim. Sci. Biotechnol. 7 (2016) 1–9.

[79] J. Hu, J. Li, L. Pan, X. Piao, L. Sui, G. Xie, S. Zhang, L. Zhang, J. Wang, Rapid determination of the content of digestible energy and metabolizable energy in sorghum fed to growing pigs by near-infrared reflectance spectroscopy, J. Anim. Sci. 97 (2019) 4855–4864.

[80] S.J. Noel, H.J.H. Jørgensen, K.E. Knudsen Bach, The use of near-infrared spectroscopy (NIRS) to determine the energy value of individual feedstuffs and mixed diets for pigs, Anim. Feed Sci. Technol. 283 (2022) 115156.

[81] L. Chen, L. Xing, L. Han, Review of the Application of Near-Infrared Spectroscopy Technology to Determine the Chemical Composition of Animal Manure, J. Environ. Qual. 42 (2013) 1015–1028.

[82] J. Wiseman, B.K. Edmunds, N. Shepperson, The apparent metabolisable energy of sunflower oil and sunflower acid oil for broiler chickens, Anim. Feed Sci. Technol. 36 (1992) 41–51.

[83] D. Bastianelli, L. Bonnal, H. Juin, S. Mignon-Grasteau, F. Davrieux, B. Carré, Prediction of the chemical composition of poultry excreta by near infrared spectroscopy, J. Near Infrared Spectrosc. 18 (2010) 69–77.

[84] K. Terpstra, N. De Hart, The estimation of urinary nitrogen and faecal nitrogen in poultry excreta, Tierphysiol. Tierenahrg. U. Futtermittelkde. 32 (1973) 306–320.

[85] T.N. Smith, G.M. Pesti, R.I. Bakalli, J. Kilburn, H.M. Edwards, The use of near-infrared reflectance spectroscopy to predict the moisture, nitrogen, calcium, total phosphorus, gross energy, and phytate phosphorus contents of broiler excreta, Poult. Sci. 80 (2001) 314–319.

[86] D. Bastianelli, N. Muley, B. Carré, L. Bonnal, F. Davrieux, Use of near infrared spectroscopy for evaluation of nutrient digestibility for genetic experiments in poultry, in: Proceedings of 11th International Conference on Near Infrared Spectroscopy, NIR Publications, Córdoba, Spain, 2004: pp. 735–738.

[87] D. Bastianelli, F. Davrieux, D. Friot, C. Hervouet, M. Lessire, Prediction of nitrogenous compounds in poultry excreta by near infrared, in: Proceedings of 11th International Conference on Near Infrared Spectroscopy, NIR Publications, Córdoba, Spain, 2004: pp. 739–742.

[88] L.J. Chen, L. Xing, L.J. Han, Influence of Data Preprocessing on the Quantitative Determination of Nutrient Content in Poultry Manure by Near Infrared Spectroscopy, J. Environ. Qual. 39 (2010) 1841–1847.

[89] B. De La Roza-Delgado, S. Modroño, F. Vicente, A. Martínez-Fernández, A. Soldado, Suitability of faecal near-infrared reflectance spectroscopy (NIRS) predictions for estimating gross calorific value, Spanish J. Agric. Res. 13 (2015) 1–7.

[90] A. Schiborra, M. Bulang, A. Berk, A. Susenbeth, E. Schlecht, Using faecal near-infrared spectroscopy (FNIRS) to estimate nutrient digestibility and chemical composition of diets and faeces of growing pigs, Anim. Feed Sci. Technol. 210 (2015) 234–242.

[91] K.G. Nirea, M.P. de Nanclares, A. Skugor, N.K. Afseth, T.H.E. Meuwissen, J. Hansen, L.T. Mydland, M. Øverland, Assessment of fecal near-infrared spectroscopy to predict feces chemical composition and apparent total-tract digestibility of nutrients in pigs, J. Anim. Sci. 96 (2018) 2826–2837.

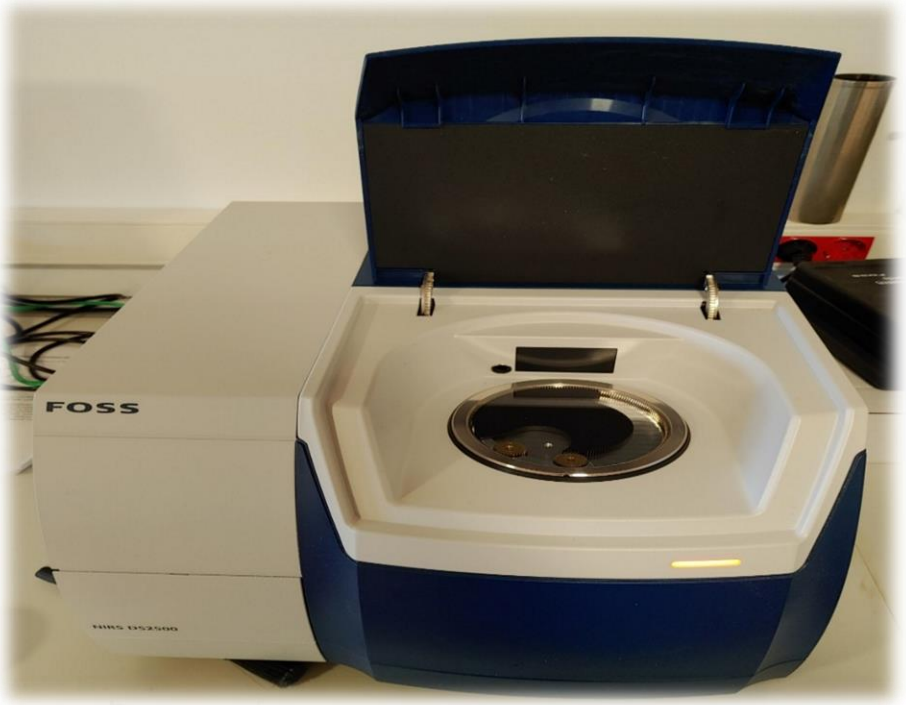
[92] D. Bastianelli, B. Carré, S. Mignon-Grasteau, L. Bonnal, F. Davrieux, Direct prediction of energy digestibility from poultry faeces using near infrared spectroscopy, in: Proceedings of 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 626–629.

[93] D. Bastianelli, L. Bonnal, Y. Jaguelin-Peyraud, J. Noblet, Predicting feed digestibility from NIRS analysis of pig faeces, Animal. 9 (2015) 781–786.

[94] V. Decruyenaere, P. Lecomte, C. Demarquilly, J. Aufrere, P. Dardenne, D. Stilmant, A. Buldgen, Evaluation of green forage intake and digestibility in ruminants using near infrared reflectance spectroscopy (NIRS): Developing a global calibration, Anim. Feed Sci. Technol. 148 (2009) 138–156.

[95] G. Meineri, P.G. Peiretti, G. Masoero, Appraisal of ingestion and digestibility in growing rabbits using near infrared reflectance spectroscopy (NIRS) of feeds and faeces, Ital. J. Anim. Sci. 8 (2009) 75–82.

Chapter 3.
Replacing wet chemistry by NIRS to improve feed formulation and reduce the costs of in-vivo assays



3.1. Introduction

The initial stage in formulating a diet involves conducting the nutritional analysis of the available ingredients. Once the compound feed has been produced it must be analysed to ensure that it has been accurately prepared at the feed mill and that the animal will receive the intended nutrients. This process is a standard practice in the majority of farms and compound feed manufacturing facilities. The agri-food industry moves tones and tones of ingredients and compound feeds and their traceability must be controlled at all times to avoid buying, producing or selling feeds with unwanted property values [1]. Research centres or companies that seek for optimal diets and conduct digestibility in-vivo assays must analyse the nutrient content of faeces or excreta and/or intestinal contents in addition to ingredients and compound feeds. Although wet chemistry methods are accurate and extensively used for the analysis of these types of samples they present disadvantages, including the high cost of some of them, the need for chemical reagents or for a tedious sample preparation. Therefore, there is a need in developing rapid, non-destructive, and chemicals-free methods, as NIRS, that can predict quantitative parameters with similar accuracy as the traditional methods [2].

IRTA is an agri-food research centre whose research line in monogastric animal nutrition conducts primarily efficacy studies of different additives and enzymes included in compound feeds. IRTA has an experimental feed factory that produces a considerable high amount of different compound feeds that must be analysed before they are given to the animals. However, IRTA is not a typical feed producer. Although the variety of ingredients used may be high, the number of batches of ingredients purchased per year is low. Because the conduction of nutritional in-vivo assays is common in IRTA, a large number of faeces, excreta and ileal digesta samples need to be also analysed.

Within the framework of this thesis, in 2020 IRTA purchased a NIRS FOSS DS2500 together with a set of commercial calibrations for moisture, CP, Fat, CF, ash and starch in raw materials and another set for the same parameters in compound feeds. As mentioned in Section 2.2, the costumer must verify that the commercial calibrations work as intended in their own type of samples and otherwise, update them or develop new calibrations with their own samples. The commercial calibrations for raw materials were tested with IRTA samples and often the predictions were unsatisfactory. Regrettably, there were not enough samples of raw materials to

develop new NIR calibrations. The dataset employed in Section 3.2 to determine the carbohydrate and lignin content in raw materials belongs to the Department of Animal and Veterinary Science of Aarhus University. For compound feeds, IRTA did have enough samples to develop new calibrations. In Section 3.3 the commercial calibrations for compound feeds and calibrations developed in-house are compared. In addition, calibrations for important parameters for IRTA research for which there were not commercial calibrations such as GE and P are presented. The use of NIRS to predict the nutrient content of poultry excreta (Section 3.4), pig faeces (Section 3.5) and intestinal contents (Section 3.6) has received relatively little attention, resulting in a lack of commercial calibrations. This may be attributed to the limited number of research centres or companies focusing on these types of samples compared to the extensive number of companies engaged in the production, sale, or purchase of raw materials and compound feeds. The possibility of determining the nutrient content of all these samples by NIRS would reduce considerably the costs of in-vivo assays, making them more accessible for who wants to obtain information about how the feed has been digested, ultimately helping to identify optimal diets for monogastric animals.

References

[1] D.I. Givens, E.R. Deaville, The current and future role of near infrared reflectance spectroscopy in animal nutrition: a review, Aust. J. Agric. Res. 50 (1999) 1131–1145.

[2] A.T. Badaró, J.P. Hebling e Tavares, J. Blasco, N. Aleixos-Borrás, D.F. Barbin, Near infrared techniques applied to analysis of wheat-based products: Recent advances and future trends, Food Control. 140 (2022).

3.2. Determination of carbohydrate and lignin content in feedstuffs for monogastric animals using near-infrared spectroscopy

Andrés Cruz-Conesa[1,2], Knud Erik Bach Knudsen[3], Itziar Ruisánchez[1], Joan Ferré[1], Anna M. Pérez-Vendrell[2] and Samantha Joan Noel[3*]

[1] Universitat Rovira i Virgili, Faculty of Chemistry, Department of Analytical and Organic Chemistry, Tarragona, Spain

*Corresponding author email address: Samantha.noel@anis.au.dk

[2] Institute of Agrifood Research and Technology, Animal Nutrition. Mas Bové, Constantí, Spain

[3] Department of Animal and Veterinary Sciences, AU Viborg Research Center Foulum, Aarhus University, DK 8830 Tjele, Denmark

Submitted to Animal Feed Science and Technology (Edited for format)

Abstract

Carbohydrates (CHO) are the principal constituents of the diets given to monogastric animals around the world. They supply the largest part of the energy to the animal, despite the fact only monosaccharides, disaccharides and starch are absorbed by the animal. Oligosaccharides (OS), resistant starch, non-starch polysaccharides (NSP) and lignin compose the so-called dietary fibre that it is not absorbed. The soluble fraction of fibre is mainly fermentable and causes benefits to the gastrointestinal tract while the insoluble is not and therefore lead to a decrease in nutrient and energy digestibility. The CHO fraction and its composition is highly variable among feed ingredients but also amongst the different varieties of the same ingredient. To improve the nutrient utilization and to formulate animal diets that are more efficient it is necessary to

characterize the CHO fraction of the ingredients. In addition, with the growing use of supplementary enzymes, such as xylanase, β-glucanase or mannanase to mitigate the anti-nutritive effect of fibre the determination of the CHO components is essential. To make a full account of the CHO and lignin constituents, several wet chemical enzymatic, chromatographic and gravimetric methods need to be employed, which are expensive and time-consuming. Near-infrared spectroscopy (NIRS) overcomes these limitations because it is a rapid, non-destructive, multi-parametric technique and requires minimal sample preparation. In this work, around 300 samples of animal feedstuffs, consisting of cereals, cereal co-products, protein-rich and fibre-rich feedstuffs were used to develop and validate near-infrared (NIR) partial least square regression (PLSR) models. Spectral pretreatment optimization, sample selection and outlier detection were considered for each model. Robust predictions (R^2_p>0.90) were obtained for digestible carbohydrates (sugars and starch) and for most of the total or insoluble components of dietary fibre (lignin, NSP, cellulose, β-glucan, non-cellulosic polysaccharides (NCP), arabinose, xylose, galactose, rhamnose and uronic acids). The OS, fructans were well predicted in cereals (R^2_p=0.94) and linearity (R^2_{CV}>0.70) was also found for α-galactosides (raffinose, stachyose and verbascose) in protein-rich feedstuffs. The obtained results were in general very satisfactory and demonstrate that NIRS is a reliable tool to characterize the carbohydrate fraction of a variety of feedstuffs.

Keywords

Cereals, protein feedstuff, multivariate calibration, dietary fibre, non-starch polysaccharides, sugars

Introduction

Carbohydrates (CHO) are the principal constituents of monogastric animal diets accounting for approximately two thirds of the dry matter consumed [1]. The CHO group includes compounds with very different chemical structures and properties that define how they are used by the animal. Only monosaccharides are directly absorbed in the small intestine, but monogastric animals have endogenous enzymes that are able to hydrolyze disaccharides and most of the starch making them available for absorption. Sugars, found in low concentration in plant materials along with abundant starch, represent the main energy source of a typical monogastric

diet [2]. However, the high specificity of the endogenous enzymes results in a significant fraction of CHO along with the non-carbohydrate component lignin not being digested by the endogenous enzymes in the small intestine. This fraction, commonly referred to as dietary fibre (DF) can be fermented to some extent by the microbiota in the large intestine and includes: oligosaccharides (OS), resistant starch (RS), non-starch polysaccharides (NSP) and lignin [3]. In plant materials, OS consist mainly of α-galactosides (raffinose, stachyose and verbascose) and fructans. RS is the non-digestible starch fraction and for most of the feedstuffs, it is present in very low concentration. The NSP group is the main constituent of the DF and includes cellulose, β-glucan and non-cellulosic polysaccharides (NCP) whose monomers are arabinose, xylose, glucose, galactose, mannose, rhamnose, fucose and uronic acids [4]. Lignin is not a carbohydrate, but it is present in most of the plant materials to a greater or lesser extent and is usually treated together with CHO because it is tightly linked to NSP and commonly used analytical methods for fibre determination include lignin [5].

The solubility of the fibre compounds affects their functionality. OS and NCP are fully or partially soluble and the microbiota inhabiting the gastrointestinal tract of monogastric animals can ferment the OS and NCP in a way that is related to how the molecules are organized. This has a positive impact on the intestinal health of the host, increasing the microbial diversity, increasing the mucin production and providing valuable metabolites with functional roles such as short-chain fatty acids [6]. β-glucan is another important partially soluble constituent, mainly found in cereal grains. It can be fermented by the microbiota and may stimulate immune activity [7]. On the other hand, insoluble NCP together with cellulose and lignin represent a fraction of the diet that is not digested or fermented and lead to less digestibility of nutrients and energy [8]. Nowadays, in order to mitigate the anti-nutritive effects of NSP, supplementing feeds with carbohydrases such as xylanase, β-glucanase or β-mannanase is a common practice in the swine and poultry industry [9-10].

The CHO fraction and its composition is highly variable between feed ingredients but also between varieties of the same ingredient. Genetic variability, environmental conditions, harvest and post-harvest processing or storage conditions can influence the CHO content [11]. To improve nutrient utilization and formulate more efficient animal diets it is necessary to characterize the CHO fraction of the ingredients. In addition, according to European Food Safety Authority the companies and research centres that test and use supplementary enzymes

such as the aforementioned carbohydrases must demonstrate their usefulness by analysing their enzymatic activity and also the presence of their substrates [12].

To make a complete account of the CHO and lignin constituents, several wet chemical enzymatic, chromatographic and gravimetric methods need to be employed, which are expensive and time-consuming. Near-infrared spectroscopy (NIRS) overcomes these limitations because it is rapid, non-destructive, multi-parametric and requires minimal sample preparation. The proven ability of NIRS to determine the nutrient content of a multitude of feedstuffs has promoted this technique for routine analysis in the agri-food sector [13-15]. However, few authors have presented NIR calibration models to determine CHO content beyond starch and most of those who did, focused on one type of feedstuff (e.g. cereals) or with limited CHO fractions (e.g. NSP) [11,16-17]. The objective of this work is to study the capability of NIRS to characterize all CHO and lignin fractions of a wide variety of feedstuffs for monogastric animals. Partial least squares (PLS) calibration models for sugars, starch, DF and its components (OS, NSP and lignin) were developed on a dataset containing cereals, cereal co-products, protein-rich feedstuffs and fibre-rich feedstuffs.

Materials and methods

Samples

The feedstuff dataset (304 samples) consists of the dataset described by Bach Knudsen 1997 [18] plus samples that have been included in other publications [19] and analysed by essentially the same analytical methods. The samples represent common and not so common feedstuffs for monogastric animals. The feed samples were collected since 1988 and come from the Danish and European feedstuffs market. Samples were grouped according to their characteristics and consisted of whole grain cereals (168 samples; barley, corn, wheat, oat, rice, rye, sorghum and triticale; cereal co-products (58 samples; barley husk, brewers spent grain, corn flour, maize bran, maize gluten, oat hull, oat meal, rice flour, rye bran, rye meal, wheat bran, wheat aleurone and wheat pericarp/testa; protein-rich feedstuffs (44 samples; soybean meal, sunflower meal, rapeseed meal and cake, fava beans, peas, lupins, red clover, rye grass, cotton seed cake, coconut cake, flaxseed cake and palm cake) and fibre-rich feedstuffs (34 samples; alfalfa, grass meal, apple pomace, pea hull, sugar beet pulp and potato pulp). A summary of the principal

constituents (dry matter, crude protein, fat, crude fibre and ash) determined for each group of samples (Table S1) is given in the Supplementary material.

Reference analysis

The CHO and lignin constituents determined with wet chemistry methods are illustrated in Figure 1. Not all the parameters were analysed in all the samples and some CHO were present at very low concentrations. For instance, β-glucan were only tested in cereals and cereal co-products while α-galactosides were only relevant in protein-rich feedstuffs. Sugars and α-galactosides were extracted together and detected using a colorimetric assay (sugars) or HPLC (α-galactosides) [20]. Fructans were determined separately by an enzymatic assay described by Starch was analysed by Larsson and Bengtsson [21] and the enzymatic-colorimetric method described by Bach Knudsen 1997 [18], resistant starch by the enzymatic-colorimetric method of McCleary and Monaghan 2002 [22] and β-glucan by the enzymatic-colorimetric method of McCleary and Glennie-Holmes 1985 [23]. Total and insoluble NSP constituents (cellulose, and NCP residues) were determined as alditol acetates by gas-liquid chromatography for neutral sugars and by colorimetry for uronic acids using a modification of the Uppsala [24] and Englyst et al. 1982 [25] procedures as described by Bach Knudsen and Laerke 2018 [2]. Lignin (Klason lignin) was determined gravimetrically as the residue resistant to hydrolysis by 2 mol/L H_2SO_4 [24].

Cellulose was calculated as difference between the determination of glucose after swelling the cellulose with 12 mol/L H_2SO_4 (cellulose and monomer glucose present in the sample) and after the hydrolysis with 2 mol/L H_2SO_4 when only the monomer glucose is present:

$$Cellulose = NSP_{glucose\,(\frac{12mol}{L}HSO4)} - NSP_{glucose\,(\frac{2mol}{L}HSO4)} \quad (1)$$

The total or insoluble non-cellulosic polysaccharides (T-NCP or I-NCP) were obtained as

$$NCP = rhamnose + arabinose + xylose + galactose + mannose + glucose + uronic\ acids \quad (2)$$

The soluble NCP was calculated as

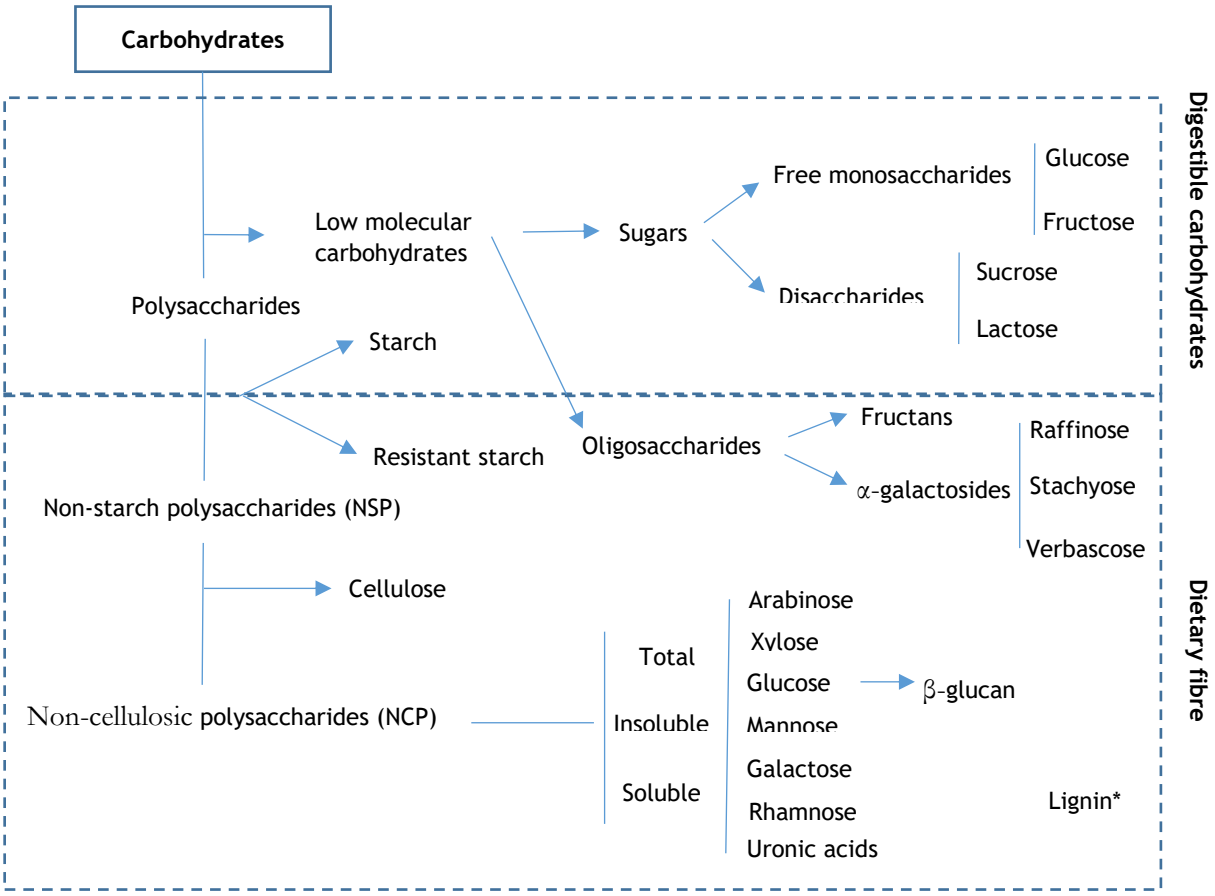
$$S\_NCP = T\_NCP - I\_NCP \quad (3)$$

and dietary fibre was calculated as

$$DF = T\_NSP + Lignin + OS + RS \quad (4)$$

with oligosaccharides calculated as

$$OS = \propto -galactosides\ (raffinose + stachyose + verbascose) + fructans \quad (5)$$



*Lignin is not a carbohydrate but it is treated together with non-starch polysaccharides because it is strongly linked to this component in the cell walls.

Figure 1. Carbohydrates in feedstuffs. Digestible carbohydrates are represented by sugars and starch. Dietary fibre is represented by non-starch polysaccharides, lignin, oligosaccharides and resistant starch.

Near-infrared analysis

Feedstuff samples were ground, passed through a 1 mm sieve and stored at -20°C in airtight containers until needed. The samples were dried at 60°C in a forced air oven for 48 hour before scanning to prevent moisture buildup in the containers that could ruin the samples on thawing. Dried and ground samples were left at room temperature for a minimum of 48 hours to reach the ambient humidity levels. Samples were scanned using a Foss NIRS DS2500 feed analyser (Foss NIR Systems, Denmark) which recorded data every 0.5 nm from 400 to 2500 nm. Samples were scanned in duplicate with a 7 cm diameter cup where 30 g approximately were introduced each time and the spectra were then averaged.

Calibration model development and validation

PLS toolbox software (PLS_Toolbox, 2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab (MATLAB, Version R2020a, The MathWorks Inc., Natick, MA, USA) was used to carry out all the chemometric calculations. Principal component analysis (PCA) was used in the first instance to study the spectral distribution of the sample set and discover possible groups of samples. Partial least squares regression (PLSR) was used to develop the calibration models for the carbohydrates and lignin. Common spectral pretreatments were tested including normalization, standard normal variate (SNV) [26], multiplicative scatter correction (MSC) [27] and 1st and 2nd order derivatives [28] using different window widths. Different wavelength ranges were evaluated for modelling. The sample set was divided into a calibration set (75% of the samples) and a validation set (25% of the samples) using the Duplex algorithm applied to the pretreated spectra, that retains the spectral distribution [29]. It should be noted that the calibration and validation sets varied from one model to another because sample selection depended on the number of reference samples available and the chosen pretreatment. A five-fold venetian blind cross validation (CV) was used to choose the optimal pretreatments, the optimal wavelength range and the optimal number of latent variables for each model. One round of outlier removal was performed where the cutoff values for the spectral outliers were high leverage (Hotelling's T^2 reduced > 3) or a high percentage of residual spectral variance (Q residuals reduced > 3). A large difference between the predicted and the reference value (Studentized residuals (t) > 3) was considered indicative of reference outliers [30].

The performance of the models was evaluated from the coefficient of determination of the regression of predicted versus reference values (R^2_p) and the root mean square error of prediction (RMSEP). To compare the calibrations developed for the different constituents and also with the results found in the literature, the coefficient of variation of prediction (CV_p) was used. CV_p allows considering the differences in the means of the data sets between different constituents and different studies [31].

Results

Carbohydrate and lignin variation in the feedstuffs

The carbohydrate distribution across the four feedstuff groupings is shown in Table 1. Sugar levels were higher in cereal co-products and protein-rich feedstuffs. Coconut cake had the highest sugar level (12.0 g/100g, as fed). Only sugar beet pulp samples contained high levels of sugars (mean=4.1 g/100g, as fed) among the fibre-rich feedstuffs. Starch was found in high concentrations in cereal samples, as expected. Among protein- and fibre-rich feedstuffs, only peas, pea hulls, and potato pulp had sugar concentrations more than 5%. Although some starch is resistant (RS) to digestion in the small intestine due to physical entrapment or steric hindrance, this amount was determined to be minimal in all samples except potato pulp (not shown).

Table 1. Carbohydrate fraction by group of feedstuffs. Number of samples (N), mean and standard deviation (SD).

	Cereals		Cereal co-products		Protein-rich feeds		Fibre-rich feeds	
	N	Mean ± SD (g/100g, as fed)	N	Mean ± SD (g/100g, as fed)	N	Mean ± SD (g/100g, as fed)	N	Mean ± SD (g/100g, as fed)
Sugars	136	2.0 ± 1.1	52	3.4 ± 2.5	42	4.0 ± 3.0	34	2.0 ± 2.0
Starch	76	66 ± 6.9	58	28 ± 25	44	10 ± 16	34	10 ± 12
Raffinose	54	0.3 ± 0.2	16	0.8 ± 0.8	31	1.0 ± 1.2	6	0.2 ± 0.2
Stachyose	21	0.2 ± 0.1	15	0.3 ± 0.2	41	1.4 ± 1.5	3	0.0 ± 0.0
Verbascose	0	-	0	-	16	1.3 ± 1.4	0	-
Fructans	136	1.2 ± 1.0	50	1.4 ± 1.1	10	1.0 ± 1.4	21	0.4 ± 0.3
β-glucan	129	1.6 ± 1.3	25	1.9 ± 1.6	-	-	-	-
T-NCP	160	10.8 ± 3.2	49	27.5 ± 10.9	42	14.8 ± 7.1	27	35.1 ± 11.0
I-NCP	160	7.4 ± 2.0	51	23.7 ± 9.6	42	10.2 ± 6.6	34	14.2 ± 4.3
Cellulose	168	2.0 ± 1.1	54	7.8 ± 5.8	37	6.1 ± 2.6	34	27.3 ± 12.2
T-NSP	164	12.7 ± 3.6	57	31.8 ± 16.0	43	18.4 ± 7.6	34	60.4 ± 16.4
I-NSP	168	9.1 ± 3.0	58	29.2 ± 16.7	44	15.8 ± 8.1	34	41.5 ± 13.4
Lignin	163	2.1 ± 1.8	52	7.2 ± 4.0	33	7.1 ± 5.1	32	4.3 ± 3.8
DF	166	15.4 ± 4.4	57	39.6 ± 19.8	44	28.7 ± 9.0	34	67.6 ± 15.4

The most abundant oligosaccharides (OS) in animal feedstuffs are α-galactosides (raffinose, stachyose, verbascose), which are present predominately in protein-rich feedstuffs, and fructans,

which have the highest concentration in cereals and cereal co-products. The number of OS values in the dataset varied because the quantity of some OS in some feedstuffs was below the quantification limit. For instance, verbascose was only found in significant amounts in peas and soybean meal, stachyose was virtually absent in triticale samples and in several wheat, maize and barley samples, and fructans were negligible in most of the protein-rich feedstuffs. Non-starch polysaccharides (NSP) are a class of non-digestible polysaccharides found mostly in plant cell walls. It is made up of the soluble and insoluble portions of non-cellulosic polysaccharides (NCP) as well as cellulose. Because it is a calculated fraction generated by subtracting the insoluble content from the total content, the soluble fraction is not presented. The highest concentration of NSP was found in fibre-rich feedstuffs due to the greater content of NCP and, in particular, the high content of cellulose (up to four times higher than in the other groups). High levels of NSP and its constituents were also found in bran and hull fractions of the cereal co-products. Protein-rich feedstuffs had nearly the same cellulose content as cereal co-products, but far fewer NCP residues. Cereals had the lowest concentrations of NSP, NCP residues, and cellulose. The β-glucan content of cereals and cereal co-products was similar. Because β-glucan is not found in protein and fibre-rich feedstuffs, they were not tested for it. Each set of feedstuffs had various levels and proportions of the NCP residues (Table 2).

Table 2. Non-cellulosic polysaccharides (NCP) residues by group of feedstuffs. Number of samples (N), mean and standard deviation (SD).

		Cereals		Cereal co-products		Protein-rich feeds		Fibre-rich feeds	
		N	Mean±SD (g/100g, as fed)	N	Mean±SD (g/100g, as fed)	N	Mean±SD (g/100g, as fed)	N	Mean±SD (g/100g, as fed)
Arabinose	Total	165	2.8 ± 0.7	58	7.0 ± 3.8	44	2.9 ± 1.5	34	8.4 ± 7.2
	Insoluble	159	2.0 ± 0.5	48	6.7 ± 2.5	42	1.9 ± 1.1	34	4.1 ± 4.3
Xylose	Total	168	4.5 ± 1.3	58	12.7 ± 6.4	44	2.3 ± 0.4	34	4.0 ± 3.2
	Insoluble	159	3.5 ± 1.1	52	12.3 ± 5.7	42	1.9 ± 1.7	34	3.7 ± 3.0
Mannose	Total	164	0.4 ± 0.1	55	0.5 ± 0.2	38	0.8 ± 0.4	34	0.8 ± 0.4
	Insoluble	146	0.3 ± 0.1	46	0.4 ± 0.2	36	0.5 ± 0.3	33	0.5 ± 0.3
Galactose	Total	168	0.4 ± 0.1	57	0.9 ± 0.4	44	2.0 ± 1.4	25	3.0 ± 1.9
	Insoluble	146	0.2 ± 0.1	48	0.8 ± 0.3	42	1.2 ± 0.8	34	1.8 ± 1.1
Glucose	Total	168	2.4 ± 1.8	58	3.5 ± 2.0	44	1.5 ± 1.9	34	1.5 ± 0.4
	Insoluble	159	1.3 ± 0.7	52	2.8 ± 1.8	34	1.4 ± 2.0	28	1.1 ± 3.2
Rhamnose	Total	113	0.0 ± 0.0	48	0.1 ± 0.0	43	0.2 ± 0.1	34	0.8 ± 0.3
	Insoluble	100	0.0 ± 0.0	38	0.0 ± 0.0	31	0.2 ± 0.1	33	0.3 ± 0.1
Uronic acids	Total	168	0.3 ± 0.1	58	1.4 ± 1.2	44	3.7 ± 1.9	34	12.8 ± 4.2
	Insoluble	159	0.2 ± 0.1	51	1.2 ± 1.0	42	2.0 ± 1.3	34	3.0 ± 0.7

For cereal and cereal co-products arabinoxylans represented 70% of the total NCP with an arabinose/xylose ratio of 0.60 and glucose was the third most abundant NCP, largely represented as β-glucan. Protein- and fibre-rich feedstuffs, had lower levels of glucose and

xylose residues. Mannose, galactose (<1 g/100g, as fed) and rhamnose (<0.1 g/100g, as fed) concentrations were very low in cereal and cereal co-products. Uronic acids were also present in low concentrations in cereals although they were higher in cereal co-products as part of the arabinoxylans complex in brans and hulls. In protein- and fibre-rich feedstuffs, uronic acids are the most abundant NCP and the NCP residue marker for pectin polysaccharides. The solubility of NCP was high only in fibre-rich feedstuffs (>50%), whereas insoluble NCP predominated (70% of the total) in cereals and protein-rich feedstuffs. The solubility of NCP in cereal co-products was even lower (<15%). Arabinose and rhamnose, being part of pectin polysaccharides, were much more soluble in fibre-rich feedstuffs, whereas xylose and glucose, markers for arabinoxylans and β-glucan respectively, were relatively more soluble in the cereals group. Lignin is a phenolic polymer, not a carbohydrate. However, it is linked to cellulose and NCP and affects the physicochemical properties and degradation of polysaccharides in the gastrointestinal tract [1]. Therefore, it is included in the calculation of the dietary fibre fractions. For some groups (e.g., protein-rich feedstuffs) the contribution of lignin and OS was considerable, reducing the contribution of NSP to the non-digestible fraction whereas for other groups (e.g., fibre-rich feedstuffs) NSP accounted for most of DF by far.

Near-infrared analysis

The raw mean Vis-NIR spectrum for each of the four groups of feedstuffs is shown in Figure 2. Despite the offset, significant differences were found in the visible region (400-800 nm) and in some NIR regions of the spectrum. Ground samples of cereals and cereal co-products were light in colour and barely absorbed in the visible region, whereas ground samples of protein-rich feedstuffs such as soy bean meal, rapeseed, or sunflower and fibre-rich feedstuffs such as sugar beet pulp or alfalfa were darker, orange or brown. In the NIR region, some spectral bands are linked to specific structures according to Osborne and Fearn (1988) [32]. As expected, protein-rich feedstuffs exhibited higher intensity in protein-related regions such as the N-H stretching first overtone (1510 nm) and the three prominent peaks of protein in the NH combination region at 1980, 2050, and 2180 nm, corresponding to the asymmetrical stretching + amide II, symmetrical stretching + amide II and 2 × amide I + amide III, respectively. Cereals and to a lesser extent cereal by-products and fibre-rich feedstuffs had well defined peaks in regions related to starch, such as the O-H stretching second overtone (990 nm), the O-H stretching first overtone (1440-1450 nm) and the combination band 2 × O-H deformation + 2 × C-O stretching (2100 nm). Fibre-rich feedstuffs showed some peaks related with cellulose

more clearly than the rest of feedstuff groups, such as the peak that encompasses the C-H stretching first overtone (1780 nm) and the combination band O-H stretching + 2 × C-O stretching (1820 nm) and the peak that encompasses the combination band C-H stretching + C-H deformation (2336 nm) and the C-H deformation second overtone (2352 nm).

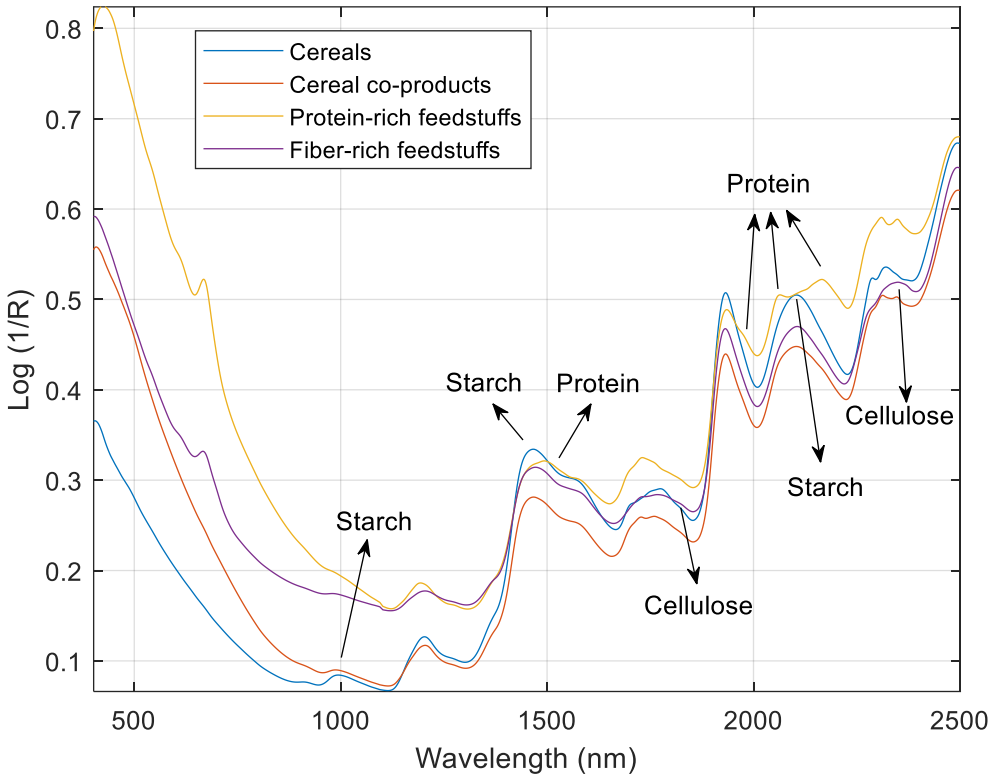


Figure 2. Mean raw spectrum of the samples grouped by type of feedstuff with band assignment for starch, protein and cellulose.

The variability of the samples can be observed in the PCA scores (Figure 3) for spectra preprocessed with the first derivative (second-order polynomial and window width of 15 points) and mean centring. The cereals formed a relatively homogeneous group whereas the other feedstuffs exhibited more variability. Many protein-rich feedstuffs, some fibre-rich feedstuffs and cereal co-products were far from the cereals and some of these were detected as outliers when developing the models. These are very heterogeneous groups with feedstuffs containing different DF and protein composition. The inclusion of all the groups of samples in the model was studied separately for each constituent and will be discussed for each calibration.

The selection of the spectral pretreatment, the number of latent variables and the removal of outliers were performed simultaneously. Pretreatment of the spectra using derivatives and mean

centring significantly improved all models. There were only minor differences in the root mean square error of cross-validation (RMSECV) when utilizing the first or the second derivative, as well as when using SNV and MSC or none of them. The selected pretreatment is specified in the model performance tables (Tables 3 and 4).

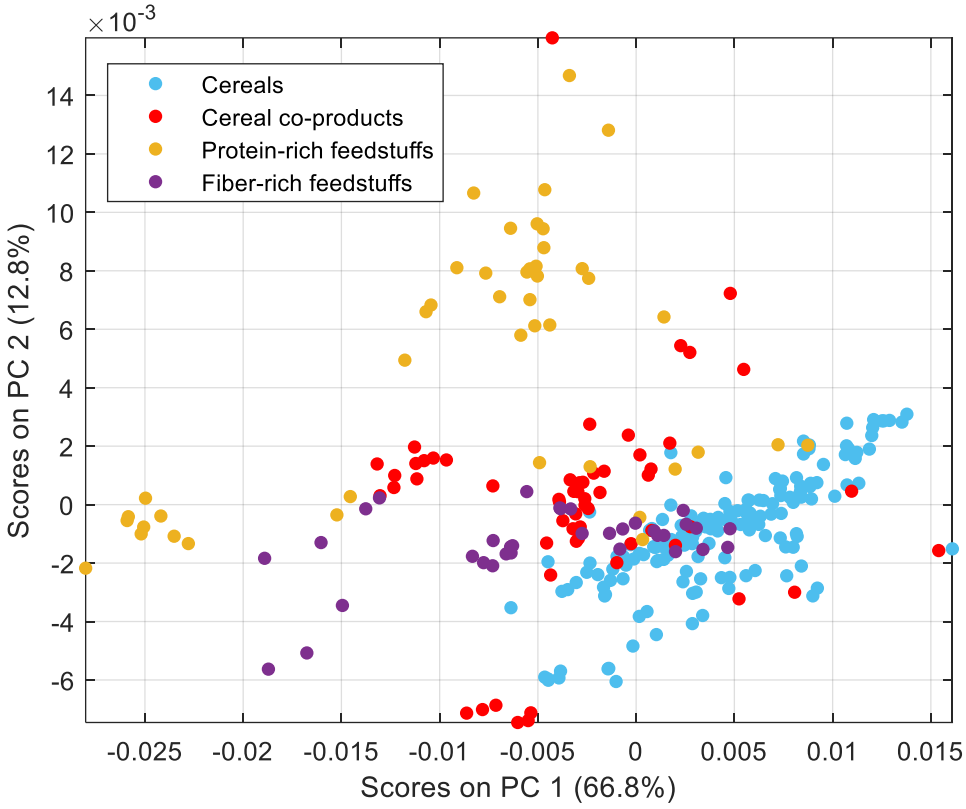


Figure 3. Score plot of the PCA model showing the four groups of feedstuffs.

For most of the modelled constituents, a considerably large number of latent variables (ranging from 10 to 16) was necessary to accommodate all the sources of variability. This was expected considering the wide variety of feedstuffs in the dataset.

In the datasets, there were some uncommon feedstuffs for which only one or very few examples were available (e.g. palm cake or red clover). Detailed information of the samples is provided in the Supplementary material (Tables S2-9). When these uncommon feedstuffs were included in the calibration set they were often flagged as spectral outliers. This is exemplified in Figure 4 that shows the spectral outliers (Hotelling's T^2 reduced > 3, Q residuals reduced > 3) found for the NSP calibration. Some protein-rich feedstuffs (palm cake, red clover (2)), fibre-rich feedstuffs (potato pulp and sugar beet pulp (2)) and cereal co-products (maize gluten) were

identified as outliers. When these unusual samples were in the validation set, they were predicted with large errors. In addition to the spectral outliers, some outlying reference values were also found when the models were developed. Both spectral and reference outliers were removed since we were unable to re-measure these reference values.

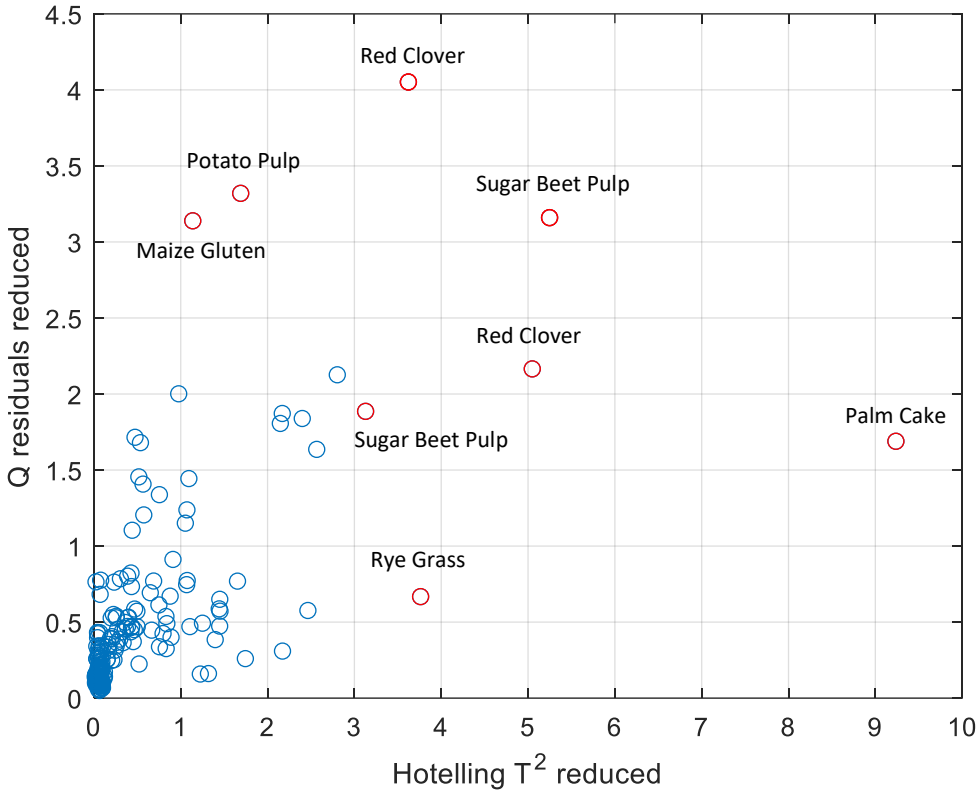


Figure 4. Q-residuals reduced versus Hotelling's T^2 reduced for total non-starch polysaccharide content (T-NSP) calibration. Calibration samples (blue) and calibration spectral outliers (red).

The performance of the NIR models predicting carbohydrates is shown in Table 3. For sugars, fibre-rich feedstuffs and the sample of coconut cake with a very high total sugar content were not included. The samples used in the final calibrations are shown in Table S2. The predictions obtained with the developed model for sugars were good (R^2_p=0.91, CV_p=23%). Starch calibration was developed mostly using cereal and cereal co-products groups, although peas, pea hulls and potato pulp samples were also included (Table S3). The model was developed with starch concentrations ranging from 5% to 90%. The variation from potato pulp and peas samples helped stabilize the model, improving the predictions that are shown in Table 3 (R^2_p=0.98, CV_p=7) by adding samples with concentrations ranging from 20 to 30 g/100g, as fed

and 40 to 50 g/100g, as fed respectively. The accuracy of this calibration is illustrated in Figure 5.A where the reference values for starch are plotted against the predicted values.

Regarding OS, α-galactosides (raffinose, stachyose and verbascose) were predicted using protein-rich feedstuffs while fructans were predicted using cereals and cereal co-products. The samples used in the final calibrations are shown in Table S4 (raffinose), Table S5 (stachyose), Table S6 (verbascose) and Table S7 (fructans). Due to the low number of reference values for α-galactosides, leave-one-out cross-validation (LOOCV) rather than a separate test set was used when evaluating the models. As expected, and as shown in Table 3, the calibration for verbascose with only 16 samples was unacceptable (R^2_{cv}=0.71, CV_{cv}=50%). The calibration models for raffinose and stachyose performed very well (R^2_c>0.98 and low RMSEC values) but the cross-validation results (R^2_{cv}<0.90 and high RMSECV) suggests that these calibrations were not robust enough and that more samples would be needed to account for the sources of variability in the different feedstuffs. As it can be observed in Table 3, the measured values for raffinose and stachyose had standard deviations that are quite similar to their mean values. Fructans were predicted with a low RMSEP and a high R^2_p. As shown in Figure 5.B two groups of samples can be distinguished: all cereals have less than 2 g/100g, as fed of fructans with the exception of rye samples which are grouped on the right-hand side of the plot.

Table 3. Statistics of the calibration models developed for the carbohydrate fraction. Dataset for sugars included cereals, cereal co-products and protein-rich feedstuffs. Dataset for starch contained cereals, cereal co-products, peas, pea hull and potato pulp samples. The dataset for α-galactosides (raffinose, stachyose, verbascose) included only protein-rich feedstuffs. Fructans and β-glucan dataset included cereals and cereal by-products. The rest of calibrations were developed with the entire dataset. Number of samples used for calibration (N_c) and validation (N_v), mean, standard deviation (SD) and range of calibration, pretreatment used (Pret.), number of latent variables (LV), coefficient of determination of calibration (R^2_c) and prediction (R^2_p), root mean square error of calibration (RMSEC) and prediction (RMSEP) and variation coefficient of prediction (CV_p).

	N_c	N_v	Mean±SD (g/100g, as fed)	Range (g/100g, as fed)	Pret.	LV	R^2_c	RMSEC (g/100g, as fed)	R^2_p	RMSEP (g/100g, as fed)	CV_p (%)
Sugars	135	49	2.7±2.1	0.5-8.9	1d	11	0.96	0.37	0.91	0.62	23
Starch	109	41	46±2.5	5.3-92	MSC 1d	16	1.00	1.61	0.98	3.03	7
Raffinose	31	-	1.0±1.2	0.2-5.4	2d	10	0.98	0.17	0.82*	0.46*	46*
Stachyose	41	-	1.4±1.5	0.0-5.0	SNV 2d	11	1.00	0.07	0.90*	0.48*	34*
Verbascose	16	-	1.3±1.4	0.0-4.2	MSC 2d	2	0.80	0.59	0.71*	0.72*	55*
Fructose	89	34	1.3±0.9	0.0-3.5	1d	13	0.98	0.15	0.94	0.23	18
β-glucan	108	37	1.6±1.4	0.0-5.3	MSC 1d	16	0.98	0.2	0.92	0.4	25
T-NCP	184	64	17±11	1-51	SNV 2d	16	0.99	0.95	0.96	2.3	14
I-NCP	186	69	12±8	0.1-36	SNV 1d	16	0.98	1.1	0.94	2.3	19
Cell	201	74	6.6±9.2	0.0-53	SNV 1d	15	0.99	0.80	0.99	1.0	15
T-NSP	194	75	23±18	5.6-85	SNV 1d	16	0.99	1.9	0.96	3.9	17
I-NSP	204	75	18±15	0.4-73	SNV 1d	16	0.99	1.4	0.97	2.7	15
Lignin	189	74	3.8±3.8	0.1-15	2d	10	0.95	0.73	0.93	1.1	29
DF	201	74	28±20	4.2-87	SNV 2d	12	0.98	2.81	0.92	6.00	21

*These values are not obtained from a test set but by leave-one-out cross-validation (LOOCV).

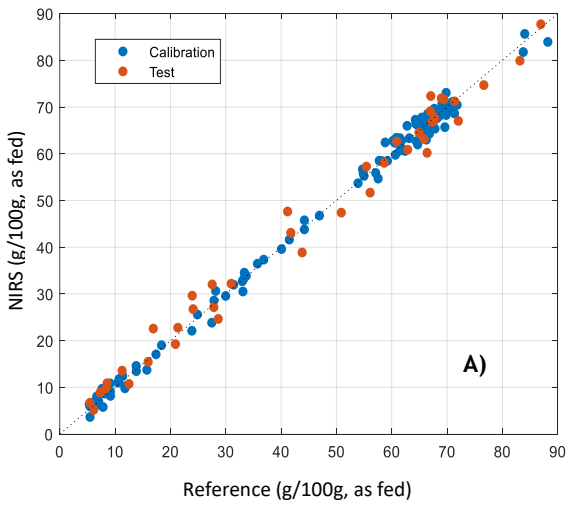
To develop the NIR calibrations for DF, lignin, total and insoluble NSP, cellulose, total and insoluble NCP and all the individual NCP residues both total and insoluble, the entire dataset was used (See Table S8 for a detailed description). The predictions for total NCP (R^2_p=0.91, CV_p=15%) and insoluble NCP (R^2_p=0.94, CV_p=17%) were satisfactory (Table 3, Figure 5.C). The calibrations for total NCP residues were generally slightly better than those for insoluble contents as shown in Table 4.

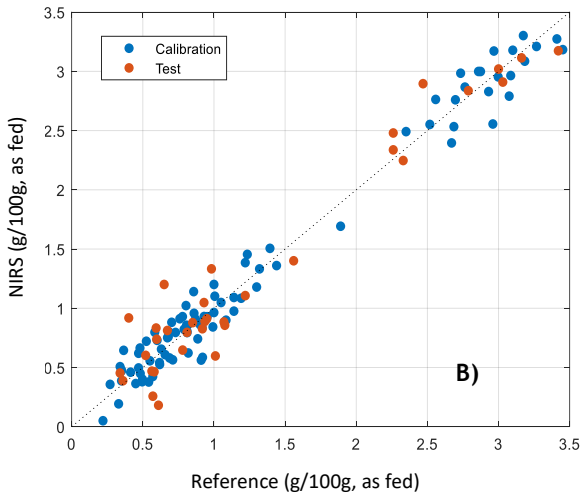
Table 4. Statistics of the calibration models developed for the non-cellulosic polysaccharides (NCP) residues. All the calibrations were developed with the entire dataset. Number of samples used for calibration (N_{cal}) and validation (N_{val}), mean, standard deviation (SD) and range of calibration, pretreatment used, number of latent variables (LV), coefficient of determination of calibration (R^2_c) and prediction (R^2_p), root mean square error of calibration (RMSEC) and prediction (RMSEP) and variation coefficient of prediction (CV_p).

		N_c	N_v	Mean±SD (g/100g, as fed)	Range (g/100g, as fed)	Pret.	LV	R^2_c	RMSEC (g/100g, as fed)	R^2_p	RMSEP (g/100g, as fed)	CV_p (%)
Arabinose	T	204	75	4.2±3.7	0.1-21	SNV 2d	13	0.98	0.47	0.95	0.93	22
	I	190	72	3.1±2.6	0.3-14	2d	13	0.99	0.48	0.96	1.00	32
Xylose	T	204	79	5.7±4.7	0.1-21	2d	16	0.99	0.51	0.95	1.2	21
	I	195	77	4.9±4.5	0.1-21	2d	16	0.99	0.48	0.96	1.53	31
Mannose	T	198	75	0.5±0.3	0.2-1.6	2d	16	0.93	0.06	0.75	0.15	30
	I	177	64	0.3±0.2	0.1-1.0	MSC 1d	11	0.80	0.07	0.77	0.09	30
Galactose	T	192	72	1.0±1.1	0.1-5.9	MSC 1d	11	0.99	0.11	0.95	0.26	26
	I	182	66	0.7±0.8	0.0-3.5	SNV 2d	12	0.96	0.12	0.90	0.22	31
Glucose	T	201	71	2.4±1.9	0.3-11	MSC 2d	15	0.86	0.63	0.81	0.87	36
	I	186	68	1.5±4.7	0.0-11	1d	12	0.79	0.51	0.62	0.77	51
Rhamose	T	159	49	0.2±0.3	0.0-1.3	SNV 1d	15	0.99	0.03	0.94	0.05	25
	I	135	44	0.1±0.1	0.0-0.6	SNV 2d	11	0.97	0.02	0.93	0.04	40
Uronic acids	T	202	68	2.4±4.2	0.1-20	MSC 2d	15	1.00	0.20	0.98	0.57	24
	I	191	68	1.0±1.2	0.0-5.2	MSC 2d	14	0.99	0.10	0.95	0.27	27

The R^2_p values of the models for insoluble contents were similar to those of total contents in most of the cases but the CV_p was lower for the total contents. Although the structural features of water extractable NCP residues are very similar to those of the non-extractable residues, they are not the same [5]. The most important NCP residues of cereals and cereal co-products (arabinose and xylose) and of protein and fibre-rich feedstuffs (uronic acids) were predicted well with high coefficients of determination (R^2_p>0.95) and relatively low prediction errors (CV_p<30%) considering the complexity of the reference analyses for these constituents, which usually involves significant analytical errors [25]. However, for glucose, another important NCP residue, the prediction was not accurate and large differences were found between the models for total (R^2_p=0.81, CV_p=36%) and insoluble fraction (R^2_p=0.62, CV_p=51%). The reason could be that NCP glucose is representing different polysaccharides in protein- and fibre-rich feedstuffs. Nevertheless, calibration models based solely on cereals, where glucose primarily represents β-glucan, did not significantly improve. The models for galactose performed well,

especially for the total content (R^2_p=0.95, CV_p=26%). Contrary to this, the model for mannose had a poor coefficient of determination (R^2_p≈0.75) and larger errors (CV_p=30%) and the rhamnose model had a R^2_p>0.90 but a large error for the insoluble fraction (CV_p=40%). NIRS appears to be not sensitive enough for modelling mannose and insoluble rhamnose contents, which are found at very low concentrations (mean<0.5 g/100g, as fed). Models were also developed to predict the soluble fraction of NCP (not shown in the table) but their performance was notably worse (R^2_p<0.80 and CV_p>30% of the mean) for most of the NCP residues. The generally higher prediction error for the soluble components was influenced by the aggregation of analytical errors deriving from total and insoluble components. Only the calibration for uronic acids content (the most abundant soluble component) exhibited a good coefficient of determination (R^2_p=0.95) although the RMSEP was high (CV_p=42%). As previously mentioned, β-glucan models were derived from only cereals and cereal co-products. The samples used in the final calibration for β-glucan are detailed in Table S9. In terms of quality, the prediction of β-glucan (R^2_p=0.92, CV_p=25%) was comparable to total and insoluble NCP residues. A separate calibration for the cluster of samples with contents between 0 and 1 % (wheat, maize, triticale, maize flour and maize gluten) (Figure 5.D) was tested, but these models did not outperform the calibration created with all the samples. Cellulose (R^2_p=0.99, CV_p=15%) (Figure 5.E), total NSP (R^2_p=0.96, CV_p=17%), insoluble NSP (R^2_p=0.97, CV_p=15%) and lignin (R^2_p=0.93, CV_p=29%) were well predicted. Like for T-NSP, DF calibration was obtained with a wide range of reference values and had high coefficients of determination (R^2_p>0.90) shown in Figure 5.F. However, the CV_p for DF was slightly higher than the one for NSP (21% vs. 15%), which could be expected since DF, unlike NSP, is not a parameter determined directly by chemical analysis.





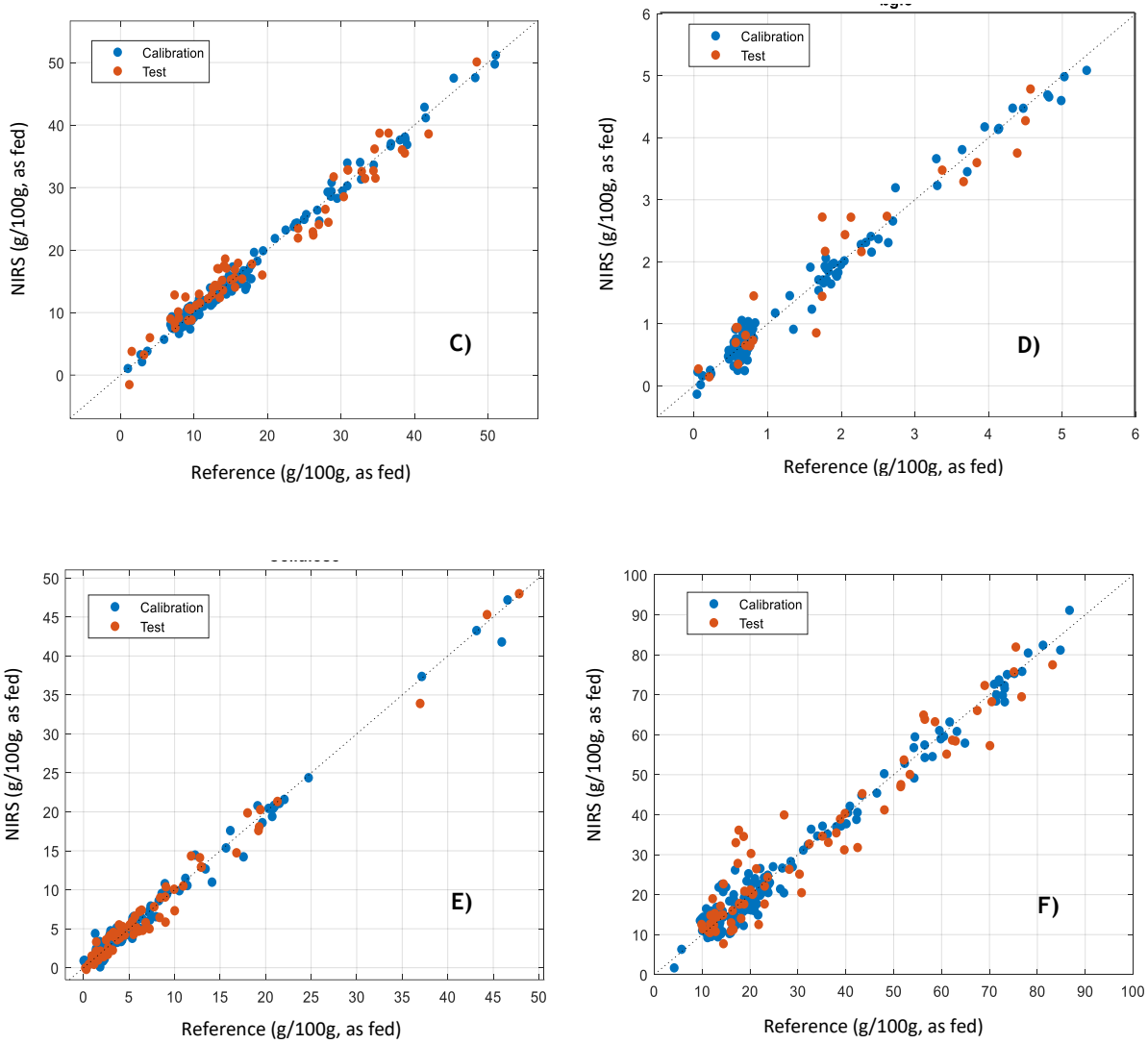


Figure 5. Predicted vs measured values for selected calibrations developed for carbohydrates: A) Starch, B) Fructans, C) Total non-cellulosic polysaccharides, D) β-glucan, E) Cellulose and F) Dietary fibre.

Discussion

The feed industry requires prompt and precise methods to determine the nutritional value of feed ingredients for creating optimal diet formulations. It is essential to lower expenses by avoiding excessive use of costly ingredients and to maintain high standards for the prepared diets. Presently, diets are designed using table values for chemical composition of the individual feedstuffs. Nonetheless, the entire carbohydrate fraction does not appear in these tables. Most tables, only refer to starch and to the traditional measurements of fibre (crude fibre (CF), neutral detergent fibre (NDF) or acid detergent fibre (ADF)). However, these parameters are not enough to understand the fibre fraction. The CF values do not represent the true fibre levels in

feedstuffs and both NDF and ADF fail to account for most, or almost all, soluble NSP in the feed. Moreover, is not totally clear what these fractions represents and hence relevance of these values for monogastric animal nutrition is called into question [33]. Examining the individual sugars, lignin and solubility of structural carbohydrates provides a more comprehensive insight into the dietary fibre makeup of feedstuffs. Unfortunately, the analysis methods to characterize the fibre fraction in this way are quite laborious, and expensive and, as a result, very few data are available in the literature [2]. In addition, as it has been shown CHO and lignin may be highly variable between and within different feedstuffs, hence it is not recommended to assume that an individual batch of feedstuff will have the same fibre profile as another.

For the current study, we have analysed sugars, starch, RS, OS (raffinose, stachyose, verbascose and fructans), cellulose, β-glucan, NSP, NCP residues, lignin and DF of around 300 samples from common and no so-common feedstuffs used to formulate monogastric diets. The distribution of these carbohydrates and lignin in the groups of feedstuffs studied (cereals, cereal co-products, protein-rich feedstuffs and fibre-rich feedstuffs) were in agreement with previous works [34-37].

Focusing on the NIR calibrations developed in this work, prior research revealed the capacity to predict certain CHO in feed components although our database contained more feedstuffs and with greater variation than the vast majority of them. Our results for sugars and starch were similar to those obtained by Losada et al. 2009 [38] who used a dataset that included different cereal and cereal co-products. As far as we know no NIR calibrations for fructans have been reported and only Hollung et al. 2005 [39] predicted α-galactosides in soybean meal where only raffinose showing a good correlation between the reference and the predicted values. Gomes et al. 2020 [17] divided a large dataset into cereals, protein- and fibre-rich feedstuffs and developed accurate, specific calibrations (R^2_{cv}>0.86) for most of the total and insoluble NCP residues. Nieto-Ortega et al. 2022 [11] used a dataset that contained cereals and cereal co-products and accurately predicted (R^2_{cv}>0.90 and CV_{cv}<25%) starch and most of the total and insoluble NCP residues, cellulose and NSP. Like in our investigation, they obtained the worst results for mannose (R^2_{cv}=0.54, CV_{cv}>100). Blakeney and Flinn 2005 [16] employed a cereals sample set and predicted cellulose, β-glucan, arabinose, xylose, glucose and total, insoluble and soluble fractions of NSP. Like in our research, the models for the soluble NSP fraction were much worse (R^2_{cv}<0.70, CV_{cv}>30) than for the total and insoluble fractions (R^2_{cv}>0.80, CV_{cv}<25). In general, the performance of our calibration models was similar to those published by Gomes et

al. 2020 [17] and Nieto-Ortega 2022 [11] for all NCP residues except for glucose but better than those reported by Blakeney and Flinn (2005) [16]. In none of these works, they developed models for rhamnose, uronic acids or NCP as the sum of all residues, which to the best of our knowledge they are for the first time described in this work. Archibald and Kays (2000) [40] predicted DF well in cereals for human consumption although the parameter might not be considered the same since the reference method used for the analysis was different and they did not account for OS and RS content in the DF calculation. For the moment, there are no calibrations reported for DF in feedstuffs used in monogastric animal diets.

The use of NIRS can help the community to access information about the CHO and lignin content of feedstuffs in a much cheaper and faster way than chemical analysis. This can lead to a deeper understanding of the fibre fraction enabling a better formulation of monogastric diets and the development of nutritional strategies such as the use of carbohydrases, which can stimulate fibre fermentation by microbiota, resulting in a positive impact on host metabolism.

Conclusions

NIR calibration models proved to be useful to predict the CHO that are digested by monogastric animals (sugars and starch) and most of those which are part of dietary fibre from a variety of feedstuffs used in animal feeding. We predicted accurately DF, lignin, T-NSP, I-NSP, cellulose, T-NCP, I-NCP, β-glucan, fructans and the total and insoluble content of the NCP residues: arabinose, xylose, galactose, rhamnose and uronic acids. The models for mannose, glucose, especially insoluble glucose and α-galactosides, especially verbascose were not good enough to consider them for routine analysis. Even though not all the feedstuffs were included in all prediction models (e.g. α-galactosides calibrations contained only protein-rich feedstuffs and fructans contained only cereals) the calibrations were developed for the feedstuffs for which the specific carbohydrates are important. The excluded samples were, in most cases, unusual feeds or feeds with very low carbohydrate content whose values are not strictly necessary to know when formulating diets. Even so, it would be convenient to expand the dataset especially with more protein- and fibre-rich samples to improve some calibrations or to develop specific calibrations for each group of feedstuffs. This may be investigated in the future when new relevant samples become available.

Acknowledgments

A. Cruz-Conesa acknowledges Universitat Rovira i Virgili (Tarragona, Spain) and the Institute of Agrifood Research and Technology (Constantí, Spain) for providing a Marti Franqués Research Fellowship (2019PMF-PIPF-62) and the Department of Animal and Veterinary Sciences at Aarhus University (Foulum, Denmark) for providing the dataset used in this study.

Founding sources

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004770.

References

[1] K.E. Bach Knudsen, H.N. Laerke, H. Jorgensen, Carbohydrates and carbohydrate utilization in swine, in: L.I. Chiba (Ed.), Sustainable swine nutrition, John Wiley & Sons Inc., Hoboken, New Jersey, 2013: pp. 109–137.

[2] K.E. Bach Knudsen, H.N. Laerke, Carbohydrates and lignin in the feed - from sugars to complex composed fibres, in: P.J. Moughan, W.H. Hendriks (Eds.), Feed evaluation science, Wageningen Academic Publishers, Wageningen, Netherlands, 2018: pp. 113–144.

[3] J.H. Cummings, A.M. Stephen, Carbohydrate terminology and classification, Eur. J. Clin. Nutr. 61 (2007) 5-18.

[4] D.M.D.L. Navarro, J.J. Abelilla, H.H. Stein, Structures and characteristics of carbohydrates in diets fed to pigs: A review, J. Anim. Sci. Biotechnol. 10 (2019) 39.

[5] K.E. Bach Knudsen, Fiber and nonstarch polysaccharide content and variation in common crops used in broiler diets, Poult. Sci. 93 (2014) 2380–2393.

[6] T.O. Adebowale, K. Yao, A.O. Oso, Major cereal carbohydrates in relation to intestinal health of monogastric animals: A review, Anim. Nutr. 5 (2019) 331–339.

[7] S. Soltanian, E. Stuyven, E. Cox, P. Sorgeloos, P. Bossier, Beta-glucans as immunostimulant in vertebrates and invertebrates, Crit. Rev. Microbiol. 35 (2009) 109–138.

[8] R. Jha, J.D. Berrocoso, Review: Dietary fiber utilization and its effects on physiological functions and gut health of swine, Animal. 9 (2015) 1441–1452.

[9] A.L. Petry, J.F. Patience, Xylanase supplementation in corn-based swine diets: A review with emphasis on potential mechanisms of action, J. Anim. Sci. 98 (2020) 1–12.

[10] N.E. Ward, Debranching enzymes in corn/soybean meal–based poultry feeds: a review, Poult. Sci. 100 (2021) 765–775.

[11] B. Nieto-Ortega, J.J. Arroyo, C. Walk, N. Castañares, E. Canet, A. Smith, Near infrared reflectance spectroscopy as a tool to predict non-starch polysaccharide composition and starch digestibility profiles in common monogastric cereal feed ingredients, Anim. Feed Sci. Technol. 285 (2022) 115214.

[12] European Food Safety Authority (EFSA), Guidance on the assessment of the efficacy of feed additives, EFSA J. 16 (2018) 5274.

[13] F. Cheli, D. Battaglia, L. Pinotti, A. Baldi, State of the art in feedstuff analysis: A technique-oriented perspective, J. Agric. Food Chem. 60 (2012) 9529–9542.

[14] F. García-Sánchez, L. Galvez-Sola, J.J. Martínez-Nicolás, R. Muelas-Domingo, M. Nieves, Using near-infrared spectroscopy in agricultural systems, in: K. Kyprianidis, J. Skvaril (Eds.), Developments in near-infrared spectroscopy, Intech-Open, London, UK, 2017: pp. 97-128.

[15] A.T. Badaró, J.P. Hebling e Tavares, J. Blasco, N. Aleixos-Borrás, D.F. Barbin, Near infrared techniques applied to analysis of wheat-based products: Recent advances and future trends, Food Control. 140 (2022) 109115.

[16] A.B. Blakeney, P.C. Flinn, Determination of non-starch polysaccharides in cereal grains with near-infared reflectance spectroscopy, Mol. Nutr. Food Res. 49 (2005) 546–550.

[17] G.A. Gomes, T.T. dos Santos, C. Piotrowski, R. García, Development of near infrared calibrations for determination of non-starch polysaccharides content in feedstuff, in: 31st Annual Australian Poultry Science Symposium., Sidney, Australia, 2020: pp. 54–57.

[18] K.E. Bach Knudsen, Carbohydrate and lignin contents of plant materials used in animal feeding, Anim. Feed Sci. Technol. 67 (1997) 319–338.

[19] A. Serena, K.E. Bach Knudsen, Chemical and physicochemical characterisation of co-products from the vegetable food and agro industries, Anim. Feed Sci. Technol. 139 (2007) 109–124.

[20] K.E. Bach Knudsen, B.W. Li, Determination of oligosaccharides in protein-rich feedstuffs by gas-liquid chromatography and high-performance liquid chromatography, J. Agric. Food Chem. 39 (1991) 689–694.

[21] K. Larsson, S. Bengtsson, Bestämning av lättilgängelig kolhydrater I växtmaterial (Determination of readily available carbohydrates in plant material), Chem. Methods Rep. 22 (1983).

[22] B. V. McCleary, D.A. Monaghan, Measurement of resistant starch, J. AOAC Int. 85 (2002) 665–675.

[23] B. V. McCleary, M. Glennie-Holmes, Enzymic quantification of (1-3), (1-4)-β-D-glucan in barley and malt, J. Inst. Brew. 91 (1985) 285–295.

[24] O. Theander, P. Aman, Studies on dietary fibres, Swed. J. Agric. Res. 9 (1979) 97–106.

[25] H. Englyst, H.S. Wiggins, J.H. Cummings, Determination of the non-starch polysaccharides in plant foods by gas - Liquid chromatography of constituent sugars as alditol acetates, Analyst. 107 (1982) 307–318.

[26] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[27] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[28] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[29] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428.

[30] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[31] A. Garrido-Varo, S. Vega, F. Maroto-Molina, M.J. De La Haba, D. Pérez-Marín, On-site quality control of processed land animal proteins using a portable micro-electro-mechanical-systems near infrared spectrometer, J. Near Infrared Spectrosc. 24 (2016) 47–58.

[32] B.G. Osborne, T. Fearn, Near Infrared Spectroscopy in Food Analysis, Longman Scientific & Technical, England, 1988.

[33] M. Choct, Feed non-starch polysaccharides for monogastric animals: Classification and function, Anim. Prod. Sci. 55 (2015) 1360–1366.

[34] N.G. Carlsson, A.S. Sandberg, H. Karlsson, Determination of Oligosaccharides in Foods, Diets, and Intestinal Contents by High-Temperature Gas Chromatography and Gas Chromatography/Mass Spectrometry, J. Agric. Food Chem. 40 (1992) 2404–2412.

[35] M.B. Pedersen, S. Dalsgaard, K.E.B. Knudsen, S. Yu, H.N. Lærke, Compositional profile and variation of Distillers Dried Grains with Solubles from various origins with focus on non-starch polysaccharides, Anim. Feed Sci. Technol. 197 (2014) 130–141.

[36] M. Rodehutscord, C. Rückert, H.P. Maurer, H. Schenkel, W. Schipprack, K.E. Bach Knudsen, M. Schollenberger, M. Laux, M. Eklund, W. Siegert, R. Mosenthin, Variation in chemical composition and physical characteristics of cereal grains from different genotypes, Arch. Anim. Nutr. 70 (2016) 87–107.

[37] A.M. Pustjens, H.A. Schols, M.A. Kabel, H. Gruppen, Characterisation of cell wall polysaccharides from rapeseed (Brassica napus) meal, Carbohydr. Polym. 98 (2013) 1650–1656.

[38] B. Losada, P. García Rebollar, P. Cachaldora, C. Álvarez, J. Méndez, C. de Blas, A comparison of the prediction of apparent metabolisable energy content of starchy grains and cereal by-products for poultry from its chemical components, in vitro analysis or near-infrared reflectance spectroscopy, Spanish J. Agric. Res. 7 (2009) 813–823.

[39] K. Hollung, M. Øverland, M. Hrustić, P. Sekulić, J. Miladinović, H. Martens, B. Narum, S. Sahlstrøm, M. Sørensen, T. Storebakken, A. Skrede, Evaluation of nonstarch polysaccharides and oligosaccharide content of different soybean varieties (Glycine max) by near-infrared spectroscopy and proteomics, J. Agric. Food Chem. 53 (2005) 9112–9121.

[40] D.D. Archibald, S.E. Kays, Determination of total dietary fiber of intact cereal food products by near-infrared reflectance, J. Agric. Food Chem. 48 (2000) 4477–4486.

Supplementary material

The supplementary material contains the results of the determination of the principal constituents (dry matter, crude protein, fat, crude fibre and ash) for each one of the four groups of samples in which the feedstuffs were classified in the current research: cereals, cereal co-products, protein-rich feedstuffs and fibre-rich feedstuffs. It also contains detailed information on the samples used to develop and validate the calibrations for carbohydrates: starch, sugars, raffinose, stachyose, verbascose, fructans, DF, lignin, T-NSP, I-NSP, T-NCP, I-NCP, cellulose, β-glucan, total and insoluble arabinose, xylose, glucose, mannose, galactose, rhamnose and uronic acids.

Table S1. Results (mean, standard deviation (SD) and range) of the determination of the principal constituents (dry matter (DM), crude protein (CP), fat, crude fibre (CF) and ash) per group of samples (cereals, cereal co-products, protein-rich feedstuffs and fibre-rich feedstuffs) obtained by wet chemistry analytical methods.

	Cereals			Cereal co-products			Protein-rich feedstuffs			Fibre-rich feedstuffs		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
DM (%)	87.0	1.5	83.8-90.7	92.1	3.2	85.8-98.4	92.5	3.2	86.5-99.4	92.6	2.3	87.5-97.9
CP (%)	12.1	2.1	7.4-16.0	18.7	5.0	6.7-25.3	36.0	10.6	17.8-65.8	9.5	3.1	5.1-18.4
Fat (%)	3.2	1.0	1.4-6.3	7.5	3.8	1.8-12.5	5.6	3.7	0.4-15.6	1.9	1.0	0.1-4.4
CF (%)	4.0	2.3	0.3-11.8	10.8	6.7	1.0-23.4	11.6	5.5	3.6-30.2	28.7	11.7	19.3-55.0
Ash (%)	2.0	0.7	0.5-3.5	4.8	1.9	0.4-11.4	8.5	6.6	2.9-30.5	4.4	2.0	2.8-11.5

Table S2. Type and number of feedstuff samples used to develop the NIR calibration models for sugars.

Sample type	Group	Number of samples
Barley	Cereal	23
Barley husk	Cereal co-products	1
Cottonseed cake	Protein-rich feedstuffs	2
Faba beans	Protein-rich feedstuffs	1
Flax seed cake	Protein-rich feedstuffs	1
Maize	Cereal	3
Maize bran	Cereal co-products	1
Maize flour	Cereal co-products	3
Maize gluten feed	Cereal co-products	2
Oat	Cereal	1
Oat hulls	Cereal co-products	1
Oat meal	Cereal co-products	1
Palm cake	Protein-rich feedstuffs	1
Peas	Protein-rich feedstuffs	6
Rapeseed cake	Protein-rich feedstuffs	1
Rapeseed meal	Protein-rich feedstuffs	8
Red clover	Protein-rich feedstuffs	2
Rye	Cereal	33
Rye bran	Cereal co-products	10
Rye grass	Protein-rich feedstuffs	2
Rye meal	Cereal co-products	1

Soybean meal	Protein-rich feedstuffs	6
Sunflower seed cake	Protein-rich feedstuffs	1
Sunflower seed meal	Protein-rich feedstuffs	2
Triticale	Cereal	31
Wheat	Cereal	43
Wheat aleurone	Cereal co-products	1
Wheat bran	Cereal co-products	16
Wheat husk	Cereal co-products	1

Table S3. Type and number of feedstuff samples used to develop the NIR calibration models for starch.

Sample type	Group	Number of samples
Barley	Cereal	13
Barley husk	Cereal co-products	1
Brewers spent grain	Cereal co-products	7
Maize	Cereal	3
Maize bran	Cereal co-products	1
Maize flour	Cereal co-products	3
Maize gluten feed	Cereal co-products	3
Oat	Cereal	1
Oat hulls	Cereal co-products	1
Oat meal	Cereal co-products	1
Peas	Protein-rich feedstuffs	6
Pea hull	Fibre-rich feedstuffs	8
Potato pulp	Fibre-rich feedstuffs	9
Rice	Cereal	1
Rice meal	Cereal co-products	1
Rye	Cereal	13
Rye bran	Cereal co-products	10
Rye meal	Cereal co-products	1
Sorghum	Cereal	1
Triticale	Cereal	12
Wheat	Cereal	24
Wheat bran	Cereal co-products	16
Wheat husk	Cereal co-products	1

Table S4. Type and number of feedstuff samples used to develop the NIR calibration models for raffinose.

Sample type	Group	Number of samples
Coconut cake	Protein-rich feedstuffs	1
Cottonseed cake	Protein-rich feedstuffs	3
Faba beans	Protein-rich feedstuffs	1
Flax seed cake	Protein-rich feedstuffs	1
Palm cake	Protein-rich feedstuffs	1
Peas	Protein-rich feedstuffs	6
Rapeseed meal	Protein-rich feedstuffs	8
Soybean meal	Protein-rich feedstuffs	6
Sunflower seed cake	Protein-rich feedstuffs	1
Sunflower seed meal	Protein-rich feedstuffs	2

Table S5. Type and number of feedstuff samples used to develop the NIR calibration models for stachyose.

Sample type	Group	Number of samples
Coconut cake	Protein-rich feedstuffs	1
Cottonseed cake	Protein-rich feedstuffs	3
Faba beans	Protein-rich feedstuffs	1
Flax seed cake	Protein-rich feedstuffs	1
Peas	Protein-rich feedstuffs	6
Rapeseed cake	Protein-rich feedstuffs	1
Rapeseed meal	Protein-rich feedstuffs	8
Red clover	Protein-rich feedstuffs	4
Rye grass	Protein-rich feedstuffs	4
Soybean meal	Protein-rich feedstuffs	7

Sunflower seed cake	Protein-rich feedstuffs	1
Sunflower seed meal	Protein-rich feedstuffs	2

Table S6. Type and number of feedstuff samples used to develop the NIR calibration models for verbascose.

Sample type	Group	Number of samples
Cottonseed cake	Protein-rich feedstuffs	2
Faba beans	Protein-rich feedstuffs	1
Peas	Protein-rich feedstuffs	6
Rapeseed meal	Protein-rich feedstuffs	2
Soybean meal	Protein-rich feedstuffs	5

Table S7. Type and number of feedstuff samples used to develop the NIR calibration models for fructans.

Sample type	Group	Number of samples
Barley	Cereal	21
Barley husk	Cereal co-products	1
Maize	Cereal	3
Maize bran	Cereal co-products	1
Maize gluten feed	Cereal co-products	1
Oat	Cereal	1
Oat hulls	Cereal co-products	1
Rye	Cereal	33
Rye bran	Cereal co-products	10
Rye meal	Cereal co-products	1
Triticale	Cereal	27
Wheat	Cereal	43
Wheat aleurone	Cereal co-products	1
Wheat bran	Cereal co-products	16
Wheat husk	Cereal co-products	1

Table S8. Type and number of feedstuff samples used to develop the NIR calibration models for dietary fibre (DF), lignin, total non-starch polysaccharides (T-NSP), insoluble non-starch polysaccharides (I-NSP), cellulose, total non-cellulosic polysaccharides (T-NCP), insoluble non-cellulosic polysaccharides (I-NCP), total and insoluble arabinose, xylose, glucose, mannose, galactose, rhamnose and uronic acids.

Sample type	Group	Number of samples
Apple pomace	Fibre-rich feedstuffs	1
Barley	Cereal	30
Barley husk	Cereal co-products	1
Brewers spent grain	Cereal co-products	10
Coconut cake	Protein-rich feedstuffs	1
Cottonseed cake	Protein-rich feedstuffs	3
Faba beans	Protein-rich feedstuffs	1
Flax seed cake	Protein-rich feedstuffs	1
Grass meal	Fibre-rich feedstuffs	2
Lucerne	Fibre-rich feedstuffs	3
Maize	Cereal	3
Maize bran	Cereal co-products	1
Maize flour	Cereal co-products	3
Maize gluten feed	Cereal co-products	3
Oat	Cereal	1
Oat hulls	Cereal co-products	1
Oat meal	Cereal co-products	1
Palm cake	Protein-rich feedstuffs	1
Peas	Protein-rich feedstuffs	6
Pea hull	Fibre-rich feedstuffs	10
Potato pulp	Fibre-rich feedstuffs	9
Rapeseed cake	Protein-rich feedstuffs	1
Rapeseed meal	Protein-rich feedstuffs	8
Red clover	Protein-rich feedstuffs	4
Rice	Cereal	1

Rice meal	Cereal co-products	1
Rye	Cereal	33
Rye bran	Cereal co-products	10
Rye grass	Protein-rich feedstuffs	4
Rye meal	Cereal co-products	1
Sorghum	Cereal	1
Soybean meal	Protein-rich feedstuffs	7
Sugar beet pulp	Fibre-rich feedstuffs	10
Sunflower seed cake	Protein-rich feedstuffs	1
Sunflower seed meal	Protein-rich feedstuffs	2
Triticale	Cereal	32
Wheat	Cereal	59
Wheat aleurone	Cereal co-products	1
Wheat bran	Cereal co-products	16
Wheat husk	Cereal co-products	1

Table S9. Type and number of feedstuff samples used to develop the NIR calibration models for β-glucan.

Sample type	Group	Number of samples
Barley	Cereal	19
Barley husk	Cereal co-products	1
Maize	Cereal	3
Maize bran	Cereal co-products	1
Maize flour	Cereal co-products	3
Maize gluten feed	Cereal co-products	3
Oat	Cereal	1
Oat hulls	Cereal co-products	1
Oat meal	Cereal co-products	1
Rice	Cereal	1
Rice meal	Cereal co-products	1
Rye	Cereal	31
Rye bran	Cereal co-products	3
Rye meal	Cereal co-products	1
Sorghum	Cereal	1
Triticale	Cereal	32
Wheat	Cereal	42
Wheat aleurone	Cereal co-products	1
Wheat bran	Cereal co-products	6
Wheat husk	Cereal co-products	1

3.3. Development of near-infrared calibration models for the nutrient content of poultry and swine compound feeds

Abstract

Animal feed manufacturers and their customers have a shared interest in checking the nutrient content of compound feeds. Near-infrared spectroscopy (NIRS) offers a more efficient and environmentally friendly alternative to traditional analytical methods. This study has found that commercial calibrations for determining crude protein, fat, crude fibre and ash in compound feeds may not be accurate enough (with RMSEP 6.4 g/kg, 5.7 g/kg, 7.9 g/kg and 11.8 g/kg, respectively) for a research institution needs and that calibrations must be developed in-house. Global calibrations were created from a calibration set that contained 1056 feed samples of broiler chickens, laying hens, broiler turkeys, pigs and sows, showing excellent performance in predicting crude protein, fat, and crude fibre (with RMSEP equal to 3.7 g/kg, 2.4 g/kg and 3.6 g/kg, respectively). Gross energy and ash predictions had acceptable prediction errors (with RMSEP of 46 kcal/kg and 4.9 g/kg, respectively) comparable to the standard error of the laboratory (SEL=40 kcal/kg and SEL=5 g/kg, respectively) but they were slightly biased when compound feeds for single species were predicted due to lower variability in reference data and indirect relationships with NIR spectra. Phosphorus prediction was unsatisfactory (R_p^2<0.50 and high RMSEP). Although specific calibrations were developed for broiler chickens and pigs, the global models were more widely applicable and are recommended.

Introduction

Both animal feed manufacturers and purchasers need to check the chemical composition of compound feeds. Compliant feed samples contribute to the health and correct development of the animals in their different stages of growth. On the contrary, noncompliant feeds can result in nutritional disorders, poor productive performance and increase the environmental impact of animal waste [1].

Animal feed inspection has been based for decades on the proximate system of feed analysis by Weende, which consists of the analysis of dry matter (DM), crude protein (CP), fat, crude fibre

(CF) and ash [2]. With the determination of these and other constituents by near-infrared spectroscopy (NIRS), the use of time-consuming, expensive and polluting wet chemical analytical methods can be reduced to a minimum. With NIRS, these methods are only needed to acquire reference values for the development of NIR models, to periodically verify that the model is working as intended and to analyse atypical feed formulations that the models fail to predict [3].

Compound feeds are spectrally complex due to the wide variety of ingredients and proportions in which they can be combined. To develop acceptable calibrations, it is necessary to use many samples with a variation of ingredients as large as possible [4]. Due to the difficulty in creating such large datasets, most feed producers or purchasers opt for using the calibration models provided by the NIR instrument manufacturers (e.g. FOSS) or specialized companies (e.g. Evonik) instead of developing their own. However, when the producers or purchasers have atypical ingredients and mixtures, which is very common in research institutions on animal nutrition, these calibrations do not always work well and must be updated with their own samples. Alternatively, the producer can opt for developing fresh in-house calibrations [5]. This latter is also the only choice for parameters that have traditionally received little attention but are gaining importance, such as phosphorus (P), gross energy (GE) or specific carbohydrates.

In this work, the performance of commercial calibrations in a NIRS DS2500 (Foss NIR Systems, Denmark) for predicting CP, fat, ash and CF in compound feeds for broilers, pigs, sows, laying hens and turkeys is evaluated and compared with models developed in-house with own samples. For other less common parameters (GE, P) only the calibrations developed with in-house samples are presented. Finally, the global models are compared with the specific calibrations trained using one animal species.

Materials and methods

Samples

The dataset contained compound feed samples produced from 2018 to 2021 at the Institute of Agrifood Research and Technology (IRTA), Constanti, Spain. The models were developed with the samples from 2018 to 2020 and validated with the samples of 2021. The calibration set

contained 1056 samples of compound feeds for broilers (618), pigs (289), sows (58), laying hens (30) and turkeys (61). The validation set contained 443 samples of compound feeds for broilers (261), pigs (126), sows (27) and laying hens (29). The validation set did not contain turkey feeds since they were not produced in 2021 in this research institute. CP, fat, CF, ash and P were determined according to the AOAC, 2016 methods 925.09, 968.06, 978.10, 942.05 and 965.17 respectively [6]. GE was determined by calorimetry using an adiabatic calorimeter (C2000, IKA, Staufen, Germany) according to the DIN 51900 (2005) norm. For the six parameters studied in this work the standard error of the laboratory (SEL) was 4 g/kg for CP, 3 g/kg for fat, 5 g/kg for CF, 5 g/kg for ash, 40 kcal/kg for GE and 0.4 g/kg for P The spectrum of approximately 100 grams of sample was measured on a NIRS DS2500 (Foss NIR Systems, Denmark) with a 10.2 cm diameter cup in reflectance mode from 800 to 2499.5 nm every 0.5 nm and averaging 8 scans per sample. Not all analytical parameters were determined in all the samples because the compound feeds spanned four years of studies, each with a particular list of required analytical parameters.

Commercial calibrations

According to the information provided by the manufacturer of NIRS DS2500, the purchased calibrations consisted of artificial neural networks (ANN) models from a sample set of around 2000 compound feeds corresponding to different animals including broiler chickens, pigs, sows, laying hens, broiler turkeys, cows and rabbits. There were two different calibrations for ash: one for compound feeds of laying hens whose ash content is usually higher (4.5-17.5 g/kg) and one for the rest whose ash content is lower (3-8.5 g/kg).

Data analysis

The spectra and the score plots of principal component analysis (PCA) were checked for sample distribution. Samples were labelled as spectral outliers and thus discarded when the leverage was too high (Hotelling's T^2reduced > 3) or the percentage of residual spectral variance of the sample was too high (Q residuals reduced > 3). Partial least squares regression (PLSR) was used to develop the in-house calibration models for CP, fat, CF, ash, GE and P with the samples produced from 2018 to 2020. Common spectral pretreatments were tested including normalization, standard normal variate (SNV) [7] multiplicative scatter correction (MSC) [8] and 1st and 2nd order derivatives using different window widths [9]. A five-fold venetian blind cross

validation was used to choose the optimal pretreatment and the optimal number of latent variables (LVs) for the models. Calibration samples were labelled as reference outliers when the difference between the predicted and the reference value was too high (|prediction error| > 3×RMSEC of the model). PLS toolbox software (2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab R2020a (The MathWorks Inc., Natick, MA, USA) was used to perform the chemometric calculations. The performance of the commercial and the in-house calibrations was compared using the test set that contained the broiler chicken, pig, sow, and laying hen compound feed samples collected during 2021. Additionally, validation was performed on compound feeds specific to each animal category by selectively extracting the respective samples from the test set. Because the sample set contained sufficient compound feeds for broilers and pigs, we studied whether the models developed for a single species could predict better than the global models that combined the compound feeds for broilers, pigs, sows, laying hens and turkeys. Three statistical tests were applied to check the validity of the predictions: a t-test on slope of the predicted versus reference regression line, a t-test on the bias of the predicted values and a F-test on standard error of prediction (SEP) according to ISO 12099:2017 procedure, whose object is to guide the development and maintenance of NIR calibrations in the agri-food sector [5].

Results

Figure 1 shows the distribution of the reference values for each parameter for each animal species. Compound feeds for turkeys contained more protein, followed by broilers and pigs. These three groups of compound feeds had also more fat and GE. Sow feeds were very rich in fibre and laying hen feeds in ash content. P was not determined in turkey feeds and was slightly higher in broiler and laying hen feeds than in pig and sow feeds. Growing animals (pigs, broiler chickens and turkeys) need more energy and protein to increase their body weight. Laying hens need large amounts of calcium for eggshell formation thus their feeds are rich in calcium carbonate, resulting in a higher ash content when the feed is analysed. Regarding fibre, several studies have shown that a high content of fibre in sow feed enhances the animal welfare by increasing satiety and improving reproductive efficiency [10].

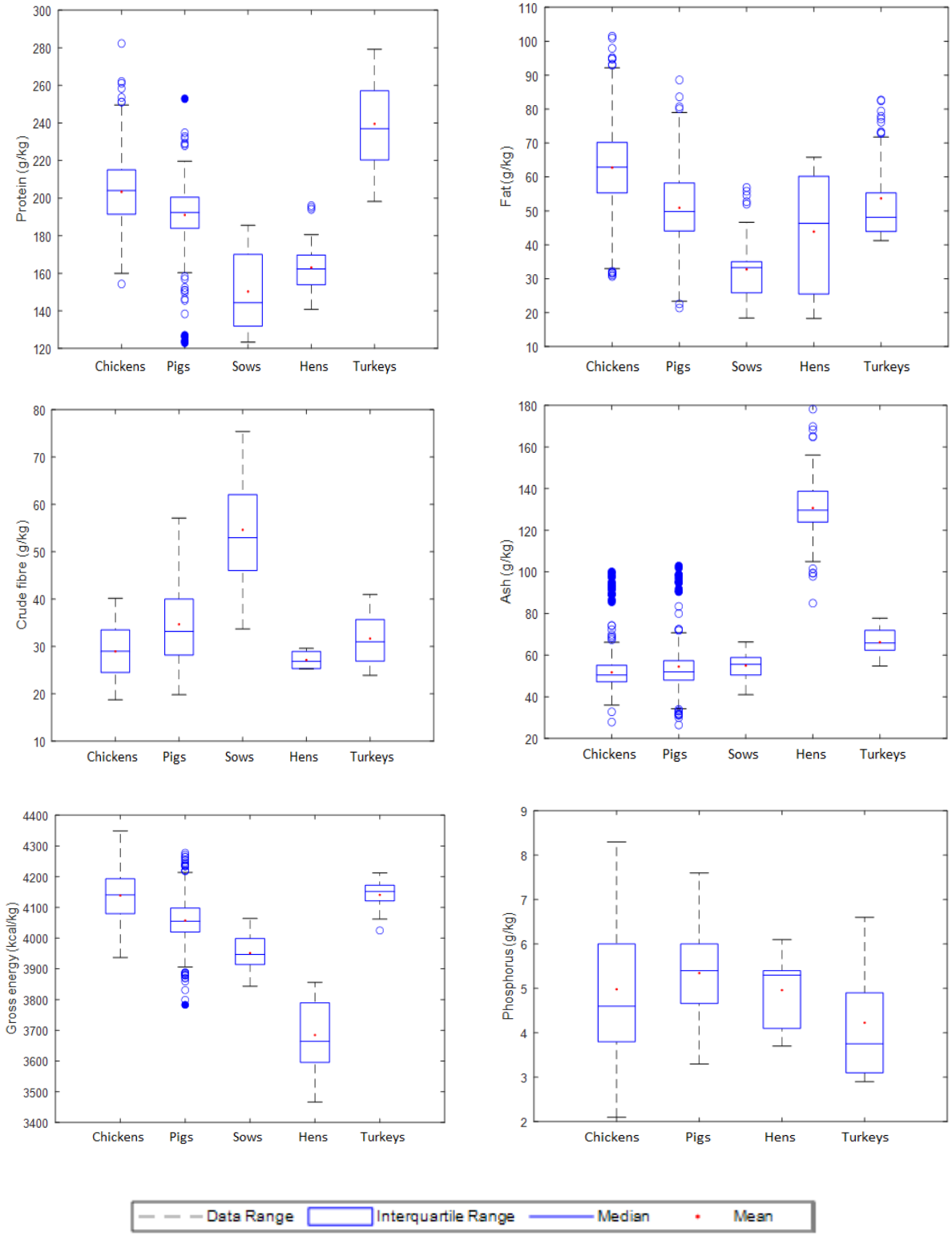


Figure 1. Box and whisker plot of the measured parameters of the five groups of compound feeds: broiler chickens, pigs, sows, laying hens and turkeys.

Figure 2.A shows the mean NIR spectra of the compound feeds of the five animal species. Spectra were all similar in shape and the intensity differences were mainly an offset. The pretreated spectra with SNV and the 1st derivative removed the offset, showing better the differences between compound feeds that cannot be observed from the raw spectra (Figure 2.B). Since the NIR bands overlap to a large extent, it is difficult to find direct correlations between the spectra and the components of the compound feeds. Protein is correlated with C-H and N-H bands. The first overtone at 1900 nm and the combination band at 2100 nm are overlapped with the O-H band. The second overtone around 1500 nm would be more easily assignable. The principal bands related to fat would be those for C-H, whose first overtone appears around 1700 nm, the second at 1200 nm and the combination bands around 2300 nm. Gross energy content correlates with fat content, so the C-H bands are also representative of this parameter. Ash and phosphorus can be determined through their association with organic components such as the phytate molecule in the case of phosphorus [11].

PCA was applied after preprocessing the spectra with SNV followed by first derivative and mean centring. 34 of the 1499 samples were considered spectral outliers because they had high Hotelling's T^2 and/or high Q spectral residuals and were removed from subsequent discussions.

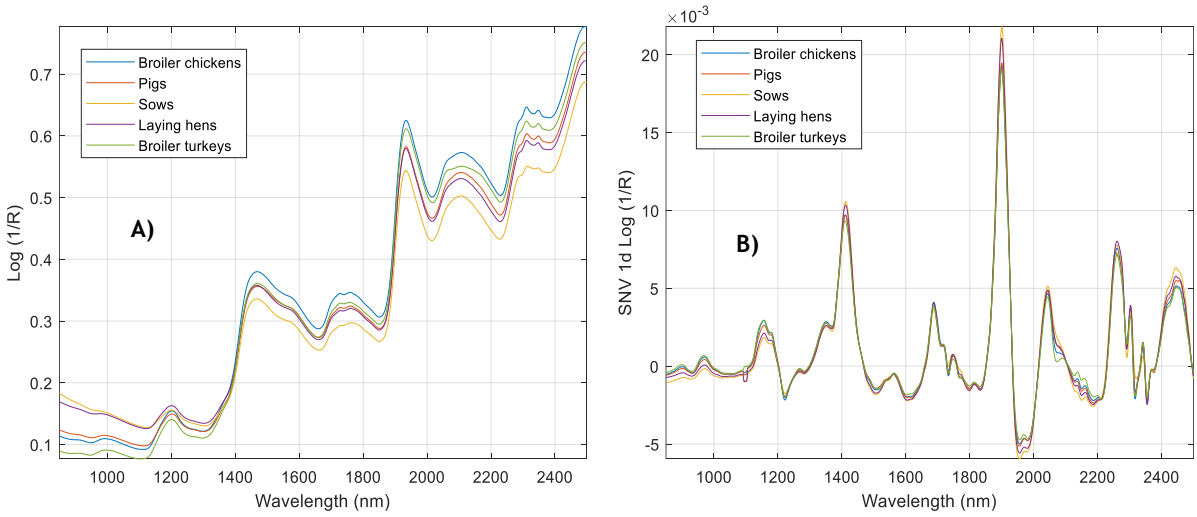


Figure 2. Mean NIR spectra of the compound feed samples for broiler chickens, pigs, sows, laying hens and broiler turkeys. A) Raw spectra. B) Spectra pretreated with standard normal variate (SNV) and the first derivative.

PCA scores (Figure 3) show the variability within and between the groups of compound feeds by animal species. Broilers and pigs are heterogeneous groups and overlap, although pig feeds are located towards the top-right of the graph while broiler feeds are at the bottom-left. As

expected, due to similarity between species, turkeys appeared together with the broiler chickens. The absence of clear separation between groups indicates that a model that combines these three groups of samples could work well. Most of the compound feeds for sows and laying hens were located far from the centre of the graph, indicating that they present different sources of variability than the compound feeds for pigs, broilers and turkeys. Feed for laying hens and sows is not intended for the growth or fattening of the animal but for reproductive efficiency. As it was shown (Figure 1), their feed contains less CP, fat and GE than those of broiler chickens, turkeys or pigs, more ash in the case of laying hens and more fibre in the case of sows. However, the low number of samples from sows and laying hens did not allow us to develop specific models for them. Therefore, so far, the only way to predict parameters for sow and laying hens is to use the commercial models or to include sow feeds and laying hen feeds along with samples from the other three groups to develop global calibrations.

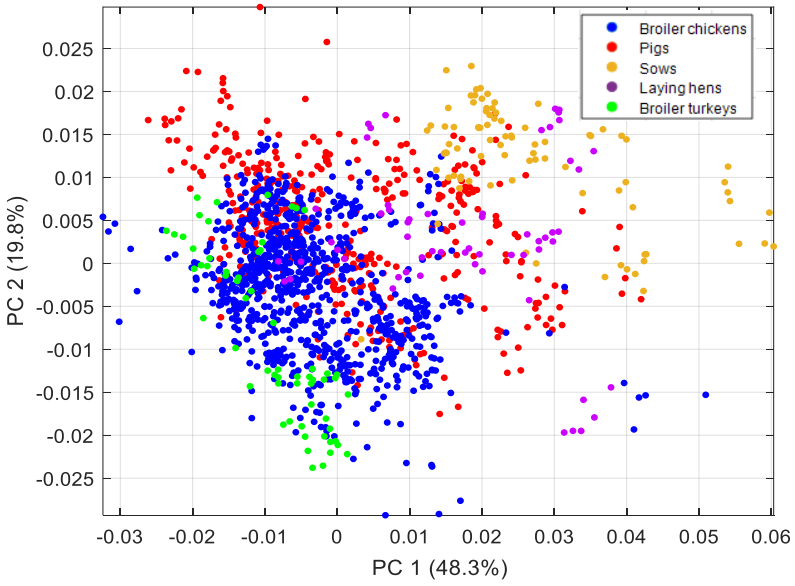


Figure 3. Score plot of the PCA model showing the five groups of compound feeds.

Performance of commercial models

The commercial models were used to predict CP, fat, CF and ash in the test set made of samples collected in 2021 (Table 1). The CP model had a prediction error for all the samples of the test set and for the samples divided by animal species (RMSEP<8 g/kg) that was lower than two times the SEL (4 g/kg), which is the common acceptance criterion [12], and the coefficients of determination were high (R_p^2>0.80). However, the predictions were largely biased according to

the t-test on bias. Predictions of fat, in addition from being biased, were also skewed as the slope of the predicted versus measured regression lines show. The CF model did not present a significant bias but the coefficients of determination and the slope of the predicted versus measured regression lines were poor. Ash was not well predicted for any animal.

Table 1. Performance of the commercial models for the test set. Number of predicted samples (N), coefficient of determination of prediction (R_p^2), root mean square error of prediction (RMSEP), bias and slope of the predicted vs measured regression line.

Test set	N	R_p^2	RMSEP	Bias	Slope
Protein (g/kg)					
All	433	0.92	6.4	-1.2	1.01
Broiler	257	0.85	7.2	-1.3	1.02
Pig	123	0.89	5.3	-1.5	0.99
Sow	27	0.97	5.2	-3.8	1.03
Laying hen	28	0.79	6.1	3.3	0.80
Fat (g/kg)					
All	386	0.90	5.7	-3.5	0.79
Broiler	228	0.86	6.5	-4.3	0.68
Pig	111	0.65	4.9	-2.4	0.79
Sow	19	0.90	2.1	-1.6	0.83
Laying hen	29	0.97	4.3	-2.7	0.85
Crude fibre (g/kg)					
All	122	0.46	7.9	0.3	0.79
Broiler	12	0.47	6.8	0.7	0.77
Pig	90	0.22	8.4	-0.5	0.66
Sow	20	0.71	5.9	1.3	0.68
Ash (g/kg)					
All	414	0.84	11.8	-6.1	0.63
Broiler	246	0.18	7.0	-3.5	0.51
Pigs	117	0.20	8.6	-4.2	0.50
Sow	17	0.88	8.6	-8.5	1.00
Laying hen	32	0.46	33.3	-31.6	0.48

Performance of the global models

Table 2 shows the descriptive and calibration statistics for the global models based on PLSR. Of the different preprocessing methods tested, SNV followed by the 1st derivative and mean centring performed the best for CP, fat, CF and ash, while MSC instead of SNV was preferred for phosphorus and the use of the 1st derivative without any scatter correction pretreatment was the best for gross energy. The optimal number of latent variables was high (14-16) for all the models, which indicates the high variability within the calibration set due to the inclusion of compound feeds for various animals (broilers, pigs, sows, hens and turkeys) and over three years (2018 to 2020). The statistics for calibration and cross-validation were very good with coefficients of determination ranging from 0.84 for phosphorus to 0.98 for CP and low RMSECV compared to the SEL of the determinations. Table 3 shows the prediction performance of the PLSR models for the 2021 samples. The global calibration predicted well

CP in the test set and in smaller test sets of specific animals (Figure 4.A). The prediction errors (RMSEP≈4g/kg) were similar to the SEL for CP (4 g/kg) and were close to the RMSEC and the RMSECV of the model (3.5 g/kg and 3.7 g/kg, respectively). Bias was negligible and the slope was comparable to 1 with a level of significance α=0.95 according to the t-tests mentioned previously. Prediction of fat was also successful (RMSEP<4g/kg, R_P^2>0.90, RMSEP<4 g/kg being SEL= 3g/kg) for all the groups of samples and all the statistical tests were passed except the one on slope when sow compound feeds were predicted (Figure 4.B). The CF global model showed good performance (R_P^2≈0.80 and RMSEP≈3.5 g/kg being SEL=5 g/kg) and all the statistical tests were passed in the prediction of the entire dataset and in the prediction of pig feed separately (Figure 4.C).

Table 2. Descriptive statistics and calibration statistics for the global models combining broiler chickens, pigs, sows, laying hens and turkeys, and the specific models for broilers and pigs. Number of samples to calibrate (N_C), pretreatment used, in brackets the window width for the derivative, number of latent variables (LV), coefficient of determination of calibration (R_C^2) and cross-validation (R_{CV}^2), root mean square error of calibration (RMSEC) and root mean square error of cross validation (RMSECV).

Model	N_C	Pretreatment	LV	R_C^2	RMSEC	R_{CV}^2	RMSECV
Protein (g/kg)							
Global	1049	SNV 1d (15)	16	0.98	3.5	0.98	3.7
Broiler	617	SNV 1d (15)	14	0.97	3.1	0.96	3.4
Pig	287	SNV 1d (15)	16	0.97	2.9	0.96	3.6
Fat (g/kg)							
Global	934	SNV 1d (15)	14	0.98	2.3	0.97	2.4
Broiler	538	MSC 1d (15)	11	0.97	2.2	0.97	2.3
Pig	268	MSC 1d (15)	13	0.98	1.9	0.97	2.3
Crude fibre (g/kg)							
Global	360	SNV 1d (15)	16	0.92	3.3	0.88	3.9
Broiler	109	MSC 1d (15)	14	0.87	2.0	0.69	3.1
Pig	188	MSC 1d (15)	16	0.93	2.2	0.88	3.0
Ash (g/kg)							
Global	993	SNV 1d (15)	16	0.93	5.7	0.90	6.1
Broiler	573	SNV 1d (15)	16	0.95	2.2	0.94	2.6
Pig	189	SNV 1d (15)	16	0.97	2.7	0.95	3.6
Gross energy (kcal/kg)							
Global	641	1d (31)	16	0.92	33	0.90	36
Broiler	341	1d (31)	14	0.90	27	0.86	32
Pig	166	1d (31)	16	0.93	22	0.88	29
Phosphorus (g/kg)							
Global	283	MSC 1d (15)	16	0.89	0.40	0.84	0.48
Broiler	165	1d (31)	16	0.97	0.24	0.94	0.34
Pig	92	1d (31)	15	0.93	0.23	0.79	0.40

CF was determined only in 12 broiler feeds and 20 sow feeds, and it was not determined in any laying hen feed. These few samples of broilers were accurately predicted (2.5 g/kg) but

presented bias, and a slope (0.78) far from 1. The ash model predicted well the entire test set. The coefficient of determination was high (R_P^2=0.95), the RMSEP (4.9 g/kg) was even lower than the RMSEC (5.7 g/kg) and the RMSECV (6.1 g/kg) of the model and all the statistical tests were successfully passed. However, when the prediction performance is studied for each animal species, low R_P^2 are found for feeds of broilers and pigs (R_P^2<0.70), bias is too high (>1 g/kg) for broilers, sows and laying hens and the slope for sows (1.49) is too different to 1 (Figure 4.D). GE models behaved similarly to ash models. Predictions looked well when the entire dataset was used but worsened when the individual species were considered (Figure 4.E). The global model for P that was promising based on calibration and cross validation statistics, failed when new samples were predicted (Figure 4.F).

Table 3 Performance of the global models that included samples from broiler, pig, sow, hen and turkey feeds. Samples in the test set (N_P), coefficient of determination (R_P^2) and root mean square error (RMSEP) of prediction, bias and slope of the predicted vs measured regression line.

Test set	N_P	R_P^2	RMSEP	Bias	Slope
Protein (g/kg)					
All	433	0.97	3.7	0.3	0.97
Broiler	257	0.94	4.0	0.3	0.97
Pig	123	0.95	3.4	-0.2	0.98
Sow	27	0.99	3.2	0.2	0.90
Laying hen	28	0.92	3.4	0.8	0.96
Fat (g/kg)					
All	386	0.97	2.4	0.2	0.96
Broiler	228	0.97	2.1	-0.1	0.93
Pig	111	0.91	2.3	0.1	0.98
Sow	19	0.93	3.3	0.3	1.25
Laying hen	29	0.97	4.0	-0.2	1.08
Crude fibre (g/kg)					
All	122	0.84	3.6	-0.3	0.91
Broiler	12	0.76	2.5	1.9	0.78
Pig	90	0.79	3.4	-0.2	0.99
Sow	20	0.62	6.7	-0.3	0.66
Ash (g/kg)					
All	414	0.95	4.9	0.1	0.92
Broiler	246	0.63	4.3	1.0	1.05
Pig	117	0.66	4.6	-0.3	0.98
Sow	17	0.78	3.6	1.8	1.49
Laying hen	32	0.88	8.8	-6.9	1.05
Gross energy (kcal/kg)					
All	254	0.91	46	6	0.87
Broiler	133	0.81	33	12	0.89
Pig	81	0.63	46	-18	0.78
Sow	15	0.42	42	-35	0.47
Laying hen	29	0.92	103	95	0.71
Phosphorus (g/kg)					
All	125	0.42	1.0	0.4	0.72
Broiler	76	0.55	1.0	0.7	1.03
Pig	42	0.43	1.0	-0.3	0.73

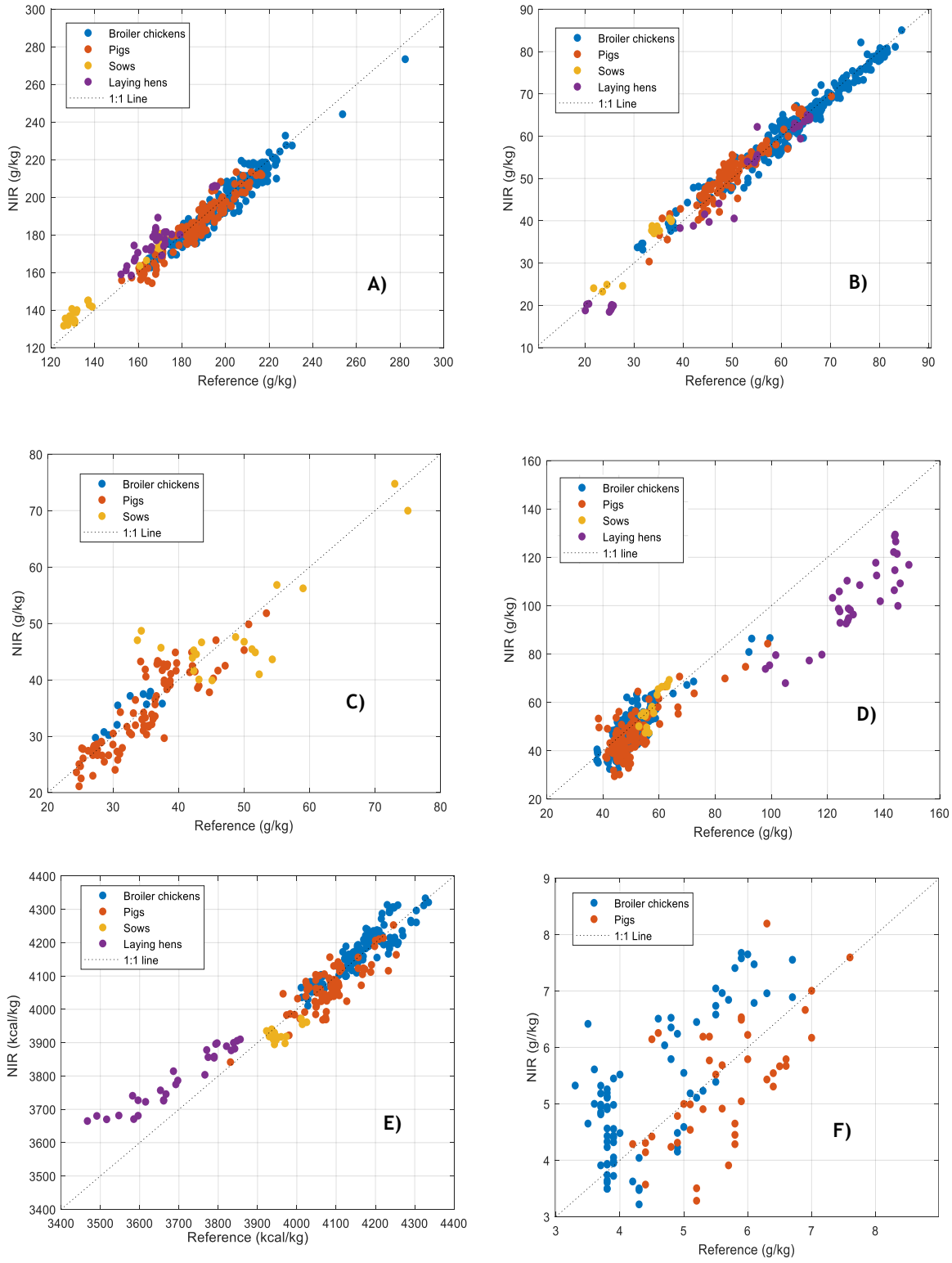


Figure 4. Predicted vs measured values of A) crude protein, B) fat, C) crude fibre, D) ash, E) gross energy and F) phosphorus obtained with the global calibrations and coloured for type of compound feed in the test set: broiler chickens (blue), pigs (red), sows (yellow) and laying hens (purple). The dashed line is the 1:1 line and the dashed red line is the fitted straight line.

Performance of specific models for broilers and pigs

Specific calibrations for broiler compound feeds and pig compound feeds were developed to check whether predictions would improve. The number of LVs for these models was lower than for the global models because sample sets had less samples and spectral variability was lower. The lower calibration and cross-validation errors obtained for the specific models suggested a better predictive ability (Table 2). However, these specific calibrations did not improve the predictions of the test set that were similar to or worse than the predictions from the global calibrations (Table 4).

Table 4. Performance of the specific models for broiler compound feeds and pig compound feeds. Number of samples in the prediction set (N_P), coefficient of determination of prediction (R_P^2), root mean square error of prediction (RMSEP), bias and slope of the predicted vs measured regression line.

Test set	N_P	R_P^2	RMSEP	Bias	Slope
Protein (g/kg)					
Broiler	257	0.94	4.2	0.5	0.95
Pig	123	0.93	4.0	-0.4	0.93
Fat (g/kg)					
Broiler	228	0.95	2.6	-0.1	0.95
Pig	111	0.92	2.0	0.5	0.94
Crude fibre (g/kg)					
Broiler	12	0.75	12.8	1.9	0.78
Pig	90	0.65	4.9	-0.2	0.99
Ash (g/kg)					
Broiler	246	0.43	4.5	1.0	1.05
Pig	117	0.68	4.9	-0.3	0.98
Gross energy (kcal/kg)					
Broiler	133	0.53	50	12	0.89
Pig	81	0.61	50	-18	0.78
Phosphorus (g/kg)					
Broiler	76	0.66	0.9	0.7	1.03
Pig	42	0.68	0.9	-0.3	0.73

Discussion

For the type of samples used in this research institution commercial calibrations were inaccurate. In the case of CP and fat the reason was a large bias that could be produced by differences between the reference method used by the commercial and the reference method used in the IRTA's laboratory (e.g. we used the Dumas method to determine the nitrogen that after is related to protein while the commercial used the Kjeldahl method that usually gives lower values of nitrogen). Although strategies to correct calibrations for bias exist [13] we accumulated

representative samples until a sufficiently large data set was available and developed in-house calibrations that proved to work better than commercials to predict CP, fat, CF and ash.

The developed global calibrations performed well for CP, fat and CF although predictions of the test samples belonging to sows and laying hens generally were predicted worse. The reason might be the lower number of compound feeds from these animals included in the calibration and prediction sets. A few calibration samples from a group are less likely to have sufficient influence during the calculation of the model that will tend to model the most representative groups better. Therefore, one may expect poorer predictions for incoming samples from these underrepresented groups. In addition, since there are also fewer samples in the prediction set, a few larger errors from a few samples significantly dominate the prediction statistics and not represent well the future performance of the model.

The global calibrations for ash and GE performed well when predicting the entire test set, although they showed lower coefficients of determination and higher bias when predicting the test sets of compound feeds of single-species. This is the consequence of two situations. Firstly, the concentration interval of the predicted property was narrower for single-species. For ash, the concentration interval for broilers, pigs and sows goes approximately from 35 to 70 g/kg while with the inclusion of laying hens it is extended to 160 g/kg. For GE, the concentration interval for broilers and pigs is only from 4000 to 4300 kcal/kg but whit the inclusion of laying hens and sows the concentration interval increased and was from 3400 to 4300 kcal/kg. It is known that, given similar prediction errors, the coefficient of determination decreases as the concentration interval decreases. The narrowness of the concentration interval should be compared with the SEL. It is generally said that to achieve acceptable coefficients of determination the range should be at least 10 times the SEL [14]. In the case of GE, the range for broiler and pig compound feeds is only 300 kcal/kg and the SEL is 40 kcal/kg.

Secondly, the fact that GE and ash (different from CP, fat and CF), are predicted based on correlations with other constituents. GE is not a chemical entity although it can be predicted from a NIR spectrum because the energy-yielding nutrients of feeds absorb in the NIR region [15]. Ash is a mineral material that does not absorb in the NIR region, and it is predicted thanks to the correlation between mineral material and organic components which form salts that affect the hydrogen bonds in the samples [11]. These models are more sensitive to new variations in the samples because any change that affects one of the several constituents from which the

property is predicted can affect the final prediction [16]. P presented these two situations (low range (4 g/kg) compared to SEL (0.4 g/kg) and indirect prediction) and others that may worsen the performance of the model such as low concentrations (P<10 g/kg) and fewer samples than in the calibrations for other parameters. Therefore, the model for P was expected to be the worst in prediction. Specific calibrations for broiler and pig compound feeds showed better calibration statistics although similar or even worse predictive ability than global models. This fact could be explained by a trade-off between accuracy and robustness [17]. Global models were built with a larger number of informative samples. Hence, to predict a certain property the PLS model coefficients must be orthogonal to a larger number of sources of variability, often by requiring more latent variables. This decreases the net analyte signal of the property to be predicted, worsening the signal-to-noise ratio and, hence, the accuracy of the model. However, at the same time, this allows the model to account for more sources of spectral variability and make the models less sensitive to changes in future samples. This translates into lower RMSEP for the test sets. Hence, there is a trade-off between including more sources of spectral variability in the model, to make it more robust, and decreasing the net signal-to-noise ratio. From the results of this work, it was found that global calibrations for compound feeds are recommended. This fact, however, may not be always generalizable. For the gain in robustness to outweigh the loss in accuracy, samples from different groups must exhibit a certain level of similarity in their characteristics. The compound feeds for some animals and others were different because the requirements of each animal species are different but they were produced in the same feed mill, using the same processing methods and essentially using the same battery of ingredients. If the samples were very different (e.g. all the compound feeds for pigs were made from corn and all for broilers from wheat) and a calibration is developed joining both types of sample the loss of accuracy would be high. Moreover, the robustness would not increase because the inclusion of samples of a type would not serve the model to better recognize samples of the second type. What is generally true is that when datasets contain only a few samples (that is, do not encompass enough variability), including different types of samples to develop global calibrations tends to work better than developing specific calibrations. However, when data sets contain a large number of samples that encompass enough variability for each of the sample types, specific calibrations may work better [18]. As can be seen (Table 3 and 4), the largest difference in prediction performance between global and specific calibrations was found for CF, whose calibration set contained in total only 360 samples compared to CP (1049), fat (934) or ash (993).

To the best of our knowledge, there are no comparisons between commercial, global, and specific calibrations in the prediction of compound feeds in the literature, although some models, both global and specific, have been presented. Fernández-Ahumada et al. 2008, [19] developed global models that included cattle, lamb, pig, poultry, rabbit and pet compound feeds. Bastianelli et al. 2005 [20] and Molognoni et al. 2020 [21] developed calibrations for poultry without specifying which poultry species included in the calibration set. Paternostre et al. 2021 [22] developed specific models for pig feeds and Khaleduzzaman et al. 2017 [23] for broiler chickens. Our coefficients of determination and prediction errors of cross-validation were in the order of the presented by these authors. In these works, for CP and fat the R^2_{CV} were higher than 0.90 and the RMSECV ranged from 3.7 to 7.0 for protein and from 2.7 to 3.9 for fat. For CF, R^2_{CV} ranged from 0.80 to 0.95 and RMSECV from 4.1 to 7.7. For ash the R^2_{CV} ranged from 0.78 to 0.81 and the RMSECV from 3.2 to 7.4. GE was only predicted by Paternostre et al. 2021 [22] with a R^2_{CV}=0.94 and a RMSECV=24 kcal/kg and P by Khaleduzzaman et al. 2017 [23] with a R^2_{CV}=0.95 and a RMSECV=0.30 g/kg. From all these works only Molognoni et al. 2020 [21] used the calibrations to predict an external dataset. Like in the current research, from the properties they studied (CP, fat, CF and ash) they obtained the worst results for ash in terms of RPD (similar interpretation to R^2).

Conclusions

We have shown that commercial calibrations to determine the nutrient content of compound feeds may not work correctly and therefore the development of calibrations using own samples may be necessary. First, global calibrations including compound feeds for different animals (broiler chickens, pigs, sows, laying hens and broiler turkeys) were developed. The validation with an external dataset revealed that these models had a great performance in the prediction of protein, fat and crude fibre. Gross energy and ash showed some bias when compound feeds for specific animals were predicted. These properties were harder to predict due to the lower variability that the reference data for each type of compound feed showed and the fact that these properties are related to the NIR spectra in an indirect way. Probably the inclusion of more representative samples will reduce bias in future predictions. Phosphorus could not be predicted satisfactory. Specific calibrations for broiler chicken and pig compound feeds were also developed but they did not outperform global models in the prediction of broiler

compound feeds and pig compound feeds, respectively. Therefore, global models that are more widely applicable are recommended from the results obtained in this work.

References

[1] M.R. Garg, P.L. Sherasia, B.M. Bhanderi, B.T. Phondba, S.K. Shelke, H.P.S. Makkar, Effects of feeding nutritionally balanced rations on animal productivity, feed conversion efficiency, feed nitrogen use efficiency, rumen microbial protein supply, parasitic load, immunity and enteric methane emissions of milking animals under field condi, Anim. Feed Sci. Technol. 179 (2013) 24–35.

[2] W. Henneberg, F. Stohmann, Über das Erhaltungsfutter volljährigen Rindviehs, J. Landwirtschaft. 3 (1859) 485–551.

[3] F. García-Sánchez, L. Galvez-Sola, J.J. Martínez-Nicolás, R. Muelas-Domingo, M. Nieves, Using near-infrared spectroscopy in agricultural systems, in: K. Kyprianidis, J. Skvaril (Eds.), Developments in near-infrared spectroscopy, Intech-Open, London, UK, 2017: pp. 97-128.

[4] D.I. Givens, J.L. De Boever, E.R. Deaville, The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans, Nutr. Res. Rev. 10 (1997) 83–114.

[5] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[6] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[7] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[8] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[9] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[10] S. Jarrett, C.J. Ashworth, The role of dietary fibre in pig production, with a particular emphasis on reproduction, J. Anim. Sci. Biotechnol. 9 (2018) 1–11.

[11] L. Chen, Z. Yang, L. Han, A review on the use of near-infrared spectroscopy for analyzing feed protein materials, Appl. Spectrosc. Rev. 48 (2013) 509–522.

[12] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[13] B.M. Wise, R.T. Roginski, A calibration model maintenance roadmap, IFAC-PapersOnLine. 48 (2015) 260–265.

[14] P.C. Williams, K. Norris, Near Infrared Technology in the Agricultural and Food Industries, 2nd Edition, American Association of Cereal Chemists, St Paul, MM, USA, 2001.

[15] E.V. Valdes, S. Leeson, Near Infrared Reflectance Analysis as a Method to Measure Metabolizable Energy in Complete Poultry Feeds, Poult. Sci. 71 (1992) 1179–1187.

[16] M.H. Rhiel, M.I. Amrhein, I.W. Marison, U. Von Stockar, The Influence of Correlated Calibration Samples on the Prediction Performance of Multivariate Models Based on Mid-Infrared Spectra of Animal Cell Cultures, Anal. Chem. 74 (2002) 5227–5236.

[17] H. Zhang, Y. Yu, J. Jiao, E.P. Xing, L. El Ghaoui, M.I. Jordan, Theoretically principled trade-off between robustness and accuracy, 36th International Conference Machine Learning ICML, CA, USA, 2019: pp. 12907–12929.

[18] J.S. Shenk, M.O. Westerhaus, Near Infrared Reflectance Analysis with Single and Multiproduct Calibrations, Crop Sci. 33 (1993) 582–584.

[19] E. Fernández-Ahumada, J.E. Guerrero-Ginel, D. Pérez-Marín, A. Garrido-Varo, Near infrared spectroscopy for control of the compound-feed manufacturing process: Mixing stage, J. Near Infrared Spectrosc. 16 (2008) 285–290.

[20] D. Bastianelli, E. Fermet-Quinet, F. Davrieux, C. Hervouet, L. Bonnal, Calibration strategies for prediction of amino acid content of poultry feeds, in: G. Burling-Claridge, S. Holroyd, R. Summer (Eds.), Proc. 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 314–318.

[21] L. Molognoni, L.A. de Sá Ploêncio, C.T.P. Deolindo, L.V.A. de Oliveira, R.B. Hoff, H. Daguer, FT-NIR combined with chemometrics versus classic chemical methods as accredited analytical support for decision-making: Application to chemical compositional compliance of feedingstuffs, Microchem. J. 158 (2020) 105126.

[22] L. Paternostre, V. Baeten, B. Ampe, S. Millet, J. De Boever, The usefulness of NIRS calibrations based on feed and feces spectra to predict nutrient content, digestibility and net energy of pig feeds, Anim. Feed Sci. Technol. 281 (2021) 115091.

[23] A.B.M. Khaleduzzaman, M.A.A. Mamun, H.M. Salim, Development of local calibrations for the nutritional evaluation of commercial poultry diets by using near infrared reflectance spectroscopy, J. Appl. Anim. Res. 45 (2017) 8–14.

.

.

3.4. Use of visible-near infrared spectroscopy to predict nutrient composition of poultry excreta

Andrés Cruz-Conesa[1,2], Joan Ferré[1*], Anna M. Pérez-Vendrell[2], M. Pilar Callao[1], Itziar Ruisánchez[1]

[1] Universitat Rovira i Virgili, Faculty of Chemistry, Department of Analytical and Organic Chemistry, Tarragona, Spain

*Corresponding author email address: joan.ferre@urv.cat

[2]Institute of Agrifood Research and Technology, Animal Nutrition. Mas Bové, Constantí, Spain

Animal Feed Science and Technology 283 (2022) 115169 (Edited for format)

Abstract

Nowadays optimal feed formulation for poultry is sought for available content, which takes into account how the nutrients are digested and metabolized by the animal. The digestibility coefficients of the nutrients are usually obtained in in-vivo trials that require feeding the birds with different diets of well-known composition and analysing a large number of excreta samples. Nutrient excreta composition is usually found by wet analytical methods. This work presents visible-near infrared (Vis-NIR) calibrations for organic matter, protein, fat, gross energy, uric acid and phosphorus in excreta from bioassays involving broiler chickens, laying hens and turkeys carried out between 2017-2020. The Vis-NIR spectra (400-2499.5 nm) were pretreated by generalized least squares weighting (GLSW) and partial least squares regression (PLSR) was used to obtain the prediction models. The six parameters were properly predicted with the values of ratio of performance of deviation (RPD) and coefficient of determination of prediction (R_P^2) of the validation set ranging from 3.7 to 4.6 and from 0.91 to 0.95 respectively. All but one of the calibrations passed the statistical tests for fit for purpose described in ISO 12099:2017. Despite the global calibrations provided satisfactory results, specific calibrations

for broiler chickens excreta and for laying hen excreta were developed to check if their predictions could be even better but the results did not improve. Finally, the root mean square error of prediction (RMSEP) of the global calibrations was compared with the standard error of the reference methods employed for the analysis of these parameters, confirming their high performance and direct applicability.

Keywords

Poultry excreta, nutrient content, Vis-NIR spectroscopy, global calibrations, specific calibrations

Introduction

Feedstuff is the largest contribution in the cost of poultry production and is a key factor in the animal growth and health. While the traditional formulation of feedstuff has been based on total content of nutrients, nowadays optimal formulations are sought for available content, which takes into account how the nutrients are digested and metabolized by the animal. There are different indicators related to feed digestion. The most important in poultry studies are the total tract digestibility coefficients of the main nutrients (protein, fat or amino acids) and the metabolizable energy (usually determined as apparent metabolizable energy, AME). Other important parameters are the assimilation of calcium and phosphorus because they are needed for the correct formation and maintenance of the skeleton.

The digestibility coefficients of the nutrients and AME are usually obtained through in-vivo assays where birds are fed with different diets of well-known composition [1]. These trials require the collection and analysis of many replicate excreta samples per dietary treatment, trying to minimize external effects such as temperature and humidity in the farm.

Near infrared spectroscopy (NIRS) is a fast analytical technique that may replace the costly wet analytical methods for determining the nutrient content of excreta. Its most important characteristics are the speed, the absence of sample treatment, the null use of solvents and no generation of waste, and the fact that is a multi-parametric technique, that is, a spectrum can be used to predict various parameters simultaneously.

NIR instrument manufacturers commercialize pre-installed "universal calibrations" that are advertised to work well for the routine analysis of the most common raw materials and compound feeds. Nevertheless, commercial calibrations for poultry excreta do not exist currently and only a few calibrations have been reported in the literature. Throughout their different works, Bastianelli et al. [5] presented calibrations for protein, gross energy, fat, uric acid and starch. Smith et al. 2001 [6] and De la Roza-Delgado et al. 2015 [7] developed also calibrations for gross energy, a key parameter that is required to calculate the metabolizable energy of the feedstuff. All these calibrations were applicable only to broiler chickens. Phosphorous has been determined by NIRS in some works focused on the use of the poultry excreta or manure as fertilizer [8-10], while Xing et al. 2008 [11] also presented a calibration model for organic matter.

In summary, there are very few published-multivariate determinations from NIR spectra for the nutrient content of poultry excreta and none of them are valid for animals other than broiler chickens. In this work, we present global calibrations for organic matter, protein, fat, gross energy, uric acid and phosphorus developed with excreta samples of broiler chickens, laying hens and turkeys. These calibrations are validated by an external dataset and also by statistical tests. Finally, they are compared with the specific calibrations developed for each poultry species.

Materials and methods

Samples and bioassays

A total of 1025 samples of poultry excreta were collected from 2017 to 2020 at the Institute of Agrifood Research and Technology (IRTA) in Constantí, Tarragona, Spain. The samples had been obtained from 31 bioassays whose objective was to measure the digestibility of diets involving different combinations of raw materials and additives (e.g., enzymes). 17 of the bioassays involved male Ross 308 broiler chickens with an age between 22 and 25 days (620 excreta samples), 11 HyLine Brown laying hens with 18 to 26 weeks of age (306 excreta samples) and two male (35 and 42 days of age) and one female (24 days) Aviagen Premium turkeys (99 excreta samples), respectively. Diets were based on soybean meal and the main cereal was corn, wheat or barley. Titanium dioxide (TiO_2) was used as indigestible marker in all the studies. The

samples of excreta were lyophilized, milled, and stored in sealed bags in a climatic chamber at 17 ºC until their analysis. The fact that the excreta samples were obtained from animals of different digestive capacity and involved very different diets provided a wide range of undigested contents of the different nutritional fractions that had to be accounted for by the calibration models.

Reference values of nutritional parameters

The excreta samples were analysed in the laboratory with validated methods. Dry matter, nitrogen, fat, ash and phosphorus were determined according to the AOAC methods 925.09, 968.06, 920.39, 942.05, 965.17 respectively [12]. Gross energy was determined by calorimetry using an adiabatic calorimeter (C2000, IKA, Staufen, Germany) according to the DIN 51900 (2005) norm [13]. Uric acid was determined by spectrophotometry following the method described by Marquardt (1983). Organic matter was calculated as the difference between dry matter and ash contents. Protein was obtained subtracting the uric acid from the total nitrogen and multiplying the result by 6.25. For the six parameters studied in this work the standard error of the laboratory (SEL) were as follows: organic matter (9 g/kg, as fed), protein (7 g/kg, as fed), fat (3 g/kg, as fed), gross energy (40 kcal/kg, as fed), uric acid (5 g/kg, as fed), phosphorus (0.4 g/kg, as fed). Since the digestibility experiments span different years and had different objectives, the number of excreta samples that could be used for modelling was not the same for all the analytical parameters.

Visible-near infrared (Vis-NIR) spectra acquisition-and data analysis

Freeze-dried excreta samples were scanned on a NIRS DS2500 (Foss NIRSystems, Denmark) in reflectance mode with a 7 cm diameter cup where 30 grams approximately are introduced each time. Spectra were collected every 0.5 nm from 400 to 2499.5 nm, thus covering the range from visible to NIR. PLS toolbox software (PLS_Toolbox, 2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab (MATLAB, Version R2020a, The MathWorks Inc., Natick, MA, USA) was used to carry out all chemometric treatments.

Partial least squares regression (PLSR) was used to develop the calibration models for organic matter, protein, fat, gross energy, uric acid and phosphorus. Common spectral pretreatments were studied including normalization, standard normal variate (SNV) [14], multiplicative scatter

correction (MSC) [15], derivatives (1st and 2nd) [16], orthogonal signal correction (OSC) [17] and generalized least squares weighting (GLSW) [18]. Cross-validation (CV) was used to choose the optimal pretreatments and the optimal number of latent variables for each model.

The spectra and especially the score plots of the PLS models were checked for sample distribution. The sample set was divided into a calibration set (75% of the samples) and validation set (25% of the samples) using the Kennard-Stone algorithm applied to the spectra, that allows to retain the spectra distribution [19]. Calibration samples were labelled as spectral outliers and thus discarded when the leverage was too high (Hotelling's T^2 reduced>3) or the percentage of residual spectral variance of the sample was too high (Q residuals reduced>3). Samples were labelled as reference outliers when the difference between the predicted and the reference value was too high (Studentized residuals (t)>3).

The validity of the predictions from the calibration models was checked with three statistical tests: a t-test on bias, a t-test on slope and an F-test on standard error of prediction (SEP) according to ISO 12099:2017, whose object is to guide the development and maintenance of NIR calibrations in the agri-food sector [20].

Finally, because the sample set contained a large number of excreta samples from broiler chickens and hens, we studied if calibrations developed for a single species could predict better than the global calibration that included the three species.

Results

Figure 1 shows the distribution of the reference values for each parameter in the poultry sample set, that included the broiler chickens dataset, the laying hens dataset and the turkeys dataset. The values for broiler chickens and laying hens only are also shown. The variability reflects the large diversity of bioassays that generated the excreta samples. The ranges and the standard deviations of the values of protein, fat and phosphorus were much larger in the broiler chickens dataset than in the laying hens dataset. Only for organic matter were clearly larger in the laying hens dataset. The mean values of the parameters were also quite different in each sample set.

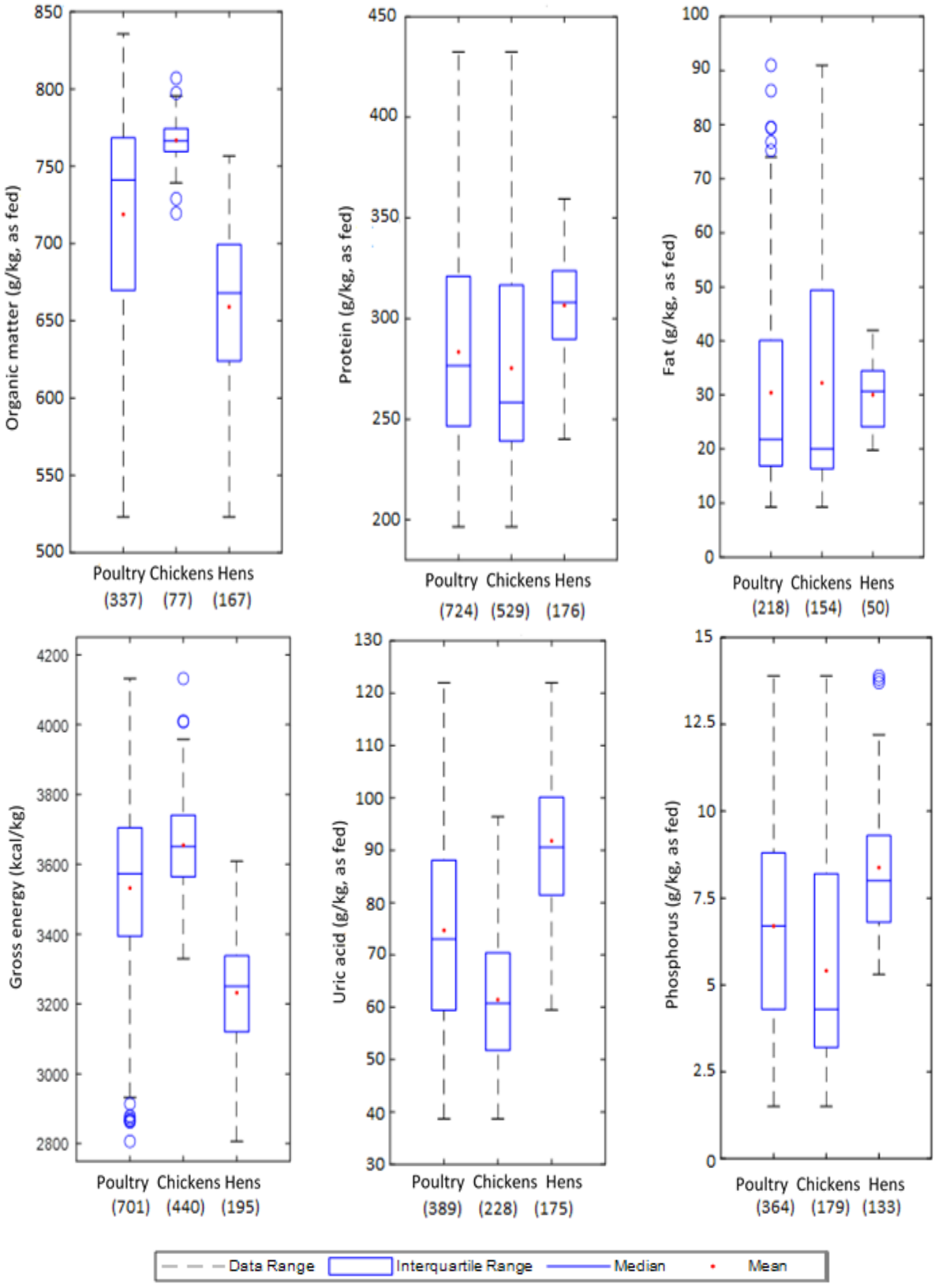


Figure 1. Box and whisker plot of the measured parameters of the three sample sets: poultry (broiler chickens, laying hens and turkeys), broiler chickens and laying hens. In parentheses the number of samples.

Figure 2 shows the mean Vis-NIR spectra of the excreta of the three poultry species. Spectra presented some differences in magnitude in the region from 400 nm to 1200 nm (which includes the visible region and the beginning of the NIR region) but they were all similar in shape and intensity in the NIR region from 1200 nm to 2499.5 where the most characteristic bands related with the nutrient content are found. Since the NIR bands are highly-overlapped it is difficult to find correlations between the spectra and the components of complex samples such as the excreta samples. Organic matter could be correlated with the intensity of the bands corresponding to the C-H, N-H and O-H bonds. The principal bands related to fat would be those for C-H, whose first overtone appears around 1700 nm, the second at 1200 nm and the combination bands around 2300 nm. Gross energy content is correlated with fat content, hence the C-H bands are also representative for this parameter. Protein and uric acid are correlated with C-H and N-H bands. The first overtone at 1900 nm and the combination band at 2100 nm are overlapped with the O-H bands. The second overtone around 1500 nm would be more easily assignable. Phosphorus can be determined through its association with organic components such as the phytate molecule. The spectral differences reflect the differences in the metabolism of the three species and the substantial differences in their diets used in the bioassays. For example, the diets given to laying hens are less energetic and have a higher inorganic content than the diets for broiler chickens. This ultimately affects the composition of the excreta.

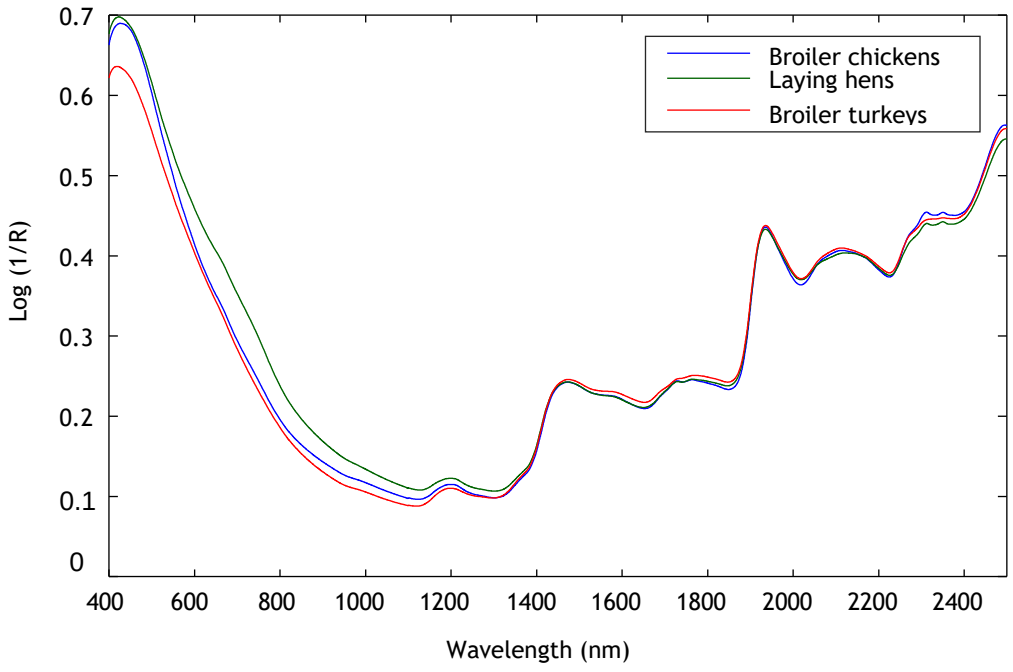


Figure 2. Mean Vis-NIR spectra of the excreta samples for broiler chickens, laying hens and broiler turkeys.

Since the spectra of the different species are similar the calibration models were first developed including the three species. Of the different preprocessing methods tested, autoscaling followed by GLSW performed the best for all the parameters. No spectral outliers were found in the dataset of NDF, ADF and lignin. For the rest of the constituents, between two and five samples were clearly identified as spectral outliers and were removed from subsequent analyses because of their high values of Hotelling's T^2 and Q residuals. Figure 3 shows the grafic of scores of the PLS model for the protein content in the excreta with code color. The scores show that there are no clear clusters of spectra due to the species, as already suggested by the spectra in Figure 2. One can also observe how the values of the property vary with the scores, as it is to be expected from a PLS model. Despite the fact that the birds received different diets and their digestive performance is different, similar score plots are obtained for the PLS models of the other parameters (not shown).

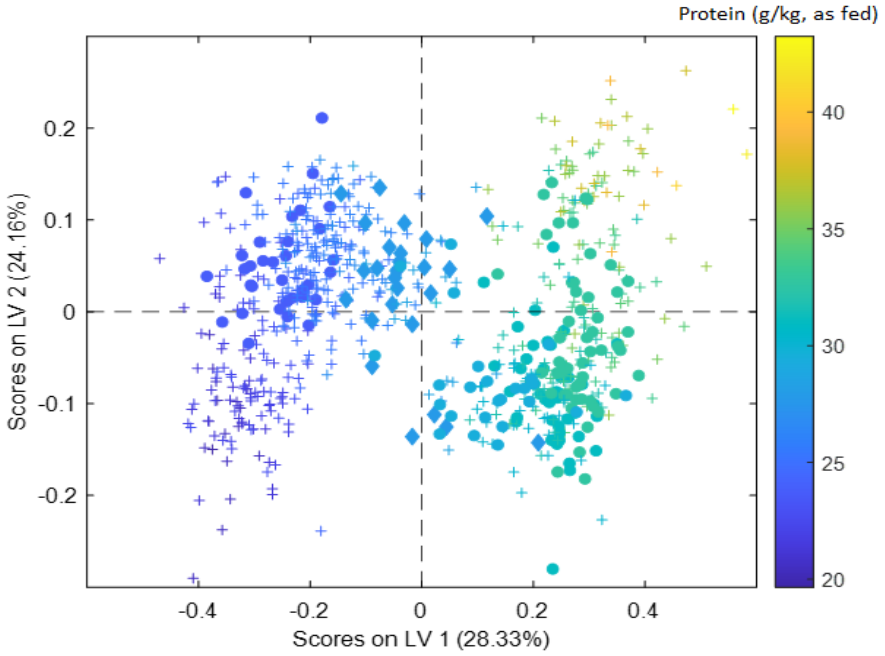


Figure 3. Score plot of the protein PLS model considering the animal species: broiler chickens (crosses), laying hens (circles), broiler turkeys (rhombus) and the concentration of protein measured in each sample (colour bar).

The optimal number of PLS latent variables ranged from 3 in the model for fat to 8 in the model for phosphorus. Once the models had been calculated, the following samples were detected as clear reference outliers and were removed from subsequent modelling: one sample for protein and uric acid, three for fat, four for phosphorus and six for gross energy. Other 11 samples, all with the largest values of protein (350-450 g/kg, as fed), were removed from the protein model

after being predicted with large errors. Removing these samples from the calibration and validation set improved greatly the performance of the calibration for all the other samples. Nevertheless, if such high values of protein are found in future assays, one may consider reintroducing those samples to expand the calibration range. Removing the part of the spectra corresponding to the visible radiation did not improve the performance of the calibration models. Therefore, the full spectral range was used in all models.

The performance of the global calibrations for poultry excreta is summarized in Table 1. The six studied parameters (organic matter, protein, fat, gross energy, uric acid and phosphorous) were well predicted.

Table 1. Statistics of the poultry excreta calibration models: number of samples used to calibrate (N_C) and validate (N_V), root mean square error of calibration (RMSEC) and prediction (RMSEP), coefficient of determination of calibration (R_C^2) and prediction (R_P^2), ratio of performance of deviation in prediction (RPDp), bias and slope of the predicted vs measured regression line.

Parameter	N_C	RMSEC	R_C^2	N_V	RMSEP	R_P^2	RPDp	Bias	Slope
Organic matter (g/kg, as fed)	255	13	0.97	85	14	0.95	4.6	-1.2	1.01
Protein (g/kg, as fed)	522	9.1	0.96	174	9.2	0.95	4.6	0.7	1.00
Fat (g/kg, as fed)	159	3.2	0.96	53	4.0	0.94	3.8	-0.4	0.95
Gross energy (kcal/kg)	513	46	0.94	171	49	0.93	3.7	2.7	0.95
Uric acid (g/kg, as fed)	294	4.9	0.93	98	5.2	0.91	4.5	0.1	0.94
Phosphorus (g/kg, as fed)	264	0.6	0.94	83	0.7	0.93	3.9	0.1	0.94

The ratio of performance of deviation (RPD) for the validation set ranged from 3.7 to 4.6 and the coefficient of determination of prediction (R_P^2) from 0.91 to 0.95. These values were good enough to accept the calibrations for routine use. The measured versus predicted values are shown in Figure 4. The t-test for the bias concluded that none of the calibrations had a significant bias. The t-test for the slope was passed by all the calibrations easily. Lastly, the results of the F-test for the standard error of prediction (SEP) suggested that there is not overfit in all the models except for the model for fat, probably due to the fact that is the model constructed with less samples.

Despite the global calibrations provided satisfactory results, specific calibrations for broiler chicken excreta and for laying hen excreta were developed to check if predictions could improve (Table 2). The statistics in this case were obtained using cross-validation (CV). In the broiler chickens sample set, four of the six studied parameters (protein, fat, gross energy, and phosphorous) were well predicted (RPDcv>3 and R_{CV}^2>0.90). However, the results of the

calibrations for laying hen excreta were poor (RPDcv<3 and R^2_{CV}<0.90) in all cases. The results of the specific calibrations comparing to the global were much worse in the case of the laying hen calibrations and similar or worse for some parameters and much worse for others in the case of broiler chickens.

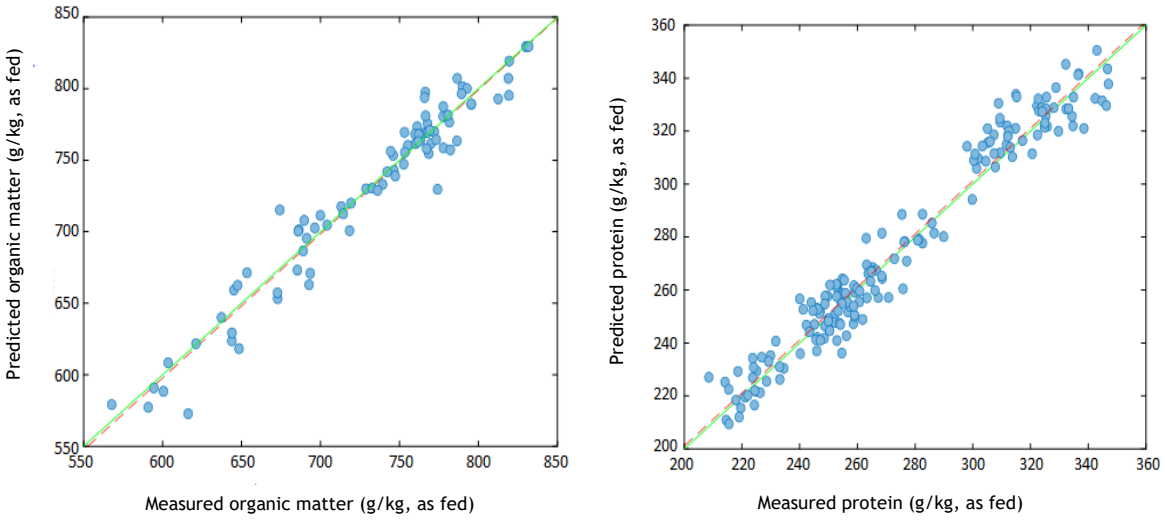
Table 2. Number of samples (*N*) and performance of the PLS models (SECV, R^2_{CV} and RPDcv) for the specific calibrations: broiler chickens dataset and laying hens dataset.

Parameter	Broiler chickens set				Laying hens set			
	N	SECV	R²cv	RPDcv	N	SECV	R²cv	RPDcv
Organic matter (g/kg, as fed)	77	7.7	0.74	1.9	167	20	0.87	2.7
Protein (g/kg, as fed)	529	10	0.95	4.8	176	10	0.87	2.8
Fat (g/kg, as fed)	154	4.1	0.96	4.9	50	5.0	0.34	1.2
Gross energy (kcal/kg)	440	38	0.90	3.1	195	80	0.73	1.9
Uric acid (g/kg, as fed)	228	4.7	0.86	2.0	175	7.0	0.74	1.9
Phosphorus (g/kg, as fed)	178	0.8	0.93	3.8	133	1.0	0.76	1.9

Discussion

Validation and quality of the global calibrations for poultry excreta

The global calibrations for poultry excreta were validated with an external sample set obtaining good results (Table 1 and Figure 4). The exception is the model for fat that will require more samples to improve the model and obtain a SEP similar to the SEC. Considering the implementation of the calibrations for routine analyses, it can be seen that the prediction errors (Table 1) are close to those of the reference methods given in the materials and methods section.



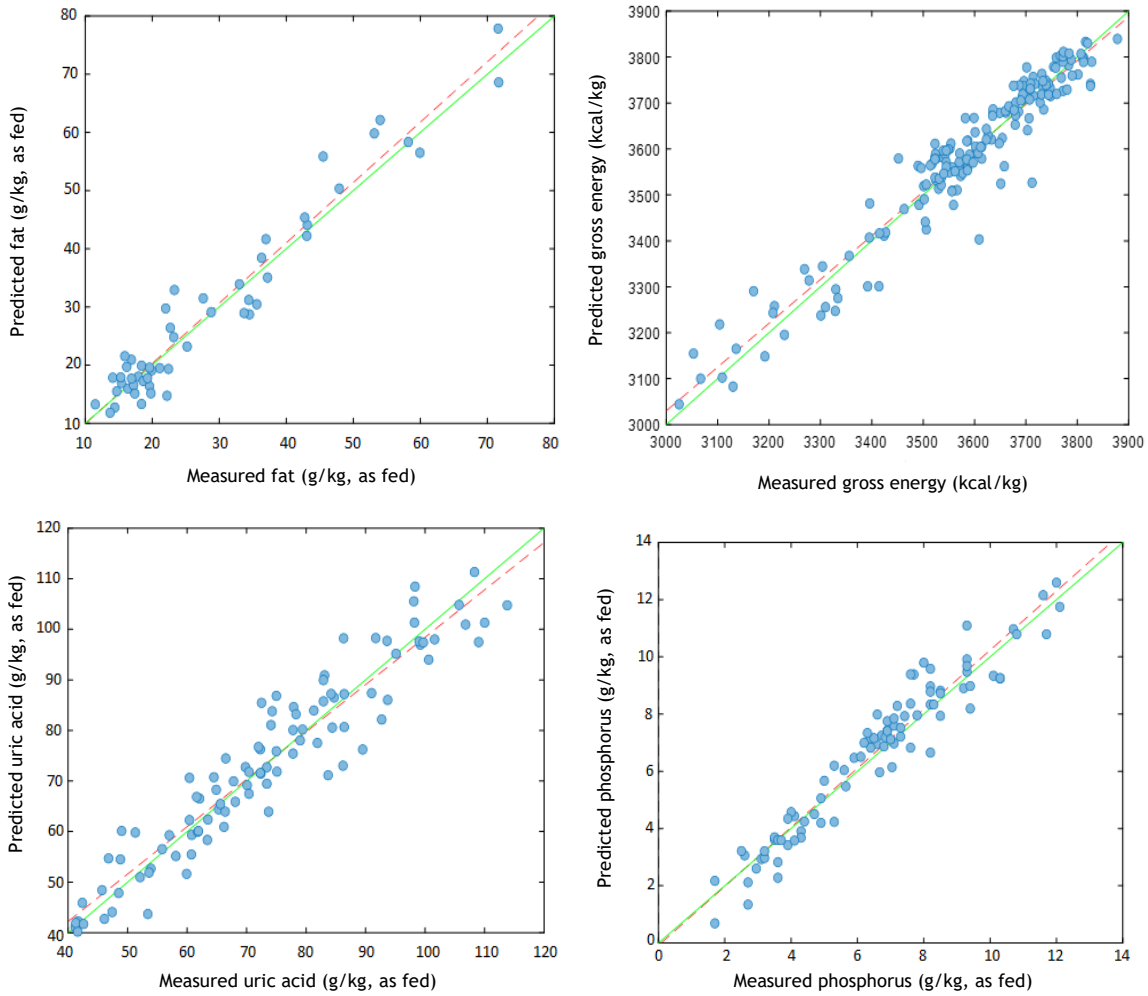


Figure 4. Predicted vs measured values of organic matter, protein, fat, gross energy, uric acid and phosphorus obtained with the poultry (broiler chickens + laying hens + turkeys) data set. The green continuous line is the 1:1 line and the dashed red line is the fitted straight line.

The calibrations were compared with other calibrations for excreta found in the literature. Bastianelli et al. 2010 [4] predicted several parameters related with the nutritional content of excreta. Unlike in our work, their sample set only contained broiler chicken excreta. The results presented in this work were similar to Bastianelli's results for protein, slightly worse for fat and gross energy and better for uric acid. These authors did not predict organic matter and phosphorous. The reason of their better performance on fat and gross energy could be due to the fact that their dataset has a larger range of fat and energy values. Smith et al. 2001 [6] predicted nitrogen, gross energy and phosphorous among other parameters in broiler chicken excreta but their models did not achieve RPD>3 in none of these three cases. De la Roza-Delgado 2015 [7] developed a gross energy calibration for poultry although no information was given about the number of samples or the poultry species in their dataset. Their prediction errors

were similar to the results presented here but their R^2 and RPD values were worse. The performance of the organic matter calibration reported in our work is better than the single study reported so far by Xing et al. 2008 [11]. The current results for phosphorous were also better than those that had been presented previously in any type of poultry excreta or manure and also better than those reported for raw materials [21] or feedstuffs [22, 23]. The explanation for these significant results could be that phosphorus in excreta is mainly in form of phytate phosphorus. An important fraction of the phosphorus that raw materials such as corn or wheat contain is in this form. Due to the poor availability of phytate in poultry intestine, the percentage of phytate phosphorus digested by the animal is small. Thus, the fraction of phosphorus, which will be more concentrate in the excreta, should be easier to predict by NIRS since phytate is an organic molecule with C-H and O-H bonds that reflect in the near infrared region.

Global calibrations versus specific calibrations

The models developed for protein, fat, gross energy, and phosphorous with the broiler chickens dataset would be acceptable (Table 2). However, the predictions of organic matter and uric acid are not good enough. In the first case, the small number of samples compared with the other parameters and the narrow range of values (Figure 1) could cause the low regression coefficient. In the second, the inaccuracy in the reference method, which is quite similar in value to the standard deviation of the values in the sample set.

The poor results for the laying hen excreta could be due to the fact that the laying hens sample set contains fewer samples and also that they belong to trials were the diets tested were very different. Nevertheless, this variability does not translate into a great standard deviation in the reference values. Moreover, the high ash content in these samples is thought that it could affect the predictions. However this higher variability in the ash content makes that the calibration for organic matter exhibited a regression coefficient higher than the one of the specific calibration for broiler chickens because the range of values is wider.

The results suggest that the global calibrations developed with the complete poultry data set, besides being the only valid ones for organic matter and uric acid, are preferable over the specific calibrations for each poultry species. On one side, global calibration models are based on a larger number of samples and include both the spectral variability and the parameter variability found in different animals and diets, and are expected to be applicable to a wider variety of future

samples than the particular models. On the other hand, using only a few global calibration models will require less effort for monitoring and updating the models than using several species-specific calibration models.

Conclusions

We have shown a new series of Vis-NIR calibration models that can replace the slow laboratory determinations of organic matter, protein, fat, gross energy, uric acid and phosphorus in poultry excreta. This will help to reduce the costs of digestibility studies and improve the discovery of optimal feedstuff diets. It was also found that, despite the fact that broiler chickens, laying hens and turkeys are different species, the spectra of their excreta could be modelled together. Joining the three datasets together increased the range of each parameter that had to be predicted and lead to models that are much widely applicable than those that are specific for each species.

Acknowledgments

A. Cruz acknowledges Universitat Rovira i Virgili (Tarragona, Spain) and the Institute of Agrifood Research and Technology (Constantí, Spain) for providing a Marti Franqués Research Fellowship (2019PMF-PIPF-62). The authors thank the various IRTA researchers that made the excreta samples available for this study, and the laboratory personnel for providing the reference values of the parameters studied in this work.

References

[1] A. Bourdillon, B. Carre, B. Leclercq, M. Lessire, L. Cannon, M. Francesch, M. Fuentes, J. Wiseman, G. Huyghebaert, W.M.M.A. Janssen, M. Rigoni, European reference method of in vivo determination of metabolisable energy in poultry: Reproducibility, effect of age, comparison with predicted values, Br. Poult. Sci. 31 (1990) 567–576.

[2] D. Bastianelli, N. Muley, B. Carré, L. Bonnal, F. Davrieux, Use of near infrared spectroscopy for evaluation of nutrient digestibility for genetic experiments in poultry, in: Proceedings of 11th International Conference on Near Infrared Spectroscopy, NIR Publications, Córdoba, Spain, 2004: pp. 735–738.

[3] D. Bastianelli, B. Carré, S. Mignon-Grasteau, L. Bonnal, F. Davrieux, Direct prediction of energy digestibility from poultry faeces using near infrared spectroscopy, in: Proceedings of 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 626–629.

[4] D. Bastianelli, L. Bonnal, H. Juin, S. Mignon-Grasteau, F. Davrieux, B. Carré, Prediction of the chemical composition of poultry excreta by near infrared spectroscopy, J. Near Infrared Spectrosc. 18 (2010) 69–77.

[5] D. Bastianelli, E. Fermet-Quinet, F. Davrieux, D. Friot, C. Hervouet, Use of near infrared spectroscopy for the prediction of quality of poultry feeds in East Africa, in: Proceedings of 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 623–625.

[6] T.N. Smith, G.M. Pesti, R.I. Bakalli, J. Kilburn, H.M. Edwards, The use of near-infrared reflectance spectroscopy to predict the moisture, nitrogen, calcium, total phosphorus, gross energy, and phytate phosphorus contents of broiler excreta, Poult. Sci. 80 (2001) 314–319.

[7] B. De La Roza-Delgado, S. Modroño, F. Vicente, A. Martínez-Fernández, A. Soldado, Suitability of faecal near-infrared reflectance spectroscopy (NIRS) predictions for estimating gross calorific value, Spanish J. Agric. Res. 13 (2015) 1–7.

[8] G.E. Aiken, D.H. Pote, S.F. Tabler, T.C. Tabler, Application of near-infrared reflectance spectroscopy to estimate chemical constituents in broiler litter, Commun. Soil Sci. Plant Anal. 36 (2005) 2529–2539.

[9] J.B. Reeves, Near-infrared diffuse reflectance spectroscopy for the analysis of poultry manures, J. Agric. Food Chem. 49 (2001) 2193–2197.

[10] W. Ye, J.C. Lorimor, C. Hurburgh, H. Zhang, J. Hattey, Application of near-infrared reflectance spectroscopy for determination of nutrient contents in liquid and solid manures, Trans. ASAE. 48 (2005) 1911–1918.

[11] L. Xing, L.J. Chen, L.J. Han, Rapid analysis of layer manure using near-infrared reflectance spectroscopy, Poult. Sci. 87 (2008) 1281–1286.

[12] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[13] DIN 51900. Testing of solid and liquid fuels-Determination of gross calorific value by the bomb calorimeter and calculation of net calorific value-Part 3. Method using adiabatic Jacket Berf., Ger, 2005.

[14] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[15] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[16] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[17] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[18] B.M. Zorzetti, J.M. Shaver, J.J. Harynuk, Estimation of the age of a weathered mixture of volatile organic compounds, Anal. Chim. Acta. 694 (2011) 31–37.

[19] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics. 11 (1969) 137–148.

[20] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[21] R. Aureli, Q. Ueberschlag, F. Klein, C. Noël, P. Guggenbuhl, Use of near infrared reflectance spectroscopy to predict phytate phosphorus, total phosphorus, and crude protein of common poultry feed ingredients, Poult. Sci. 96 (2017) 160–168.

[22] E. Swart, T. Brand, J. Engelbrecht, The use of near infrared spectroscopy (NIRS) to predict the chemical composition of feed samples used in ostrich total mixed rations, S. Afr. J. Anim. Sci. 42 (2012) 550–554.

[23] A.B.M. Khaleduzzaman, M.A.A. Mamun, H.M. Salim, Development of local calibrations for the nutritional evaluation of commercial poultry diets by using near infrared reflectance spectroscopy, J. Appl. Anim. Res. 45 (2017) 8–14.

3.5. Prediction of pig faeces nutrient content by visible-near infrared spectroscopy

Abstract

Nowadays there is a growing interest in optimizing pig feed formulations based on the available nutrient content (that is, how these nutrients are digested and metabolized by the animal). To determine the digestibility coefficients of nutrients, in-vivo assays are conducted where pigs are fed with different diets of known composition and the faeces are analysed, traditionally, with wet analytical methods. This study studies the development of visible-near infrared (Vis-NIR) calibrations for various parameters, namely organic matter, protein, fat, gross energy, crude fibre, neutral detergent fibre, acid detergent fibre, and phosphorus . The samples were collected in in-vivo assays conducted between 2017 and 2020. Vis-NIR spectra were measured between 400 nm and 2499.5 nm and were preprocessed using generalized least squares weighting (GLSW). Partial least squares regression (PLSR) was used to establish the prediction models. Protein, gross energy, crude fibre, acid detergent fibre and P were properly predicted with a coefficient of determination of prediction (R_P^2) ranging from 0.88 to 0.94. These five parameters passed the statistical test for fit for purpose described in ISO 12099:2017 for bias and slope, but did not pass the F-test that determines whether the root mean square error of prediction (RMSEP) is close enough to the root mean square error of calibration (RMSEC). This indicates that the models could not be considered robust enough and that they require further improvement with new representative samples. Organic matter and lignin (R_P^2=0.90) failed also the t-test for slope and the t-test for bias, respectively. Fat and neutral detergent fibre showed worse predictive statistics (R_P^2=0.84) and did not pass any statistical test. These results represent a significant step forward in harnessing Vis-NIRS to assess a wide range of nutritional parameters in pig faeces.

Introduction

Feed accounts for the largest share of swine production costs and is a key factor in animal growth and health [1]. While traditionally feed formulation was based on total nutrient content, nowadays formulations are designed for available content, which takes into account how

nutrients are digested and metabolized by the animal. There are several indicators of feed digestion. The most important in studies with pigs are the total tract digestibility coefficients for the main nutrients (protein, fat or fibre) and digestible energy (DE). Other important parameters are the intake of calcium and phosphorus, since these are needed for the correct formation and maintenance of the skeleton. Digestibility coefficients of nutrients and DE are usually determined by in-vivo assays in which pigs are fed different diets of known composition [2]. These trials require the collection and analysis of many replicate faecal samples per feed treatment, while attempting to minimize external influences such as temperature and humidity on the farm. Near-infrared spectroscopy (NIRS) is a rapid analytical technique that can replace costly wet analytical methods for determining the nutrient content of faeces. Its main features are speed, sample preparation is not necessary (except milling), solvents are not required and, therefore, there is no waste. It is also a multiparametric technique since different parameters can be predicted simultaneously from one spectrum. Manufacturers of NIR instruments offer commercial "universal calibrations" that are expected to be suitable for routine analysis of the most common raw materials and compound feeds. Nevertheless, there are currently no commercial calibration models for swine faeces, and only a few calibrations for the nutrient content have been reported in the literature [3–5]. Other authors, focusing on the use of manure as fertilizer, have reported calibrations for, among others, organic matter, nitrogen (total, organic, and ammoniacal), and phosphorus [6].

In this work, we developed calibration models for vis-NIR spectra of pig faeces to predict organic matter (OM), crude protein (CP), fat, gross energy (GE), crude fibre (CF), acid detergent fibre (ADF), neutral detergent fibre (NDF), lignin and phosphorus (P). The calibrations have been tested with an external dataset and their validity was evaluated by means of statistical tests and by comparison with the aforementioned bibliography.

Materials and methods

Samples

924 samples of pig faeces were collected from 2017 to 2021 at the Institute of Agri-Food Research and Technology (IRTA) in Constantí, Tarragona, Spain. The samples were from 18 in-vivo assays with pigs aimed at measuring the digestibility of feeds containing different

combinations of raw materials and additives (e.g., enzymes). The feeds were based on soybean meal and the main cereal was corn, wheat or barley. Titanium dioxide (TiO_2) was used as an indigestible marker in all studies. Faecal samples were freeze-dried, ground, and stored in sealed bags in a climate chamber at 17 ºC until analysis. The fact that the faecal samples were from animals with different digestive capacities and different diets resulted in a wide range of undigested contents of the different dietary fractions that had to be accounted for by the calibration models.

Reference values of nutritional parameters

The faeces samples were analysed in the laboratory with validated methods. DM, CP, fat, ash, CF and P were determined according to the AOAC (2016) methods 925.09, 968.06, 920.39, 942.05, 978.10 and 965.17 respectively [7]. GE was determined by calorimetry using an adiabatic calorimeter (C2000, IKA, Staufen, Germany) according to the DIN 51900 (2005) norm [8]. NDF, ADF and lignin were determined sequentially by gravimetry using detergents following the Van Soest method [9]. OM was calculated as the difference between DM and ash contents. Because the digestibility trials spanned different years and had different objectives, the number of faecal samples that could be used for modelling was not the same for all analytical parameters.

Visible-near infrared spectra acquisition-and data analysis

Freeze-dried faecal samples were scanned using a NIRS DS2500 (Foss NIRSystems, Denmark) in reflectance mode with a 7-cm-diameter cup into which approximately 30 grams were introduced each time. Spectra were acquired every 0.5 nm from 400 to 2499.5 nm, covering the range from visible to NIR. PLS Toolbox software (PLS_Toolbox, 2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab (MATLAB, version R2020a, The MathWorks Inc., Natick, MA, USA) was used to perform all chemometric treatments.

Partial least squares regression (PLSR) was used to develop calibration models for OM, CP, fat, GE, CF, ADF, NDF and P. Common spectral pretreatments were studied including normalization, standard normal variate (SNV) [10], multiplicative scatter correction (MSC) [11] 1st and 2nd derivatives [12] orthogonal signal correction (OSC) [13] and generalized least squares weighting (GLSW) [14]. A five-fold venetian blind cross validation (CV) was used to choose the optimal pretreatments and the optimal number of latent variables for each model. The Duplex

algorithm was applied to the spectra to divide the data set into a calibration set (75% of the samples) and validation set (25% of the samples) [15].

Calibration samples were flagged as spectral outliers and thus discarded if the leverage was too high (Hotelling's T^2 reduced>3) or the percentage of the sample's residual spectral variance was too high (Q residuals reduced>3). Samples were classified as reference outliers if the difference between the predicted and reference values was too large (Studentized Residuals (t) > 3).

The validity of the predictions from the calibration models was checked with three statistical tests: a t-test for bias, a t-test for slope, and an F-test for the standard error of prediction (SEP) according to ISO 12099:2017, whose aim is to guide the development and maintenance of NIR calibrations in the agri-food sector [16].

Results and discussion

Table 1 summarizes the results of the reference analyses. The faeces had little variability in OM and GE (relative standard deviation (RSD) = 5 and 7%, respectively). The highest variability was that of P, followed by CF and ADF. Most samples in which P was determined were from in-vivo assays in which the pigs were fed diets containing varying amounts of phytase, which have been shown to affect the digestibility of P. Similarly, dietary fibre parameters were obtained from in-vivo assays that studied carbohydrases, which have been shown to improve dietary fibre digestibility.

Table 1. Statistical overview of reference analysis. Number of samples (N), mean value for the dataset Mean (g/kg, as fed), standard deviation (SD) (g/kg, as fed), relative standard deviation (RSD) (%), minimum (Min) (g/kg, as fed) and maximum (Max) (g/kg, as fed), standard error of the laboratory (SEL) (g/kg, as fed) and relative standard error of the laboratory (RSEL) (%).

Property	N	Mean	SD	RSD	Min	Max	SEL	RSEL
Protein	845	232	32	14	131	321	5	2
Fat	309	72	16	22	31	116	3	4
Gross energy*	704	4208	277	7	3616	5228	40	1
Crude fibre	290	135	35	26	67	214	7	5
ADF	266	216	53	25	100	320	10	5
NDF	266	381	69	18	226	593	10	3
Organic matter	716	749	35	5	644	827	9	1
Lignin	208	89	28	31	42	151	8	9
Phosphorus	421	17	6	35	6	34	1	6

*The unit is kcal/kg, as fed instead of g/kg, as fed.

The relative standard error of the laboratory (RSEL) was low for OM, GE and CP. The remaining analyses were more complex (involving extractions and/or digestions) and had higher errors. The highest error was expected for lignin, as it is the last product determined in the Van Soest method [9]. Figure 1 shows the mean Vis-NIR spectrum of the pig faeces samples.

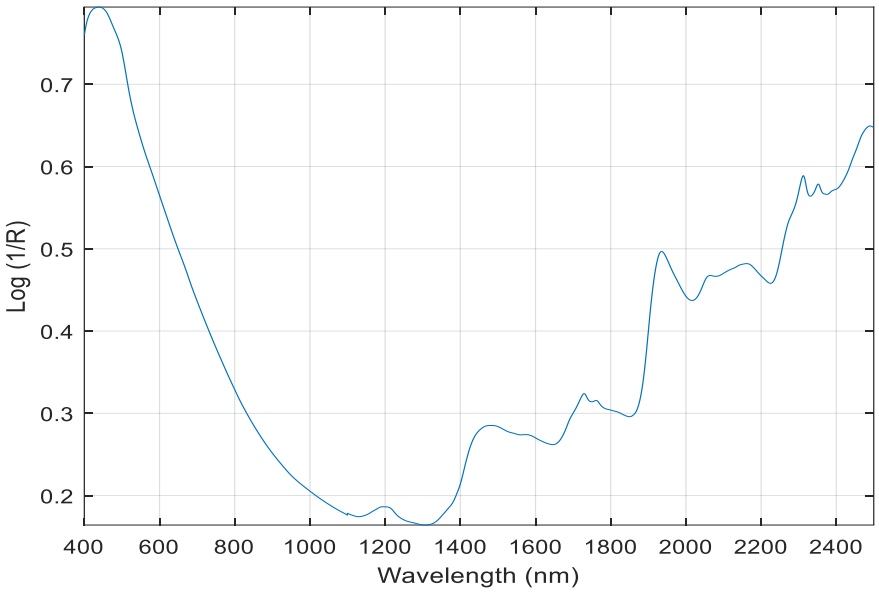


Figure 1. Mean Vis-NIR spectrum of the pig faeces samples.

The assignment of spectral bands to the constituents of complex samples such as the faecal samples is not straightforward. Nevertheless, OM is expected to be related to the bands corresponding to C-H, N-H and O-H bonds. CF, NDF and ADF are carbohydrates that, along with lignin, are poorly digested by the animal and are related to the C-H and O-H bands. The most important bands related to fat are the C-H bands, whose first overtone appears at 1700 nm, the second at 1200 nm, and the combination bands at 2300 nm. GE content is correlated with fat content or starch hence the C-H bands may be informative of this parameter. CP is correlated with C-H and N-H bands. The first overtone at 1900 nm and the combination band at 2100 nm overlap with the O-H bands. The second overtone at 1500 nm would be easier to assign. P does not absorb at vis-NIR and may be determined by its association with organic components such as the phytate molecule.

Of the various preprocessing methods tested, autoscaling followed by GLSW gave the best results for CP, fat, GE, CF, ADF, NDF, and OM. For lignin, the best results were obtained with SNV followed by Savitzky-Golay first derivative, interpolating with a second order

polynomial and a window width of 15 points, and mean centring. The model for P required Savitzky-Golay first derivative, interpolating with a second order polynomial and a window width of 17 points. Spectral outliers were not found in the data sets of NDF, ADF, and lignin. For the remaining constituents, from two to five samples were clearly identified as spectral outliers and removed from subsequent analyses due to their high values of Hotelling's T^2 and Q residuals. The optimal number of latent variables of the PLS models ranged from 3 in the model for fat to 14 in the model for P. After the models were calculated, some samples were identified as unique reference outliers and removed from further modelling. Concretely, four samples for CP and OM, three for fat, GE, and P, and two samples for CF, ADF, NDF and lignin. Removing the visible region of the spectra did not improve the calibration models. Therefore, the entire spectral range was used in all models.

Table 2 shows the performance of the predictive models for pig faeces. CP, GE, CF, ADF, and P were predicted well. The coefficient of determination of prediction (R^2_{p}) ranged from 0.88 to 0.94 and the RMSEP values were within two times the value of the standard error of the laboratory (SEL), which is commonly used as a rule of thumb for the acceptability of NIR calibration models [17]. The t-test for bias showed that none of the calibrations had significant bias, and the t-test for slope was easily passed by all calibrations. However, the results of the F-test for the root mean square error of prediction (RMSEP) suggested that these calibrations required further improvement by including future representative samples. The calibration models can be considered stable if the RMSEP values are close to the corresponding RMSEC values [18].

Table 2. Statistics of the calibration models for pig faeces: number of calibration samples (N_c) and prediction (N_v), coefficient of determination of calibration (R^2_{c}) and prediction (R^2_{p}), root mean square error of calibration RMSEC (g/kg, as fed) and prediction RMSEP (g/kg, as fed), bias in the prediction, and slope of the predicted versus measured regression line.

Property	N_c	N_v	R^2_{c}	RMSEC	R^2_{p}	RMSEP	Bias	Slope
Protein	584	251	0.96	6.11	0.94	7.93	-0.14	0.97
Fat	213	92	0.91	4.98	0.84	6.67	-0.13	0.83
Gross energy*	486	210	0.96	53	0.94	74	-1.0	0.97
Crude fibre	200	86	0.94	8.31	0.88	12.39	-0.29	1.00
ADF	184	79	0.96	10.17	0.89	16.52	-0.41	0.98
NDF	184	78	0.91	20.44	0.84	25.93	-0.78	0.90
Organic matter	499	214	0.90	10.82	0.88	12.41	-0.19	0.91
Lignin	145	60	0.87	10.17	0.88	9.99	3.02	0.93
Phosphorus	246	83	0.95	1.45	0.94	1.54	-0.05	0.96

*The unit is kcal/kg, as fed instead of g/kg, as fed.

The measured and predicted values of these five parameters (CP, GE, CF, ADF and P) are shown in Figure 2. OM and lignin had R^2_p=0.90 and low RMSEP for these determinations compared to SEL. However, they did not pass the t-test for slope, and lignin neither passed the t-test for bias. Fat and NDF showed worse predictive statistics (R^2_p=0.84 and RMSEP > 2×SEL) and did not pass any statistical test, so these two models cannot be used with the dataset available.

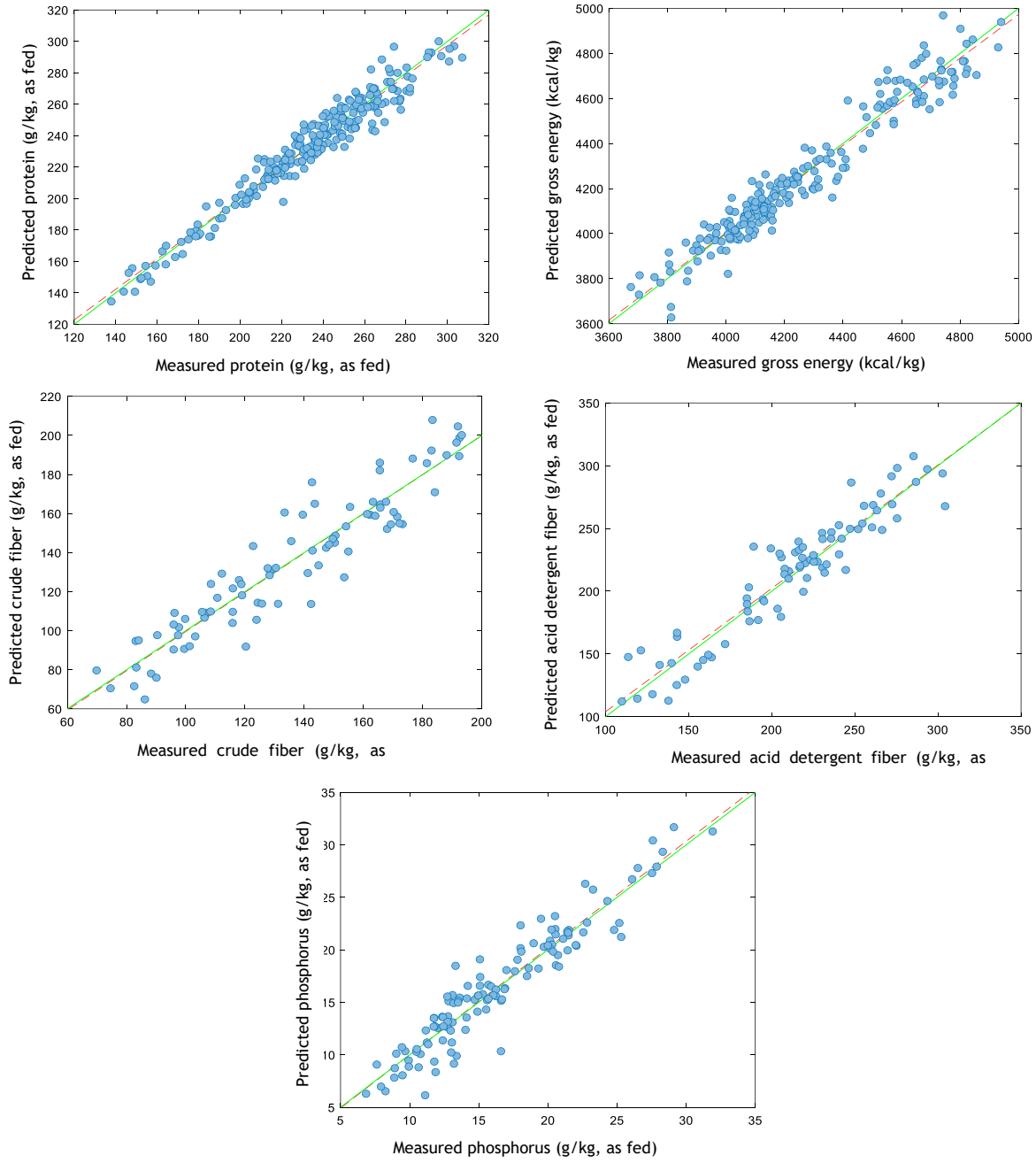


Figure 2. Predicted vs measured values of crude protein, gross energy, crude fibre, acid detergent fibre and phosphorus. The green continuous line is the 1:1 line and the dashed red line is the fitted straight line.

Calibrations were compared to other reported calibration results for pig faeces. The performance of our calibrations for OM, CP, fat, and GE was better than the performance reported by Schiborra et al. 2015 [3] and by Nirea et al. 2018 [4] and similar to that of Paternostre et al. 2021 [5]. GE was also predicted with similar success by De la Roza et al. 2015 [19]. For fibre fractions, CF was predicted with similar accuracy to Schiborra et al. 2015 and slightly better than Paternostre et al. 2021. Predictions for NDF and ADF were of the same order of magnitude as those of Schiborra et al. 2015. Prediction of lignin in pig faeces are reported for the first time in this work. Predictions for P have appeared in some reports where pig faeces were studied as manure [20–22]. Our results were much better than those in the literature, probably because our samples were faeces from in-vivo assays where P digestibility was the target and the P content had a much larger variability than the variability indicated in the literature.

Conclusions

We have studied the potential of using Vis-NIR spectroscopy to determine several nutritional parameters in pig faeces, namely organic matter, crude protein, fat, gross energy, crude fibre, acid detergent fibre, neutral detergent fibre, lignin and phosphorus (P). Although the coefficients of determination and the prediction errors were satisfactory for all the parameters except fat and neutral detergent fibre, all the calibrations developed failed at least one of the three statistical test for fit for purpose described in ISO 12099:2017. This indicates that the models, although promising, still need to be improved to replace the laboratory determinations of these parameters.

References

[1] J.F. Patience, M.C. Rossoni-Serão, N.A. Gutiérrez, A review of feed efficiency in swine: Biology and application, J. Anim. Sci. Biotechnol. 6 (2015) 1–9.

[2] J. Noblet, S.B. Wu, M. Choct, Methodologies for energy evaluation of pig and poultry feeds: A review, Anim. Nutr. 8 (2022) 185–203.

[3] A. Schiborra, M. Bulang, A. Berk, A. Susenbeth, E. Schlecht, Using faecal near-infrared spectroscopy (FNIRS) to estimate nutrient digestibility and chemical composition of diets and faeces of growing pigs, Anim. Feed Sci. Technol. 210 (2015) 234–242.

[4] K.G. Nirea, M.P. de Nanclares, A. Skugor, N.K. Afseth, T.H.E. Meuwissen, J. Hansen, L.T. Mydland, M. Øverland, Assessment of fecal near-infrared spectroscopy to predict feces chemical composition and apparent total-tract digestibility of nutrients in pigs, J. Anim. Sci. 96 (2018) 2826–2837.

[5] L. Paternostre, V. Baeten, B. Ampe, S. Millet, J. De Boever, The usefulness of NIRS calibrations based on feed and feces spectra to predict nutrient content, digestibility and net energy of pig feeds, Anim. Feed Sci. Technol. 281 (2021) 115091.

[6] L. Chen, L. Xing, L. Han, Review of the Application of Near-Infrared Spectroscopy Technology to Determine the Chemical Composition of Animal Manure, J. Environ. Qual. 42 (2013) 1015–1028.

[7] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[8] DIN 51900. Testing of solid and liquid fuels-Determination of gross calorific value by the bomb calorimeter and calculation of net calorific value-Part 3. Method using adiabatic Jacket Berf., Ger, 2005.

[9] P.J. Van Soest, Use of Detergents in the Analysis of Fibrous Feeds. II. A Rapid Method for the Determination of Fiber and Lignin, J. Assoc. Off. Agric. Chem. 46 (1963) 829–835.

[10] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[11] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[12] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[13] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[14] B.M. Zorzetti, J.M. Shaver, J.J. Harynuk, Estimation of the age of a weathered mixture of volatile organic compounds, Anal. Chim. Acta. 694 (2011) 31–37.

[15] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428.

[16] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[17] P. Dardenne, G. Sinnaeve, V. Baeten, Multivariate calibration and chemometrics for near infrared spectroscopy: Which method?, J. Near Infrared Spectrosc. 8 (2000) 229–237.

[18] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[19] B. De La Roza-Delgado, S. Modroño, F. Vicente, A. Martínez-Fernández, A. Soldado, Suitability of faecal near-infrared reflectance spectroscopy (NIRS) predictions for estimating gross calorific value, Spanish J. Agric. Res. 13 (2015) 1–7.

[20] W. Saeys, P. Darius, H. Ramon, Potential for on-site analysis of hog manure using a visual and near infrared diode array reflectance spectrometer, J. Near Infrared Spectrosc. 12 (2004) 299–309.

[21] Z. Yang, L. Han, X. Fan, Rapidly estimating nutrient contents of fattening pig manure from floor scrapings by near infrared reflectance spectroscopy, J. Near Infrared Spectrosc. 14 (2006) 261–268.

[22] D.F. Malley, L. Yesmin, R.G. Eilers, Rapid Analysis of Hog Manure and Manure-amended Soils Using Near-infrared Spectroscopy, Soil Sci. Soc. Am. J. 66 (2002) 1677–1686.

3.6. Determination of protein and amino acid composition of poultry ileal digesta by near-infrared spectroscopy

Abstract

This work presents the use of near-infrared spectroscopy (NIRS) and multivariate calibration to determine protein content and 17 amino acids in poultry ileal digesta. The NIR spectra were registered between 850 and 2499.5 nm and pretreated by the first derivative. Partial least squares regression (PLSR) was used to obtain the prediction models. Protein, 11 out of 17 amino acids (glycine, alanine, valine, leucine, isoleucine, phenylalanine, tyrosine, aspartic acid, glutamic acid, threonine, proline) and the total of amino acids were well predicted with a coefficient of determination of prediction (R^2_p) higher than 0.80 and a coefficient of variation of prediction (CV_p) lower than 10%. For lysine and methionine, proper coefficients of determination ($R^2_p \approx 0.82$) but high prediction errors ($CV_p \approx 15\%$) were obtained. Finally, 4 out of 17 amino acids (cysteine, arginine, serine and histidine) were poorly predicted ($R^2_p < 0.70$, $CV_p > 10\%$). The best predicted amino acids were those with the highest correlation with protein. The models accounted for the covariance structure of the amino acids content in the calibration samples, arising from the nature of the ileal samples. Therefore, although valid for the existing data, the models may require updates to predict accurately new samples with a different covariance structure. The results demonstrated that NIRS is a reliable technique for assessing the protein and amino acids content in poultry ileal digesta.

Introduction

Feed is the factor that more critically affects the growth and health of poultry species. Industry and research centres are continuously investigating the formulations that provide the animal with the right amount of nutrients while avoiding any excesses that would lead to unnecessary costs and environmental problems.

In this sense, modern animal feed formulations tend to be based on amino acid (AA) content instead of protein content. One of the benefits of supplementing feed-grade AAs is that low-

protein diets can be used while continuing to cover animal feed requirements. The benefits of low-protein diets are clear from the environmental point of view. For example, nitrogen excretion is reduced and less-soybean meal is used. This crop is the most important source of protein for poultry production and has undesirable environmental impacts such as energy consumption, the emission of nitrous oxide from the fields and land-use change [1]. For the correct growth of the animals, a diet must contain an adequate AA profile. Some diets, especially those with a low-protein profile, may not fully cover the nutritional requirements of the animal, if the content of some AAs is limited. If so, they should be supplemented. In current soy-based diets used for poultry production, the limiting AAs are lysine, methionine, cysteine, and threonine [2]. Tryptophan could also be limiting when corn is the major cereal in the diet and valine when it is wheat. A further step to improve diets is to formulate them based on available AA contents rather than total AA content.

Traditionally, the parameters related to feed digestion are determined in the animal excreta obtained from in-vivo assays. However, the excreta contain urinary nitrogen and AAs that should not be taken into account for the calculation of AA digestibility. Moreover, the microbiota present in the hindgut influences the AA composition of the excreta due to the absorption or synthesis of some of them. Therefore, digestibility calculated from the ileal digesta is a better indicator of the real use of protein and AAs by poultry [3]. To obtain the values of ileal digestibility, in-vivo assays are conducted as follows: birds are fed with the test diet and, after euthanasia, the ileal digesta are collected, processed, and finally analysed, together with the diet [4]. Those analyses are currently done in the laboratory using costly reference methods, which are mainly based on chromatography.

Near-infrared spectroscopy (NIRS) is a well stablished technique for the routine determination of constituents in ingredients and diets [5–7]. Its advantages are that it is fast, reliable and non-destructive. The determination of protein in cereals was one of the first NIRS practical applications [8,9]. Shortly after, the first NIR calibrations to predict AAs appeared. Rubenthaler and Bruinsma 1978 [10] predicted lysine in a variety of cereals and Gill et al. 1979 [11] in barley of different breeds. Williams et al. 1984 [12] predicted some limiting AAs (lysine, methionine, threonine and tryptophan) in wheat and barley. Since then, several authors have published methods for most of the AAs in a variety of raw materials [13–22] and also in compound feeds [18,23,24]. However, there are very few publications about the use of NIRS to predict constituents in digestive contents. Noah et al. 1997 [25] predicted starch in pig digestive contents

and Lebzien and Paul 1997 [26] estimated the microbial portion of non-ammonia-nitrogen in the duodenum of cows. Nevertheless, there is none about poultry digestive contents, neither protein nor AA predictions.

The purpose of this study is to evaluate the potential of NIRS and multivariate partial least squares regression (PLSR) to determinate total protein content, 17 AAs including alanine (Ala), arginine (Arg), aspartic acid (Asp), cysteine (Cys), glutamic acid (Glu), glycine (Gly), histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), proline (Pro), serine (Ser), threonine (Thr), tyrosine (Tyr), and valine (Val) and the total of AAs (TAA) in the ileal digesta of poultry.

Materials and methods

Samples and in-vivo assays

A total of 179 samples of poultry ileal digesta were collected from 2018 to 2020 at the Institute of Agrifood Research and Technology (IRTA) in Constantí, Tarragona, Spain. The samples had been obtained from four in-vivo assays. Two of them aimed at measuring the digestibility of wheat-based diets containing different doses of xylanase. Both involved male Ross 308 broiler chickens. The other two consisted of measuring the digestibility of corn-based diets containing different doses of manannase. One involved male Ross 308 broiler chickens and the other female Aviagen Premium broiler turkeys. At the end of the in-vivo assays (day 24 or 25) both types of broilers were euthanized, the intestine removed and the lower half of the ileum was gently flushed to remove the ileal contents for digestibility measurements. The ileum is the portion of the small intestine extending from the Meckel's diverticulum to a point 40 mm proximal to the ileocecal junction [27]. Ileal digesta samples were freeze-dried, equilibrated at ambient temperature for 1 day, weighed, ground and stored prior to analysis.

Reference analyses

Ileal digesta samples were analysed in the laboratory using reference methods. Protein was determined by the Dumas procedure using a nitrogen/protein analyser FP-528 (LECO Corporation, United States) according to the AOAC (2016) method 925.09 [28]. AAs were

liberated from protein by hydrolysis with HCl. To determine Met and Cys, a peroxidation with performic acid was done prior to hydrolysis. Sodium metabisulfite was added to decompose performic acid. These procedures were carried out according to the AOAC (2016) method 994.12 [28]. Once the protein hydrolysates were obtained, the AAs were determined by precolumn derivatisation with ophthaldehyde (OPA) and 9-fluorenylmethyl chloroformate (FMOC) [29] using a HPLC Agilent 1100 coupled with diode array detector (DAD) (Agilent Technologies, United States). Norvaline was used as internal standard and the chromatographic peaks were integrated manually. Trp, which must be treated separately by an alkaline instead of acid hydrolysis, was not determined in this work. TAA was calculated as the sum of the 17 AAs determined in each sample.

Near-infrared (NIR) spectra acquisition and data analysis

Freeze-dried and grounded ileal digesta samples were analysed on a NIRS DS2500 (Foss NIRSystems, Denmark) in reflectance mode with a 1.8 cm diameter cup where 1 g approximately is introduced each time. NIR spectra were collected every 0.5 nm from 850 to 2499.5 nm. PLS toolbox software (PLS_Toolbox, 2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab (MATLAB, Version R2020a, The MathWorks Inc., Natick, MA, USA) was used to carry out the chemometric treatments.

Partial least squares regression (PLSR) was used to develop the calibration models for protein and AAs. Common spectral pretreatments were studied including normalization, standard normal variate (SNV) [30], multiplicative scatter correction (MSC) [31] and derivatives (1st and 2nd) [32] using different window widths. Finally, the first derivative was applied to the raw spectra interpolating with a second-order polynomial and an optimized window width for each calibration model that varied from 9 to 31 points. A five-fold Venetian blind cross validation (CV) was used to choose the optimal pretreatments and the optimal number of latent variables for each model. Duplex algorithm was applied to the spectra set to divided into calibration set (75% of the samples) and validation set (25% of the samples) allowing to retain the spectral distribution [33]. Samples were identified as spectral outliers and therefore discarded when they presented large leverage (Hotelling's T^2 reduced > 3) or a large percentage of residual spectral variance (Q residuals reduced > 3). Samples were labelled as reference outliers when the difference between the predicted and the reference value was too high (Studentized residuals (t) > 3).

Results

Reference analyses

Figure 1A shows the chromatogram obtained after acid hydrolysis of a randomly selected sample. All the peaks were well resolved. Only alanine exhibited a small signal interference. To determine Met and Cys, sample peroxidation and acid hydrolysis were done. The chromatogram of the mentioned sample is shown in Figure 1B. The Met and Cys peaks were not as well resolved as the AAs shown in Figure 1A. Like in this research, Fontaine et al. 2001 [14] found baseline interferences for Met and Cys peaks. In both chromatographic analyses, the norvaline peak was used as internal standard (Figure 1).

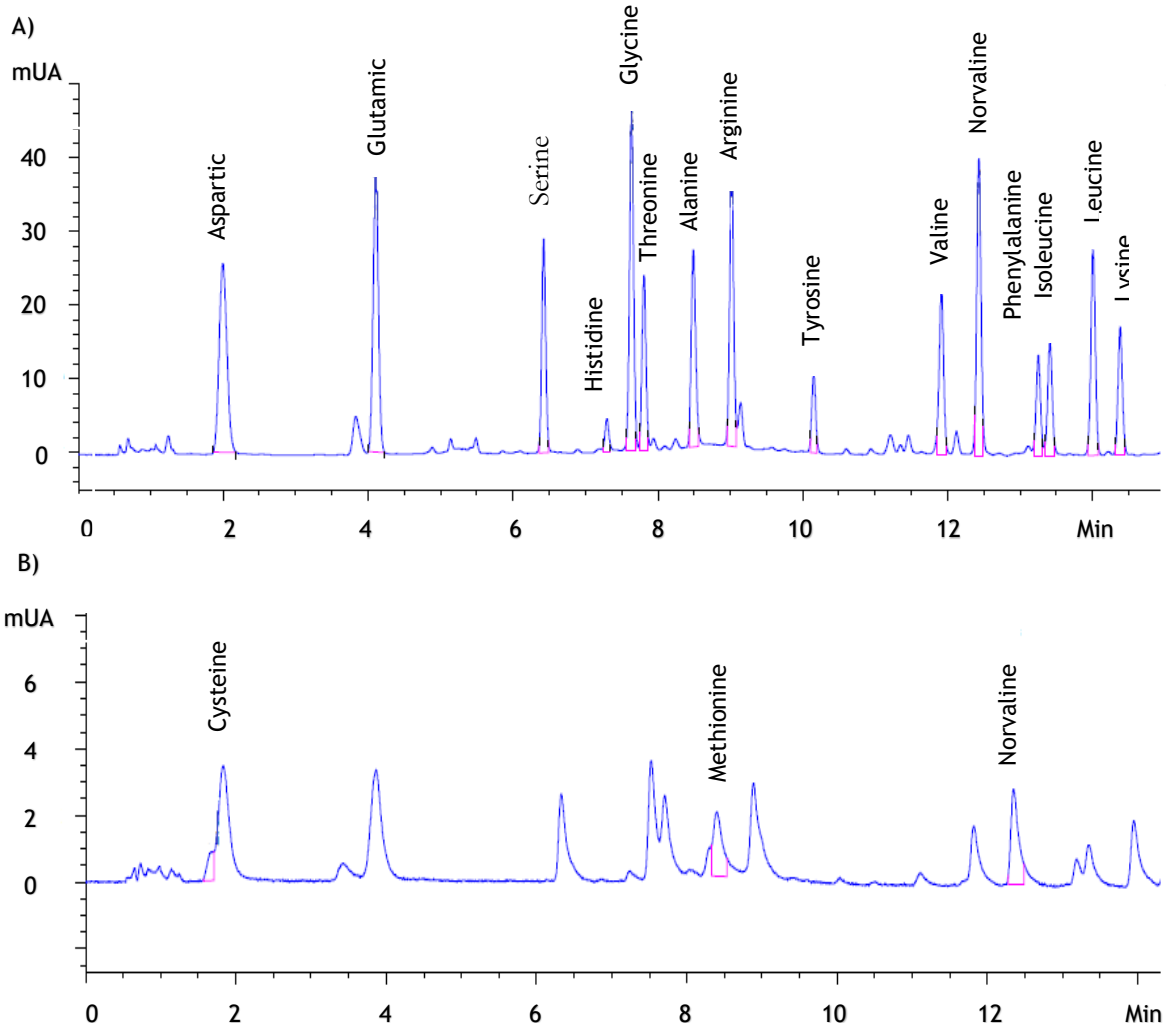


Figure 1. Chromatograms obtained for one of the analysed samples. A) Sample treated with only acid hydrolysis. B) Sample treated with peroxidation and after that acid hydrolysis.

The standard error of the laboratory (SEL (%)) obtained by dividing the standard error by the mean value of the analyte was 3% for protein and less than 5% for the AAs shown in Figure 1A. The baseline interferences for Met and Cys peaks (Figure 1B) together with the peroxidation extra step, increased the SEL for these two determinations.

Table 1 shows several statistics related to the protein and AA content in the ileal digesta. The number of samples was slightly lower for the determination of Met and Cys because in some cases there were not enough amount of sample to perform both chromatographic procedures. In those cases, only the determination with acid hydrolysis was made. Asp and Glu were the most concentrated AAs in the ileal digesta samples (13.1 and 14.8 g/kg, as fed respectively), while His, Met and Cys were the least concentrated (2.3, 1.8 and 2.6 g/kg, as fed respectively). The others ranged from 3 to 8 g/kg, as fed. The variability of the AA content, evaluated by the relative standard deviation (RSD), depends on the type of AA considered due to the different ingredients and enzymes used to make the nutrients more digestible in some diets. It was especially high for Met and Lys (44% and 33%) which are limiting AAs typically supplemented in the diet. For the rest, the RSD ranged from 10% to 20%.

Table 1. Statistical overview of chemical analysis. Number of samples (N), mean value for the dataset (Mean), standard deviation (SD), relative standard deviation (RSD), minimum (Min) and maximum (Max) values.

	N	Mean (g/kg, as fed)	SD (g/kg, as fed)	RSD (%)	Min (g/kg, as fed)	Max (g/kg, as fed)
Protein	179	131	17	13	87	203
Aspartic acid	179	13.1	1.6	12	9.2	19.4
Glutamic acid	179	14.8	2.6	18	9.5	21.6
Serine	179	6.4	1.1	17	3.6	10.2
Glycine	179	6.3	0.7	11	4.5	9.0
Threonine	179	5.2	0.7	13	3.7	8.1
Arginine	179	7.5	1.6	21	3.6	11.8
Alanine	179	5.8	1.0	17	4.0	9.9
Tirosine	179	3.4	0.5	15	2.2	5.6
Cysteine	157	2.6	0.6	23	1.2	4.3
Valine	179	4.6	0.8	17	2.9	7.6
Methionine	157	1.8	0.8	44	0.7	5.8
Phenylalanine	179	4.2	0.7	17	2.5	7.2
Isoleucine	179	3.8	0.6	16	2.5	6.9
Leucine	179	7.3	1.2	16	4.8	12.1
Lysine	179	5.1	1.7	33	1.9	9.6
Proline	179	6.5	1.5	23	3.1	10.8
Histidine	179	2.3	0.4	17	1.2	3.6
Total AA	179	100	14	14	67	149

Near-infrared spectroscopy and multivariate calibration

Figure 2 shows the raw NIR spectra of the ileal digesta samples (Figure 2A) and the NIR spectra after applying the first derivative and mean centring (Figure 2B). The NIR spectra of this kind of samples are complex. In addition to protein and AAs, other constituents such as fat, sugars or water have bonds that absorb in the NIR region. Protein is correlated with C-H and N-H bands. AAs are also associated with the O-H bands due to the carboxyl group (-COOH) and the -OH group in some of them (Ser, Thr or Phe). The first overtone of the C-H bond appears around 1700 nm, the second at 1200 nm and the combination bands from 2300 to 2500 nm although the vibration frequencies can vary a little depending on the type of C (primary, secondary, tertiary, aromatic…) that is linked to H. Because of that, it is possible to observe different peaks overlapped at the mentioned wavelengths. For N-H, the most intense band corresponding to the first overtone is placed at 1900 nm but it coincides with the water band. The second overtone at 1550 nm or the combination bands around 2150 nm are usually related to molecules with N-H bonds. For O-H, in addition to the first overtone (1900 nm) it is possible to identify the second around 1400 nm and the combination bands around 2050 nm. The S-H bond that is present in Met and Cys also absorbs in the NIR region, but the absorptions are weak and only the first overtone that appears typically at 1750 nm and probably overlapped with the C-H bond (1700 nm) can be identified.

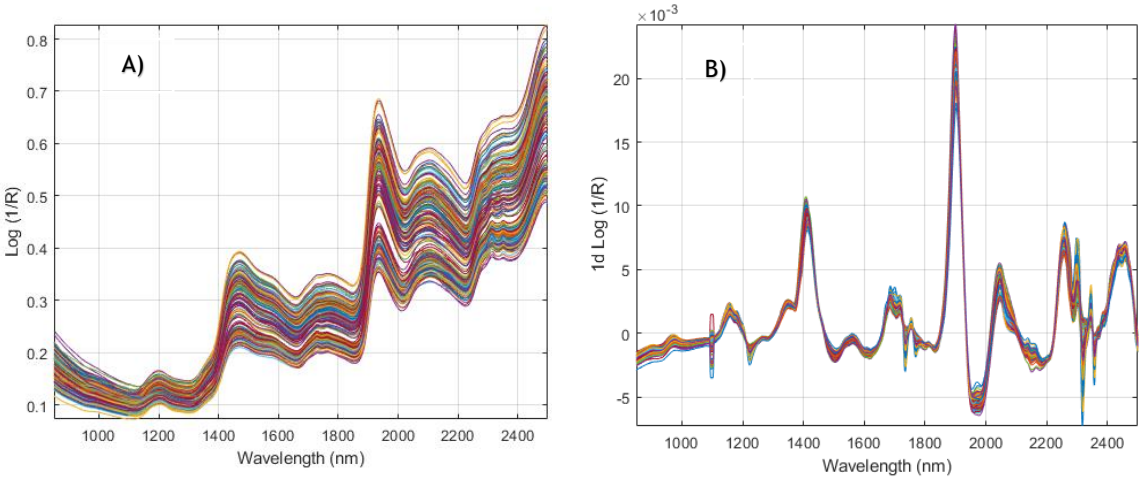


Figure 2. NIR spectra of the ileal digesta samples: A) raw and B) after first derivative.

PLS calibration models were developed with 127 calibration samples for all constituents except for Met and Cys that were developed with 113 samples. A test set of 52 samples (44 samples for Met and Cys) was used to validate the models. During the model optimization process, two

types of outliers were identified. A sample was identified as a spectral outlier for all the constituents except for Lys, Cys and Pro due to its high residual spectral variance (Q_{res}>3). For protein and some AAs, one and sometimes two reference outliers were also found and removed.

Table 2 shows the main PLS performance statistics. The optimal number of latent variables ranged from 8 to 12 for protein and all the AAs except Lys (4), Cys (5) and Ser (6). R^2 is an indicator of the quality for the predictions. According to Shenk and Westerhaus 1995 [34] the fit of a PLS calibration may be considered as: excellent if R^2>0.90, good if R^2=0.70-0.90 and insufficient if R^2<0.70. It can be seen that the coefficient of determination of calibration (R_c^2), cross validation (R_{cv}^2) and prediction (R_p^2) were higher than 0.90 or than 0.70 for all the constituents except Arg, Ser and His (R_{cv}^2 and R_p^2 lower than 0.70). This means that, those last three PLS models could not explain the variance of the prediction samples as well as the variance of the calibration samples. Finally, Cys had just the R_c^2 lower than 0.70.

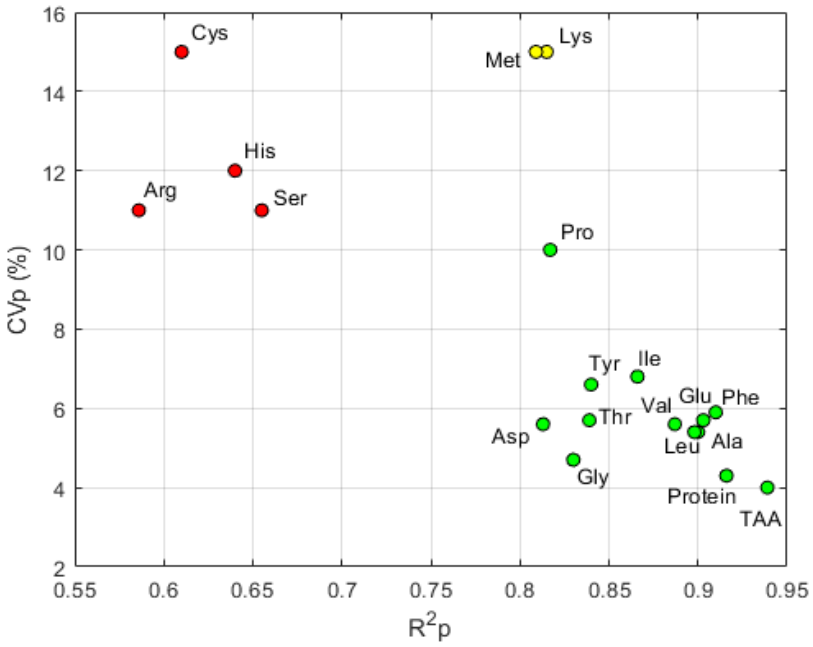
Table 2. Statistics of the ileal digesta calibration models: number of calibration samples (N_C) and prediction (N_V), latent variables (LV), coefficient of determination of calibration (R_c^2), cross-validation (R_{cv}^2) and prediction (R_p^2), root mean square error of calibration RMSEC, cross-validation RMSECV and prediction RMSEP, bias in the prediction, and coefficient of variation of prediction (CV_p).

	N_C	N_V	LV	R_c^2	RMSEC (g/kg, as fed)	R_{cv}^2	RMSECV (g/kg, as fed)	R_p^2	RMSEP (g/kg, as fed)	Bias (g/kg, as fed)	CV_p (%)
Protein	125	51	11	0.95	3.4	0.92	4.5	0.92	5.9	0.13	4.3
Aspartic acid	126	51	10	0.92	0.42	0.87	0.54	0.81	0.74	0.01	5.6
Glutamic acid	125	51	10	0.95	0.60	0.91	0.76	0.90	0.86	-0.04	5.7
Serine	126	52	6	0.73	0.54	0.64	0.62	0.66	0.73	-0.11	11
Glycine	126	51	9	0.92	0.18	0.89	0.21	0.83	0.30	-0.04	4.7
Threonine	125	52	9	0.93	0.17	0.88	0.23	0.84	0.31	-0.03	5.7
Arginine	125	51	12	0.96	0.26	0.87	0.46	0.59	0.78	0.16	11
Alanine	124	52	8	0.93	0.24	0.90	0.29	0.90	0.33	-0.04	5.4
Tyrosine	125	52	9	0.91	0.15	0.87	0.18	0.84	0.23	-0.02	6.6
Cysteine	111	43	5	0.77	0.32	0.70	0.36	0.61	0.37	-0.10	15
Valine	125	51	10	0.93	0.18	0.88	0.24	0.89	0.27	-0.03	5.6
Methionine	111	43	8	0.94	0.18	0.91	0.22	0.81	0.29	0.03	15
Phenylalanine	125	51	10	0.94	0.17	0.89	0.23	0.91	0.25	-0.00	5.9
Isoleucine	124	51	11	0.95	0.12	0.86	0.20	0.87	0.27	-0.03	6.8
Leucine	125	51	10	0.94	0.25	0.90	0.32	0.90	0.41	-0.05	5.4
Lysine	126	51	4	0.83	0.67	0.81	0.72	0.82	0.74	-0.10	15
Proline	125	52	9	0.88	0.51	0.81	0.63	0.82	0.68	0.13	9.8
Histidine	125	51	9	0.79	0.19	0.66	0.24	0.64	0.28	-0.03	12
Total AA	124	51	10	0.95	2.9	0.92	3.7	0.94	4.1	-0.52	4.0

As a general trend, the calibration errors were slightly lower than the prediction errors. According to Marten et al. 1989 [35] the RMSEP (or the CV_p if it is scaled) should not be more than twice the SEL (or the SEL (%) if it is scaled). As it was stated in the reference analyses section, SEL (%) for the current research was 3% for protein and less than 5% for the AAs.

Hence, in this particular case, the CV_p should be smaller than 6 for protein and 10 for AAs. To clarify the results, Figure 3 shows the CV_p against the R^2_p for protein and AAs. Three groups can be distinguished in the graph. A group of well-predicted constituents (coloured in green) that includes those with R^2_p>0.80 and CV_p<10%. There are the protein, all the aliphatic AAs (Gly, Ala, Val, Leu and Ile), the aromatic ones (Phe and Tyr), the acids (Asp and Glu), Thr, Pro and TAA. A second group coloured in yellow that is made of Lys and Met with proper correlations (R^2_p≈0.82) but high prediction errors (CV_p≈15%). These models are not accurate enough for quantitative analysis but could be useful for semi-quantitative or screening analyses. Finally, a third group (AAs coloured in red) made of Cys, Arg, Ser and His that were not predicted well enough (R^2_p<0.70, CV_p>10%). Their calibration models are not good enough for analytical use.

Figure 3. Coefficient of variation of prediction (CV_p) vs coefficient of determination of prediction (R^2_p). Colored code: green for well predicted, yellow for fair predicted and red for bad predicted variables.



Additional insight into the PLSR models can be obtained by plotting the reference values obtained by chromatography with respect to the NIR predictions. For the sake of simplicity, Figure 4 shows the graph for three constituents: protein, one well-predicted AAs (Val and Ile) and one no-so-well predicted (Lys). The graphs for the rest of AAs are presented in the supplementary material (Figure 1S). Protein and most of the AAs showed high correlations between reference and predicted values (see Table 2). The fitted line in this graph should ideally be a line with slope 1 and offset 0. The joint test of offset and slope was carried out [36]. The values of F for all the parameters together with the statistics used to calculate them (slope, offset

and residual variance) are presented as supplementary material (Table 1S). The critical *F* value ranged from 3.18 to 3.23 due to the slight difference in the number of samples of the validation set for each parameter. The *F* statistic was lower than 3 for all the parameters except for Cys (3.36), Arg (3.87), Ser (6.44) and His (3.46). As expected, these four AAs (the worst predicted) did not pass the test. For the rest, the predictions obtained with the calibration models were comparable to the reference values with a level of significance of 0.05.

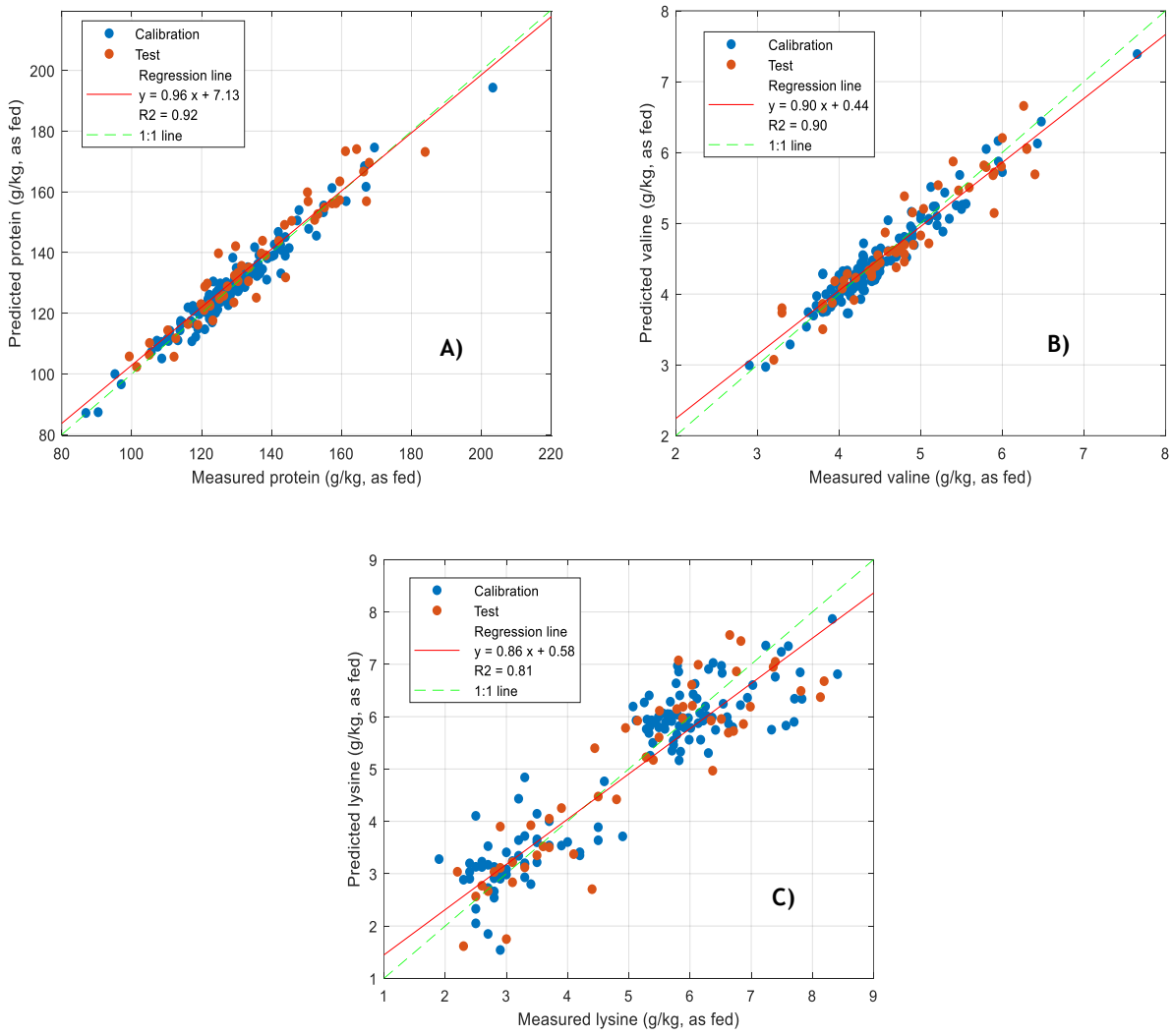


Figure 4. Predicted vs measured values of: A) protein, B) valine and C) lysine.

Discussion

According to the National Research Council (NRC) of the United States (National Research Council, 1994), protein and the digestibility of eight AAs (Met, Cys, Lys, Thr, Trp, Ile, Arg and Val) should be considered in poultry diet formulation. Other institutions such as the Spanish Foundation for the Development of Animal Nutrition (FEDNA) [38] include two additional AAs (Leu and Gly). The NIR calibrations developed for the prediction of protein, Thr, Ile, Val, Leu and Gly performed well and can be used in routine laboratory. Those of Lys and Met could still be used for semi-quantitative or screening analyses. Arg and Cys were poorly predicted, and a calibration model was not developed for Trp because there were no reference data available. The total AA content (TAA) and other AAs (Ala, Phe, Tyr, Pro, Glu and Asp) that might be less important for feed formulation but that could be interesting for other kind of studies were also well predicted. Total AA content (TAA) can be used to calculate relative AAs contents and TAA/protein ratio in samples, which are important indexes for estimating the quality of diets and raw materials [13].

Protein concentration is usually high in feedstuffs and in samples with an animal origin such as ileal digesta and the Dumas method to determine protein is reliable and accurate. The concentration of a specific AA is much lower than protein. Little concentration changes in some AA might be difficult to model if the differences produced in the spectra due to these changes are minimal. The three AA less concentrated (Met, Cys and His) were some of the worst predicted. The accuracy of the HPLC determination of AAs is significantly worse than the one of Dumas method for protein, mainly due to the sample treatments that are necessary to obtain the liquid aliquot that is injected in the chromatograph. Usually, the laboratory wants to determine all or most of the AAs at the same time, and therefore the sample treatments are a compromise approach for all of them and not the optimal one for the determination of each one separately. This could mean higher analytical errors for some AAs and in consequence worse NIR predictions. The method followed in this work consisted of an acid hydrolysis for all AAs except for Met and Cys and a peroxidation followed by an acid hydrolysis for those two. Met and Cys were two of the worst predicted AAs. This peroxidation extra step makes the analytical error of Met and Cys higher. In addition as Figure 1 showed Met and Cys chromatographic peaks suffered from baseline interferences that made difficult the integration of the peak. Higher than expected analytical errors could also have caused that other AAs were

not well predicted. Under the aggressive conditions of the hydrolysis, some AAs may undergo from degradation, which is the main reason for incomplete recovery [39]. For instance, Ser can be converted into α-keto acids by β-elimination of water and Arg could be partially degraded to ornithine. Agents present in the sample solution or in impure hydrolyzing reagents can react or catalyse reactions that affect the quantification of some AAs, including Met and His [40]. Aliphatic AA (Gly, Ala, Val, Leu and Ile) that do not have functional groups and are much more stable during the hydrolysis were all accurately predicted.

The AAs calibrations reported in the literature have very diverse quality. Fontaine et al. 2001, 2002 [14,15] developed specific calibrations for several raw materials, including wheat, corn, barley, rapeseed and other less common ones. They predicted the most important AAs in poultry nutrition according to the aforementioned NRC, obtaining very good correlations (R^2_{cv}>0.90) and low errors (CV_{cv}<5%) in most cases. Overall, the worst results were obtained for Met and Cys followed by Lys and Trp. Kovalenko et al. 2006 [17] predicted the same AAs as in the current research plus Trp in soybean meals. They obtained correlations (R^2_{cv}) ranging from 0.70 to 0.90 for all except Trp, Cys, Met, and Ser (R^2_{cv}<0.70). Zhou et al. 2012 [16] developed calibrations for distillers dried grains with solubles (DDGS). Unlike most of the revised works they obtained a very good calibration model for Met (R^2_{cv}=0.92, CV_{cv}=2.6%). They obtained R^2_{cv}=0.82-0.89 for His, Ile, Lys, Phe and Thr and R^2_{cv}<0.70 for Arg, Trp and Val. Recently, Noel et al. 2021 [18] developed global calibrations with a sample set containing cereals, cereal co-products, cereal substitutes, protein concentrate, fibre-rich by-products and others. Due to the high variability in AA composition that the data set involved they obtained very high correlations (R^2_p>0.90) but also high errors (CV_p>10%). Whereas with a dataset containing only cereals, the correlations decreased (R^2_p=0.70-0.80 for Cys, Thr, Ile, Leu, His, Val, Arg and R^2_p<0.70 for Lys, Met, Leu and Trp). From these works, it can be concluded that sulphured AAs (Met and Cys) and Trp are usually the worst predicted AAs by NIRS. As in the current research, some of them also found difficulty in the prediction of Lys, Ser and Arg.

The results presented for some AA may be worse than others presented previously for raw materials but it should be considered that the current dataset contained more sources of variation that need to be modelled when developing the PLSR models if these have to be generally applicable. Although the composition of the ileal digesta is feed-dependent, it is not only affected by the diet given to the animal but also by the type of animal. In particular, the

present study contains samples from broiler chickens and broiler turkeys of different digestive capacity. Actually, the digestive capacity can even differ among animals of the same species and breed. Second, farm conditions such as humidity or temperature affect the feed intake and energy consumption or sample extraction. Finally, the storage and presentation that is more complex than in the case of raw materials.

Cage of covariance in the prediction of amino acids

Feedstuffs and by extension intestinal contents are multicomponent samples consisting mostly of water, protein, fat, carbohydrates and other minor constituents. In this type of samples, like in other complex biological samples, the amounts of the specific constituents that make up the main constituent (e.g., AAs in "protein", or fatty acids in "fat") can be highly collinear. When several constituents covary, their spectral contributions also vary at the same time and the spectrum behaves as if there was only one main constituent. This has three related consequences. Firstly, the model of one particular constituent is using the signal from that constituent and from the others that covary with it. Thus, to obtain accurate predictions of this one constituent, the other constituents must also be present in the sample and in the same concentration ratios as in the training set. This phenomenon has been called cage of covariance [41]. Secondly, since the models developed for each constituent will have very similar coefficients, they make use of almost the same spectral variations and the predictions of the different constituents will be correlated. This may not be worrisome as long as the concentrations of the constituents in the future samples to be predicted fall in the cage of covariance (i.e., have the same covariance pattern), but it is a limitation if one is mostly interested in which samples break the correlation pattern. The correlated predictions will tend to make these differences less obvious. Finally, the models developed for each specific constituent may not be based on the direct relationship between the constituent and the spectrum, but on indirect relationships with the "main" constituent. The use of models based partially or totally on indirect relationships easily fails if the correlation structures of new samples do not match the ones exhibited by the calibration samples [42].

Eskildsen et al. 2014, 2016 studied the cage of covariance in the predictive ability and the robustness of PLS calibration models in the determination of fatty acids [43] and caseins [44] in milk. They found that the individual constituents were predicted relying on indirect covariance structures, based on covariation with fat and protein respectively. The models worked well for

samples whose correlation structure was similar to those of the training set but failed for samples that had different correlation structures. A similar situation happens for AAs. Figure 5 shows that the correlation coefficient among AAs and between some AAs and protein can be as high as 0.90. The reason behind some AAs (Arg, Met, Ser, Cys, Lys, Pro) present correlation coefficients lower than 0.70 is not clear. The aforementioned possible unreality of the reference values due to the reference method could affect the correlations in the dataset but also the difference in how the AAs interact and are digested in the small intestine of the animal.

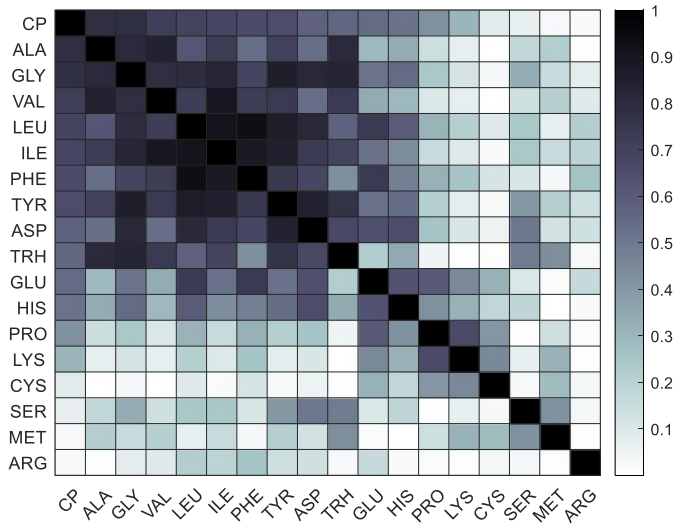


Figure 5. Heat map of correlation coefficients between protein and amino acids.

The strong correlations found for several AAs rise questions about whether the prediction of the highly correlated AAs is driven by a direct relationship with the spectrum or if it originates from indirect correlations caused by the collinearity among the training set constituents. Figure 6 shows the results of PCA of the reference concentration of AAs. The first principal component (PC) explains c.a. 60% of the variance, and 95% of the variance is explained with the first four PCs, in agreement with the observed correlation among the reference values. The correlations are even higher in the matrix of predicted concentrations that make the PCs account for a higher amount of variance than the reference data case (95% of explained variance is reached with only three PCs). This indicates that the covariances are stronger in the predictions than in the reference values and that predictions tend to be dependent on each other as a result of the lack of orthogonality in the AA concentrations in the training samples. A similar behaviour was reported by Eskildsen et al. 2021 [41] in the NIR prediction of fatty acids in bovine milk, Atlantic salmon muscle and porcine adipose tissue.

The PLS regression vectors of protein, Ala, Gly, Val, Leu, Ile, Phe, Tyr, Asp, Thr and Glu were similar (Figure 7), which indicates that the models used very similar spectral patterns. The models with the most distinct regression coefficients (His, Pro, Lys, Cys, Ser, Met, and Arg) had the worst predictive ability. Covalence et al. 2006 [17] also found that most of the PLS regression vectors of NIR models for AAs in soybean meals followed the same trend and suggested that the calibrations mostly predicted protein.

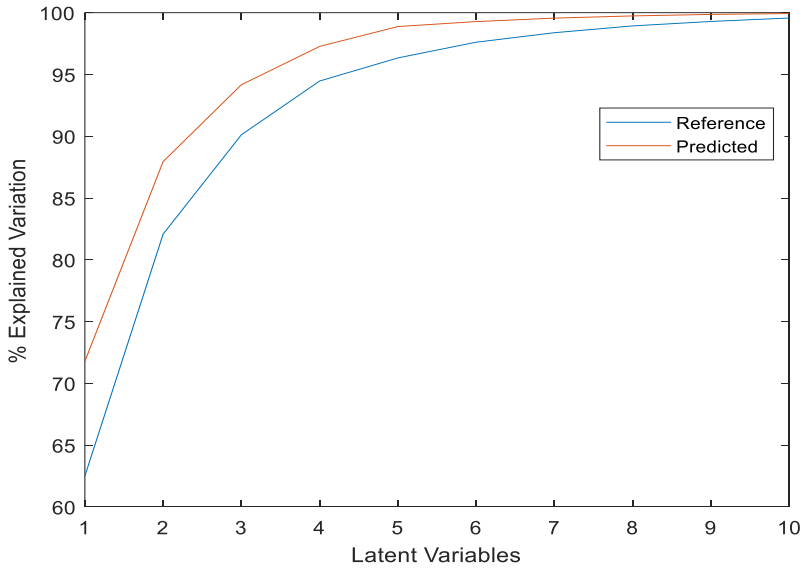


Figure 6. Cumulative (%) explained variance as a function of the number of PCs performed on the reference AA concentrations and the predicted concentrations.

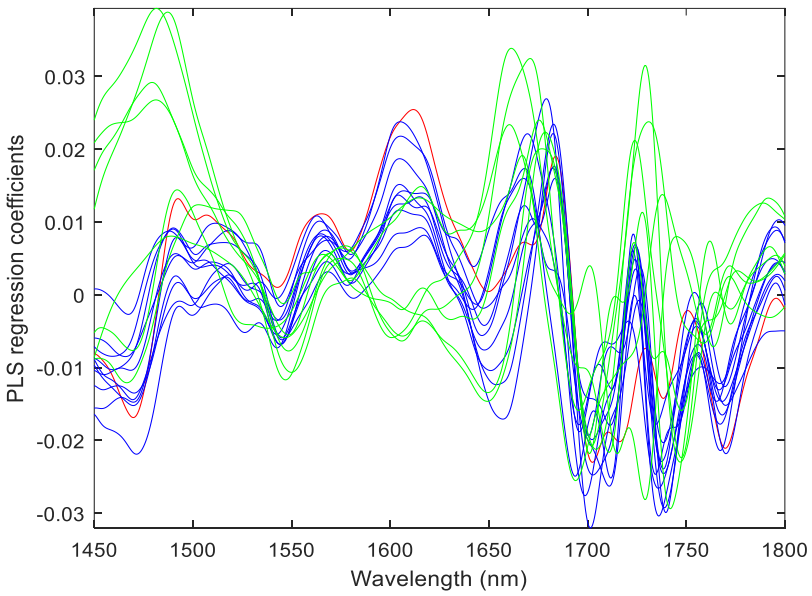


Figure 7. Normalized PLS regression vectors of protein (red), amino acids with correlation coefficients with protein higher than 0.70 (blue) and amino acids with correlation coefficients with protein lower than 0.70 (green).

The AAs (Ala, Gly, Val, Leu, Ile, Phe, Tyr, Asp, Thr and Glu) that are highly correlated with protein (R>0.75) were all well predicted while the ones showing lower correlations (His, Pro, Lys, Cys, Ser, Met, and Arg) were predicted less well (Figure 8). Similar results were found by Kovalenko et al. 2006 [17] and by Fontaine et al. 2001, 2002 [14,15] during the AAs prediction of a variety of feedstuffs by NIRS and PLSR.

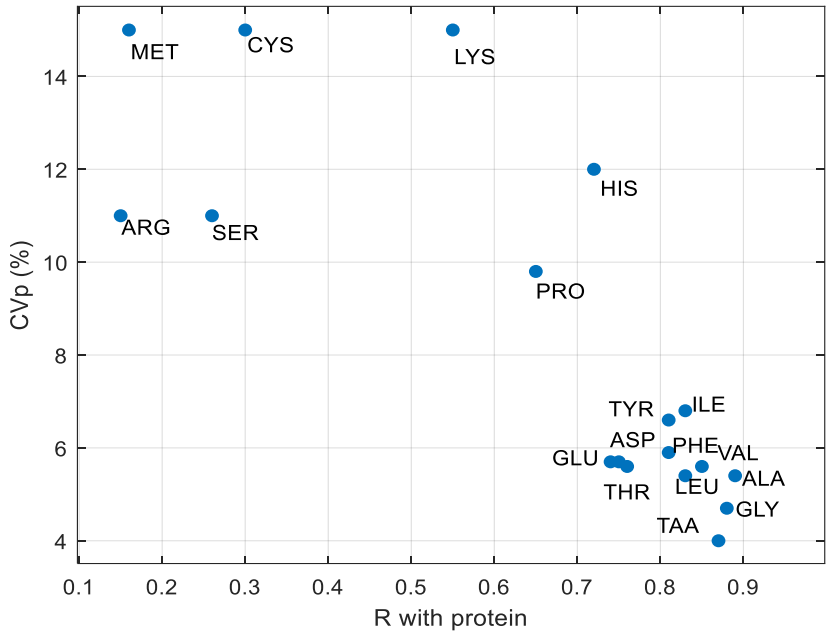


Figure 8. Coefficient of variation of prediction (CVp) versus correlation coefficient (R) between amino acids and protein.

The above results suggest that, as a consequence of the natural correlation between the concentrations of AAs in the calibration samples, the AAs are predicted based on indirect relationships between them and protein or combined signals from AAs that form the so-called "cage of covariance". Ultimately, this means that the models will be valid for new samples whose ratio of concentrations of AAs are very similar to those used for training, but one must be aware that the results may be misleading if the concentration ratios in the new samples change. Since the correlations are the result of natural process in biological systems, it is difficult to break the covariance structure between the AAs in the intestinal content by designing the experiments and therefore one may expect that the models, while valid for the present samples, may need timely revaluation with new samples until the covariance pattern diminishes. Extending the dataset with future ileal samples from animals of different digestive capacity (broilers and turkeys) and involving very different diets may contribute to widen the "cage of covariance" and make the models more apt for a variety of concentration ratios.

Conclusions

NIR calibration models for poultry ileal digesta have been developed for the first time. Protein, several amino acids (Gly, Ala, Val, Leu, Ile, Phe, Tyr, Asp, Glu, Thr and Pro) and the total content of amino acids (TAA) were well predicted. The NIR determination of all these constituents could reduce the cost of digestibility studies and contributing in this way to expedite the obtaining of ileal digestibility coefficients of protein and amino acids of feedstuffs and diets. The fact that the calibration samples originated from natural processes introduced a covariance structure in the concentration ratios of the samples. As a result, the majority of the amino acids concentrations were correlated among them and with the protein content. This affected the independence of the predicted concentration of amino acids. As a consequence, new samples may not be predicted well unless they have the same amino acid profiles as the training data. Future research will seek the widening of the cage of covariance by measuring samples in more diverse in-vivo assays and update the models accordingly, to make them usable for more diverse compositions.

References

[1] L. Cappelaere, J. Le Cour Grandmaison, N. Martin, W. Lambert, Amino Acid Supplementation to Reduce Environmental Impacts of Broiler and Pig Production: A Review, Front. Vet. Sci. 8 (2021) 1–14.

[2] C.W. Maynard, M.T. Kidd, P. V. Chrystal, L.R. McQuade, B. V. McInerney, P.H. Selle, S.Y. Liu, Assessment of limiting dietary amino acids in broiler chickens offered reduced crude protein diets, Anim. Nutr. 10 (2022) 1–11.

[3] P.J. Moughan, V. Ravindran, J.O.B. Sorbara, Possible Substrates for Exogenous Enzymes Dietary protein and amino acids — Consideration of the undigestible fraction 1, Poult. Sci. (2014) 1–11.

[4] V. Ravindran, O. Adeola, M. Rodehutscord, H. Kluth, J.D. van der Klis, E. van Eerden, A. Helmbrecht, Determination of ileal digestibility of amino acids in raw materials for broiler chickens – Results of collaborative studies and assay recommendations, Anim. Feed Sci. Technol. 225 (2017) 62–72.

[5] F. Cheli, D. Battaglia, L. Pinotti, A. Baldi, State of the art in feedstuff analysis: A technique-oriented perspective, J. Agric. Food Chem. 60 (2012) 9529–9542.

[6] L. Chen, L. Xing, L. Han, Review of the Application of Near-Infrared Spectroscopy Technology to Determine the Chemical Composition of Animal Manure, J. Environ. Qual. 42 (2013) 1015–1028.

[7] F. García-Sánchez, L. Galvez-Sola, J.J. Martínez-Nicolás, R. Muelas-Domingo, M. Nieves, Using near-infrared spectroscopy in agricultural systems, in: K. Kyprianidis, J. Skvaril (Eds.), Developments in near-infrared spectroscopy, Intech-Open, London, UK, 2017: pp. 97-128.

[8] W. Hunt, D. Fulk, B. Elder, K. Norris, Collaborative study on infrared reflectance devices for determination of protein in hard red winter wheat, and for protein and oil in soybeans., Cereal. Foods World. 22 (1977) 534–536.

[9] P.C. Williams, Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds, Cereal Chem. 52 (1975) 561–576.

[10] G. Rubenthaler, B. Bruinsma, Lysine Estimation in Cereals by Near-Infrared Reflectance, Crop Sci. 18 (1978) 1039–1042.

[11] A.A. Gill, C. Starr, D.B. Smith, Lysine and nitrogen measurement by infra-red reflectance analysis as an aid to barley breeding, J. Agric. Sci. 93 (1979) 727–733.

[12] P.C. Williams, K.R. Preston, K.H. Norris, P.M. Starkey, Determination of Amino Acids in Wheat and Barley by Near-Infrared Reflectance Spectroscopy, J. Food Sci. 49 (1984) 17–20.

[13] G.L. Chen, B. Zhang, J.G. Wu, C.H. Shi, Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 165 (2011) 111–119.

[14] J. Fontaine, J. Hörr, B. Schirmer, Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal, J. Agric. Food Chem. 49 (2001) 57–66.

[15] J. Fontaine, B. Schirmer, J. Horr, Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum, J. Agric. Food Chem. 50 (2002) 3902–3911.

[16] L.J. Zhou, L.Y. Zhang, E.X. Zhang, J.T. Li, W.J. Yang, Z.Y. Wang, Rapid determination of swine available energy and amino acids in corn distillers dried grains with solubles by near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 175 (2012) 198–202.

[17] I. V. Kovalenko, G.R. Rippke, C.R. Hurburgh, Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy, J. Agric. Food Chem. 54 (2006) 3485–3491.

[18] S.J. Noel, H.J.H. Jørgensen, K.E. Bach Knudsen, Prediction of protein and amino acid composition and digestibility in individual feedstuffs and mixed diets for pigs using near-infrared spectroscopy, Anim. Nutr. 7 (2021) 1242–1252.

[19] D.L. Pazdernik, A.S. Killam, J.H. Orf, Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy, Agron. J. 89 (1997) 679–685.

[20] S. Pujol, A.M. Pérez-Vendrell, D. Torrallardona, Evaluation of prediction of barley digestible nutrient content with near-infrared reflectance spectroscopy (NIRS), Livest. Sci. 109 (2007) 189–192.

[21] E.O. Alamu, A. Menkir, M. Adesokan, S. Fawole, B. Maziya-dixon, Near-Infrared Reflectance Spectrophotometry (NIRS) Application in the Amino Acid Profiling of Quality Protein Maize (QPM), Foods, 11 (2022) 2779.

[22] D. Shi, J. Hang, J. Neufeld, S. Zhao, J.D. House, Estimation of crude protein and amino acid contents in whole, ground and defatted ground soybeans by different types of near-infrared (NIR) reflectance spectroscopy, J. Food Compos. Anal. 111 (2022) 104601.

[23] D. Bastianelli, E. Fermet-Quinet, F. Davrieux, C. Hervouet, L. Bonnal, Calibration strategies for prediction of amino acid content of poultry feeds, in Proceedings of the 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 314–318.

[24] I. González-Martín, N. Álvarez-García, J.M. González-Cabrera, Near-infrared spectroscopy (NIRS) with a fibre-optic probe for the prediction of the amino acid composition in animal feeds, Talanta. 69 (2006) 706–710.

[25] L. Noah, P. Robert, S. Millar, M. Champ, Near-Infrared Spectroscopy As Applied to Starch Analysis of Digestive Contents, J. Agric. Food Chem. 45 (1997) 2593–2597.

[26] P. Lebzien, C. Paul, Use of near-infrared reflectance spectroscopy for the estimation of the microbial portion of non-ammonia-nitrogen in the duodenum of dairy cows, Anim. Feed Sci. Technol. 68 (1997) 225–233.

[27] V. Ravindran, W.L. Bryden, Amino acid availability in poultry-in vitro and in vivo measurements, Aust. J. Agric. Res. 50 (1999) 889–908.

[28] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[29] U. Bütikofer, D. Fuchs, J.O. Bosset, W. Gmür, Automated HPLC-amino acid determination of protein hydrolysates by precolumn derivatization with OPA and FMOC and comparison with classical ion exchange chromatography, Chromatographia. 31 (1991) 441–447.

[30] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[31] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[32] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[33] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428.

[34] J. Shenk, M. Westerhaus, Routine Operation, Calibration, Development and Network System Management Manual. NIRSystems, Silver Spring, MD, 1995.

[35] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[36] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1998.

[37] Council National Research, Nutrient Requirements of Poultry: Ninth Revised Edition, The National Academic Press, Washington, DC, 1994.

[38] G. Santomá, G.G. Mateos, Necesidades nutricionales en avicultura. Normas FEDNA, 2a, Improtalia S.L., Madrid, Spain, 2018.

[39] A. Lamp, M. Kaltschmitt, O. Lüdtke, Improved HPLC-method for estimation and correction of amino acid losses during hydrolysis of unknown samples, Anal. Biochem. 543 (2018) 140–145.

[40] M. Fountoulakis, H.W. Lahm, Hydrolysis and amino acid composition analysis of proteins, J. Chromatogr. A. 826 (1998) 109–134.

[41] C.E. Eskildsen, T. Næs, P.B. Skou, L.E. Solberg, K.R. Dankel, S.A. Basmoen, J.P. Wold, S.S. Horn, B. Hillestad, N.A. Poulsen, M. Christensen, T. Pieper, N.K. Afseth, S.B. Engelsen, Cage of covariance in calibration modeling: Regressing multiple and strongly correlated response variables onto a low rank subspace of explanatory variables, Chemom. Intell. Lab. Syst. 213 (2021) 104311.

[42] M.H. Rhiel, M.I. Amrhein, I.W. Marison, U. Von Stockar, The Influence of Correlated Calibration Samples on the Prediction Performance of Multivariate Models Based on Mid-Infrared Spectra of Animal Cell Cultures, Anal. Chem. 74 (2002) 5227–5236.

[43] C.E. Eskildsen, M.A. Rasmussen, S.B. Engelsen, L.B. Larsen, N.A. Poulsen, T. Skov, Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables, J. Dairy Sci. 97 (2014) 7940–7951.

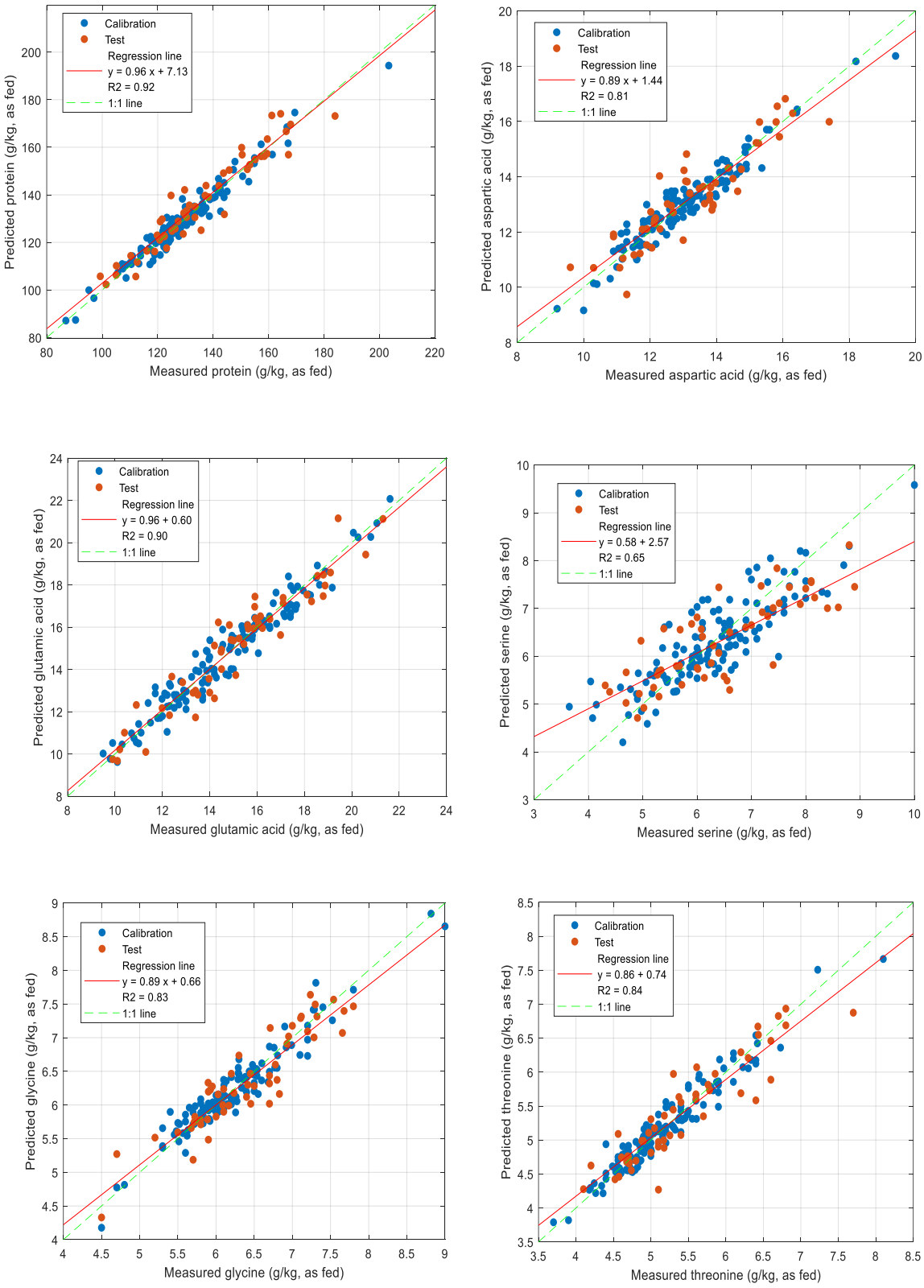
[44] C.E. Eskildsen, T. Skov, M.S. Hansen, L.B. Larsen, N.A. Poulsen, Quantification of bovine milk protein composition and coagulation properties using infrared spectroscopy and chemometrics: A result of collinearity among reference variables, J. Dairy Sci. 99 (2016) 8178–8186.

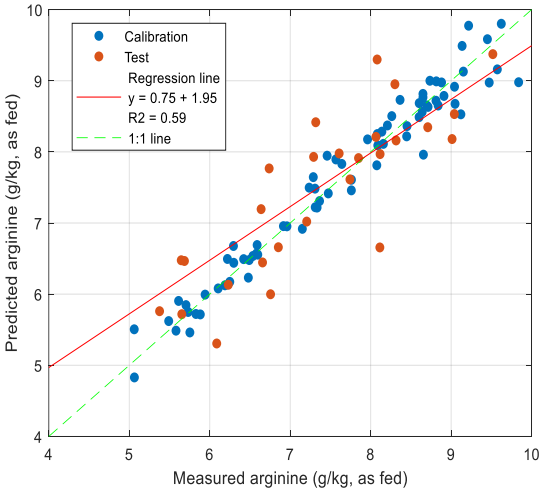
Supplementary material

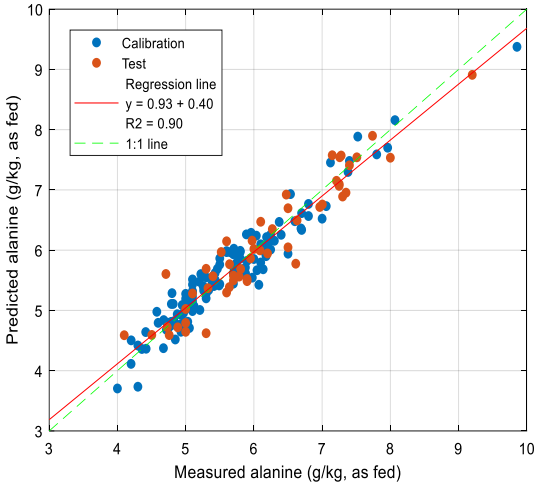
Table S1. Results of the join test of offset and slope. Slope, ofsett, number of samples (N), residual variance (S_e), F statistic (F_{stat}) and critical F (F_{crit}) with a level of significance of 0.05.

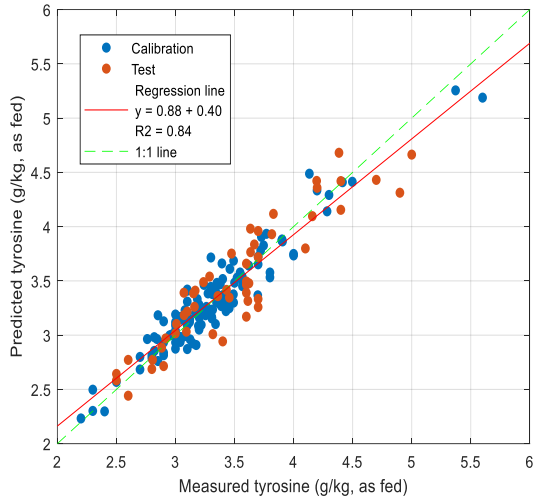
	Slope	Offset	N	S_e	F_{stat}	F_{crit}
Protein	0.96	4.61	51	5.99	1.68	3.19
Aspartic acid	0.89	1.44	51	0.76	1.44	3.19
Glutamic acid	0.96	0.60	51	0.87	0.50	3.19
Serine	0.58	2.57	52	0.74	6.44	3.18
Glycine	0.89	0.66	51	0.30	2.01	3.19
Threonine	0.86	0.73	52	0.32	2.83	3.19
Arginine	0.75	1.95	51	0.81	3.87	3.19
Alanine	0.93	0.40	52	0.34	1.50	3.18
Tyrosine	0.88	0.40	52	0.24	2.12	3.18
Cysteine	0.75	0.54	43	0.38	3.36	3.23
Valine	0.90	0.44	51	0.27	2.19	3.19
Methionine	0.87	0.28	43	0.30	1.73	3.23
Phenylalanine	0.99	0.03	51	0.26	0.01	3.19
Isoleucine	0.94	0.22	51	0.27	0.99	3.19
Leucine	0.91	0.60	51	0.42	2.11	3.18
Lysine	0.86	0.58	51	0.76	2.48	3.19
Proline	0.86	1.01	52	0.69	2.85	3.19
Histidine	0.76	0.53	51	0.29	3.46	3.19
Total of amino acids	0.92	7.50	51	4.23	2.50	3.19

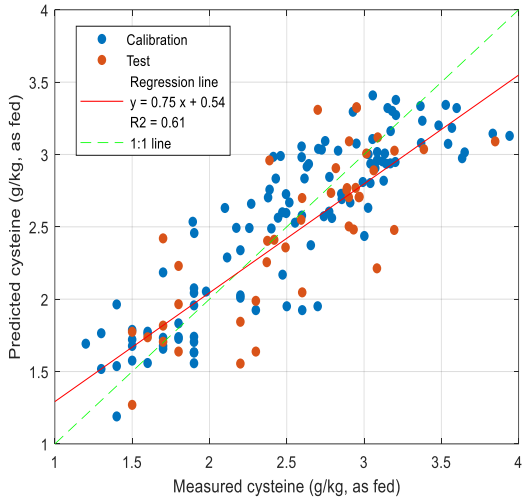
Figure S1. Predicted vs measured values of protein and amino acids.

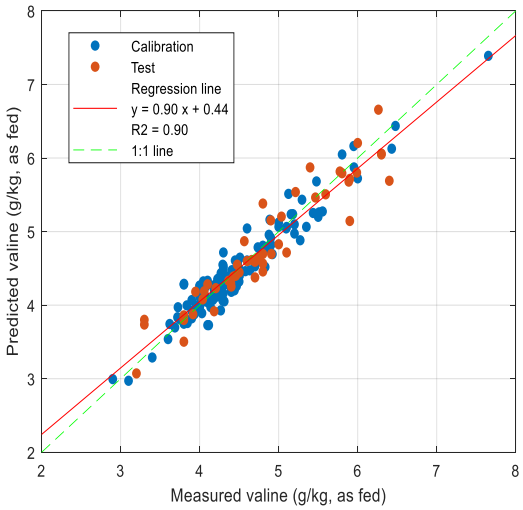


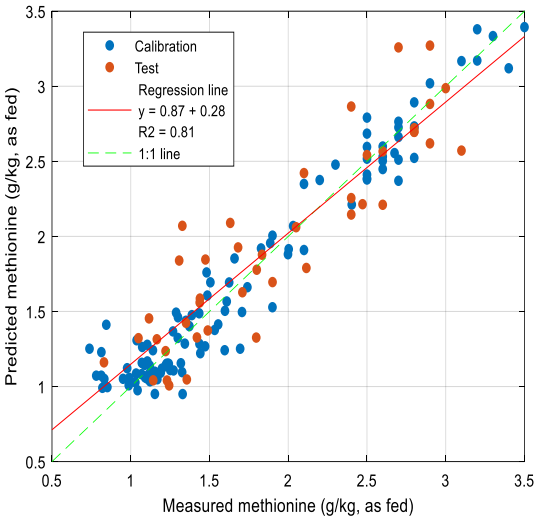

Calibration
Test
Regression line
y = 0.75 + 1.95
R2 = 0.59
1:1 line
Predicted arginine (g/kg, as fed)
Measured arginine (g/kg, as fed)

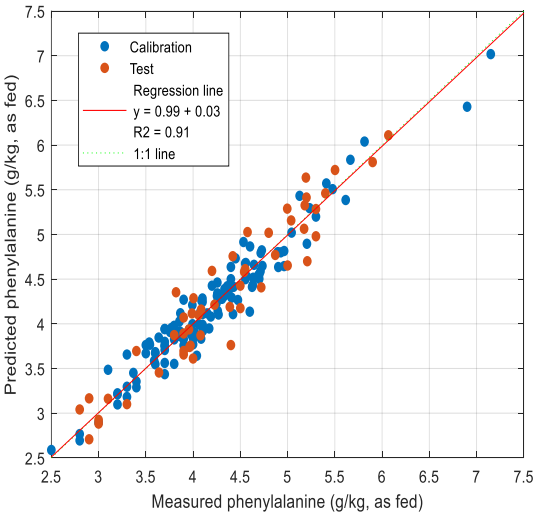

Calibration
Test
Regression line
y = 0.93 + 0.40
R2 = 0.90
1:1 line
Predicted alanine (g/kg, as fed)
Measured alanine (g/kg, as fed)

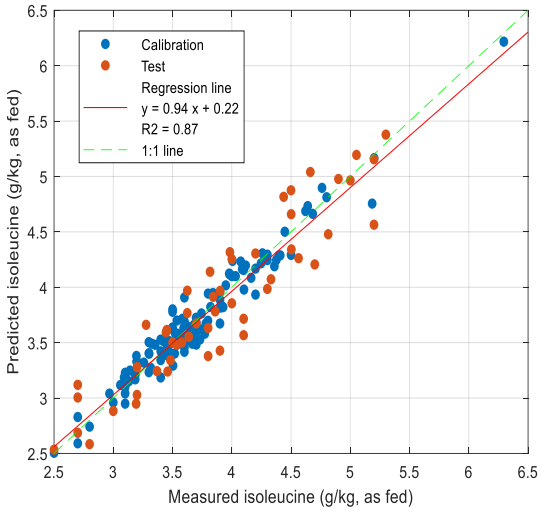

Calibration
Test
Regression line
y = 0.88 + 0.40
R2 = 0.84
1:1 line
Predicted tyrosine (g/kg, as fed)
Measured tyrosine (g/kg, as fed)

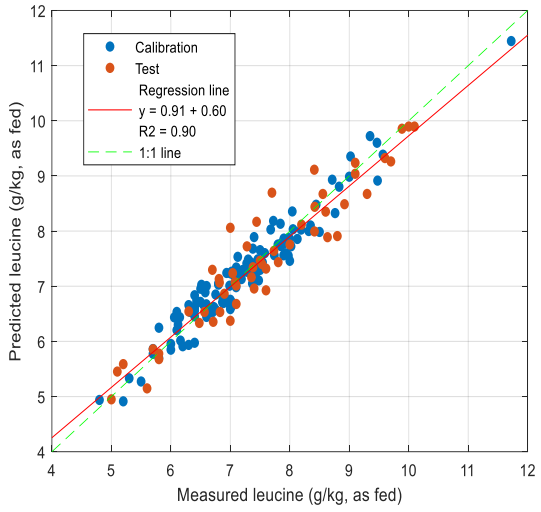

Calibration
Test
Regression line
y = 0.75 x + 0.54
R2 = 0.61
1:1 line
Predicted cysteine (g/kg, as fed)
Measured cysteine (g/kg, as fed)

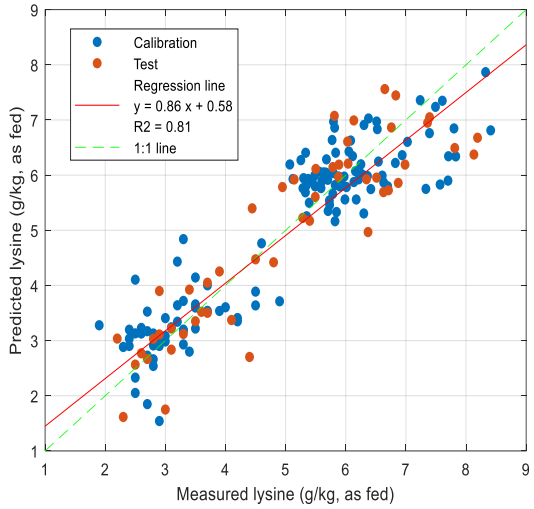

Calibration
Test
Regression line
y = 0.90 x + 0.44
R2 = 0.90
1:1 line
Predicted valine (g/kg, as fed)
Measured valine (g/kg, as fed)

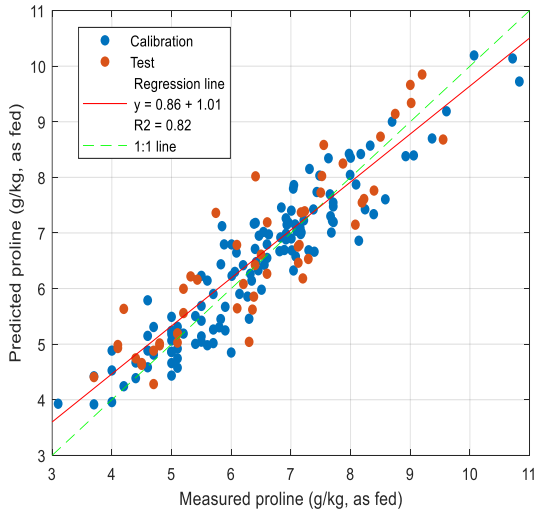

Calibration
Test
Regression line
y = 0.87 + 0.28
R2 = 0.81
1:1 line
Predicted methionine (g/kg, as fed)
Measured methionine (g/kg, as fed)

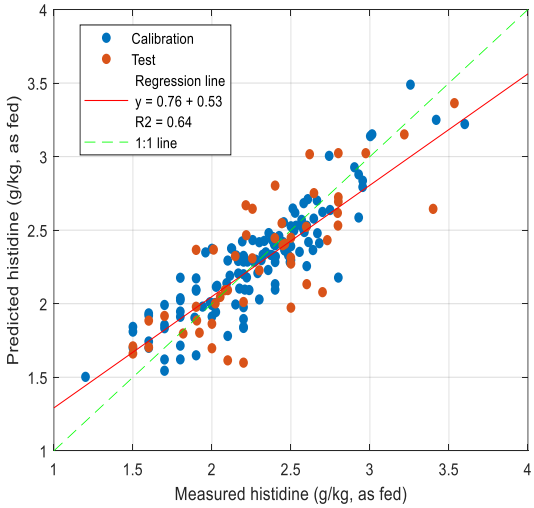

Calibration
Test
Regression line
y = 0.99 + 0.03
R2 = 0.91
1:1 line
Predicted phenylalanine (g/kg, as fed)
Measured phenylalanine (g/kg, as fed)

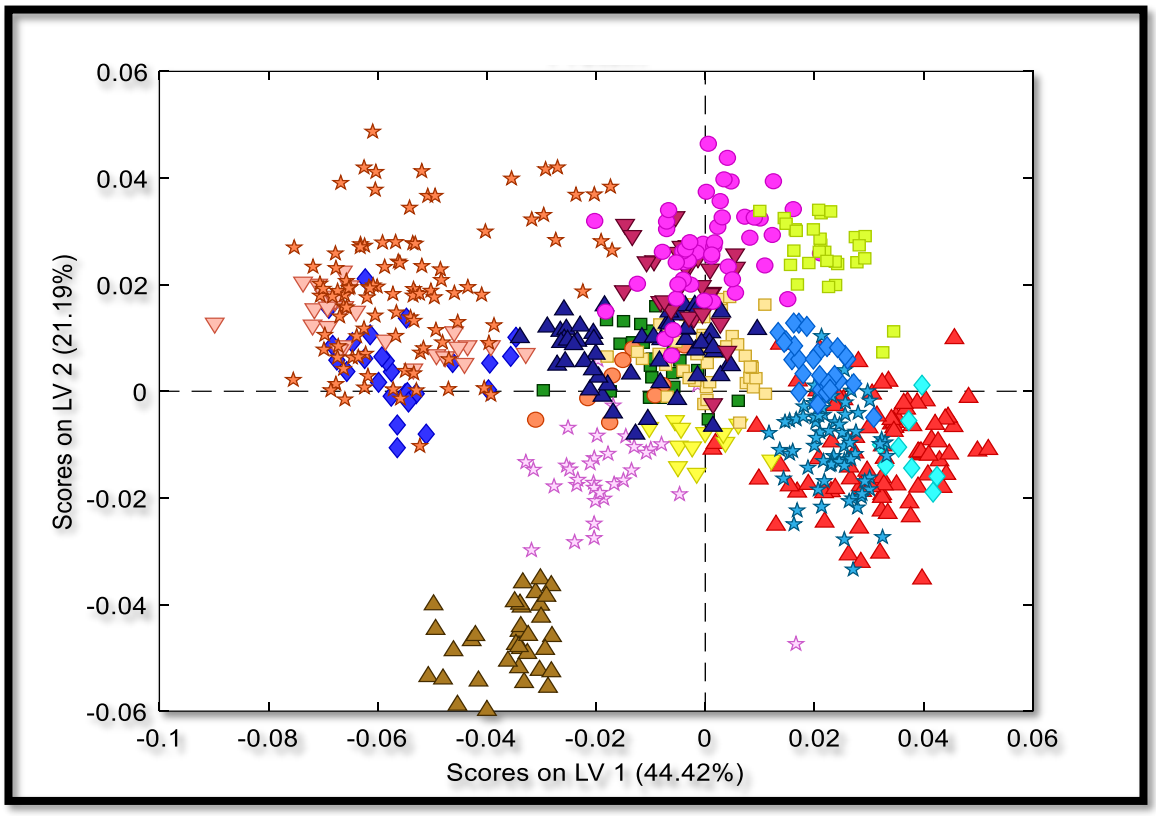

Calibration
Test
Regression line
y = 0.94 x + 0.22
R2 = 0.87
1:1 line
Predicted isoleucine (g/kg, as fed)
Measured isoleucine (g/kg, as fed)


Calibration
Test
Regression line
y = 0.91 + 0.60
R2 = 0.90
1:1 line
Predicted leucine (g/kg, as fed)
Measured leucine (g/kg, as fed)

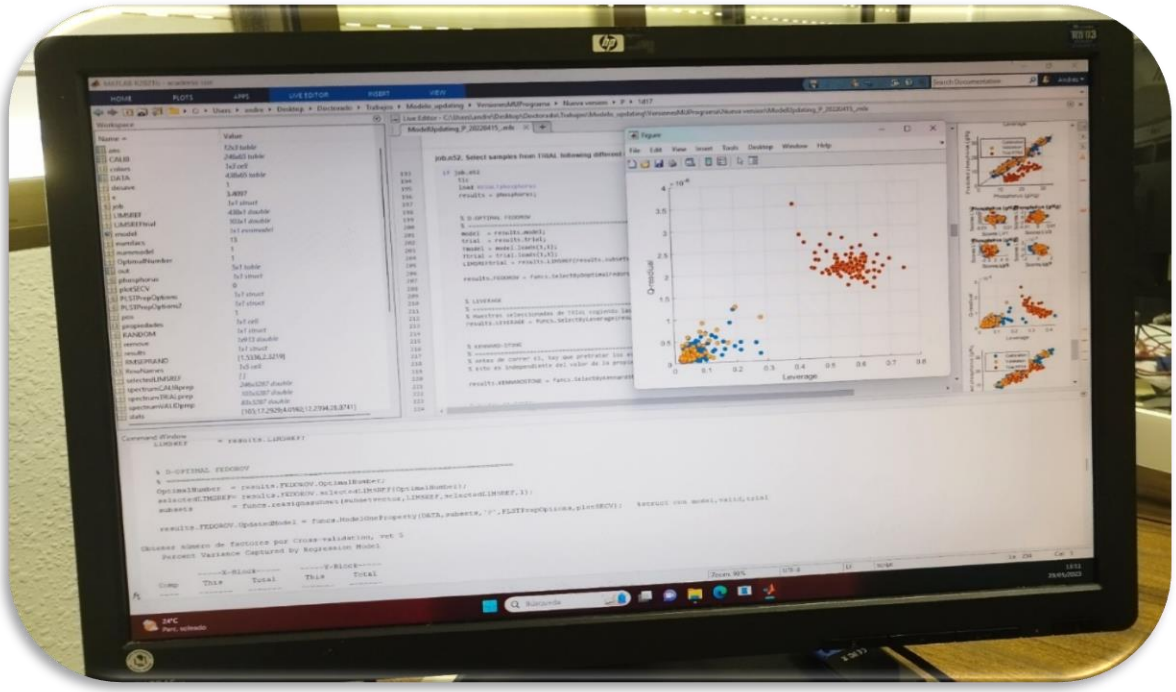

Calibration
Test
Regression line
y = 0.86 x + 0.58
R2 = 0.81
1:1 line
Predicted lysine (g/kg, as fed)
Measured lysine (g/kg, as fed)


Calibration
Test
Regression line
y = 0.86 + 1.01
R2 = 0.82
1:1 line
Predicted proline (g/kg, as fed)
Measured proline (g/kg, as fed)


Calibration
Test
Regression line
y = 0.76 + 0.53
R2 = 0.64
1:1 line
Predicted histidine (g/kg, as fed)
Measured histidine (g/kg, as fed)



Chapter 4.
Maintenance of NIR models



4.1. Introduction

As it was mentioned in Section 2.2, it is important to adapt calibration models to maintain prediction performance despite possible changes. There are two types of changes, either caused by 'product drift' or caused by 'instrumental drift' [1]. In some cases, the boundary of the sample population for future analyses may not be known or may change due to controlled or uncontrolled factors. An example is the change of supplier in the case of a calibration model used for raw materials in a feed factory, so that the raw materials are from a different region and have new features. Changes in the instrumental response can be caused by ageing, repair or replacement of the instrument used to develop the original calibrations. Generally, instrumental drifts are more problematic than product drifts because the first case affects all future predictions while the second case will only affect the predictions of the novel samples.

To detect instrumental drifts the ISO 12099:2017 norm [2] recommends using control charts. Samples that span the calibration range and the sources of variability included in the calibration model are selected (based on the knowledge of the analyst about the samples or using some sample selection method such as Kennard Stone), analysed by reference methods, stored and measured periodically. Prediction errors out from the alert limit (generally, 3 × RMSEP), biased predictions or an excessive variation in results would indicate changes or degradation of the instrument performance. Several calibration transfer methods have been proposed to correct instrumental drifts [3]. This topic is outside the scope of this thesis since instrumental drifts were not detected during the period in which it was carried out.

To detect product drifts, the knowledge of the analyst about unusual characteristics of the new samples can warn that a calibration is about to fail although sometimes the factors that affect the predictions are not obvious. Conversely, the calibration model could withstand changes that one would initially expect to negatively affect prediction performance. A common practice to identify unusual characteristics of the samples is the use of outlier diagnostic measures such as Hotelling's T^2 and Q residuals. Samples with high T^2 value might have extreme concentrations or unusual combinations of concentrations but usually retain the same composition as the calibration data while samples with high Q value are often unique samples with new analytes or new variations [4]. Samples with T^2 or Q values beyond the limits flag extrapolation out of the calibration and hence predictions cannot be trusted. If only a small number of samples per batch

or season exceed the limits, these samples would be considered outliers and would be analysed by the reference method. If the new samples present consistently high T^2 or Q values beyond the limits this would indicate a product drift (i.e., a novel product) as long as a sudden instrument malfunction has been discarded through the control charts [3]. There exist different methods to adapt the calibration to the new situation avoiding a full recalibration. Among them, model updating using new samples to expand the domain is probably the most used [1]. Usually, new samples may have new redundant information, so the addition of all the new samples to the model is not necessary to account for the new sources of variability. In order to reduce costs and effort it is convenient to have a joint strategy to monitor and update the model analysing the same selected samples.

Strategies to maintain the calibration models for raw materials and compound feeds have been presented [5,6]. In Section 4.2 we apply a strategy to maintain compound feed calibrations that consists of including the incoming samples that had T^2 or Q values beyond the limits [7]. In Section 4.3 we show a strategy to select reference samples for updating calibration models used in the analysis of pig faeces. Both strategies improve the performance of a calibration model for new samples by incorporating new variations into the model, but they are applied differently. The first strategy is useful when the samples to be measured arrive successively or in small batches while the second is applied when the samples come in large batches.

References

[1] X. Capron, B. Walczak, O.E. De Noord, D.L. Massart, Selection and weighting of samples in multivariate regression model updating, Chemom. Intell. Lab. Syst. 76 (2005) 205–214.

[2] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[3] J.J. Workman, A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy, Appl. Spectrosc. 72 (2018) 340–365.

[4] B.M. Wise, R.T. Roginski, A calibration model maintenance roadmap, IFAC-PapersOnLine. 48 (2015) 260–265.

[5] P. Dardenne, G. Sinnaeve, V. Baeten, Multivariate calibration and chemometrics for near infrared spectroscopy: Which method?, J. Near Infrared Spectrosc. 8 (2000) 229–237.

[6] M. Schoot, C. Kapper, G. van Kessel, G. Postma, L.M. Buydens, J.J. Jansen, Cost-benefit analysis of calibration model maintenance strategies for process monitoring, Anal. Chim. Acta. 1180 (2021) 338890.

[7] S.K. Setarehdan, J.J. Soraghan, D. Littlejohn, D.A. Sadler, Maintenance of a calibration model for near infrared spectrometry by a combined principal component analysis-partial least squares approach, Anal. Chim. Acta. 452 (2002) 35–45.

4.2. Maintenance of near-infrared calibration models for the nutrient content of compound feeds

Abstract

Monitoring and updating calibration models are common maintenance tasks in analytical methods based on near-infrared spectroscopy. This work shows the over time performance of PLSR models for key constituents in compound feeds, including crude protein, fat, crude fiber, and ash. The dataset used encompasses poultry and swine compound feeds produced over multiple years. A strategy involving the addition of outlier samples is applied to periodically update the models. The performance of the models was evaluated using new samples from each subsequent period. By incorporating samples exhibiting new or extreme sources of variations (those with Hotelling's T^2 and/or Q residuals exceeding the established limits for these statistics) the predictive capability of the crude protein, fat, crude fiber, and ash models for incoming samples was enhanced. As a result, with the updated models, the presence of outliers and the root mean square error of prediction (RMSEP) decreased. This updating approach was proven effective in keeping the prediction errors for crude protein and fat within the limits established by the laboratory (RMSEP<8 g/kg and <6 g/kg, respectively). In the case of crude fiber, it took a longer time to develop a reliable model (RMSEP<10 g/kg) that accurately predicts incoming samples. This was primarily due to the smaller number of samples available with crude fiber content compared to the other constituents studied. For the ash model, the improvements made through the updating process were insufficient to achieve prediction errors lower than the limits set by the laboratory (RMSEP>4 g/kg). One possible explanation is that the information captured in the NIR spectra does not sufficiently represent the values of ash. Overall, the findings of this study highlighted the importance of regular assessment and optimization of calibration models to ensure reliable and accurate results in compound feed analysis

Introduction

Compound feed manufacturers, farms, public or private control laboratories and agri-food research centres analyse compound feeds on a routine basis. NIRS determinations are fast and cheap and for years they have been replacing wet chemistry methods in the nutritional analysis

of agricultural and food products [1]. Before a prediction model is implemented, it must be validated. Although validation methods such as cross-validation (CV) can be used, the use of an independent subset of the data is preferable [2]. Even if the independent test set is representative of future samples, ensuring long-term validation remains challenging due to the influence of controlled or uncontrolled factors that can affect samples over time. Therefore, it is very important to have strategies to ensure confidence in NIR predictions for incoming samples and to take appropriate action in case outliers are detected. One approach to verify that NIR measurements fall within an ordinary or expected range of variation, accounting for changes in the environment or instrument, is by implementing control charts. The spectra of a reference sample is measured over time and its prediction value is controlled [3]. Prediction errors out from the alert limit (generally, 3 × RMSEP), biased predictions or an excessive variation in results indicate the presence of unusual variability in the NIR measurements. Another situation concerns when an unknown sample is predicted and its model diagnostic measures (Hotelling's T^2, Q residuals or similar) warn that it is not within the limits of the calibration model [4]. High T^2 samples have the same directions of variation as the calibration samples although more extreme, whereas high Q samples exhibit new variations. Samples with higher values of these statistics than the model confidence limits are identified as spectral outliers, and therefore they cannot be predicted with the validated model [5]. These samples should be analysed by reference methods and can be added to the calibration model to expand the multivariate calibration space. This strategy can enhance the predictive ability of the model for samples with new variations [6,7]. Depending on the number of samples and sources of variability covered by the model and the possible variability of the incoming samples, more effort will be required to add samples until a stable calibration (few outliers and accurate predictions for future samples) is achieved.

This study presents the over time performance of partial least squares regression (PLSR) models for key constituents in compound feeds, namely crude protein (CP), fat, crude fiber (CF), and ash. The dataset used in this study consists of swine and poultry compound feeds produced over multiple years. The effectiveness of a model maintenance strategy involving the addition of outlying samples is applied periodically to update the models. The efficacy of this strategy is compared across the four constituents mentioned.

Materials and methods

Samples

The dataset used contained historical data of compound feed samples for swine and poultry produced from 2018 to 2021 in the Institute of Agrifood Research and Technology (IRTA, Cosntanti, Spain). All the samples were analysed by reference methods. CP, fat, CF and ash were determined according to the AOAC, 2016 methods 925.09, 968.06, 978.10, 942.05 respectively [8]. For the four parameters studied in this work the standard error of the laboratory (SEL) was as follows: CP (4 g/kg), fat (3 g/kg), CF (5 g/kg) and ash (2 g/kg). The spectrum of approximately 100 grams of sample was measured on a NIRS DS2500 (Foss NIR Systems, Denmark) with a 10.2 cm diameter cup in reflectance mode from 800 to 2499.5 nm every 0.5 nm.

To compare the performance of calibration models over time without and with maintenance the dataset was split. Initial models were developed using only the samples produced in 2018. The remaining samples, those produced in 2019, 2020 and 2021, were treated as incoming samples. We decided that two stages of maintenance per year would be appropriate considering the rate of compound feed production per year (around 500 samples). Therefore, the samples produced each year were split into two halves in the order they were produced so there were 6 sets of incoming samples: 2019a, 2019b, 2020a, 2020b, 2021a and 2021b. We retain the order in which the samples were produced to examine the impact of emergent variations on prediction performance. In the Supplementary Material the descriptive statistics of CP, fat, CF and ash for each dataset are shown (Table S1-4).

Initial models

Partial least squares regression (PLSR) was used to develop the initial calibration models for CP, fat, CF and ash using the samples produced in 2018. Common spectral pretreatments were tested including normalization, standard normal variate (SNV) [9] multiplicative scatter correction (MSC) [10] and 1st and 2nd order derivatives [11] using different window widths. A five-fold venetian blind cross validation was used to choose the optimal number of latent variables (LVs) for the models. Samples were labelled as reference outliers and removed from

the calibration set when their prediction error of cross validation was higher than 3 times the RMSECV of the calibration set.

PLS toolbox software (2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab R2020a (The MathWorks Inc., Natick, MA, USA) was used for the development of the prediction models and all the chemometric treatments.

Performance of calibration models without maintenance

The 6 sets of incoming samples (2019a, 2019b, 2020a, 2020b, 2021a and 2021b) were predicted with the initial prediction models. The coefficient of determination of calibration (R_C^2) and cross-validation (R_{CV}^2) and the root mean square error of calibration (RMSEC) and cross-validation (RMSECV) of the initial models were compared with the coefficient of determination of prediction (R_P^2) and the root mean square error of prediction (RMSEP) for each set of incoming samples. The percentage of spectral outliers found for each dataset was also computed but spectral outliers were not removed in order to calculate R_P^2 and RMSEP. The samples considered spectral outliers were those that had T^2 and/or Q values greater than T_{lim}^2 and/or Q_{lim}. The limits were calculated as the 95% confidence interval for $T^2$and Q based on the calibration data. It should be considered that T^2 and Q are reported in units which are sensitive to the number of samples, factors and preprocessing. Therefore, in order to compare values reported by different models and setting a standard alarm level for all the models, the statistics were normalized by dividing them by the confidence limit at 95% for each model [12]. Hence, samples with T^2 reduced and/or Q reduced values higher than 1 were considered spectral outliers.

Maintenance strategy

The set of incoming samples from the first half of 2019 (dataset 2019a) was predicted with the initial model. The spectral outliers found in the dataset were considered as indicators of variability sources not adequately represented by the initial model. The update consisted of developing a new model since the beginning including these samples in the initial dataset. This means that the selection of the pretreatment, the selection of the number of latent variables and the deletion of reference outliers was revaluated. New values of R_C^2, RMSEC, R_{CV}^2 and RMSECV

were computed. After the first update, the updated model was used to predict the next incoming samples, that is those from the second half of 2019 (dataset 2019b) and the spectral outlying samples from these period were used to update the model for second time. The procedure was repeated with the samples from 2020 (datasets 2020a and 2020b) and 2021 (datasets 2021a and 2021b). When each dataset was predicted the R_P^2, the RMSEP and the percentage of outliers were computed and compared with those provided when the models are not maintained.

In a real scenario with incoming samples, as the reference values are not known the flow followed to maintain the models would be the shown in Figure 1. This strategy is similar to the proposal by Setarehdan et al. 2002 [6] with the difference that we used the PLS score space instead of the PCA score space to detect the outliers. According to Dardenne 2010 [13] the PLS score space is better than PCA score space to study if a certain model is suitable for the predicted samples.

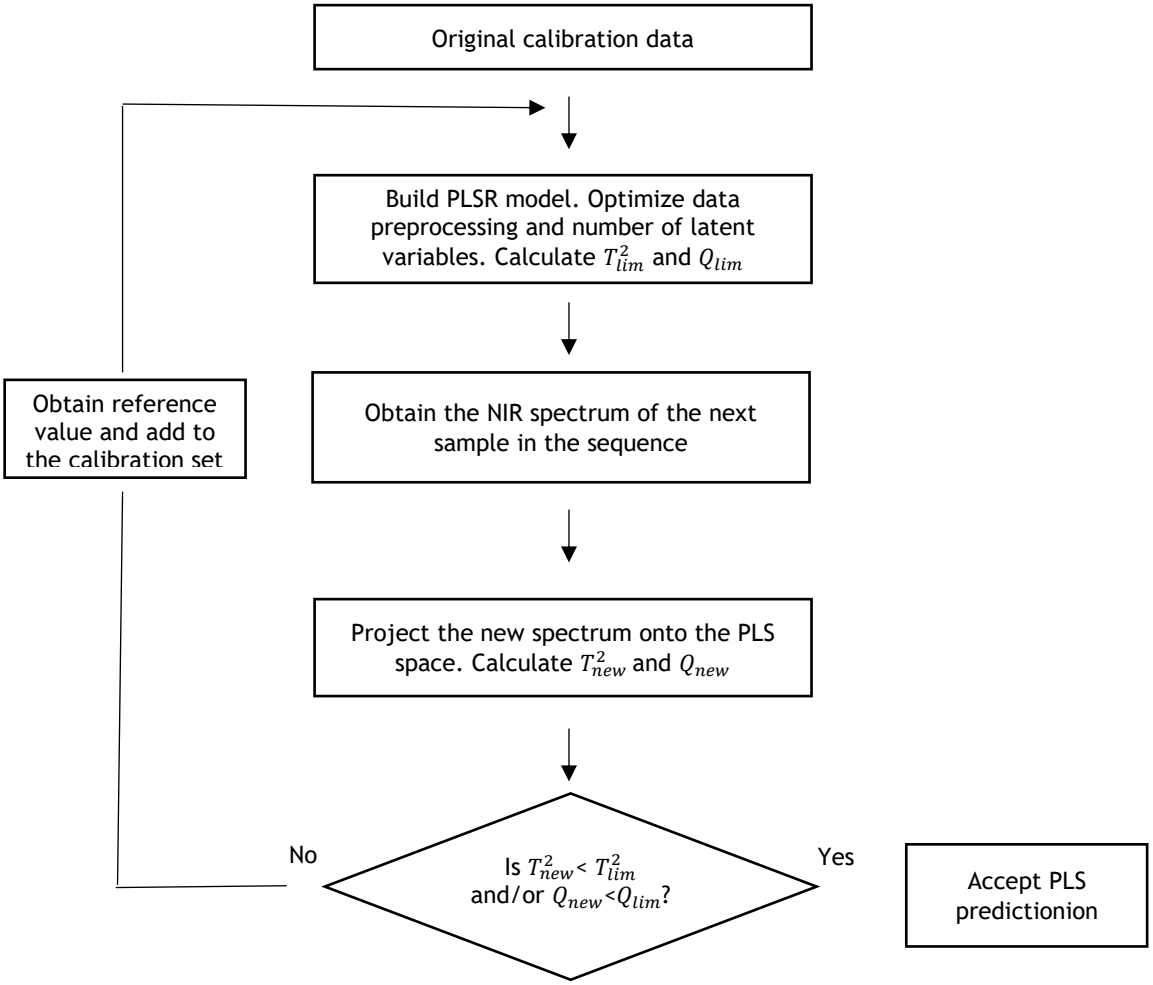


Figure 1. Block diagram of the maintenance strategy.

Results and discussion

Initial models

Table 1 shows the performance statistics of the initial PLSR models for CP, fat, CF, and ash in compound feeds. These models were developed with the samples from 2018 and included approximately 160 samples for CP, fat and ash and 70 for CF. The models involved 14 latent variables for CP, 6 for fat, 7 for CF and 16 for ash as determined by cross-validation. Overall, the models performed well attending to their high R^2_{cv} (>0.90) and the fact that the RMSECV were within two times the SEL [14].

Table 1. Characteristics of the initial and updated calibration models developed for crude protein, fat, crude fibre and ash. Initial models were developed with the 2018 dataset. The statistics of prediction are shown for the 2021b dataset. Updated models in addition to the 2018 samples include also the added samples from the datasets 2019a, 2019b, 2020a, 2020b and 2021a. Number of samples used for calibration (N_c) and prediction (N_V). Optimal pretreatment: multiplicative scatter correction (MSC), standard normal variate (SNV), derivative (d) and derivative window width in brackets. Number of latent variables (LV). Coefficient of determination of calibration (R^2_C), cross-validation (R^2_{CV}) and prediction (R^2_p). Root mean square error of calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP). Percentage of outliers in the prediction set.

	Crude protein		Fat		Crude fibre		Ash	
	Initial	Updated	Initial	Updated	Initial	Updated	Initial	Updated
N_c	161	541	160	447	66	263	154	537
N_v	263	263	230	230	86	86	235	235
Pretreatment	MSC 1d (31)	SNV 1d (15)	SNV 1d (15)	SNV 2d (15)	SNV 1d (15)	SNV 1d (15)	1d (31)	1d (31)
LV	14	16	6	12	7	8	16	16
R^2_c	0.99	0.98	0.97	0.98	0.98	0.88	0.98	0.89
RMSEC (g/kg)	2.0	3.7	2.7	2.4	2.4	4.4	1.1	2.5
R^2_{cv}	0.98	0.97	0.96	0.97	0.95	0.86	0.93	0.84
RMSECV (g/kg)	3.7	4.6	3.1	2.9	3.8	4.9	2.0	3.0
R^2_p	0.84	0.96	0.86	0.97	0.45	0.89	0.37	0.52
RMSEP (g/kg)	9.0	4.4	6.8	3.3	11.6	5.3	6.2	4.1
Outliers (%)	94	11	76	20	33	13	93	27

Performance of calibration models without maintenance

Figure 2 shows the performance of the CP model over time in the absence of maintenance. Similar behaviour was observed for the other constituents predicted (fat, CF and ash). Since the beginning (prediction of dataset 2019a) a lack of model robustness was evident. The RMSEP was three times the RMSEC and nearly two times the RMSECV although it remained below two times the SEL.

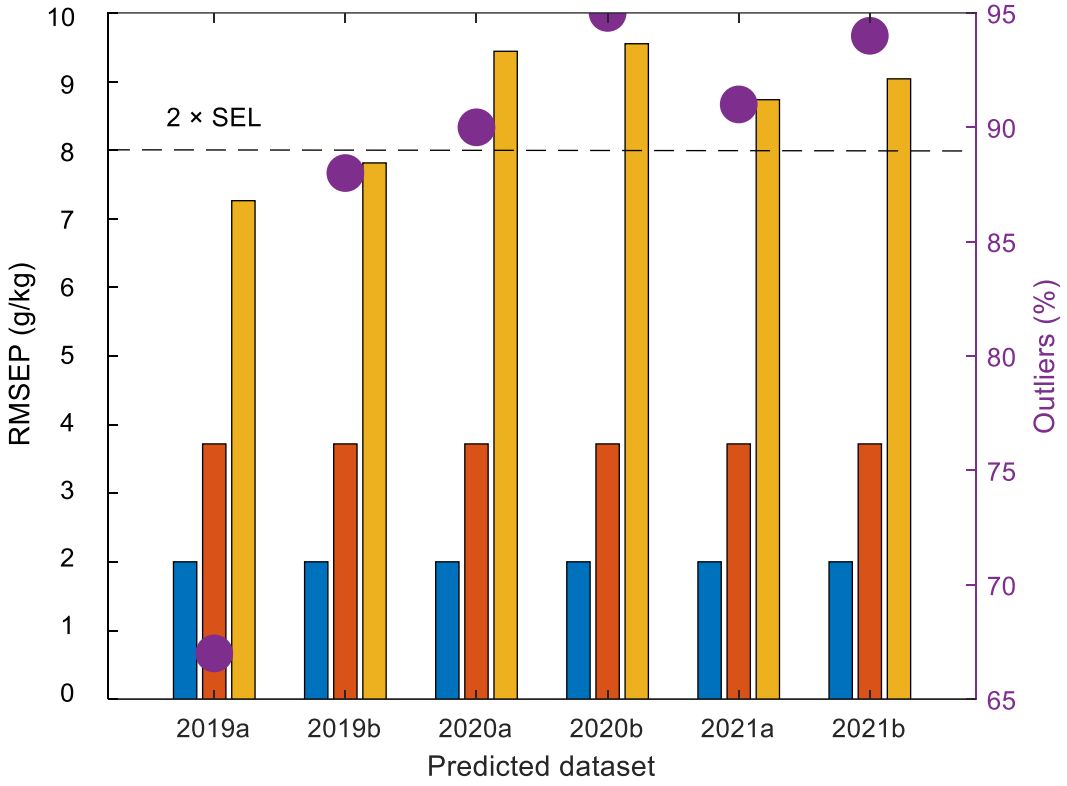


Figure 2. Performance of the CP original model with new data set recorded in different years. Root mean square error of calibration (blue bar), cross-validation (orange bar) and prediction (yellow bar). Percentage of outliers (purple circle) for the predicted dataset.

The RMSEC and the RMSECV gave an over-optimistic idea of the actual performance of the model. Although the samples are different, they may have some common aspects during a specified period. For example, compound feeds that have been made from the same batch of soybean meal, or the fact that some of the compound feed samples were ground, treated, and measured on the same day by the same operator. Or a more extreme case: compound feeds that are produced for an in-vivo assay whose objective is the evaluation of an additive and therefore are very similar in ingredients and nutrient composition. The spectral information carried by the samples is not completely independent of each other. Therefore, when cross-validation is performed, some predicted samples have a spectrum similar to that of some calibration samples, and consequently the prediction error is low. The presence of similarity among the calibration samples also impacts the robustness of the model. Although having more than 150 samples (such as CP, fat, and ash) is typically considered a sufficient amount of samples for calibration development, if many of them share information, it is more likely that future samples will introduce sources of variability unknown to the model. This is revealed by the high percentage

of samples (64%) that were identified as outliers during the prediction of the 2019a dataset because they fell outside T^2_{lim} and/or Q_{lim} (Figure 3). Logically, the RMSEP and the percentage of outliers increased over time since as time passes the samples that are produced have less in common with the calibration samples. From 2020, the RMSEP is higher than two times the SEL and therefore the model would be useless.

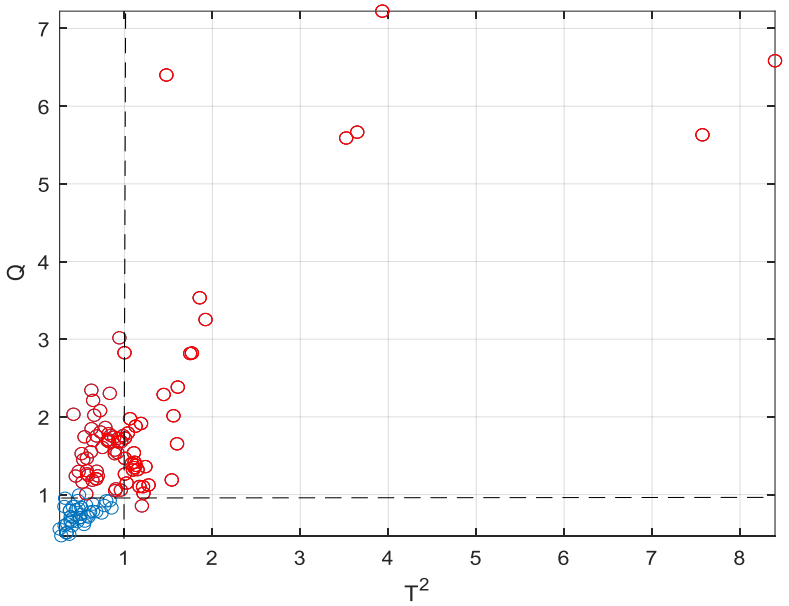


Figure 3. Q residuals reduced versus Hotelling's T^2. Samples from the 2019a dataset that stayed within the limits at 95% of confidence when they were predicted with the original model (blue). Samples that fell out (red).

Maintenance strategy

The presented strategy was applied to update the models. Each semester the models were updated by adding the samples that were considered outliers by the previous model during this period. Figure 4 shows how the CP model predicts incoming samples after successive updates. Similar behaviour was observed for the other constituents predicted (fat, CF and ash). The spectral outliers and the RMSEP decreased with the updates while the RMSEC and the RMSECV of the updated models increased. This pattern was already observed by Dardenne et al. 2000 [15] when samples from new harvests (all the available samples) were used to update PLSR models for CP, CF, starch and organic matter in plant maize samples. Each time new sources of variability are included, the PLSR model forces the spectral projection of the analyte to be orthogonal to this new variability. Therefore, the net analyte signal decreases leading to higher uncertainty [16]. At the same time, the inclusion of emerging sources of variability allows for better recognition of newer samples, resulting in fewer spectral outliers and improved

predictions for these samples. Calibration models can be considered stable when the RMSEP values are close to the corresponding SECV and SEC values as happens for CP after the last update (2021a).

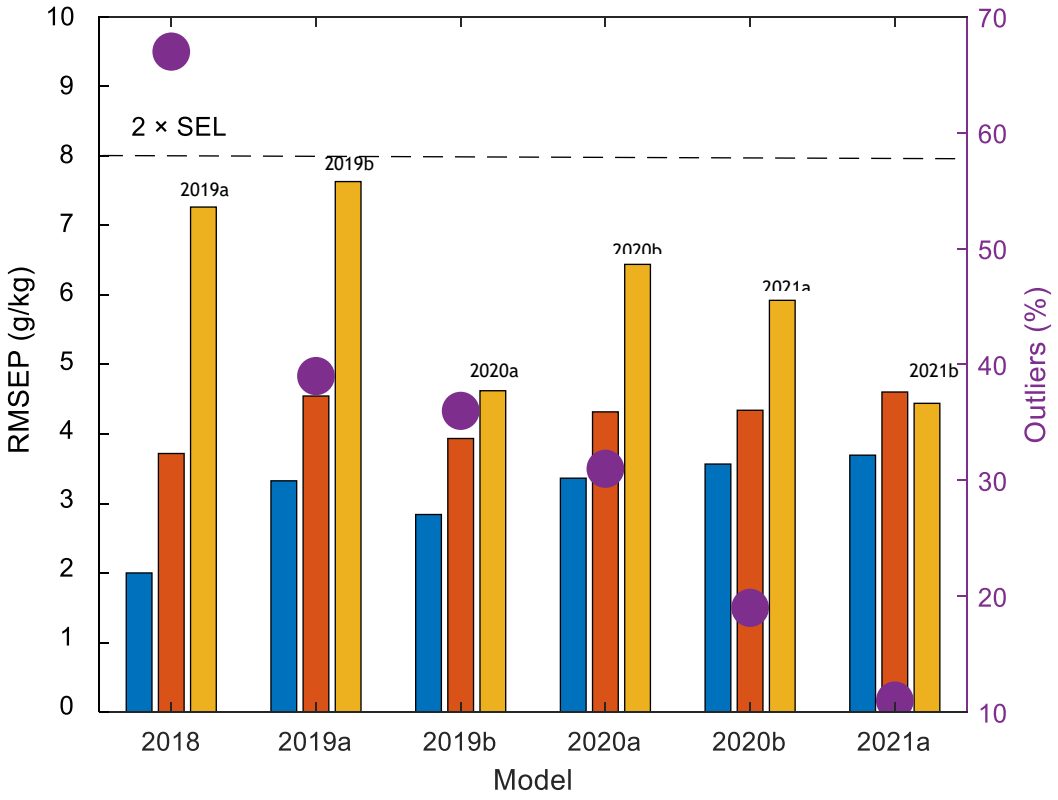


Figure 4. Performance of the successive models after following the model maintenance strategy. Root mean square error of calibration (blue bar) and cross-validation (orange bar) of the model and prediction of the test set corresponding to the next dataset (yellow bar). Percentage of outliers (purple circle) for the predicted dataset.

Figure 5.A shows the predicted CP content of dataset 2021b with the initial model and Figure 5.B with the updated model against the reference values, highlighting the importance and efficacy of following the maintenance strategy. These results agree with Schoot et al. 2021 [7] who also demonstrated that adding only the samples with new variations (those that can be considered outliers) the improvement and future performance of the updated models is similar to the performance of models updated by adding all the available samples. Therefore, the presented strategy can be considered much more cost-effective since only a reduced number of samples (and as has been shown, decreasing over time) must be analysed by reference methods and used to update the models.

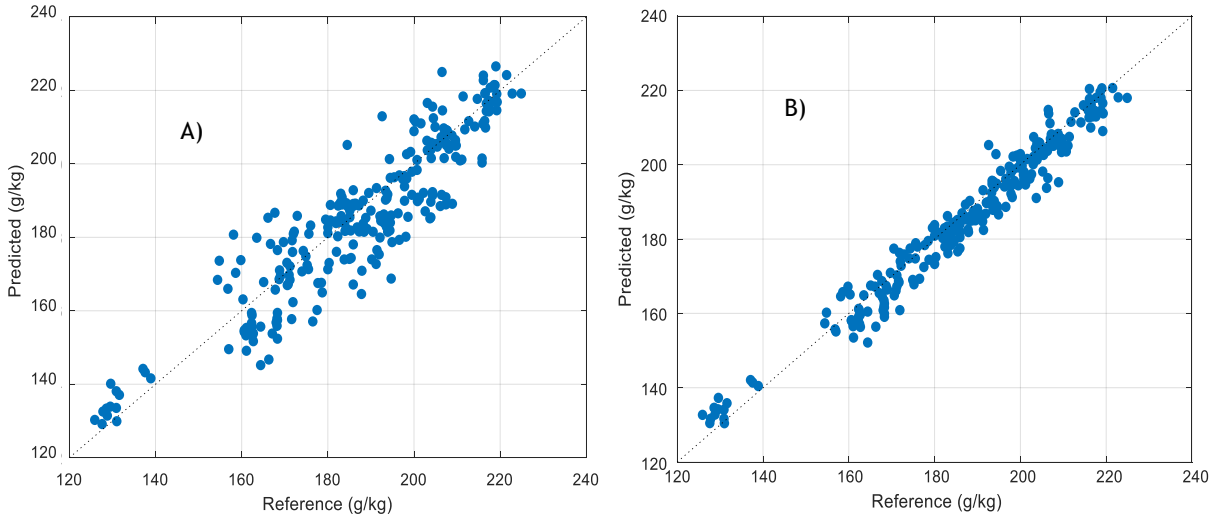


Figure 5. Predicted vs measured values of CP content for the samples of the 2021b dataset. A) Using the original model. B) Using the model updated following the model maintenance strategy.

Table 1 displays the performance of the models in the prediction of incoming samples before and after being updated following the proposed strategy. The statistics for prediction R_P^2, RMSEP and Outliers (%) were calculated for the furthest dataset (2021b). CF was the model that most improved after the successive updates since it was the model that included initially less samples and hence was less robust than the others. All the models except the one for ash become stable reaching low RMSEP values similar to the RMSEC and the RMSECV. Ash does not have characteristic NIR absorption bands and it is predicted indirectly thanks to correlation between mineral material and organic components which form salts that affect the hydrogen bonds in the samples [17]. The use of models based partially or totally in indirect relationships easily fails if the correlation structures of new samples do not match the ones exhibited by the calibration samples [18]. To obtain accurate predictions of ash the other constituents must also be present in the sample and in the same concentration ratios as in the training set. We obtained very good R^2 and RMSE for calibration and cross-validation for the initial model because the samples of 2018 were quite similar but the accuracy in the prediction of new samples was very poor. Without update, ash model predicted much worse new samples (RMSECV=2.0 g/kg and RMSEP=6.2 g/kg) than CP model (RMSECV=3.7 g/kg and RMSEP=9.0 g/kg) and fat model (RMSECV=3.1 g/kg and RMSEP=6.8 g/kg) including all of them a similar number of samples. The increase of the RMSEC (1.1 g/kg to 2.5 g/kg) and RMSECV (2.0 g/kg to 3.0 g/kg) due to the update was also more pronounced for the ash model. Samples having different correlation

structures are likely to be predicted with large errors and when included in the model, contribute to breaking the covariance between ash and organics in the calibration set samples. This provides more robustness to the model because now it is not necessary to have a rigid covariance structure to predict properly the new samples. However, at the same time the accuracy of the model is reduced considerably because precisely this correlation was exploited by the model to predict ash. It is not possible to predict an analyte if it does not absorb and no correlations exist in the calibration samples.

Another aspect that critically affects the performance of the models over time is the number of latent variables. Models with many LVs predict well samples similar to those of the calibration set but tend to be brittle, and predictions suffer considerably even when samples have minimal changes. In comparison, models with fewer LVs tend to predict less well the samples that are similar to the calibration samples but are more robust to new samples with slightly different characteristics [5]. The initial CP model involved 14 LVs while the fat model that was developed with a similar number of samples involved only 6 LVs. In the case of CP, the RMSEP for the dataset 2021b was four and a half times the RMSEC of the initial model (0.90% against 0.20%) while in the fat case, the RMSEP was less than three times the RMSEC (0.27% against 0.68%). The percentage of outliers found for each model was very different, while for CP the 94% of the samples of the 2021b dataset were detected as outliers, for fat only the 33% was detected. After the update, the model for CP involved 2 LVs more and the model of fat 6 LVs more. As the CP model initially involved many LVs, the new variance carried by the incoming samples can be partially distributed over them and not many additional LVs were needed. Note that the optimal pretreatments have also changed after the update for these two models. Scoot et al. 2020 [19] proved that a chosen pretreatment can be suboptimal after the introduction of a new variation to the calibration data.

Conclusions

In this work, it was shown how prediction models for compound feeds do not predict well samples with new sources of variability, as is the case of samples from upcoming years. The addition of these samples improved the predictive ability for new samples of the crude protein, fat, crude fibre and ash models. Over time, as the size of the calibration dataset increased and more variability was included in the models, the number of new samples with unmodelled

variations decreased resulting in fewer and fewer samples that needed to be analyzed with the reference method. With the model maintenance strategy the crude protein and fat errors obtained for new samples were always adequate. In the case of crude fibre, it took longer to develop a stable model that predicted well new samples due to the fact that the initial model very limited and had been developed with fewer samples. The model for ash, probably due to the fact that ash is predicted indirectly from the spectrum could not achieve prediction errors below the limits imposed by the laboratory during the studied period.

References

[1] F. García-Sánchez, L. Galvez-Sola, J.J. Martínez-Nicolás, R. Muelas-Domingo, M. Nieves, Using Near-Infrared Spectroscopy in Agricultural Systems, Dev. Near-Infrared Spectrosc. (2017).

[2] F. Westad, F. Marini, Validation of chemometric models - A tutorial, Anal. Chim. Acta. 893 (2015) 14–24.

[3] International Organization for Standardization (ISO), Animal feeding stuffs, cereals and milled cereal products-Guidelines for the application of near infrared spectrometry (ISO 12099). Vernier, Geneva, Switzerland, 2017.

[4] ASTM E1655-17, Standard Practices for Infrared Multivariate Quantitative Analysis, American Society for Testing and Materials. ASTM Internacional, West Conshohocken, PA, 2017.

[5] B.M. Wise, R.T. Roginski, A calibration model maintenance roadmap, IFAC-PapersOnLine. 48 (2015) 260–265.

[6] S.K. Setarehdan, J.J. Soraghan, D. Littlejohn, D.A. Sadler, Maintenance of a calibration model for near infrared spectrometry by a combined principal component analysis-partial least squares approach, Anal. Chim. Acta. 452 (2002) 35–45.

[7] M. Schoot, C. Kapper, G. van Kessel, G. Postma, L.M. Buydens, J.J. Jansen, Cost-benefit analysis of calibration model maintenance strategies for process monitoring, Anal. Chim. Acta. 1180 (2021) 338890.

[8] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[9] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[10] D. MacDougall, H. Martens, P. Geladi, Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat, Appl. Spectrosc. 39 (1985) 491–500.

[11] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[12] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, Chemometrics tutorial for PLS_Toolbos and Solo, Eigenvector Research Inc., Wenatchee, WA, USA, 2019.

[13] P. Dardenne, Some Considerations about NIR Spectroscopy: Closing Speech at NIR-2009, NIR News. 21 (2010) 8–14.

[14] G.C. Marten, J.S. Shenk, F.E. Barton II, Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, United States Department of Agriculture, Springfield, 1989.

[15] P. Dardenne, G. Sinnaeve, V. Baeten, Multivariate calibration and chemometrics for near infrared spectroscopy: Which method?, J. Near Infrared Spectrosc. 8 (2000) 229–237.

[16] J. Ferré, N.M. Faber, Net analyte signal calculation for multivariate calibration, Chemom. Intell. Lab. Syst. 69 (2003) 123–136.

[17] L. Chen, Z. Yang, L. Han, A review on the use of near-infrared spectroscopy for analyzing feed protein materials, Appl. Spectrosc. Rev. 48 (2013) 509–522.

[18] M.H. Rhiel, M.I. Amrhein, I.W. Marison, U. Von Stockar, The Influence of Correlated Calibration Samples on the Prediction Performance of Multivariate Models Based on Mid-Infrared Spectra of Animal Cell Cultures, Anal. Chem. 74 (2002) 5227–5236.

[19] M. Schoot, C. Kapper, G.H. van Kollenburg, G.J. Postma, G. van Kessel, L.M.C. Buydens, J.J. Jansen, Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size, Chemom. Intell. Lab. Syst. 204 (2020) 104105.

Supplementary material

Table S1. Descriptive statistics of the datasets for crude protein. Number of samples (N), mean, standard deviation (SD), minimum (Min) and maximum (Max).

	N	Mean (g/kg)	SD (g/kg)	Min (g/kg)	Max (g/kg)
2018	161	199	25	138	272
2019a	123	211	23	170	279
2019b	124	198	28	123	253
2020a	272	199	23	137	271
2020b	274	196	27	126	279
2021a	262	192	23	126	282
2021b	263	187	22	126	225

Table S2. Descriptive statistics of the datasets for fat. Number of samples (N), mean, standard deviation (SD), minimum (Min) and maximum (Max).

	N	Mean (g/kg)	SD (g/kg)	Min (g/kg)	Max (g/kg)
2018	160	57	15	18	85
2019a	119	58	15	32	98
2019b	120	55	18	18	102
2020a	231	58	15	19	82
2020b	232	57	14	21	100
2021a	228	60	14	22	101
2021b	230	50	12	20	77

Table S3. Descriptive statistics of the datasets for crude fibre. Number of samples (N), mean, standard deviation (SD), minimum (Min) and maximum (Max).

	N	Mean (g/kg)	SD (g/kg)	Min (g/kg)	Max (g/kg)
2018	66	36	17	22	75
2019a	41	32	6	19	44
2019b	43	42	12	23	63
2020a	84	40	14	21	77
2020b	86	35	11	20	77
2021a	85	36	10	24	75
2021b	86	37	6	24	54

Table S4. Descriptive statistics of the datasets for ash. Number of samples (N), mean, standard deviation (SD), minimum (Min) and maximum (Max).

	N	Mean (g/kg)	SD (g/kg)	Min (g/kg)	Max (g/kg)
2018	124	56	8	40	78
2019a	112	55	7	44	74
2019b	114	54	4	45	63
2020a	217	51	5	33	66
2020b	218	53	7	38	74
2021a	233	62	6	40	77
2021b	235	48	5	38	62

4.3. Selection of reference samples for updating multivariate calibration models used in the analysis of pig faeces

Andrés Cruz-Conesa[1,2], Joan Ferré[1*], Itziar Ruisánchez[1], Anna M. Pérez-Vendrell[2]

[1] Universitat Rovira i Virgili, Faculty of Chemistry, Department of Analytical and Organic Chemistry, Tarragona, Spain

*Corresponding author email address: joan.ferre@urv.cat

[2] Institute of Agrifood Research and Technology, Animal Nutrition. Mas Bové, Constantí, Spain

Chemometrics and Intelligent Laboratory Systems 234 (2023) 104749 (Edited for format)

Abstract

Monitoring and updating calibration models are common tasks when analytical methods are based on near-infrared spectroscopy. This work describes a situation in which a PLS calibration model that is used routinely for the determination of phosphorus content in pig faeces in digestibility studies had to be updated in order to be used with the faeces collected in a new trial with phytases. An approach based on D-optimality is presented that selects a reduced number of the new samples to be analysed with the reference analytical method so that the small set is used to confirm the need to update the model and validate it. The rest of the new samples that had not been selected by the algorithm were accurately predicted with the updated model. The updated model maintained its previous performance for the samples in the validation set (an RMSEP of 1.58 g kg^{-1} compared with an RMSEP of 1.54 g kg^{-1} before the update) and the prediction error for the new samples was RMSECV = 1.95 g kg^{-1}, much lower than the RMSEP = 11.38 g kg^{-1} obtained before the model update. In addition, the predictive ability of the updated PLS model was significantly better than updated models selecting the reduced dataset using other sample selection methods such as Kennard-Stone, a leverage-based selection method and random selection.

Keywords

Sample selection, model updating, near-infrared spectroscopy, D-optimal, phosphorus, PLS

Introduction

Animal nutrition research is a wide field aimed at the efficient and sustainable production of food. A large branch of animal nutrition research is devoted to finding optimal formulations for the diets of farm animals at the different growth stages and understanding how the ingredients interact and enhance nutrient digestibility. This valuable information is obtained in *in-vivo* trials in which ingredients, feeds and faeces must be analysed. Over the years, the traditional time-consuming analytical methods used to analyse these samples have been replaced by rapid, reagent-free, waste-free determinations based on near-infrared spectroscopy (NIRS) and multivariate calibration. These models have been shown to predict accurately the nutritional content of a variety of ingredients [1-4], feeds [5,6] and faeces [7,8] so NIRS is now widely used, not only in animal research but also as a routine analytical technique and legal feed labelling by feed producers in the agri-food sector.

To provide accurate predictions, NIRS-based models must be trained on representative samples for which the reference parameters have been determined with validated analytical methods, usually official methods. This is the longest part of method development since it involves collecting and analysing samples from different sources, over long periods of time until all probable future sources of spectral variations have been taken into account. In the feed-production sector, for example, this implies including different raw materials from different origins, from various harvests and stored in a variety of conditions. Since model predictions are only reliable for samples obtained under the same conditions (within the limits) as the samples from the calibration set, the performance of these models must be monitored to uncover unmodelled spectral variability. If it exists, then the model must be adapted to the new situation. To do that, a lot of calibration transfer or domain adaptation methods have been proposed in the literature [9,10]. Among them, the present study focuses on model update with new samples to expand the domain [10,11]. How often an update is needed depends on how universal the model is. It is less frequent in feed production facilities, which use stable sources of raw materials, than in animal research studies, where new combinations of ingredients are regularly

being tested. This work describes an example of the latter, in which NIRS-based multivariate models are used to predict organic matter, crude protein, fat, neutral detergent fibre, acid detergent fibre and phosphorus in pig faeces during digestibility studies carried out at the Institute of Agrifood Research and Technology (IRTA) in Constantí, Tarragona, Spain. The composition of the studied faeces depends on the weight, age, and genetic background of the animal [12] but especially on the type and digestibility of the diet. Therefore, it is not uncommon for faeces spectra from a new digestibility trial to show variations not recognized by the current model. If that were the case, the model would produce unreliable predictions for these samples and therefore they would have to be analysed with the slower analytical method. Alternatively, one wishes not to analyse the whole batch of samples but just a reduced representative number of them. This subset would be then used to update the calibration model so that the new model can be used to predict the rest of the batch as well as all future samples of the same type.

The selection of representative samples is a recurrent topic in the NIRS literature, ranging from dividing a dataset into training, validation and test sets [13], to selecting a subset of samples for model transfer between instruments or in new conditions [14], as well as the selection of samples for updating running models [11]. This work focuses on the latter case. The simplest selection method in model updating is to randomly select samples from the new batch. While this method is statistically sound, the main drawback is that the selected subset may not expand to the limits of the new spectral domain. Hence, some of the samples to be predicted may still appear extreme to the updated model, leading to extrapolation problems [15]. To better ensure the representativeness of the selected subset and the coverage of the spectral domain, the selection can be made with specific algorithms.

Algorithms can be grouped into those that make use of the spectra (X) and the reference values (y) and those that only use the spectra. Sample set partitioning based on the joint X-y distances (SPXY) is an example of the first group [16]. This algorithm has been shown to select more representative subsets than those based solely on the spectra [17]. Other examples are the successive projections algorithm (SPA) [18] and the reference value (YR)-based sample selection algorithm [19]. The main drawback of these algorithms is that they require analysing all the new samples with the reference method. This means that the updated model will only be useful for future samples but the batch that just arrived cannot benefit from the update. A more interesting approach is to select samples based on the spectra only, which is particularly useful when the new batch is large. Once the outlier diagnostics have warned about the likely inaccuracy of the

predictions, a few of these samples are selected based on their spectra, are analysed by the reference method and are used to update the model. The updated model can be then used to predict the rest of the batch. A variety of algorithms can select the samples based solely on the spectra. Two popular ones are the Kennard-Stone [20] and the duplex [21] algorithms that try to uniformly cover the multidimensional spectral space by selecting the samples with the maximum distance (commonly Euclidean distance or Mahalanobis distance) between the selected samples. Other algorithms are those based on leverage or Mahalanobis distance [22,23]. More recently, Xu et al. [24] proposed to use the simple interval calculation (SIC) leverage as the criterion and Chen et al. [25] developed an algorithm based on isolation forests for outlier detection and subset selection (IOS).

The cited algorithms rank the samples by their importance but there is no criterion that indicates what is a sufficient number of samples. An optimal subset size can be provided by criteria from the field of optimal design of experiments. To decide optimal sets of experimental conditions, criteria such as the D-criterion, the G-efficiency criterion and the A-criterion [26] provide optimal sets on the basis of Multiple Linear Regression (MLR) models. Ferré and Rius used the D-optimality criterion to select samples based on their spectra for different types of spectroscopies [27,28]. In their work, the samples were selected from an initial set for building a model, but not to update an existing one.

This work presents as a case study the selection of samples for the updating of a partial least squares (PLS) model for the prediction of phosphorus content in pig faeces from their NIR spectra. The need arose when the spectra of the pig faeces from a new digestibility trial were flagged as outliers for the running model. Thus, a reduced subset of samples from the new trial was selected, analysed with the reference method and used to update the model. The sample selection approach is inspired by the D-optimality criterion used in optimal experimental design. Its performance is compared with that of random selection, the Kennard-Stone algorithm and based on the leverage.

Materials and methods

Samples

The existing model for phosphorus content in pig faeces was built with pig faeces samples collected during digestibility studies from 2018 to 2020 at the Institute of Agrifood Research and Technology (IRTA) in Constantí, Spain. The faeces samples were lyophilized, ground and stored in sealed bags in the refrigerator until analysis. Phosphorus content (expressed as g kg^{-1} related to raw product) was determined by UV-VIS spectroscopy using molybdovanadate reagent according to AOAC Official Method 965.17 [29]. The spectrum of approximately 30 grams of sample was measured on a NIRS DS2500 (Foss NIR Systems, Denmark) with a 7 cm diameter cup in reflectance mode from 800 to 2499.5 nm every 0.5 nm. The data set was randomly divided into 246 training samples and 83 validation samples that were used to develop the model running in the laboratory. In 2021, a new batch of 103 samples was collected in a digestibility study that investigated the efficacy of different phytases on the performance of weaned piglets fed with a complex diet based on wheat-corn and soybean meal. The determination of phosphorus in these samples from their spectra using the existing calibration model produced outlier detection warnings and hence, an update of the current model was required. For this process, the phosphorus content in some selected samples of the new batch was needed. This was found with the reference method for phosphorus content mentioned above.

Data analysis

Partial least squares regression (PLSR) was used to develop the calibration model for phosphorus content. The spectra were pretreated with Savitzky–Golay first derivative interpolating with a second-order polynomial and a window width of 17 points [30] and mean-centered. A five-fold venetian blind cross validation was used to choose the optimal number of latent variables (LVs) for the model. PLS toolbox software (2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab R2020a (The MathWorks Inc., Natick, MA, USA) was used to develop the calibration models. The Fedorov algorithm used for sample selection was programmed in-house.

Selection of samples for model updating

D-optimality is a numerical criterion used in the field of optimal experimental design that defines the quality of an experimental design. Standard experimental designs, such as the full factorial design, are D-optimal. D-optimality is also used to create experimental designs when the experimental domain is irregular, in which case an algorithm such as Fedorov's algorithm [31] or genetic algorithms [32] is used to choose from a list of candidate examples which ones should be selected, under the criterion that they should maximize the determinant of the information matrix in an MLR model. In the context of multivariate calibration, this idea was used to select subsets of calibration samples to obtain models using fewer samples. It was found to perform as well as or even better than Kennard-Stone algorithm or random sampling [28,33,34]. Different from the Kennard–Stone algorithm, which seeks to maximize the distance between the selected samples to uniformly cover the calibration domain, the samples selected with the D-optimal criterion tend to be located at the edges of the domain in lineal models where the most influential samples are found.D-optimality is also used in experimental design to repair designs when some of the planned experiments cannot be executed, or the domain must be extended. In that case, the algorithm searches which examples from a list of candidates should be added to the existing ones in order to maximize the determinant of the information matrix. This use of D-optimality is exploited in this work, where a procedure using Fedorov's algorithm is presented. This algorithm will be used to find which samples from an external trial should be added to the existing calibration set to update the original PLS model.

For a working PLS model, let $\mathbf{T}$ be the $N \times P$ matrix of scores of the calibration samples where each row corresponds to a calibration sample and each column corresponds to a latent variable of the model. Let $\mathbf{T_B}$ be $I \times P$ matrix of the scores of the new batch of samples projected onto the latent variable space of the current model. These samples are candidates to being analysed and used for model updating. The algorithm starts by creating the matrix $\mathbf{T}_n$ by randomly selecting n rows of $\mathbf{T_B}$. Then, the determinant $\mathrm{Det}(\mathbf{T_E^T T_E})$ is evaluated, where $\mathbf{T_E}$ is the matrix $\begin{bmatrix}\mathbf{T}\\ \mathbf{T}_n\end{bmatrix}$ and T indicates transpose. Next, one of the rows of $\mathbf{T}_n$ is exchanged for one of the remaining rows of $\mathbf{T_B}$ so that the determinant increases as much as possible. The exchanged rows are decided following Fedorov's algorithm to find D-optimal subsets and can be found elsewhere [26-28]. This step is repeated iteratively until the determinant no longer improves. Since the algorithm can find local maxima, it can be restarted multiple times with a new random

set of samples $\mathbf{T}_n$ and the set of n samples with the largest determinant is kept as optimal. The whole procedure is then repeated to select subsets with a different number of samples n and the one with a maximum $\mathrm{Det}\left(\frac{\mathbf{T_E^T T_E}}{n+N}\right)^{\frac{1}{P}}$, which is a measure of the information content per sample, is finally selected as the optimal subset to be used for updating the model.

Results and discussion

Detecting the need for model updating

At IRTA, a PLSR model was used to predict the phosphorus content in faeces. Table 1 shows the performance measures for this model. The model involved 14 LVs as determined by cross-validation. This relatively high number was attributed to the fact that the training set included faeces from diverse research studies from 2018 to 2020, which involved different diets fed to pigs of different ages, sexes and weights. Also notice that different from other nutrients such as protein or fat, phosphorus does not present specific absorption bands in the studied spectral range. Its prediction is possible thanks to the correlations between this constituent and the absorbance of some organic molecules of the sample such as phytate that it is abundant in plant materials and hence in pig diets and faeces. This makes the prediction of phosphorus usually less accurate than the predictions for main nutrients that have stronger contributions in NIRS. Overall, the model performed well with a coefficient of determination of prediction (R_p^2) of 0.94 and a root mean square error of prediction (RMSEP) of 1.54 g kg^{-1} for the validation set. Thus, the model was considered valid for routine use. The routine use of the model included checking the sum of squares of the spectral residuals (Q-residuals) and the leverage of the new samples to be predicted [35]. These are common diagnostics to flag unmodeled spectral variability in the new spectra. Q-residuals and/or leverage larger than those of the training and validation samples warn of unreliable predictions. The detected outliers can be the result of erroneous measurements but can also indicate unique samples. The analyst's decision on how to proceed next depends on knowledge about the unusual characteristics of the new samples and whether they are occasional (so they should simply be analysed with the reference method) or whether more samples of the same type are likely to arrive in the future, in which case it may be worth updating the model [36].

Table 1. Characteristics of the previous and updated calibration models developed for phosphorus content in pig faeces (g·kg^{-1}). Number of samples used for calibration (N_c) and validation (N_v), number of P704 trial samples in the D-optimal selected subset ($N_{\mathrm{D-sel}}$), number of samples used to validate the methodology ($N_{D-\mathrm{val}}$), number of latent variables (LV), coefficient of determination of calibration (R_c^2) and prediction (R_p^2), root mean square error of calibration (RMSEC) and prediction (RMSEP), bias and slope of the predicted vs measured regression line, root mean square error of prediction of the P704 samples included in the subset ($RMSEP_{\mathrm{D-sel}}$) and used to validate the methodology ($RMSEP_{\mathrm{D-val}}$).

	Previous model	Updated model
N_c	246	271
N_v	83	83
$N_{\mathrm{D-sel}}$	25	25
$N_{D-\mathrm{val}}$	78	78
LV	14	14
R_c^2	0.95	0.93
RMSEC (g·kg^{-1})	1.45	1.63
R_p^2	0.94	0.93
RMSEP (g·kg^{-1})	1.54	1.58
Bias	0.24	0.25
Slope	0.96	0.96
$RMSEP_{\mathrm{D-sel}}$ (g·kg^{-1})	11.38	1.95 [1]
$RMSEP_{\mathrm{D-val}}$ (g·kg^{-1})	10.45	1.66

[1] RMSECV calculated from the cross-validation results that considered only the prediction errors of the P704 samples in the subset.

In this work, a new batch of pig faeces came from a trial coded P704 that tested a complex diet with ingredients that had not been used before. This included rapeseed meal, rice bran and sunflower seeds, but also different phytases, which could affect the digestion of the nutrients in the feed, mainly phosphorus digestion. Therefore, it was likely that the faeces from this trial could be outliers for the current model for phosphorus content. As expected, the faeces spectra had significant differences in signal intensity from those used to train and validate the PLS model (Figure 1).

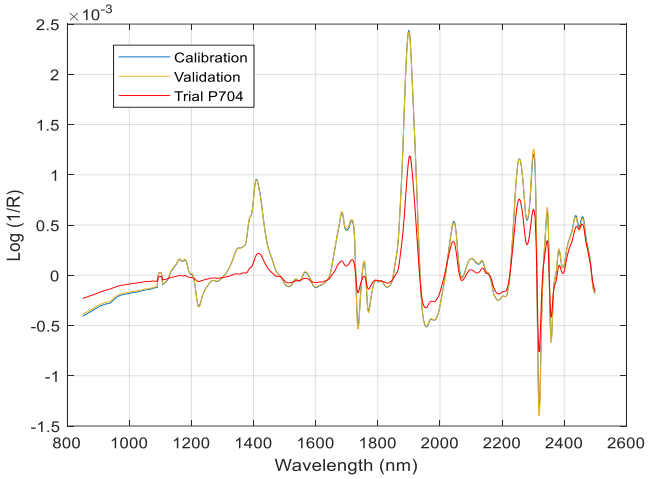


Figure 1. Mean spectrum (after 1stderivative) of the calibration and validation sets, and that of the new batch of samples.

In addition, the entire batch was positioned in the upper right quadrant of the Q-residuals versus leverage plot (Figure 2). The unmodeled variability in the new spectra could lead to large errors in the predicted phosphorus content. Note that although limits for the leverage and Q-residuals can be defined from the training and validation data [35,37], in this case visual inspection of this plot was sufficient to reveal the abnormal behavior of the new batch.

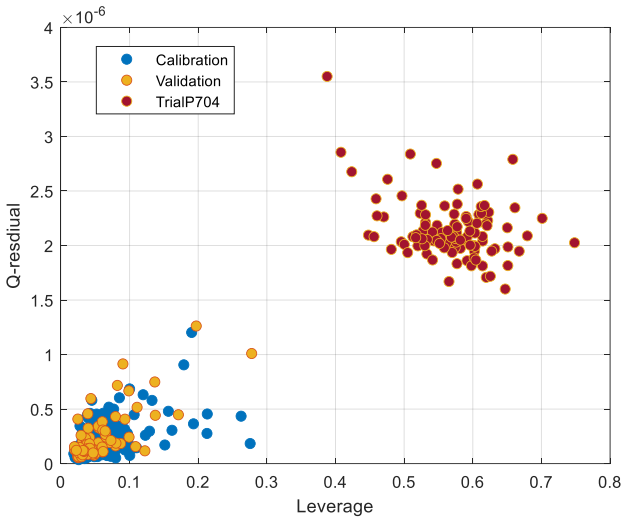


Figure 2. Q-residuals versus leverage. Calibration samples of the stablished models (blue), validation samples of the stablished models (orange) and samples of the new batch (red).

The expected large errors were confirmed by analyzing a subset of the samples of trial P704 with the reference method. As expected, (Figure 3), the errors were unacceptably large, with an RMSEP of 11.38 g kg^{-1}, much higher than the 1.54 g kg^{-1} that had been accepted for the current model.

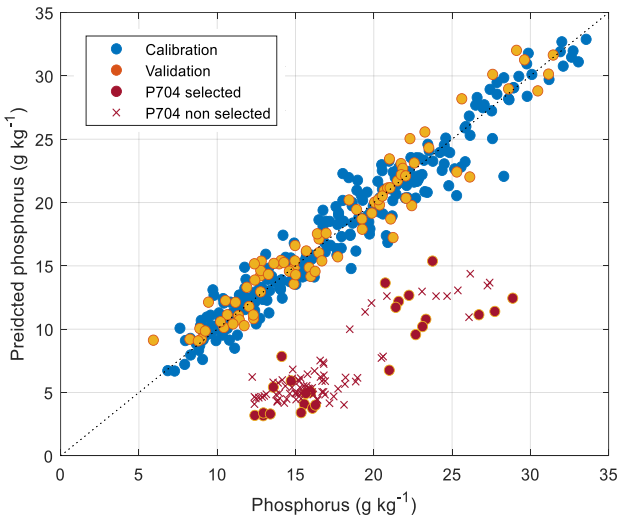


Figure 3. Predicted vs measured values of phosphorus content with the current model for the calibration set, validation set, the selected samples of Trial P704 and the non-selected samples of Trial P704.

Since more digestibility studies of the same type were expected in the future, it was more convenient to update the model than to exceptionally analyse all the new batch of samples with the reference method. Therefore, those same selected samples that had been used to confirm that the predictions errors for this batch were unacceptable, were also used to update the model. The following sections describe the selection of the subset of samples to be analysed and the validation of the updated model.

Subset selection

Once the leverage vs Q-residuals plot flagged all 103 samples of trial P704 as outliers, the model scores (**T**) and the scores of the new trial spectra ($\mathbf{T_B}$) were submitted to the selection algorithm to select which samples should be analysed with the reference method. The algorithm returned the subsets of size n =1,…103 that maximized $\mathrm{Det}\left(\frac{\mathbf{T_E^T T_E}}{n+N}\right)^{\frac{1}{P}}$. Although subsets with very low n are expected to be useless, they serve to understand the evolution of the optimization criterion when n varies. Figure 4 shows the value of this determinant against the number of samples in the subset n.

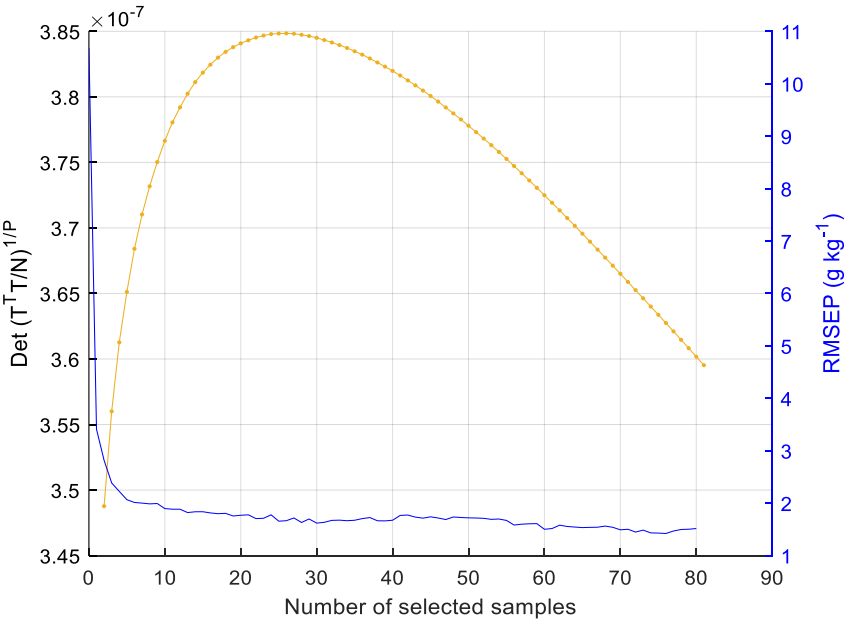


Figure 4. Number of selected samples used to update the model against the D-criterion. Root mean square error of prediction (RMSEP) of the non-selected samples of trial P704.

The large increase in the determinant on the left of the graph indicates that the subsets with low n incorporate informative samples that can improve the model. Adding more samples continues to increase the information content of the subset, but the improvement is less and less because

the newly selected samples were less unique (that is, their spectra were like the spectra of the already selected samples). A plateau is reached for subsets containing 20 to 30 samples, obtaining the maximum determinant for all subset sizes with the 25-sample subset. For subsets of more than 25 samples, the additional samples did not contribute significant new information per sample and the determinant begins to decrease. Therefore, those samples in the subset of 25 were the most informative and were analysed with the reference method. They were first used to confirm the need to update the model (as discussed in the previous section) and to update the model.

Model update and validation

The model was recalculated by adding the 25 selected samples from the trial P704 into the existing training set. The optimal number of LVs obtained by cross-validation was 14, as in the original model. The fact that the inclusion of new sources of spectral variance did not increase the number of LVs means that the variance was distributed over the many LVs of the model. A similar behavior had been observed by Capron et al. [11], who noted that only one out of four models that had been updated required an additional factor while the rest used the same number.

The validation of the updated model was as follows. Commonly, the original data sets are divided into training, validation and test sets that are used to compute the model (training set), select model parameters such as the number of factors or guide wavelength selection (validation set) and verify the actual performance of the final model (test set). When the number of samples is low, the validation set (and sometimes the test set as well) is replaced by some alternative validation method such as cross-validation. In the case of updating a model, the validation of the model with an independent set of samples would require analyzing with the reference method not only those necessary to update the model (as indicated by the selection algorithm) but also a sufficient number to verify that the model can predict the new samples correctly. However, the analysis of too many samples reduced the benefit of the presented approach and it was decided that no additional samples would be analysed in addition to those selected by the algorithm. Therefore, validation was carried out by predicting the existing validation set, to confirm that the model maintained the prediction ability for the previous samples, and by cross-validation. The common cross-validation returns an average prediction error over all the samples included in the model (in the form of the root-mean-squared error of cross-validation

RMSECV). This obscures the performance of the model for the newly included samples. To focus only on the new samples, a variation of the common RMSECV was calculated from the cross-validation results that considered only the prediction errors of the samples of the new trial. As it can be seen in Table 1, the updated model maintained its previous performance for the samples in the validation set (an RMSEP of 1.58 g kg^{-1} compared with an RMSEP of 1.54 g kg^{-1} before the update) and the prediction error for the new samples was RMSECV = 1.95 g kg^{-1}, much lower than the RMSEP = 11.38 g kg^{-1} obtained before the model update. As discussed, a fairer comparison of prediction errors could be obtained by analysing extra samples of the new trial. However, this additional effort was not considered to be necessary since the improvement after the update was significant enough to conclude that the update was successful. The updated model was still valid for predicting the original type of samples and also for predicting the rest of the samples of the P704 trial. This finished the model update procedure.

Performance of the subset selection approach

Exclusively for this work, the 78 not selected samples from the P704 trial were also analysed with the reference method to study the performance of the presented approach. It should be clear that this is not part of the model update strategy but was necessary to compare the selection strategy with other options.

As it has been shown previously, the need to model was discovered from the outlier diagnostics (Figure 2) and confirmed by the prediction errors of a subset of selected samples from the P704 trial. Figure 3 shows the predicted phosphorus content with the original model for the selected subset of samples (25 red dots) but also of the unselected samples (78 red crosses) against their reference values. As expected from Figure 2, the prediction errors were large for all samples in the P704 trial and confirm the conclusions obtained from only the selected samples.

Figure 4 shows the evolution of the RMSEP of the unselected samples of P704 trial when updated models with an increasing number of optimal samples were used to predict them. The RMSEP remained high when only a few new samples were used to update the model, since the influence of these samples in the calculated model was diluted by the many others in the training set. As expected, the RMSEP decreased as more samples of the P704 trial were included in the model until it reached a stable value of 1.66 g kg^{-1} when the 25 new subset samples were added.

This means that the model finally accounted for the spectral variability in the new samples and was able to accurately predict the rest of the samples of the new batch. This model performance correlated well with the information content per sample that was used as a criterion to decide the optimal number of samples. The larger reduction of RMSEP occurred when the increments of the determinant were large which happens with the small size subsets. The RMSEP then kept decreasing as the subset sizes increased while the determinant increased, and the stable RMSEP was reached for the subset with the maximum determinant, at 25 samples. This behavior suggested that the D-optimal criterion helps to select informative samples to update the model.

Figure 5 shows the predictions of the training samples of the updated model, showing those that were part of the original training set, and those that were added from the selection algorithm.

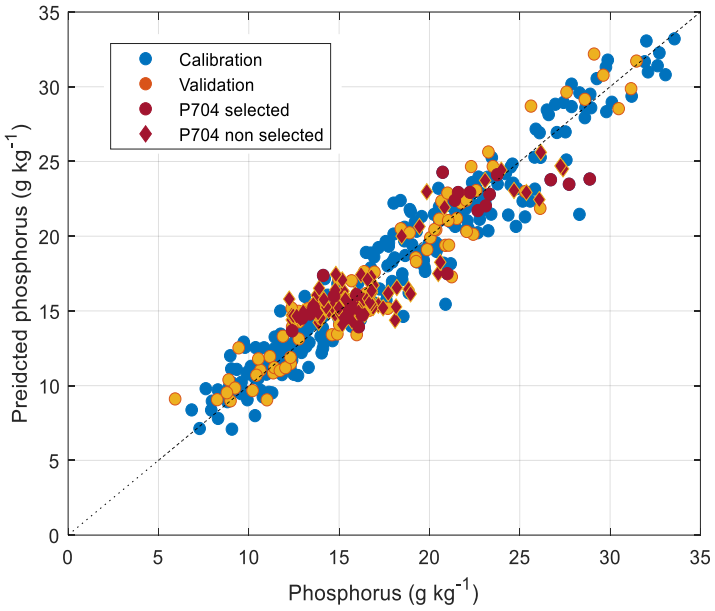


Figure 5. Predicted vs measured values of phosphorus content with the updated model for the calibration set, validation set, the selected samples of Trial P704 and the non-selected samples of Trial P704.

It can be observed that the selected samples, which were predicted with large errors before the update, are now predicted with low errors. This was to be expected, as these samples have now been used in the training step. The most relevant results are those of the unselected samples (red diamonds). As Table 1 shows, these samples had large prediction errors in the original model (RMSEP of 10.45 g kg^{-1}) although this value would never actually be known in the proposed approach because the reference values of these samples would not be known. Now they are predicted well, with an RMSEP of 1.66 g kg^{-1}, similar to the 1.95 g kg^{-1} value obtained by cross-validation of the updated model and lower than the maximum accepted error for this

model (2 g kg^{-1}). The prediction of the samples that had not been used for the model update was one of the objectives of the update.

Figure 6 compares the effectiveness of the proposed approach with other sample selection methods, namely random selection, the Kennard-Stone algorithm and the selection of the samples with the highest leverage, fixing to 25 the number of samples selected by each method.

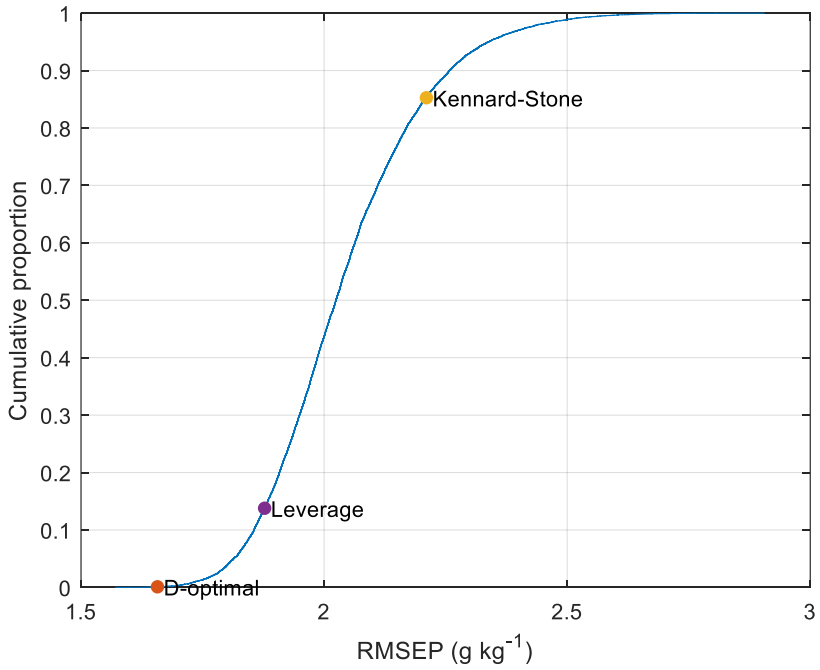


Figure 6. Cumulative distribution of the RMSEP of 10000 phosphorus models updated using a subset of samples selected at random. The RMSEP obtained selecting the samples by the D-optimal criterion, Kennard-Stone and sorted leverage are indicated.

To select the samples randomly, it must be considered that any selected subset of 25 samples out one of the many possible subsets (5.6×10^{23} combinations) that can be created from 103 samples. The performance of these subsets can be ranked to see how one approach compares to the others. For this purpose, 10.000 models were calculated with 25 randomly selected samples from the P704 trial, with the number of LVs selected by cross-validation, and the RMSEP of the unselected samples was plotted as a cumulative distribution in Figure 6. On the curve one can read the proportion of models that gave an RMSEP less than or equal to a given value. It is seen that random sampling can provide updated models with an RMSEP as low as 1.57 g kg^{-1} but also as high as 2.91 g kg^{-1}. Considering that the acceptable RMSEP for the determination of phosphorus content in the digestion studies is 2 g kg^{-1}, it is seen that the effort of updating a model with 25 random samples can sometimes be useless and can end-up with

updated models that predict the remaining samples very poorly (although better than without updating the model). To ensure the efficiency of the experimental effort, a criterion must be sought to guide the selection of the subset so that the RMSEP is as low as possible. The proposed algorithm based on the D-optimality criterion selected a subset whose updated model was better than 99.5 % of the randomly subsets selected. The subset selected by sorting the leverage also performed better than most random selections but worse than the D-criterion. In this dataset, the Kennard-Stone algorithm performed the worst. We could not find an explanation for such a bad performance, and the usual performance of the Kennard Stone algorithm found with models of other properties (results not shown) is usually better than the average of random sampling, although still worse than that of D-optimality.

Conclusions

A new sample selection algorithm inspired by the D-optimality criterion was successfully used to update a functional PLS model that predicts the phosphorus content of pig faeces from their NIR spectra. The selection algorithm used only the information already available, that is, the complete set of spectra from a new batch. Those spectra had already been measured since the samples were intended to be predicted by the current model. Once outlier detection diagnostics had shown the failure of the model for these samples, Fedorov's algorithm was used to select a subset of the batch that was used to update the model. The new model was validated by cross-validation. The optimal selection of additional samples to be used as a test set to validate the updated model was not considered and is the subject of current research. The results showed that the predictive ability of the updated model was significantly better than the prediction ability of other updated models after selecting the subset by other sample selection methods such as random selection, Kennard-Stone and leverage-based selection.

Acknowledgments

A. Cruz-Conesa acknowledges Universitat Rovira i Virgili (Tarragona, Spain) and the Institute of Agrifood Research and Technology (Constantí, Spain) for providing a Marti Franqués Research Fellowship (2019PMF-PIPF-62). The authors thank the various IRTA researchers that made the faeces samples available for this study, and the laboratory personnel for providing the reference values.

References

[1] P.C. Garnsworthy, J. Wiseman, K. Fegeros, Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy, J. Agric. Sci. 135 (2000) 409–417.

[2] D. Cozzolino, A. Fassio, E. Fernández, E. Restaino, A. La Manna, Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 129 (2006) 329–336.

[3] M.R. Akkaya, Prediction of fatty acid composition of sunflower seeds by near-infrared reflectance spectroscopy, J. Food Sci. Technol. 55 (2018) 2318–2325.

[4] B. Carbas, N. Machado, D. Oppolzer, L. Ferreira, C. Brites, E.A.S. Rosa, A.I.R.N.A. Barros, Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for the determination of nutritional and antinutritional parameters in common beans, Food Chem. 306 (2020) 125509.

[5] D.C. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero-Ginel, A. Gómez-Cabrera, Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: Chemical composition and open-declaration, Anim. Feed Sci. Technol. 116 (2004) 333–349.

[6] E. Fernández-Ahumada, J.E. Guerrero-Ginel, D. Pérez-Marín, A. Garrido-Varo, Near infrared spectroscopy for control of the compound-feed manufacturing process: Mixing stage, J. Near Infrared Spectrosc. 16 (2008) 285–290.

[7] L. Paternostre, V. Baeten, B. Ampe, S. Millet, J. De Boever, The usefulness of NIRS calibrations based on feed and feces spectra to predict nutrient content, digestibility and net energy of pig feeds, Anim. Feed Sci. Technol. 281 (2021) 115091.

[8] A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, M.P. Callao, I. Ruisánchez, Use of visible-near infrared spectroscopy to predict nutrient composition of poultry excreta, Anim. Feed Sci. Technol. 283 (2022) 115169.

[9] J.J. Workman, A Review of Calibration Transfer Practices and Instrument Differences in Spectroscopy, Appl. Spectrosc. 72 (2018) 340–365.

[10] P. Shahbazikhah, J.H. Kalivas, A consensus modeling approach to update a spectroscopic calibration, Chemom. Intell. Lab. Syst. 120 (2013) 142–153.

[11] X. Capron, B. Walczak, O.E. De Noord, D.L. Massart, Selection and weighting of samples in multivariate regression model updating, Chemom. Intell. Lab. Syst. 76 (2005) 205–214.

[12] D. Bastianelli, L. Bonnal, Y. Jaguelin-Peyraud, J. Noblet, Predicting feed digestibility from NIRS analysis of pig faeces, Animal. 9 (2015) 781–786.

[13] F. Westad, F. Marini, Validation of chemometric models - A tutorial, Anal. Chim. Acta. 893 (2015) 14–24.

[14] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: A review, Chemom. Intell. Lab. Syst. 64 (2002) 181–192.

[15] K. Rajer-Kanduč, J. Zupan, N. Majcen, Separation of data on the training and test set for modelling: A case study for modelling of five colour properties of a white pigment, Chemom. Intell. Lab. Syst. 65 (2003) 221–229.

[16] R.K.H. Galvão, M.C.U. Araujo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, A method for calibration and validation subset partitioning, Talanta. 67 (2005) 736–740.

[17] Z.D. Lin, Y.B. Wang, R.J. Wang, L.S. Wang, C.P. Lu, Z.Y. Zhang, L.T. Song, Y. Liu, Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization, J. Appl. Spectrosc. 84 (2017) 529–534.

[18] H.A. Dantas Filho, R.K. Harrop Galvão, M.C. Ugulino Araújo, E.C. Da Silva, T.C. Bezerra Saldanha, G.E. José, C. Pasquini, I.M. Raimundo, J.J. Rodrigues Rohwedder, A strategy for selecting calibration samples for multivariate modelling, Chemom. Intell. Lab. Syst. 72 (2004) 83–91.

[19] Z. He, M. Li, Z. Ma, Design of a reference value-based sample-selection method and evaluation of its prediction capability, Chemom. Intell. Lab. Syst. 148 (2015) 72–76.

[20] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics. 11 (1969) 137–148.

[21] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. 19 (1977) 415–428.

[22] G. Puchwein, Selection of Calibration Samples for Near-Infrared Spectrometry by Factor Analysis of Spectra, Anal. Chem. 60 (1988) 569–573.

[23] J.S. Shenk, M.O. Westerhaus, Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy, Crop Sci. 31 (1991) 469–474.

[24] B. Xu, Z. Wu, Z. Lin, C. Sui, X. Shi, Y. Qiao, NIR analysis for batch process of ethanol precipitation coupled with a new calibration model updating strategy, Anal. Chim. Acta. 720 (2012) 22–28.

[25] W.R. Chen, Y.H. Yun, M. Wen, H.M. Lu, Z.M. Zhang, Y.Z. Liang, Representative subset selection and outlier detection: Via isolation forest, Anal. Methods. 8 (2016) 7225–7231.

[26] P.F. de Aguiar, B. Bourguignon, M.S. Khots, D.L. Massart, R. Phan-Than-Luu, D-optimal designs, Chemom. Intell. Lab. Syst. 30 (1995) 199–210.

[27] J. Ferré, F. Xavier Rius, Selection of the best calibration sample subset for multivariate regression, Anal. Chem. 68 (1996) 1565–1571.

[28] J. Ferré, F.X. Rius, Constructing D-optimal designs from a list of candidate samples, TrAC - Trends Anal. Chem. 16 (1997) 70–73.

[29] Official Methods of Analysis of AOAC International. 20th Ed., AOAC International, Gaithersburg, MD, USA, 2016.

[30] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[31] V. Fedorov, Theory of optimal experiments, translated and edited by W.J. Studden and E.M. Klimko, (1972).

[32] A. Broudiscou, R. Leardi, R. Phan-Tan-Luu, Genetic algorithm as a tool for selection of D-optimal design, Chemom. Intell. Lab. Syst. 35 (1996) 105–116.

[33] W. Wu, B. Walczak, D.L. Massart, S. Heuerding, F. Erni, I.R. Last, K.A. Prebble, Artificial neural networks in classification of NIR spectral data: Design of the training set, Chemom. Intell. Lab. Syst. 33 (1996) 35–46.

[34] O.Y. Rodionova, A.L. Pomerantsev, Subset selection strategy, J. Chemom. 22 (2008) 674–685.

[35] ASTM E1655-17, Standard Practices for Infrared Multivariate Quantitative Analysis, American Society for Testing and Materials. ASTM Internacional, West Conshohocken, PA, 2017.

[36] S.K. Setarehdan, J.J. Soraghan, D. Littlejohn, D.A. Sadler, Maintenance of a calibration model for near infrared spectrometry by a combined principal component analysis-partial least squares approach, Anal. Chim. Acta. 452 (2002) 35–45.

[37] T.W. Anderson, An introduction to multivariate statistical analysis, Third edit, Jhon WIley & Sons, Inc., Hoboken, New Jersey, 2003.

Chapter 5.
Strategies based on NIRS to predict digestibility of monogastric animal diets

5. Strategies based on near-infrared spectroscopy to predict digestibility of monogastric animal diets

Abstract

Precise feed formulation is essential for optimizing farm animal productivity while avoiding nutritional imbalances. By formulating feeds based on parameters that reflect the digestive process of the animal, it is possible to enhance their growth without causing nutritional excesses or deficiencies. Digestible or metabolizable energy, as a driving force for animal growth, is often considered the primary factor in feed evaluation and diet cost determination. However, the digestibility of specific nutrients, such as phosphorus (P), also plays a crucial role in diet formulation. Traditionally, digestible energy and the digestibility of nutrients has been assessed through in-vivo assays, which involve feeding animals a test diet and collecting a digestion product (excreta/faces, and/or ileal digesta) for analysis. However, in-vivo assays are controversial due to the use of animals and can be expensive, partly due to the large number of samples that need to be analysed. In this work, we evaluate several strategies involving NIRS to determine apparent metabolizable energy (AME) for broilers, apparent digestible energy for pigs (ADE) and apparent ileal digestibility of phosphorus (AIDP). Five modelling strategies using partial least squares regression (PLSR) were compared: A) the use of NIR predictions coming from models for feeds and for the digestion product, B) using the feed spectra, C) using the digestion product spectra, D) combining feed and digestion product spectra and E) including the phytase concentration as an additional predictor variable in strategies B, C and D when AIDP is predicted. The coefficient of determination R^2_{CV} and the root mean standard error of leave one-out cross validation (RMSECV) were used to compare the models. Strategy A performed well (R^2_{CV} for AME=0.98, for ADE=1.00 and for AIDP=0.82) enabling the calculation of the parameters without the need for wet chemistry analysis of feed and digestion product nutrient contents. ADE for pigs was accurately predicted (R^2_{CV}=0.96, RMSECV=56 kcal/kg) from the faeces spectra (strategy C) and AIDP for broilers (R^2_{CV}=0.80, RMSECV=5.7 %) from the ileal digesta spectra modelled together with the feed phytase content (strategy E). These models simplify in-vivo assays further by eliminating the need for starvation periods, indigestible marker use, or precise measurement of the amount of feed intake and the amount of digestion product generated. AME for broilers was predicted well (R^2_{CV}=0.91, RMSECV=61

kcal/kg) from the feed spectra (strategy B) potentially allowing for the future reduction of animal usage to determine this parameter. Combination of spectra (strategy D) did not yield significant improvements for any of the studied parameters.

Introduction

Precise evaluation of feed quality holds significant importance, considering that feed constitutes two-thirds or even more of the expenses incurred in livestock production and plays a crucial role in improving performance and reducing the environmental impact resulting from this industry [1].

In the past, feedstuff formulation was conventionally based on the total nutrient content. However, modern-day practices prioritize formulations based on the available nutrient content, which considers how the animal digests the nutrients. The feed has chemical and physical characteristics that give it potential digestibility, while the animal will use it more or less efficiently depending on its species, genetic potential, age, etc.

Energy is the primary factor in feed evaluation since it ultimately drives the growth of the animal and determines the cost of the diet [2]. Digestible energy (DE) is the difference between the gross energy (GE) in the feed, defined as the total chemical energy measured upon complete combustion of the feed in a calorimeter bomb and the energy losses found in the faeces by the same method. Metabolizable energy (ME) also considers the losses in urine and gases from digestive fermentation. In poultry, since faeces and urine are excreted together, ME is the most determined energy form while DE is the most common for swine. DE and ME are also referred to as apparent DE (ADE) and apparent ME (AME), respectively [3]. In recent years and due to environmental concerns, there has been a growing interest in measuring the utilization of phosphorus (P), which is the third most expensive nutrient in poultry feed [4]. In poultry, both apparent ileal digestibility (AIDP) and apparent total tract digestibility (ATTDP) have been common methods for assessing P bioavailability [5]. Nowadays, AIDP is generally preferred because the linear relationship between dietary P and AIDP can be observed over a wider range compared to ATTDP because of the urinary P present in the excreta [6].

The use of in-vivo assays to obtain AME, ADE or AIDP is controversial due to the use of animals and expensive partly because a large number of samples must be analysed, that is the feed fed to the animal and the digestion product collected (excreta/faces, and/or ileal digesta). Therefore, it is important to establish reliable and more rapid methods to assess the nutritional quality of feed ingredients or compound feeds [2]. Near-infrared spectroscopy (NIRS) has been widely used to evaluate the nutritional quality of agricultural products [7] and to a lesser extent the nutritional quality of pig faeces [8] and poultry excreta [9].

In this work, we evaluate the use of NIRS to determine digestibility using three different datasets: 1) AME of broiler diets, 2) ADE of pig diets and 3) AIDP of broiler diets. Four modelling strategies using partial least squares regression (PLSR) were compared to predict AME of broiler diets and ADE of pig diets: A) the use of NIR predictions coming from models for feeds and for the digestion product, B) using the feed spectra, C) using the digestion product spectra and D) combining feed and digestion product spectra. Digestibility is affected by various factors beyond feed chemical composition, such as physical form, enzymes, and technological treatments. However, NIRS can only detect factors with a spectral fingerprint, such as chemical composition. The feeds in dataset 3 contained different amounts of phytase that has been proved to affect the P digestibility [10]. Therefore, an additional strategy is evaluated to predict AIDP: E) including the enzyme concentration as an additional predictor variable in strategies B), C) and D).

For the moment some studies have pointed out the potential of NIRS to predict AME for broilers and ADE for pigs from the feed spectra [11,12] (strategy B), from the the digestion product spectra [13,14] (strategy C) and by combining the spectra from feed and digestion product (strategy D) [15,16]. However, the use of NIR predictions coming from models for feeds and for the digestion product (strategy A) has not been studied and in this work more approaches to fuse the spectra of feed and digestion product are tested (strategy C). The feasibility of NIRS to predict AIDP has not been explored before, nor has the inclusion of enzyme concentration as an additional predictor variable. Hence, the prediction of AIDP by NIRS and the use of this strategy are first appearing in this work.

Materials and methods

Datasets

Three datasets were generated from in-vivo digestibility studies performed between 2018 and 2022 at the experimental farms of the Institute of Agrifood Research and Technology (IRTA, Constanti, Spain).

Dataset 1 consisted of feed and broiler excreta samples from 8 in-vivo assays, which were carried out with 2 to 12 diets resulting in a total of 56 diets whose AME values were known. Titanium dioxide (TiO_2) was added as non-digestible marker to all diets in a concentration varying from 4 to 5 g/kg to allow spot sampling of the excreta instead of total collection for estimation of digestibility. Diets were fed to male broiler chickens of the Ross 308 strain housed in cages in a room under controlled environmental conditions (temperature and ventilation rate). Overall, each diet was administered to 8 cages at the rate of two chickens per cage. Therefore, 8 excreta samples were collected per diet. At the end of the in-vivo assay, excreta samples were lyophilized, milled, and stored in sealed bags in a climatic chamber at 17 ºC until their analysis.

Dataset 2 consisted of feed and pig faeces samples from 8 in-vivo assays, which were carried out with 2-10 diets resulting in a total of 41 diets with known ADE values. Titanium dioxide (TiO_2) was added as non-digestible marker to all diets in a concentration varying from 3 to 5 g/kg. Diets were fed to growing pigs ([Large White × Landrace] × Pietrain, mixed sexes) distributed by initial weight into blocks that consisted of a certain number of pens (the same number as number of diets tested in the in-vivo assay) of 3-4 animals each. Within each block, test diets were randomly distributed among the pens. Overall, each diet was administered to 12 pens. Therefore, 12 pig faeces samples were collected per diet and lyophilized, milled, and stored in sealed bags in a climatic chamber at 17 ºC until their analysis.

Dataset 3 consisted of feed and broiler ileal digesta samples from 7 in-vivo assays, which were carried out with 6-16 diets resulting in a total of 72 diets with AIDP values. These in-vivo assays investigated the efficacy of including in the diets different amounts of phytases from different brands and companies. Hence, diets belonging to the same in-vivo assay showed low variability. TiO_2 was added as non-digestible marker to all diets in a concentration varying from 4 to 5

g/kg. Diets were fed to male broiler chickens of the Ross 308 strain housed in cages in a room under controlled environmental conditions (temperature and ventilation rate). Overall, each diet was administered to 8 cages at the rate of two-three chickens per cage. Therefore, 8 ileal digesta samples were collected per diet. At the end of the in-vivo assays, broilers were euthanized, the intestine removed, and the lower half of the ileum was gently flushed to remove the ileal contents for digestibility measurements. The ileum is the portion of the small intestine extending from the Meckel's diverticulum to a point 40 mm proximal to the ileocecal junction [17]. Ileal digesta samples were freeze-dried, equilibrated at ambient temperature for 1 day, weighed, ground and stored prior to analysis.

Reference values

Feed, excreta, faeces and ileal digesta samples were analysed in the laboratory with validated methods. Gross energy was determined by calorimetry using an adiabatic calorimeter (C2000, IKA, Staufen, Germany) according to the DIN 51900 (2005) norm [18]. Phosphorus content was determined by UV-VIS spectroscopy using molybdovanadate reagent according to AOAC Official Method 965.17 [19]. TiO_2 was determined by UV-Vis spectroscopy according to the method developed by Short et al. 1996 [20].

AME for broilers was calculated as follows:

$$AME\left(\frac{kcal}{kg}\right) = GE_{feed} - \frac{GE_{excreta} \times [TiO_2]_{feed}}{[TiO_2]_{excreta}} \quad (1)$$

where GE_{feed} is the gross energy in the feed, $GE_{excreta}$ is the gross energy in the excreta, $[TiO_2]_{feed}$ the concentration of TiO_2 in the feed and $[TiO_2]_{excreta}$ the concentration of TiO_2 in the excreta.

ADE for pigs was calculated similar to AME by means of Eq. 1 using GE_{faeces} instead of $GE_{excreta}$ and $[TiO_2]_{faeces}$ instead of $[TiO_2]_{excreta}$.

AIDP for broilers (Dataset 3) was obtained as follows:

$$AIDP(\%) = 100 \times \left(1 - \frac{[P]_{id} \times [TiO_2]_{feed}}{[P]_{feed} \times [TiO_2]_{id}}\right) \quad (2)$$

where $[P]_{feed}$ is the concentration of P in the feed, $[P]_{id}$ the concentration of P in the ileal digesta, $[TiO_2]_{feed}$ the concentration of TiO_2 in the feed and $[TiO_2]_{id}$ the concentration of TiO_2 in the ileal digesta.

The digestibility values calculated from the same feed and replicate digestion product samples should be similar although some outliers were found. Outliers were defined as values larger than three times the scaled median absolute deviation (MAD) [21]. In Dataset 1, of the 444 excreta samples, 44 outliers were removed from subsequent calculations of AME. In the dataset 2, of the 480 faeces samples, 20 were removed from the calculations of ADE. In Dataset 3, of the 439 ileal digesta samples, 15 outliers were removed from the calculations of AIDP. Most of the outliers were due to out-of-range values of TiO_2 in the digestion product samples. When a non-digestible marker is used it is assumed that the marker is completely indigestible, that it distributes evenly in the feed and in the digestion product and that can be measured accurately even at low concentrations [3]. However, sometimes some of these assumptions, mainly the evenly distribution, can fail yielding to unreliable TiO_2 values.

NIR spectra acquisition and data analysis

Ground samples of feed, excreta, faeces and ileal digesta were scanned on a NIRS DS2500 (Foss NIRSystems, Denmark) in reflectance mode in the range from 800 to 2499.5 nm with steps of 0.5 nm. The size of the cup chosen to scan the samples depended on the amount of sample available. Feed samples were scanned with the large cup (10.2 cm diameter, approx. 100 g of sample), excreta and faeces samples with the small cup (7 cm diameter, approx. 30 g of sample) and ileal digesta contents with the ring cup (1.8 cm diameter, approx. 1 g of sample). PLS toolbox software (PLS_Toolbox, 2016, Eigenvector Research, Inc., Manson, WA, USA) running in Matlab (MATLAB, Version R2020a, The MathWorks Inc., Natick, MA, USA) was used to carry out all chemometric treatments.

PCA of the mean-centred spectra was carried out after preprocessing the spectra with Savitzky–Golay first derivative with a second-order polynomial and a window width of 17 points [22].

Samples were discarded as spectral outliers when the leverage was too high (Hotelling's T^2 reduced > 3) or the percentage of residual spectral variance of the sample was too high (Q residuals reduced > 3). In Dataset 1, none of the feeds was spectral outlier and 6 excreta samples were removed from subsequent analysis. In Dataset 2, one feed sample and 15 ileal digesta samples were spectral outliers. In Dataset 3, two feed samples and 20 faeces samples were discarded.

After removing reference and spectral outliers, the spectra and the digestibility values corresponding to the same diets were averaged to obtain a single digestion product spectrum and a single digestibility value per diet. Finally, Dataset 1 contained 56 feed spectra, 56 excreta spectra and 56 AME values; Dataset 2 contained 39 feed spectra, 39 faeces spectra and 39 ADE values and Dataset 3 contained 71 feed spectra, 71 ileal digesta spectra and 71 AIDP values.

Because datasets contained a few samples (N<80) leave-one-out cross-validation was used to select model parameters, the optimal pretreatment and as validation method.

Predictions based on models for feed and for digestion product

Partial least squares regression (PLSR) models were developed to predict GE in the feed and in the excreta of Dataset 1. The predicted values of GE were used in Eq. 1 together with the reference values of TiO_2 to calculate AME. It was impossible to predict TiO_2 because this inorganic compound does not absorb in the NIR region. It cannot be determined indirectly either from the NIR spectrum because it does not interact or correlate with the organic matrix of the samples. Similarly, ADE was calculated from predicted GE values of feeds and faeces from Dataset 2. To calculate AIDP (Dataset 3), PLSR models for P were developed in feed and ileal digesta and inserted together with the TiO_2 reference values in Eq. 2.

Calibrations based on feed spectra alone

PLSR calibration models based only on the feed spectra were developed to predict AME for broilers (Dataset 1), ADE for pigs (Dataset 2) and AIDP for broilers (Dataset 3).

Calibrations based on digestion product spectra alone

PLSR calibration models were developed to predict AME for broilers from the excreta spectra (Dataset 1), ADE for pigs from the faeces spectra (Dataset 2) and AIDP for broilers from the ileal digesta spectra (Dataset 3).

Data fusion of spectra from feed and digestion product

Feed spectra and excreta spectra were combined to predict AME for broilers (Dataset 1), feed spectra and faeces to predict ADE for pigs (Dataset 2) and feed spectra and ileal digesta to predict AIDP for broilers (Dataset 3). Data fusion provides a combined response from spectra or data from different analytical techniques, but it can be also applied for a single technique to combine spectra or data of different type of samples that are related to each other. The fusion can be performed at three levels [23]. For this work, we compared several low-level and mid-level options to combine the data. In the low-level, the spectra of the different type of samples were combined sample-wise into a single matrix, from which the prediction models were developed. Three combinations of spectra were tested: A) merging the spectra of feed and digestion product, B) subtracting the spectra of the digestion product from the feed spectra and C) averaging both types of spectra. It was also studied whether pretreating both spectra separately before combining them was better than pretreating the joint spectra. Different block scaling methods were tested to properly weight the spectra blocks before combining them and developing the PLSR models [24]. Mid-level data fusion methods are not affected by the scaling problem because the relevant features from each block (e.g. the scores of PCA or PLS) are extracted and then combined to a single matrix that is used to develop the prediction models. When both, the decomposition of each block and the modelling of the combined scores matrix is done by PLSR it has been called Hierarchical PLS (HPLS) method [25]. The principal question when these methods are used is how to select the adequate number of factors in which each block is reduced. Different approaches can be used but dealing only with two blocks we decided to test all the possible combinations between 1 and 20 factors for each block of spectra.

Modelling enzyme content to improve the prediction models

The feeds and ileal digesta from Dataset 3 were obtained from studies of the effect of phytase in P digestibility. If the enzyme changes significantly the digestibility of the nutrients, then it could be interesting to include the enzyme concentration as an additional predictor variable since the very small concentration of the enzyme (20-200 ppm) would not cause sufficient changes in the NIR spectrum. The scale and size of the enzyme variable (a $n\ x\ 1$ vector) is different to the scale and size of the spectra matrix ($n\ x\ 3287$) therefore scaling methods or feature extraction were needed to give the enzyme variable an appropriate influence in the model. Similar to the fusion of spectra, low-level and mid-level strategies were tested.

Results

Figure 1 shows the mean raw spectra of the samples scanned for this work. Despite the offset, all the spectra were similar in shape and intensity for most of the bands. This was expected considering that excreta, faeces or intestinal contents are the resulting products of the digestion of the feed and hence are constituted by essentially the same components.

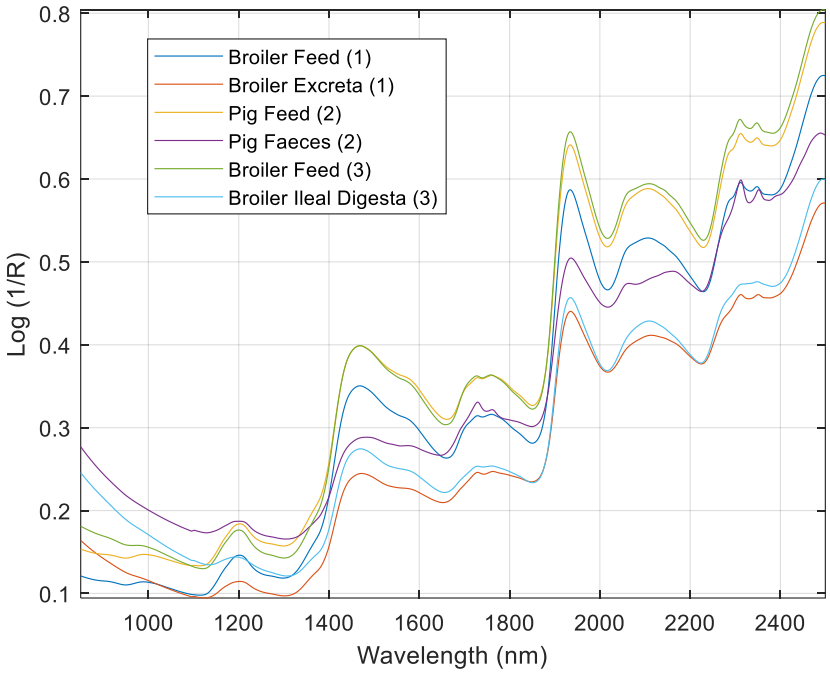


Figure 1. Mean raw spectra for the samples that conform the Dataset 1 (broiler feed and excreta), the Dataset 2 (pig feed and faeces) and the Dataset 3 (broiler feed and ileal digesta).

Variances in nutrient digestibility can either concentrate or disperse the nutrient in the digestion product, resulting in subtle variations in intensity across certain spectral bands. For instance, from the carbohydrate fraction, fibre that is mainly non-digestible by monogastric animals will be more concentrated in the digestion product than in the feed while starch that is the principal source of energy for monogastric animals will be much more concentrated in the feed comparing to the digestion product. The fact that absorbance is of the same order of magnitude in the various types of samples suggest that when fusing feed spectra with the spectra of the digestion product it may not be necessary to apply hard scaling methods prior to combine the spectra or make use of mid-level data fusion.

Dataset 1. AME for broilers

Table 1 shows that the mean values and variability of GE in feed were not significantly different from those in excreta. This could mean that the variations in the digestive capacity of broilers were minimal. On the other hand, there was a higher concentration of TiO_2 in the excreta compared to the feed leading to increased variability of this parameter in the excreta.

Table 1. Descriptive statistics of AME and the properties from which it is calculated: gross energy (GE) in the feed and in the excreta and TiO_2 in the feed and in the excreta. Mean, standard deviation (SD), coefficient of variation (CV), minimum (Min) and maximum (Max) values.

	Mean	SD	CV (%)	Min	Max
GE feed (kcal/kg)	4048	147	3.6	3504	4174
GE excreta (kcal/kg)	3631	175	4.8	3192	4074
TiO_2 feed (g/kg)	4.78	0.26	5.4	4.16	5.74
TiO_2 excreta (g/kg)	15.56	1.46	9.4	12.37	18.43
AME (kcal/kg)	2921	200	6.8	2321	3183

Table 2 shows the results of applying different strategies to predict AME. The models developed to predict GE in feed (15 LV) and in excreta (16 LV) performed well (R^2_{CV}=0.96, RMSECV=27 kcal/kg and R^2_{CV}=0.97, RMSECV=28 kcal/kg, respectively) using first derivative, mean centred spectra. The predicted GE values by these models were used together with the reference TiO_2 values to calculate AME, obtaining very good results (Table 2, Figure 2). The calibration based on feed spectra alone to predict AME performed better than the calibration based on excreta spectra. This indicates that although the composition of the excreta is feed-dependent, not all the information necessary to predict AME in the feed was present in the excreta. However, the fact that the predictions improve when both spectra are combined proved that there were sources of variability in the excreta spectra that were not present in the feed spectra. The

combination of both spectra assures that the composition of the feed is being considered as well as the way in which the animal has digested the feed. The best combination strategy consisted of pretreating the spectra separately using the first derivative and mean centring the spectra. After that, pretreated spectra were auto scaled together and averaged. The 2nd Level data fusion of spectra did not overperformed 1st Level approaches.

Table 2. Performance of the different strategies applied to predict AME (kcal/kg). Latent variables of the models (LV), coefficient of determination of calibration (R_C^2) and cross-validation (R_{CV}^2), root mean square error of calibration (RMSEC) and cross-validation (RMSEC).

Strategy	LV	R_C^2	RMSEC	R_{CV}^2	RMSECV
Calculated from NIR predictions	15 + 16	1.00	11	0.98	27
Predicted from feed spectra	14	0.98	28	0.91	61
Predicted from excreta spectra	13	0.95	46	0.82	86
1st Level data fusion of spectra	19	0.99	15	0.92	55
2nd Level data fusion of spectra	13+8	0.98	27	0.92	57

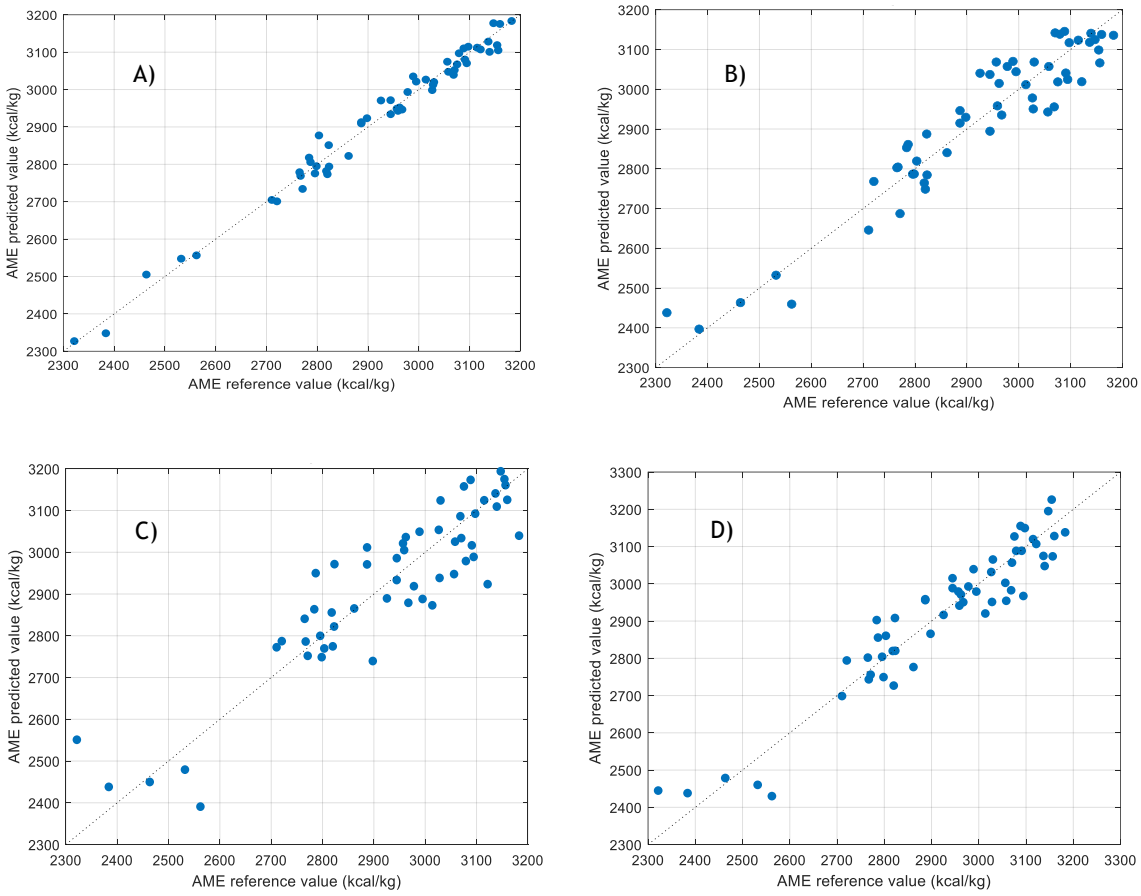


Figure 2. Predicted versus measured regression line for the following strategies to predict AME: A) the use of GE predictions coming from NIR models for feeds and excreta; B) using the feed spectra; C) using the excreta spectra and D) combining feed and excreta spectra.

Dataset 2. ADE for pigs

As Table 3 shows, the range of values and the coefficient of variation of GE in the feed were very low. The GE of the faeces and the ADE varied much more. Similar feed samples in terms of energy have led to faeces that were highly variable in energy. This suggest that in this dataset there were important digestive differences between individuals.

Table 3. Descriptive statistics of ADE and the properties from which it is calculated: gross energy (GE) in the feed and in the excreta and TiO_2 in the feed and in the excreta.

	Mean	SD	CV (%)	Min	Max
GE feed (kcal/kg)	4021	64	1.6	3855	4095
GE faeces (kcal/kg)	4198	268	6.4	3806	4731
TiO_2 feed (g/kg)	4.23	0.80	18.9	2.62	5.23
TiO_2 faeces (g/kg)	26.55	8.46	31.9	15.41	40.57
ADE (kcal/kg)	3281	276	8.4	2554	3601

The models developed to predict GE in feed (11 LV) and in faeces (13 LV) performed well (R^2_{CV}=0.91, RMSECV=11 kcal/kg and R^2_{CV}=0.98, RMSECV=42 kcal/kg, respectively) applying the first derivative and mean centring the spectra. The predicted GE values of these models were used together with the reference TiO_2 values to calculate ADE, obtaining very good results (Table 4, Figure 3). The calibration based on faeces spectra to predict ADE performed better than the calibration based on feed spectra. This is another indicator that the variability in ADE was mainly originated by differences in the digestive capacity of the pigs. Faeces reflected composition of the feed consumed by pigs as well as the physiological status of the animal. The best 1st level data fusion strategy was concatenating the spectra after being pretreated separately with the first derivative and mean centring. Combining both spectra improved the results obtained from the feed spectra alone but no from the faeces spectra alone. This means that the spectral information present in faeces predominated over the spectral information contained in feeds. The sources of variability with respect to ADE represented by the feed spectra are also represented by the faeces spectra. When the feed spectra are added to the faeces spectra, instead of including new sources of variability necessary to predict AME the relevant information is diluted. In fact, when the 2nd Level data fusion strategy was performed the best prediction was obtained using only 1 PLS factor for the feed spectra and 11 for the faeces spectra.

Table 4. Performance of the different strategies applied to predict ADE (kcal/kg).

Strategy	LV	R_C^2	RMSEC	R_{CV}^2	RMSECV
Calculated from NIR predictions	11 + 13	1.00	10	1.00	17
Predicted from feed spectra	8	0.90	65	0.90	86
Predicted from faeces spectra	12	0.99	28	0.96	56
1st Level data fusion of spectra	7	0.97	48	0.94	68
2nd Level data fusion of spectra	1+11	0.99	28	0.96	57

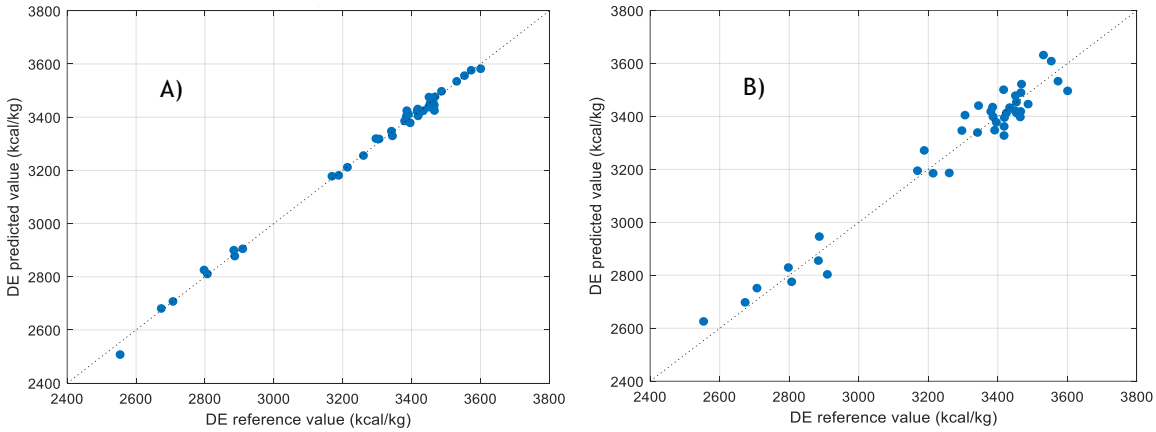


Figure 3. Predicted versus measured regression line for the following strategies to predict ADE: A) the use of GE predictions coming from NIR models for feeds and faeces; B) using the faeces spectra.

Dataset 3. AIDP for broilers

As Table 5 shows, the range of values and the variability of the P content in the feed was low comparing to those in the ileal digesta. This might be due to differences in the digestive capacity of broilers within and between in-vivo assays but it is more likely that is due to the different concentration of phytases in the feeds.

Table 5. Descriptive statistics of AIDP and the properties from which it is calculated: P in the feed and in the ileal digesta and TiO_2 in the feed and in the ileal digesta.

	Mean	SD	CV (%)	Min	Max
P feed (g/kg)	4.0	0.55	1.4	3.3	6.7
P ileal digesta (g/kg)	3.7	1.8	4.9	0.7	10.2
TiO_2 feed (g/kg)	4.71	0.31	6.6	4.1	5.6
TiO_2 faeces (g/kg)	15.0	2.0	13.3	10.9	20.2
AIDP (%)	70.6	12.5	17.7	41.0	95.6

As Figure 4 shows there was a logarithmic relationship between AIDP and phytase activity where the greatest improvements in digestibility are seen at low levels of phytase activity, and the rate of improvement slows as phytase activity increases further.

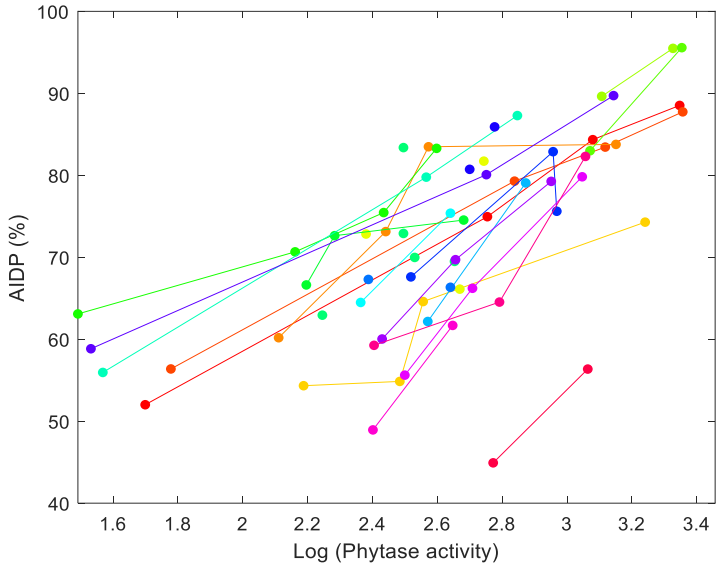


Figure 4. Variation in the AIDP of the feed when the phytase activity of the enzyme increases. Points connected by a line of the same colour correspond to feeds that were formulated to have the same nutrient content and different contents of phytase.

The model developed to predict P in feed (16 LV) showed a low coefficient of determination (R^2_{CV}=0.68) but an acceptable prediction error (RMSECV=0.33 g/kg) compared to the standard error of the laboratory for P in feed (0.2 g/kg). The model for ileal digesta (16 LV) had a higher coefficient of determination (R^2_{CV}=0.87) and a higher prediction error (RMSECV=0.67 g/kg) due to the larger variability of P in the ileal digesta. Before developing both models, the spectra were pretreated by the first derivative and mean centred. The use of the predicted P values and the reference TiO_2 values to calculate AIDP gave rise quite good predictions (R^2_{CV}=0.82, RMSECV=5.5%) considering that P is an inorganic species that is predicted indirectly by NIRS (Table 6, Figure 5).

Table 6. Performance of the different strategies applied to predict AIDP (%).

Strategy	LV	R^2_C	RMSEC	R^2_{CV}	RMSECV
Calculated from NIR predictions	16+16	0.97	2.1	0.82	5.5
Predicted from feed spectra	5	0.22	11.0	0.03	12.9
Predicted from ileal digesta spectra	11	0.89	4.2	0.73	6.5
1st Level data fusion of spectra	12	0.91	3.7	0.74	6.5
2nd Level data fusion of spectra	3+11	0.89	4.1	0.73	6.5
From feed spectra and phytase	16	0.95	2.6	0.40	10
From ileal digesta spectra and phytase	10	0.90	3.9	0.80	5.7

It was not possible to predict AIDP from the feed spectra (Table 6). This was expected since most of the diets contained similar contents of P and the largest differences in AIDP were

caused by phytase that does not produce changes in the spectra. The predictions from the ileal digesta spectra were much better although they cannot be considered well enough due to the low coefficient of determination (R^2_{CV}<0.80). The improvement by combining feed and ileal digesta spectra was negligible, highlighting the scarce information about AIDP that is found in the feed spectra. Since the phytase concentration in the feed affects AIDP values, including this variable into the feed and the ileal digesta models was studied. The best option to merge the spectra and the phytase concentration was to log10 transform the phytase concentration and then column autoscale the transformed phytase variable and the spectra. The improvement of the model based on feed spectra was notable although insufficient (R^2_{CV}=0.40). As it has been showed, feed spectra explains very little variation related to AIDP, therefore the performance of the combined model with phytase is similar to the univariate model that would result from regressing AIDP from the feed enzyme content. The model based on ileal digesta spectra improved including phytase and reached acceptable coefficient of determination (R^2_{CV}=0.80) and prediction error (RMSECV=5.7 %).

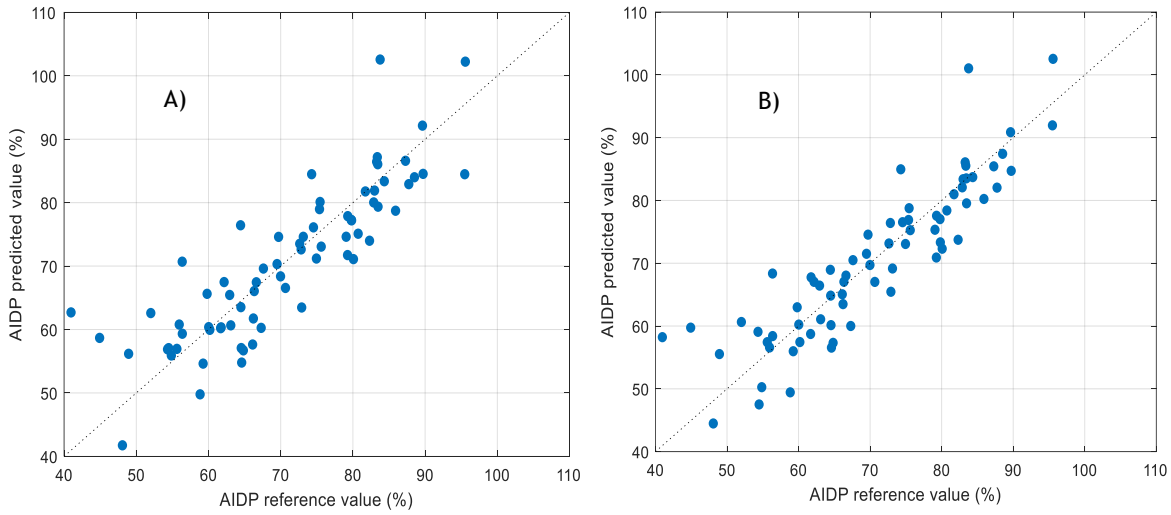


Figure 5. Predicted versus measured regression line for the following strategies to predict AIDP: A) the use of P predictions coming from NIR models for feeds and ileal digesta; B) using the ileal digesta and including the phytase concentration as an additional predictor variable.

Discussion

The feed industry requires fast and accurate methods for determining the nutritional value of diets. This is essential for reducing costs by preventing over-formulation with expensive ingredients and ensuring the quality control of the formulated diets. Different approaches have been studied to replace in-vivo assays. Most commonly, the nutrient content of diets is

calculated from tabulated values containing the chemical composition and digestibility of feed ingredients, but tabulated values might exhibit a lack of precision due to the inherent characteristics and variability of each specific batch of ingredients. Moreover, tabulated values might not exist for no so-common ingredients or for the digestibility of specific nutrients such as AIDP. Mathematical models or equations to relate the nutrient content of the feedstuffs and diets with its ADE value have been developed for pigs [26] and with its AME value for poultry [27]. The disadvantage of these models is that the nutrient content of each sample must be determined by wet chemical analysis, which may be costly or time consuming. Other approach to estimate digestibility is the in vitro experiments [2,28]. This option is cheaper and faster than in-vivo assays but still need the use of expensive equipment and reagents as well as long times of reaction. NIRS provides a considerably quicker alternative to in vitro experiments and equations based on nutrient content and generally provides more precise results than table values. NIRS and the aforementioned methods have been compared in some works [11,29,30] obtaining variable results.

From the NIR-based strategies presented in this work, only the prediction based on feed spectra alone or modelling the enzyme amount together with the feed spectra would reduce the use of animals. From the three datasets tested, only AME for broilers was accurately predicted from the feed spectra (Table 2). The model to predict AME in broiler diets performed similar to the model developed by Valdes et al. 1992 [31] and better than the one by Coulibaly et al. 2013 [16].

ADE for pigs was predicted well from the faeces spectra (Table 4). These results were better than those obtained by other authors [8,13,32]. This approach needs in-vivo assays for faeces collection, making it impractical for routine analysis in feed mills. However, it can be a valuable tool in research, enabling easy and cost-effective measurement of digestibility, which facilitates the investigation of factors that affect digestion (e.g., age, climate, and genetics) and their interplay. The measurements could be conducted on a large number of animals, eliminating the need for starvation periods, the use of an indigestible marker or precisely measure feed intake and faecal output [33].

Good AIDP predictions were obtained by modelling the enzyme content of the feed together with the ileal digesta spectra (Table 6). There are no precedents about the prediction of this parameter using NIRS. The advantages of this model would be the same as those mentioned

earlier for prediction from faeces spectra. However, since the spectra of ileal digesta must be collected, broilers should be sacrificed like in the in-vivo assay.

The combination of the spectra resulted in only a slight improvement over direct predictions from feed spectra or excreta spectra in the case of AME. For the other two datasets, there was no improvement. These results do not agree with previous works [15,16]. Coulibaly et al. 2013 [16] obtained a RMSECV value for poultry AME two times lower after concatenating the spectra of feed and excreta and Paternostre et al. 2021 [15] observed an important improvement when spectra are concatenated in the prediction of net energy in pigs. The variability in the datasets seems to govern the predictions and it is the key to obtain better results using a strategy or another. When the diets vary and the factors affecting digestion such as genetics, age, etc. are maintained (the case of AME for broilers), the predictions from the feed spectra are viable and the excreta do not contain much additional information describing the digestibility of the feed. On the contrary, when the diets show minimal variation (ADE for pigs and specially AIDP for broilers) and digestibility is principally affected by factors beyond feed chemical composition (animal physiology in the case of ADE and enzymes in the case of AIDP) the feed spectra is not suitable to predict the digestibility of the property. Much more useful information to predict digestibility was found in the digestion product spectrum than in the feed spectrum in these two particular cases.

The best predictions of AME, ADE and AIDP were obtained modelling the property from which they come from (GE for AME, and ADE and P for AIDP) in the feed and in the digestion product and using the predicted values together with the TiO_2 concentrations to calculate digestibility. During in-vivo assays, in addition to the feeds it is usual to analyse by wet chemistry methods many replicate digestion product samples per diet trying to minimize external effects such as temperature and humidity in the farm or the particular digestion differences between animals of the same species, age and gender. With the success of this strategy, the determination of GE or P in the feed and in the digestion product by wet chemistry methods would no longer needed.

Due to the limited size of the datasets (<70 samples) and consequently, the use of cross-validation to test the strategies, the obtained results could be overly optimistic and may not be generalizable. More samples and an independent test set would be necessary to draw general conclusions about what strategy one must use for each case. Furthermore, even with a larger

dataset, the effectiveness of a particular strategy could vary depending on the construction of the dataset, including factors such as feed variability, animal physiology, and enzyme usage, among others. Throughout this work, we wanted to show the different possibilities to approach digestibility prediction from NIRS using as experimental case three important properties (AME for broilers, ADE for pigs and AIDP for broilers) trying to exemplified in which situations and for which purposes a strategy, or another may fit better.

Conclusions

The use of NIRS can reduce the costs of the in-vivo assays directed to obtain AME for broilers, ADE for pigs and AIDP for broilers. It has been found that NIRS can accurately predict GE and P in the digestion product (excreta, faeces or ileal digesta) that, combined with the predictions of these parameters in the diets, makes it possible the calculation of AME, ADE and AIDP without analysing these parameters by wet chemistry methods. ADE for pigs and AIDP for broilers were well predicted from the faeces spectra and the ileal digesta modelled together with the feed phytase content, respectively. These models simplify more the in-vivo assays than the previous strategy since in addition to avoiding the use of wet chemistry methods to determine GE and P, the starvation periods, the use of an indigestible marker or precisely measure feed intake and output are no longer needed. AME for broilers was predicted well from the feed spectra, hence this is an important step towards reducing the use of animals to obtain this property in the future. Combination of spectra did not lead to significant improvements for any of the studied properties.

References

[1] L.F. Wang, R.T. Zijlstra, Prediction of bioavailable nutrients and energy, in: P.J. Moughan, W.H. Hendricks (Eds.), Feed evaluation science, Wageningen Academic Publishers, Wageningen, Netherlands, 2018: pp. 337–386.

[2] F. Zaefarian, A.J. Cowieson, K. Pontoppidan, M.R. Abdollahi, V. Ravindran, Trends in feed evaluation for poultry with emphasis on in vitro techniques, Anim. Nutr. 7 (2021) 268–281.

[3] C. Kong, O. Adeola, Evaluation of amino acid and energy utilization in feedstuff for swine and poultry diets, Asian-Australasian J. Anim. Sci. 27 (2014) 917–925.

[4] R.K. Mutucumarana, V. Ravindran, G. Ravindran, A.J. Cowieson, Measurement of true ileal digestibility of phosphorus in some feed ingredients for broiler chickens, J. Anim. Sci. 92 (2014) 5520–5529.

[5] M. Rodehutscord, A. Dieckmann, M. Witzig, Y. Shastak, A note on sampling digesta from the ileum of broilers in phosphorus digestibility studies, Poult. Sci. 91 (2012) 965–971.

[6] S.H. An, J.Y. Sung, C. Kong, Ileal digestibility and total tract retention of phosphorus in inorganic phosphates fed to broiler chickens using the direct method, Animals. 10 (2020) 1–9.

[7] V. Baeten, P. Dardenne, Application of NIR in Agriculture, in: Y. Ozaki, C. Huck, S. Tsuchikawa, S.B. Engelsen (Eds.), Near-Infrared Spectroscopy., Springer, Singapore, 2021: pp. 347–360.

[8] K.G. Nirea, M.P. de Nanclares, A. Skugor, N.K. Afseth, T.H.E. Meuwissen, J. Hansen, L.T. Mydland, M. Øverland, Assessment of fecal near-infrared spectroscopy to predict feces chemical composition and apparent total-tract digestibility of nutrients in pigs, J. Anim. Sci. 96 (2018) 2826–2837.

[9] A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, M.P. Callao, I. Ruisánchez, Use of visible-near infrared spectroscopy to predict nutrient composition of poultry excreta, Anim. Feed Sci. Technol. 283 (2022) 115169.

[10] Y. Dersjant-Li, A. Awati, H. Schulze, G. Partridge, Phytase in non-ruminant animal nutrition: A critical review on phytase activities in the gastrointestinal tract and influencing factors, J. Sci. Food Agric. 95 (2015) 878–896.

[11] B. Losada, P. García-Rebollar, C. Álvarez, P. Cachaldora, M.A. Ibáñez, J. Méndez, J.C. De Blas, The prediction of apparent metabolisable energy content of oil seeds and oil seed by-products for poultry from its chemical components, in vitro analysis or near-infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 160 (2010) 62–72.

[12] S.J. Noel, H.J.H. Jørgensen, K.E. Knudsen Bach, The use of near-infrared spectroscopy (NIRS) to determine the energy value of individual feedstuffs and mixed diets for pigs, Anim. Feed Sci. Technol. 283 (2022) 115156.

[13] D. Bastianelli, L. Bonnal, Y. Jaguelin-Peyraud, J. Noblet, Predicting feed digestibility from NIRS analysis of pig faeces, Animal. 9 (2015) 781–786.

[14] D. Bastianelli, B. Carré, S. Mignon-Grasteau, L. Bonnal, F. Davrieux, Direct prediction of energy digestibility from poultry faeces using near infrared spectroscopy, in: Proceedings of the 12th International Conference on Near Infrared Spectroscopy, IM Publications, Auckland, New Zealand, 2005: pp. 626–629.

[15] L. Paternostre, V. Baeten, B. Ampe, S. Millet, J. De Boever, The usefulness of NIRS calibrations based on feed and feces spectra to predict nutrient content, digestibility and net energy of pig feeds, Anim. Feed Sci. Technol. 281 (2021) 115091.

[16] I. Coulibaly, J.P. Métayer, C. Pascal, B. Mahaut, I. Bouvarel, P. Hogrel, D. Bastianelli, La combinaison des infromations issues des aliments et des fientes ameliore la prediction par spir de la digestibilite chez le poulet, in: 10th Poultry Research Days, La Rochelle, France, 2013: pp. 640–644.

[17] V. Ravindran, W.L. Bryden, Amino acid availability in poultry-in vitro and in vivo measurements, Aust. J. Agric. Res. 50 (1999) 889–908.

[18] DIN 51900, 2005. Testing of solid and liquid fuels-Determination of gross calorific value by the bomb calorimeter and calculation of net calorific value-Part 3. Method using adiabatic Jacket Berf., Ger., 2005.

[19] Official Methods of Analysis of AOAC International. 20th Ed., (2016).

[20] F.J. Short, P. Gorton, J. Wiseman, K.N. Boorman, Determination of titanium dioxide added as an inert marker in chicken digestibility studies, Anim. Feed Sci. Technol. 59 (1996) 215–221.

[21] P.L.S Toolbox, Q. Card, PLS_Toolbox 7.0 quick reference card, 2012.

[22] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem. 36 (1964) 1627–1639.

[23] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment - A review, Anal. Chim. Acta. 891 (2015) 1–14.

[24] M.P. Campos, M.S. Reis, Data preprocessing for multiblock modelling – A systematization with new methods, Chemom. Intell. Lab. Syst. 199 (2020) 103959.

[25] S. Wold, N. Kettaneh, K. Tjessem, Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection, J. Chemom. 10 (1996) 463–482.

[26] J.Y. Sung, B.G. Kim, Prediction equations for digestible and metabolizable energy concentrations in feed ingredients and diets for pigs based on chemical composition, Anim. Biosci. 34 (2021) 306–311.

[27] S. Barzegar, S.B. Wu, J. Noblet, R.A. Swick, Metabolizable energy of corn, soybean meal and wheat for laying hens, Poult. Sci. 98 (2019) 5876–5882.

[28] T.A. Woyengo, R. Jha, E. Beltranena, R.T. Zijlstra, In vitro digestion and fermentation characteristics of canola co-products simulate their digestion in the pig intestine, Animal. 10 (2016) 911–918.

[29] E.R. Deaville, D.J. Humphries, D.I. Givens, Whole crop cereals. 2. Prediction of apparent digestibility and energy value from in vitro digestion techniques and near infrared reflectance spectroscopy and of chemical composition by near infrared reflectance spectroscopy, Anim. Feed Sci. Technol. 149 (2009) 114–124.

[30] L. Paternostre, S. Millet, J. De Boever, Comparison of feed tables , empirical models and near-infrared spectroscopy to predict chemical composition and net energy of pelleted pig feeds, Anim. Feed Sci. Technol. 297 (2023) 115578.

[31] E.V. Valdes, S. Leeson, Near Infrared Reflectance Analysis as a Method to Measure Metabolizable Energy in Complete Poultry Feeds, Poult. Sci. 71 (1992) 1179–1187.

[32] A. Schiborra, M. Bulang, A. Berk, A. Susenbeth, E. Schlecht, Using faecal near-infrared spectroscopy (FNIRS) to estimate nutrient digestibility and chemical composition of diets and faeces of growing pigs, Anim. Feed Sci. Technol. 210 (2015) 234–242.

[33] D. Bastianelli, NIRS as a Tool to Assess Digestibility of Feeds and Feedstuffs, in: International Congress of Advanced Poultry Production on Middle East African States, World Poultry Science Assosciation (WPSA), Antalya, Turkey, 2013.

Chapter 6.
Conclusions and perspectives

6.1. Conclusions

This chapter summarizes the main findings of this doctoral thesis and puts them in perspective.

The following conclusions can be drawn from Chapter 3. *Replacing wet chemistry by NIRS to improve feed formulation and reduce the costs of in-vivo assays*:

1. It was possible to predict, from a NIR spectrum, most of the carbohydrate and lignin fractions in a large variety of feedstuffs (cereals, cereals co-products, protein-rich and fibre-rich feedstuffs) that are used to formulate monogastric diets. Predictions were accurate for sugars, starch, fructans, dietary fibre, lignin, cellulose, β-glucans, total and insoluble non-starch polysaccharides (NSP) and some of the total and insoluble non-cellulosic polysaccharide (NCP) residues (arabinose, xylose, rhamnose, galactose and uronic acids). The predictions for α-galactosides (raffinose, stachyose and verbascose) and some NCP residues (glucose and mannose) were not accurate enough. The number of samples with values of α-galactosides were very low to be representative and mannose showed low concentrations and a narrow concentration interval.

2. Multivariate calibration models for NIR spectra developed in-house performed better than commercial models in the prediction of crude protein (CP), fat, crude fibre (CF) and ash in compound feeds for poultry and swine. In our case, commercial models were not ready for the specific type of samples that are handled in research institutes such as IRTA, where the samples tested are sometimes not common, with novel combinations of ingredients. Global models that combined compound feeds for broilers, pigs, sows, laying hens and turkeys performed better than specific models of compound feeds for a single species, for the prediction of CP, fat, CF, gross energy (GE), ash and phosphorus (P). The calibrations of CP, fat and CF had excellent prediction performance when an external dataset was predicted. Those for GE and ash were less accurate and P was not predicted satisfactorily. The inorganic nature of ash and P and the fact that GE is not a chemical entity but it is a property derived from chemical composition could be the cause.

3. It was possible to predict accurately the nutrient content (organic matter, CP fat, GE, uric acid and P) in poultry excreta from its NIR spectrum. Global calibrations (those that combined excreta from broilers, hens and turkeys) performed better than specific models that were training with excreta from a single species.

4. Promising results were obtained in the prediction of organic matter, CP, GE, CF, acid detergent fibre, lignin and P in pig faeces from the NIR spectrum. Fat and neutral detergent fibre were predicted less well. The complexity of the reference analysis of these constituents could be the cause.

5. It was possible to predict well CP and several amino acids (Gly, Ala, Val, Leu, Ile, Phe, Tyr, Asp, Glu, Thr and Pro) in poultry ileal digesta from the NIR spectrum. Met, Lys and specially Ser, His, Cys and Arg were predicted poorly. These AAs exhibited weaker correlations with protein and the other AAs.

The models developed in this chapter provide evidence that NIRS can to some extent replace wet chemistry methods for the determination of the nutrient content of samples used in in-vivo assays. The substitution of wet chemistry by NIRS offers significant cost reduction in digestibility studies while enhancing the understanding of how the nutritional fractions are digested by the animal. This can enable a better formulation of monogastric diets and the development of nutritional strategies such as the use of enzymes to enhance the digestibility of nutritional fractions.

The following conclusions can be drawn from Chapter 4. *Maintenance of NIR calibration models*:

1. Prediction models for compound feeds exhibited poor performance when they were used to predict incoming samples of feeds manufactured for example, with a different combination of ingredients. Prediction models for pig faeces failed when the models were used to predict batches of samples coming from in-vivo assays where for example, new enzymes that affect the digestibility of the diets were tested.

2. The predictive accuracy of the CP, fat, CF, and ash models in compound feeds notably improved after the inclusion of compound feed samples manufactured with a different combination of ingredients or using different batches of ingredients. These samples

were identified through outlier diagnostic measures (Hotelling's T^2 and Q statistic derived from spectral residuals) and analysed by reference methods. As the calibration dataset was expanded and encompassed more variability of compound feed compositions, the occurrence of new samples with unmodelled variations diminished. Consequently, the need for testing with the reference method progressively decreased over time.

3. The implementation of a novel sample selection algorithm, inspired by the D-optimality criterion, led to a successful update of a functional PLSR model to predict P content in pig faeces. The model was updated with a reduced number of samples selected by the algorithm from a new batch of samples and was able to predict accurately the rest of the samples from the batch. Selecting the samples with the new algorithm lead to an updated model with superior predictive ability than alternative models updated by selecting samples randomly or based on Kennard-Stone algorithm or a leverage-based algorithm.

The proposed strategies offer effective means for updating models over time with minimal costs and effort. As elucidated in Chapter 4, both strategies enhance the performance of calibration models in predicting new samples by incorporating novel variations. However, the selection of samples, as it has been explained, differ between the two strategies and also their applicability. The first strategy, discussed in Section 4.2 for compound feeds, is designed for situations where new samples are introduced successively or in small batches while the second is tailored for scenarios where samples arrive in large batches.

The following conclusions can be drawn from Chapter 5. *Strategies based on near-infrared spectroscopy to predict digestibility of monogastric animal diets*.

1. It was possible to predict, from a NIR spectrum, the GE in excreta and faeces, and the P content in ileal digesta. By complementing the predictions of GE and P in diets, it becomes possible to calculate the apparent metabolizable energy (AME) in broiler diets, the apparent digestible energy (ADE) in pig diets and the apparent ileal digestibility of phosphorus (AIDP) in broiler diets without the need of conducting wet chemistry methods to determine GE in the case of AME and ADE and P in the case of AIDP.

2. Faecal spectra was highly successful to accurately predict ADE in pig diets. Additionally, the utilization of ileal digesta spectra, in combination with feed phytase content, enabled accurate prediction of AIDP in broilers diets. These models, when confirmed and extensively validated offer a simplified approach to in-vivo assays eliminating the need for conducting wet chemistry methods to determine GE in the case of ADE and P in the case of AIDP, the application of starvation periods, the use of indigestible markers, or the precise measurements of feed intake and output.

3. The feed spectrum served to accurately predict AME in broiler diets. Consequently, the utilization of animals to determine this property can be reduced in the future by employing this model, once confirmed and extensively validated.

4. The fusion of the feed and the digestion product (faeces, excreta or ileal digesta) spectra did not result in noteworthy enhancements for any of the properties (AME, ADE and AIDP) investigated.

6.2. Perspectives

Looking ahead, improvements in the models can be expected as new samples become available. This is particularly important in the calibrations for AME, ADE and AIDP that were based on a small number of diets and models of the same kind that were based on in-vivo assays that tested the efficacy of the inclusion of enzymes in the diets.

New models could be developed as more samples become available. It is the case of AAs and carbohydrates in compound feeds, fat, GE and P in broiler ileal digesta and CP and GE in pig ileal digesta. Phytate P, for which calibrations have not been presented in in this thesis could be also an interesting parameter to model due to the increasing use of phytases in poultry and swine nutrition.

The strategy used to update the calibration models for the nutrient content in compound feeds could be also applied to update the models of carbohydrates and lignin content for feedstuffs. The developed algorithm used to update the P model for pig faeces could be also applied to update other models in pig faeces and also in excreta and ileal digesta since these samples are also typically obtained in large batches at the end of in-vivo assays.

The strategies based on NIRS to predict AME, ADE and AIDP could be also applied to predict other digestibility parameters such as the apparent total tract coefficient of digestibility of CP, OM and fat or the apparent ileal digestibility coefficient of AAs, both for broilers and pigs. The success of the inclusion of phytase as an additional predictor in the prediction of AIDP suggest that the inclusion of other enzymes could be worth testing. For example the inclusion of carbohydrase as an additional predictor in the prediction of the digestibility of fibre.

The prediction of in-vitro digestibilities is another subject whose study deserves attention. Although in-vitro methods are not as reliable as in-vivo to determine digestibility, they are more and more accurate and their use is growing. In-vitro data are easier and cheaper to obtain than in-vivo. This would facilitate the construction of large databases to develop NIR prediction models, which could be validated with the growing availability of new samples.

Appendix A

List of papers by the author presented in this thesis

1. A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, M.P. Callao, I. Ruisánchez, Use of visible-near infrared spectroscopy to predict nutrient composition of poultry excreta. Anim. Feed Sci. Technol. 283 (2022) 115169.

2. A. Cruz-Conesa, J. Ferré, I. Ruisánchez, A.M. Pérez-Vendrell, Selection of reference samples for updating multivariate calibration models used in the analysis of pig faeces. Chemometr. Intell. Lab. Syst. 234 (2023) 104749.

3. A. Cruz-Conesa, K.E. Back Knudsen, I. Ruisánchez, J. Ferré, A.M Pérez-Vendrell, S.J. Noel, Determination of carbohydrate and lignin content in feedstuffs for monogastric animals using near-infrared spectroscopy. Anim. Feed Sci. Technol. Submitted.

Meeting contributions

Oral communications

A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, M.P. Callao, I. Ruisánchez. Simultaneous updating of near-infrared calibration models using a sample selection based on D-optimal criterion. 12ª Trobada de Joves Investigadors dels Països Catalans, Girona (Spain, 2022).

A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, I. Ruisánchez. Novel applications of NIR spectroscopy for the optimization of monogastric animal feed. IRTA PhD Candidate Annual Seminar, Caldes de Montbuí (Spain, 2022).

A. Cruz-Conesa, A.M. Pérez-Vendrell, J. Ferré, I. Ruisánchez. Predicción de energía metabolizable (EM) en piensos para broilers y de energía digestible (ED) en piensos para cerdos mediante espectroscopía de infrarrojo cercano (NIRS). XX Jornadas sobre Producción Animal, Zaragoza, (Spain, 2023).

Poster communications

A. Cruz-Conesa, I. Ruisánchez, M.P. Callao, A.M. Pérez-Vendrell, J. Ferré. Use of Vis-NIR spectroscopy to predict nutrient composition of poultry excreta. 20th International Conference on Near Infrared Spectroscopy, Beijing (China, 2021).

A. Cruz-Conesa, J. Ferré, M.P. Callao, A.M. Pérez-Vendrell, I. Ruisánchez. Simultaneous updating of NIR calibration models to predict protein, gross energy, fat and fiber in pig feces using a sample selection algorithm based on D-optimal criterion. 20th International Conference on Near Infrared Spectroscopy, Beijing (China, 2021).

A. Cruz-Conesa, K.E. Bach Knudsen, S.J. Noel, A.M. Pérez-Vendrell, I. Ruisánchez, J. Ferré. Near-infrared spectroscopy to predict the carbohydrate and lignin contents of a variety of feedstuffs included in poultry diets. 23rd European Symposium on Poultry Nutrition, Rimini (Italy, 2023).

A. Cruz-Conesa, I. Ruisánchez, A.M. Pérez-Vendrell, J. Ferré. Determination of protein and amino acid composition of poultry ileal digesta by near-infrared spectroscopy. 23rd European Symposium on Poultry Nutrition, Rimini (Italy, 2023).

A. Cruz-Conesa, J. Ferré, A.M. Pérez-Vendrell, I. Ruisánchez. Near-Infrared spectroscopy (NIRS) to determine apparent metabolizable energy (AME) and apparent ileal phosphorus digestibility (AIDP) in broiler diets. 11th Colloquium Chemiometricum Mediterraneum, Padova (Italy, 2023).

Appendix B

Research stay and training courses

Research stay

Viborg (Abril-Julio 2022)
Objective: Development of prediction models for carbohydrate and lignin content in monogastric animal feeds.
Department of Animal and Veterinary Sciences, AU Viborg research center Foulum, Aarhus University, DK 8830 Tjele, Denmark.

Training courses

Fundamentals and Applications of Near Infrared Spectroscopy, June-August 2021
University of Cordoba, Spain.
Practical training in monogastric nutrition, November 2021-April 2022
Institute of Agri-food Research and Technology, Spain


UNIVERSITAT
ROVIRA i VIRGILI


IRTA
Institute
of Agrifood Research
and Technology