Animal Nutrition 7 (2021) 1242–1252

Contents lists available at ScienceDirect

Animal Nutrition

journal homepage: http://www.keaipublishing.com/en/journals/aninu/

Original Research Article

Prediction of protein and amino acid composition and digestibility in individual feedstuffs and mixed diets for pigs using near-infrared spectroscopy

Samantha J. Noel*, Henry J.Høgh Jørgensen, Knud E. Bach Knudsen

Department of Animal Science, AU Foulum, Aarhus University, DK 8830, Tjele, Denmark

ARTICLE INFO

Article history:
Received 17 September 2020
Received in revised form
30 June 2021
Accepted 8 July 2021
Available online 27 September 2021

Keywords:
Near-infrared spectroscopy
Amino acids
Protein
Apparent ileal digestibility
Total tract digestibility
Swine

ABSTRACT

Knowledge of the amounts and digestibility of amino acids in pig feedstuffs is essential for calculating the appropriate inclusion level in a complete diet. Wet chemical analysis and in vivo digestibility trials are time-consuming and costly and cannot be used for routine assessment. Near-infrared spectroscopy (NIRS) offers a rapid, cost effective and environmentally friendly method for evaluating feedstuffs. Calibrations models were developed using NIRS to predict the content of crude protein and 18 amino acids from a wide range of feedstuffs used in pig production ($n = 607$). The samples ranged from single feed ingredients (containing amino acids from 0.3 to 129.8 g/kg of dry matter) to feed mixtures (containing amino acids from 1.2 to 53.2 g/kg of dry matter). The predictive ability of the calibrations was tested with an independent dataset ($n = 150$) and with cross-validation. Furthermore, we compare these calibrations with calibrations developed on more narrowly defined groups of samples and with predictions from regression analysis of crude protein. The models were able to predict the concentrations of crude protein and 18 amino acids with good levels of precision and high coefficients of determination for calibration (RSQ^{CAL}) from 0.91 to 0.99 and validation (RSQ^{VAL}) from 0.87 to 0.97. Calibration models were able to predict all amino acids except tryptophan and valine with greater accuracy than those from protein regression. We also developed calibration models to predict the apparent ileal and total tract digestibility of protein and amino acids. With the exception of tryptophan, RSQ values (>0.7) and standard error of cross validation (SECV) values (<5%) were obtained for the digestibility of most of the amino acids. In conclusion, NIRS can be used to predict crude protein and amino acid concentrations from a wide range of single ingredients and feed mixtures used for pig diets without separate models for each feedstuff. The digestibility of protein and amino acids can be predicted with an acceptable accuracy to be useful in formulating pig diets.

1. Introduction

A correct supply of dietary protein and amino acids is important for optimal growth and protein accretion in pigs (Moughan et al., 2018a) as undersupply will have a strong negative impact on animal performance and oversupply a negative impact on the environment in terms of nitrogen leaching to the aquatic environment, to drinking water, and nitrogen fallout from evaporation from pig housing and slurry storage facilities (Millet et al., 2018). Protein quality evaluation aims to determine the capacity of the feedstuffs and diets to meet the protein and essential amino acid requirement, which is determined by the absolute and relative quantities of dietary indispensable amino acids in feeds, the digestibility of protein

* Corresponding author.
E-mail address: Samantha.Noel@anis.au.dk (S.J. Noel).
Peer review under responsibility of Chinese Association of Animal Science and Veterinary Medicine.

Production and Hosting by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.aninu.2021.07.004

in the gastrointestinal tract, and the bioavailability of amino acids (Moughan et al., 2018a; Moughan et al., 2018b). Cereals are in general low in lysine whereas legumes are low in sulphur containing amino acids, methionine and cysteine, but combined they may provide a better balance (Jezierny et al., 2010; Rosenfelder et al., 2015; Spindler et al., 2016). However, because of a general wish to reduce the protein content of diets it is often necessary to supplement diets with synthetic amino acids, the most important being lysine, methionine, threonine, tryptophan and valine (Han et al., 2001). Optimizing the amino acid composition is mostly done on the basis of table values with average values from each feed type, but crops will vary in their amino acid content due to growing conditions, harvest year and processing (van Barneveld, 1999; Rosenfelder et al., 2015; Spindler et al., 2016). Measurement of the amino acid profiles by chemical methods is time-consuming and expensive for routine use and obtaining information of the nutritional value of feedstuffs by measuring the digestibility of the protein and amino acids are even more difficult to obtain, as it requires lengthy and expensive digestibility experiments (Just et al., 1985; Rosenfelder et al., 2015; Spindler et al., 2016).

The use of near-infrared spectroscopy (NIRS) to predict amino acid composition in cereals and feedstuffs has been widely accepted as an alternative to wet chemistry methods (Chen et al., 2013). Near-infrared spectroscopy has the benefit of being cheaper, faster and non-destructive and with a potential of being implemented in real-time for precision livestock feeding. A limitation, however, is that NIRS cannot directly measure crude protein or amino acid concentrations and therefore is a secondary method relying on calibration with a database of samples with known values (Shenk and Westerhaus, 1991). Another limitation is that the prediction models are only valid for the types of samples that have been used in the development of the models and these samples must represent the range of diversity available to be able to make robust predictions. Today, many models exist for different types of samples for example; sunflower meal (Fontaine et al., 2001), peas (Fontaine et al., 2001), cereal ingredients (Fontaine et al., 2002; Hoehler et al., 2005), millet (Yang et al., 2013), dried distillers grains with solubles (DDGS) (Zhou et al., 2012), processed animal proteins or meals (Fontaine et al., 2001; Hoehler et al., 2005; De La Haba et al., 2006), rice (Fontaine et al., 2002; Wu et al., 2002; Zhang et al., 2011), rapeseed (Fontaine et al., 2001; Chen et al., 2011), quinoa (Escuredo et al., 2014), animal feed (Gonzalez-Martin et al., 2006), peanuts (Wang et al., 2013) and soybean (Pazdernik et al., 1997; Fontaine et al., 2001). However, in this project we wished to assess whether one model could be developed to predict feed ingredients and mixed diets used in Danish pig production. It is to be expected that models built with similar types of samples will be more accurate (lower prediction error) but less robust (valid with a smaller range of sample types and values) than models built with all the samples types together (Shenk et al., 2001; Perez-Marin et al., 2007). Therefore, we tested models built with all feed types against those built from groups of similar sample types. On a subset of samples, we also investigated the possibility of developing a model for predicting ileal and total tract digestibility of protein and amino acids. Due to the lower number of feedstuffs with measured digestibility, models were only built on the total dataset.

2. Materials and methods

The animal experiments were performed in concordance with The Animal Experiments Inspectorate, Danish Veterinary and Food Administration, Ministry of Environment and Food of Denmark.

2.1. Sample description

We assessed both conventional and unconventional feedstuffs that have been evaluated for nutritive value in the Danish Feed Evaluation system (Just, 1982; Just et al., 1983). The feedstuffs representing a diverse range of feed types have been collected since 1975 and onwards and stored at −20 °C. Additionally, 353 common and experimental feed mixture samples used in experiments with pigs were also included. Samples were classified into groups of similar types including: cereals, supplemental feed ingredients, which included all the remaining plant-based single ingredient feedstuffs (i.e. cereal co-products, cereal substitutes, protein concentrate, grass meal, fiber-rich by-products and other) and feed mixtures (diets). The cereal group included oats, maize, wheat, barley, rice and triticale and were all but maize and rice of Danish origin. The supplemental ingredient group included cereal co-products: corn gluten feed, malt sprouts, maize middlings, maize bran, wheat middlings, wheat bran, rice middlings, rice bran, barley protein, and barley groats; protein concentrates: soybean meal, linseed meal, peas, faba beans, cottonseed meal, cottonseed cake, coconut cake, palm cake, lupin, potato protein, sunflower seed, rapeseed meal and rapeseed cake; and miscellaneous: tapioca, citrus pulp, apple pomace, maize silage, lucerne, guar meal, and grass meal. All samples in the supplemental ingredient group were imports except co-products from barley and wheat, peas, faba beans, lucerne and grass meals which originated in Denmark. The feed mixtures group included many formulations of pig diets including commercial diets and balanced diets formulated for use in different experiments (Just, 1982; Just et al., 1983a, 1983b, 1985; Fernandez et al., 1986).

2.2. Sample selection for predicting protein and amino acid content

Sample spectra were assessed with principal component analysis (PCA) plots to determine the spectral outliers and assess the diversity of spectra within sample groups. Outlier samples were determined as one or more of the following criteria; spectral outliers, the normalized difference between the reference and the predicted value (critical T) was high (>2.5) in most of the constituents (>40%) indicating an error in sample classification and sample types that were underrepresented or too different from the total sample pool (e.g., nitrogen-free-diets, sodium hydroxide treated barley straw, pectin and brewer's yeast). From 791 samples with measured chemical composition in at least 1 of the constituents (protein or amino acid), 34 samples were removed as outliers leaving 757 samples to build and test the models. Samples were then randomly divided into samples for calibration and samples for validation (80:20 split). This data set was labelled the total data set. As not all samples had measured values for each constituent (protein or amino acid), each model was built with different numbers of observations. The actual number of samples used for model development and validation is shown in Table 1. The validation set contained similar proportions of the 3 sample groups and similar ranges of values for the constituents thereby representing the range of variation seen in the samples. The samples were then further divided into 3 data sets (each with a calibration and validation set each, 80:20 split) based on the sample types, cereals, supplemental ingredients and feed mixtures. The actual number of samples used for model development and validation of cereals, supplemental ingredients and feed mixtures are shown in Tables 2–4 respectively.

2.3. Sample selection for predicting protein and amino acid digestibility

On a subset of samples, ileal and total tract digestibility of protein and amino acids were studied. The number of samples

Table 1
Summary statistics for calibration (CAL) and validation (VAL) of crude protein and the amino acid composition (g/kg of dry matter) of total (individual ingredient and feed mixtures) pig feedstuffs.

Item	CAL statistics							VAL statistics				
	N ^{CAL}	Range ^{CAL}	Mean	SD	Factors	RSQ ^{CAL}	SECV	N ^{Val}	Range ^{VAL}	RSQ ^{VAL}	SEP	RPD
CP	607	27.7 - 708.2	180.8	86.2	13	0.98	15.73	150	88.6 - 506.3	0.95	16.87	5.11
Indispensable amino acids												
LYS	395	0.9 - 44.4	7.9	6.7	14	0.98	1.30	94	2.8 - 28.3	0.96	1.05	6.39
MET	395	0.3 - 9.7	2.8	1.6	13	0.97	0.41	94	1.3 - 9.7	0.94	0.42	3.84
CYS	395	0.3 - 11.1	3.3	1.6	8	0.95	0.44	94	1.1 - 11.8	0.94	0.44	3.58
THR	395	0.8 - 27.9	6.1	4.1	16	0.99	0.71	94	3.2 - 21.8	0.96	0.68	6.09
ILE	387	0.8 - 34.6	6.8	4.9	8	0.98	0.84	94	3.2 - 22.6	0.97	0.71	6.94
LEU	387	1.4 - 55.4	12.1	7.7	8	0.98	1.43	94	5.4 - 38.9	0.95	1.42	5.43
HIS	387	0.4 - 19.3	4.0	2.7	7	0.97	0.54	94	2.0 - 14.3	0.97	0.44	6.26
PHE	387	0.6 - 36.6	8.1	4.8	5	0.97	0.95	94	4.1 - 22.1	0.96	0.85	5.69
VAL	387	1.0 - 36.2	8.5	5.2	8	0.98	0.84	95	1.1 - 28.3	0.94	1.13	4.59
ARG	387	1.2 - 52.7	10.5	8.8	15	0.99	1.89	94	2.8 - 46.6	0.97	1.57	5.62
TRP	163	0.2 - 9.1	2.2	1.6	2	0.93	0.45	37	0.8 - 6.0	0.87	0.45	3.62
Dispensable amino acids												
ALA	387	1.1 - 30.0	7.3	4.7	13	0.98	1.01	94	4.0 - 23.5	0.96	0.84	5.58
ASP	387	1.8 - 79.4	13.2	12.4	14	0.99	1.85	94	5.6 - 39.2	0.97	1.69	7.30
GLU	387	3.4 - 129.8	37.4	18.3	10	0.97	4.22	94	12.8 - 98.9	0.96	3.30	5.56
GLY	387	1.0 - 29.2	7.5	5.0	5	0.97	0.93	94	3.6 - 27.6	0.96	0.96	5.21
PRO	387	1.0 - 36.1	13.5	5.2	10	0.91	1.91	94	6.3 - 32.5	0.92	1.36	3.80
SER	387	0.9 - 38.3	8.0	5.2	7	0.98	0.94	94	4.0 - 23.5	0.97	0.78	6.65
TYR	369	0.5 - 26.5	5.5	3.9	8	0.98	0.72	92	2.6 - 16.5	0.96	0.58	6.70

N = the number of samples; RSQ = regression coefficient; SECV = standard error of cross validation; SEP = standard error of prediction; RPD = ratio of performance deviation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

Table 2
Summary statistics for calibration (CAL) and validation (VAL) of crude protein and amino acid composition (g/kg of dry matter) of cereals.

	CAL statistics							VAL statistics				
	N ^{CAL}	Range ^{CAL}	Mean	SD	Factors	RSQ ^{CAL}	SECV	N ^{Val}	Range ^{VAL}	RSQ ^{VAL}	SEP	RPD
CP	211	90.6 - 176.7	127.0	17.8	9	0.95	5.47	59	88.6 - 161.1	0.91	5.47	3.26
Indispensable amino acids												
LYS	194	2.5 - 8.8	4.3	0.7	8	0.83	0.37	55	2.8 - 6.3	0.67	0.35	1.97
MET	192	1.4 - 3.1	2.1	0.3	7	0.84	0.14	55	1.3 - 2.6	0.59	0.19	1.45
CYS	194	2.0 - 4.6	2.7	0.4	8	0.88	0.18	55	2.0 - 4.5	0.73	0.25	1.57
THR	194	3.0 - 6.4	4.0	0.5	10	0.92	0.25	55	3.2 - 5.0	0.72	0.27	2.02
ILE	194	1.4 - 6.8	4.6	0.7	7	0.88	0.30	55	3.2 - 6.0	0.77	0.35	2.00
LEU	191	5.6 - 13.5	8.7	1.4	11	0.96	0.49	55	5.4 - 13.8	0.69	0.93	1.47
HIS	193	2.0 - 4.0	2.8	0.4	5	0.76	0.23	55	2.0 - 3.8	0.72	0.24	1.71
PHE	194	3.8 - 9.1	6.1	1.1	9	0.93	0.39	55	4.1 - 8.5	0.87	0.40	2.64
VAL	195	4.4 - 9.4	6.2	0.9	7	0.87	0.42	55	4.8 - 8.2	0.73	0.49	1.89
ARG	192	4.4 - 12.1	6.3	1.0	8	0.89	0.41	55	4.4 - 10.2	0.78	0.53	1.81
TRP	64	0.8 - 2.1	1.5	0.3	3	0.65	0.18	15	0.8 - 1.9	0.52	0.20	1.29
Dispensable amino acids												
ALA	192	2.5 - 8.4	5.0	0.7	9	0.90	0.32	55	4.0 - 8.0	0.76	0.42	1.70
ASP	192	5.1 - 14.1	7.2	1.1	8	0.87	0.51	55	5.6 - 11.7	0.76	0.57	1.87
GLU	193	13.7 - 53.3	31.7	7.7	11	0.98	1.93	55	17.4 - 44.4	0.88	2.32	3.31
GLY	192	3.7 - 8.2	5.1	0.7	8	0.88	0.32	55	3.9 - 7.4	0.80	0.34	2.07
PRO	194	5.7 - 28.7	12.7	2.8	10	0.96	0.88	55	6.3 - 18.9	0.89	0.93	3.00
SER	192	3.9 - 8.0	5.7	0.9	10	0.94	0.34	55	4.0 - 7.7	0.82	0.41	2.14
TYR	193	2.4 - 5.8	3.9	0.7	9	0.93	0.25	55	2.6 - 5.7	0.81	0.31	2.16

N = the number of samples; RSQ = regression coefficient; SECV = standard error of cross validation; SEP = standard error of prediction; RPD = ratio of performance deviation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

was too low to split the samples into groups or an external validation and calibration set. Therefore, for these models we used only cross-validation to validate the models. The 2 data sets are referred to as the ileal digestibility data set and the total tract digestibility data set. The proportions of sample types with ileal digestibility measurements were 15% cereals, 17% supplemental feed ingredients and 68% feed mixtures. The proportion of sample types with total tract digestibility measurements were 13% cereals, 14% supplemental feed ingredients and 73% feed mixtures.

2.4. Chemical analysis

The chemical analysis of the samples was undertaken at the time of the sample collection i.e. from 1975 onwards. Crude protein (N × 6.25) was determined by the Kjeldahl method using a Kjell-Foss 16,200 autoanalyzer and recorded as grams per kilogram of dry matter (DM). Twenty selected samples with the original crude protein measurement made from 1975 to 1996 were reanalyzed by the Kjeldahl method. Amino acids were analyzed as described by Mason et al. (1980) and recorded as g/kg of DM. This method is

Table 3
Summary statistics for calibration (CAL) and validation (VAL) of crude protein and amino acid composition (g/kg of dry matter) of supplemental feed ingredients.

Item	CAL statistics							VAL statistics				
	N ^{CAL}	Range ^{CAL}	Mean	SD	Factors	RSQ ^{CAL}	SECV	N ^{Val}	Range ^{VAL}	RSQ ^{VAL}	SEP	RPD
CP	101	27.7 - 708.2	269.3	155.3	7	0.99	18.46	23	103.1 - 506.3	0.99	16.93	9.17
Indispensable amino acids												
LYS	92	0.9 - 44.4	14.2	10.5	7	0.98	2.33	20	2.8 - 28.3	0.97	1.27	8.29
MET	92	0.3 - 9.7	4.2	2.7	8	0.98	0.60	20	1.8 - 9.7	0.98	0.43	6.20
CYS	92	0.3 - 11.1	4.6	2.7	7	0.97	0.60	20	1.1 - 11.8	0.96	0.63	4.33
THR	92	0.8 - 27.9	10.2	6.4	6	0.99	1.07	20	3.8 - 21.8	0.97	1.00	6.44
ILE	92	0.8 - 34.6	11.5	7.7	6	0.99	1.09	20	3.7 - 22.6	0.96	1.24	6.20
LEU	92	1.4 - 55.4	19.7	12.0	9	0.99	1.88	20	8.2 - 38.9	0.96	1.74	6.87
HIS	92	0.4 - 19.3	6.7	4.3	6	0.98	0.83	20	2.5 - 14.3	0.96	0.75	5.71
PHE	92	0.6 - 36.6	12.5	7.7	5	0.98	1.41	20	4.7 - 22.1	0.94	1.43	5.37
VAL	92	1.0 - 36.2	13.7	8.0	5	0.99	1.15	20	5.4 - 28.3	0.93	1.85	4.36
ARG	92	1.2 - 52.7	19.9	13.3	9	0.99	3.45	20	2.8 - 46.6	0.96	2.64	5.03
TRP	47	0.2 - 9.1	3.6	2.4	1	0.93	0.66	9	2.0 - 6.0	0.85	0.52	4.63
Dispensable amino acids												
ALA	92	1.1 - 30.0	12.4	6.7	7	0.98	1.58	20	4.8 - 23.5	0.94	1.20	5.63
ASP	92	1.8 - 79.4	26.2	18.4	7	0.99	3.28	20	6.8 - 39.2	0.92	3.84	4.80
GLU	92	3.4 - 129.8	50.1	30.6	8	0.98	6.26	20	12.8 - 98.9	0.98	4.12	7.43
GLY	92	1.0 - 29.2	12.7	7.4	6	0.98	1.54	20	3.6 - 27.6	0.95	1.67	4.41
PRO	92	1.0 - 36.1	15.1	8.7	9	0.98	2.51	20	7.2 - 32.5	0.93	2.13	4.10
SER	92	0.9 - 38.3	12.9	8.2	8	0.99	1.28	20	4.7 - 23.5	0.97	1.16	7.04
TYR	92	0.5 - 26.5	9.1	5.9	4	0.97	1.23	20	3.4 - 16.5	0.97	0.76	7.82

N = the number of samples; RSQ = regression coefficient; SECV = standard error of cross validation; SEP = standard error of prediction; RPD = ratio of performance deviation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

Table 4
Summary statistics for calibration (CAL) and validation (VAL) of crude protein and amino acid composition (g/kg of dry matter) of pig feed mixtures.

Item	CAL statistics							VAL statistics				
	N ^{CAL}	Range ^{CAL}	Mean	SD	Factors	RSQ ^{CAL}	SECV	N ^{Val}	Range ^{VAL}	RSQ ^{VAL}	SEP	RPD
CP	284	102.6 - 299.3	189.2	40	7	0.92	12.82	69	108.3 - 271.3	0.78	17.89	2.24
Indispensable amino acids												
LYS	98	3.2 - 15.5	9.0	2.3	8	0.87	1.26	20	7.4 - 13.0	0.30	1.66	1.41
MET	100	1.2 - 4.5	2.8	0.7	9	0.88	0.42	20	1.6 - 3.9	0.16	0.84	0.78
CYS	100	2.1 - 4.7	3.1	0.6	5	0.83	0.29	20	2.3 - 4.2	0.75	0.27	2.10
THR	101	3.3 - 10.4	6.3	1.5	2	0.60	0.97	20	3.7 - 9.5	0.43	0.88	1.64
ILE	91	3.7 - 12.5	6.6	2.1	3	0.94	0.63	20	3.7 - 11.4	0.84	0.84	2.56
LEU	91	6.8 - 20.3	11.7	3.4	6	0.96	0.93	20	6.9 - 18.7	0.84	1.34	2.57
HIS	93	2.2 - 8.1	4.0	1.1	7	0.93	0.42	20	2.4 - 6.3	0.61	0.68	1.64
PHE	90	4.7 - 14.1	7.9	2.2	5	0.95	0.62	20	4.9 - 13.3	0.84	0.88	2.54
VAL	91	5.1 - 14.2	8.1	2.3	8	0.97	0.63	20	5.2 - 13.1	0.83	0.91	2.54
ARG	93	5.0 - 20.1	9.7	3.9	8	0.97	1.09	20	5.0 - 17.0	0.68	2.44	1.60
TRP	45	1.3 - 3.3	1.9	0.6	7	0.97	0.18	14	1.3 - 3.1	0.81	0.26	2.15
Dispensable amino acids												
ALA	92	4.1 - 13.6	6.8	2.2	9	0.97	0.65	20	4.2 - 10.7	0.88	0.72	3.06
ASP	92	5.9 - 26.5	12.5	5.6	9	0.98	1.24	20	5.9 - 23.2	0.81	2.54	2.23
GLU	93	22.7 - 53.2	37.3	7.9	3	0.86	3.52	20	23.4 - 55.0	0.79	4.00	1.97
GLY	92	4.0 - 16.8	7.0	2.3	4	0.92	0.75	20	4.1 - 10.9	0.83	0.95	2.40
PRO	93	6.5 - 21.6	13.5	2.9	3	0.75	1.66	20	8.5 - 20.5	0.75	1.45	2.00
SER	93	4.4 - 14.7	7.9	2.3	5	0.91	0.87	20	4.6 - 13.1	0.77	1.04	2.21
TYR	74	3.1 - 10.4	5.3	1.8	8	0.98	0.50	18	3.2 - 9.2	0.68	1.01	1.76

N = the number of samples; RSQ = regression coefficient; SECV = standard error of cross validation; SEP = standard error of prediction; RPD = ratio of performance deviation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

based on the same principles in term of hydrolysis, detection and instrumentation as described by the European Commission (1998; 2009).

2.5. Determination of ileal and fecal digestibility of amino acids

The experiments for the determination of the ileal and fecal digestibility of protein and amino acids have been performed over time but comprise of 6 (40 to 50 kg) pigs per group, cannulated at the end of the small intestine as described by Just at al (1985). In brief, surgery was performed at 30 to 35 kg and a T-cannula was placed in the ileum approximately 150 mm anterior to the ileo-caecal junction. The animals were fed 3 times daily at 07:00, 15:00 and 23:00 with 1.6 to 2.0 kg/d of diet adjusted to give the same amount of net energy per day (Just et al., 1983). The feed was thoroughly mixed with water before feeding. After a 7-d adaptation period, feces were collected quantitatively on d 7 to 11 and ileal digesta on d 12 to14. Ileal digesta were collected for a total period of 12 h, on d 12 from 09:00 to 11:00 and 13:00 to 15:00, on d 13 from 08:00 to 10:00 and 12:00 to 14:00 and on d 14 from 07:00 to 09:00 and 11:00 to 13:00. Feces were collected twice daily, frozen and stored at −20 °C. At the end of the experiment, the feces was mixed

before sampling for analysis. The ileal digesta were collected on ice, frozen immediately after collection, stored at − 20 °C and mixed thoroughly before samples were taken for analysis. Three samples of each diet, ileal digesta and feces samples were subjected to a chemical analysis of crude protein and amino acids. Two samples of each diet were analyzed for DM.

2.6. NIRS analysis

Feed samples were stored at −20 °C in airtight containers until needed. To ensure there was no moisture build-up in the containers that could ruin the samples upon defrosting, samples were dried at 60 °C in an air-forced oven for 48 h. Over 50% of the samples were stored already ground (1 mm) and the remaining samples were milled to a 1-mm particle size in an ultra-Centrifugal Mill ZM 200 (Retsch, Haan, Germany). Dried and ground feed samples were left to equilibrate to ambient moisture levels at room temperature for a minimum of 48 h prior to scanning. Ground samples were packed into a sample cup with quartz window and scanned using a Foss NIRS DS2500 feed analyzer (FOSS Analytical A/S, Silver Springs, MD, USA). Each scan was the average of 32 scans from various positions on the sample cup using the wavelength range from 400 to 2,500 nm with data recorded every 0.5 nm. Each sample was scanned in duplicate from 2 separate samplings and the duplicate spectra averaged.

2.7. Spectral pretreatment and calibration development

Calibrations for crude protein, amino acids and digestibility of protein and amino acids were developed with WinISI version 4.9.0 (FOSS Analytical A/S, Hillerød, Denmark). Different spectral preprocessing methods and model types (including non-linear) were investigated (data not shown) and the methods that produced the best models are described here. Sample spectra were mathematically preprocessed and the spectral range reduced prior to model development. The spectra were preprocessed using the standard-normal-variate (SNV) method along with detrending (D) (Barnes et al., 1989) to minimize baseline offset and reduce scatter. A second order derivative with a gap of 8 and 4 points of smoothing was then applied (math treatment 2, 8, 4, 1). The spectral range was reduced to remove spectra in the visible light region to include wavelengths between 780 and 2,500 nm with data points every 0.5 nm resulting in 1,698 data points per scan. Calibration models were built with the modified partial least squares method (mPLS). No further outlier removal was used for digestibility models or the total and supplemental ingredients datasets as these included a very diverse range of samples and the outlier samples were identified in the prescreening procedure. For the cereals and feed mixture, datasets models were improved with 1 round of outlier removal a conservative critical T value of 3 as the cutoff (0 to 3 outliers removed per constituent). Cross-validation was performed to determine the number of factors to include in the model and to validate the models in the case of the digestibility models. Cross-validation was performed by dividing the calibration samples that were ranked on their values into groups of 8 and building successive models with 1 group left out. Each group is then evaluated using the model developed on the other samples. The number of factors to be included in the models were chosen to include as much information as possible without overfitting by assessing when the standard error of cross validation (SECV) reached its lowest value. For amino acid content predictions, calibration models were built using the total dataset as well as the 3 groups of similar sample types i.e. cereal, supplemental ingredients and feed mixtures. To assess whether using the entire quite dissimilar sample types together could make a stronger model the equations developed with the total dataset was then evaluated with four validation sets, total, cereal, supplemental ingredients and feed mixtures. The models developed for the 3 groups (cereals, supplemental ingredients and feed mixtures) were evaluated with their corresponding validation set.

2.8. Equation evaluation

The regression coefficient (RSQ), standard error of calibration (SEC), standard error of cross validation (SECV), standard error of prediction corrected for bias (SEP) and ratio of performance deviation (RPD) were used to evaluate calibration performance (Sapienza et al., 2008; ISO, 2017). The RSQ describes the fit when the reference values are plotted against the predicted values. The higher the RSQ value the better the fit; 1 equals a perfect fit. Values for RQS were determined for the calibration and the external validation samples. The SECV shows how well the calibration model predicts the reference values when some samples are selectively removed. Lower SECV values indicate higher precision in the models accuracy. The SEP evaluates the performance of the model on a set of independent samples. This is the most important indication of the precision of a calibration model to predict new samples. The international standard (ISO 12099:2017) recommends that there should be at least 20 samples in a validation set. The RPD was calculated by dividing the standard deviation (SD) by the SEP. The RPD gives an indication on whether the SEP values are low enough in comparison to the variation seen in the population used to make the model. Relative SEP was also calculated by dividing the SEP by the mean of the lab values for the measured amino acid and multiplying by 100. Coefficient of variation (CV) was calculated for the lab values by dividing the SD by the mean and multiplying by 100. The CV is an independent measure of the variation that enables the different amino acids with different means to be compared.

3. Results and discussion

3.1. Sample variability for the total, cereal, supplemental ingredient and mixtures data sets

The crude protein and amino acid compositions for all pig feed samples are shown in Table 1, for cereals in Table 2, for supplemental ingredients in Table 3 and feed mixtures in Table 4. The crude protein content in all the samples ranged from 27.7 to 708.2 g/kg of DM and for the individual groups; cereals 90.6 to 176.7 g/kg of DM, supplemental ingredients 27.7 to 708.2 g/kg of DM and feed mixtures 102.6 to 299.3 g/kg of DM. Overall the values for the 18 amino acids had the lowest range and mean values for the cereal samples followed by the feed mixtures and the supplemental ingredients had by far the largest range also the highest mean values reflecting the diversity of samples in this group. The cereal and feed mixtures groups also showed a high level of diversity albeit less than the supplemental ingredients as they contained broad variability in species and cultivar, harvest year and formulation (for the mixtures) which is important for developing calibration equations to predict future samples. Furthermore, the small differences in range between the calibration and validation set indicate that both sets are representative of the overall variation in the samples making them suitable for NIRS calibrations.

3.2. Method development

To determine the best mathematical treatment and spectral pretreatment to develop equations on the total, cereal, supplemental ingredient and feed mixture data sets, many different

combinations were tested (not shown). The best combination of SNV + D with wavelength reduction to 780 to 2,500 nm and mathematical treatment of 2, 8, 4, 1 was chosen as this produced equations with overall the highest RSQ and lowest SECV and SEP values. In addition to this, different model methods were tested. The mPLS method produced the best performing models with this data. Indeed, PLS regression methods are the most widely used for NIR models for agro-food applications (Perez-Marin et al., 2007).

3.3. Calibration equations with the total samples

Calibration equations were developed using 607 samples of all samples types, and the remainder were used to test the model in the validation set. The statistics describing the calibration model and validation of the total samples are shown in Table 1. Additionally, the equations developed from the total dataset were tested with the separate validation sets used for the cereal, supplemental feed ingredients and feed mixtures datasets to make a comparison of the total model with the models of the separate groups (Table 5). The equations for crude protein and 18 amino acids showed high coefficients of determination for calibration (RSQ^{CAL} 0.91 to 0.99) and almost as good for validation (RSQ^{VAL} 0.87 to 0.97). It is normal to find the RSQ values found by validation a little lower than those obtained by calibration (Fearn, 2014). The SECV and SEP were low and in good agreement with each other. The RSQ values for the total validation set were much better than the cereal and feed mixtures validation sets and similar to the supplemental ingredients (Tables 1–4). This can be partly explained by the lower number of samples in the individual group validation datasets. The comparison of the performance of the models developed on the total dataset with the four different validation sets in Table 5 shows how the validation set affects the perceived model performance. Relative SEP (SEP/mean × 100) puts the SEP in context with the mean value of the amino acid being estimated, with larger SEP values expected for higher measured values and makes it easier to compare between studies that may have used different units to report amino acid values. The relative SEP, shown for the indispensable amino acids in Fig. 1., ranges between 9.3 and 20.5% with a mean of 12.8%. Arginine and methionine have higher relative SEP of 15% and tryptophan has the highest of 20.5%. Overall the relative SEP values are higher than those found in calibration derived from single sample types; wheat or corn (3% to 6.7%) (Fontaine et al., 2002), soybean/soybean meal (1.75% to 4.38%) (Fontaine et al., 2001), brown rice flour (3% to 15%) (Zhang et al., 2011), soybean (FOSS instrument, PLS model 2% to 16%) (Kovalenko et al., 2006). However, our calibrations were made with many different samples and the amino acids were not all analyzed in the same laboratory which could have increased the variability in the measurements and thereby the SEP. Our relative SEP, however, compared favorably with the reproducibility of the method for determining 16 amino acids in various chicken feed and corn which varied in the range between 6.2% and 23.3% (ISO, 2005) but higher than the reproducibility of 2% to 5% for amino acids of the reference method reported by Fontaine et al. (2001). The RPD and the ratio of the SD of the amino acid in the calibration population to the SEP, were all high (>3) indicating that the models are adequate for screening for improved selection and 14 amino acids (all excluding cysteine, methionine, valine, proline and tryptophan) as well as crude protein had values > 5 indicating the model is very precise (Sapienza et al., 2008). Together, the results indicate the values obtained by NIR calibrations are meaningful and useful for the total dataset (Fearn, 2002).

3.4. Calibration equations with the cereal samples

Calibrations equations were developed using only the cereal samples including varieties of oats, maize, wheat, barley, rice and triticale (Table 2). The equations for crude protein and 18 amino acids showed high coefficients of determination for calibration (RSQ^{CAL} 0.76 to 0.98) and low standard errors of cross validation SECV and prediction SEP except for tryptophan. Tryptophan had a lower RSQ^{CAL} of 0.65 indicating a poorer fit of the predicted values. The SECV and SEP of the amino acids were comparable to those from NIRS prediction models for brown rice (Zhang et al., 2011) but

Table 5
Summary statistics for equations built on the all sample types with independent validation samples representing the total samples, cereals, supplemental ingredients or feed mixtures (g/kg of dry matter).

Item	Total			Cereals			Supplemental ingredients			Feed mixtures		
	N	RSQ	SEP	N	RSQ	SEP	N	RSQ	SEP	N	RSQ	SEP
CP	150	0.95	16.87	59	0.85	7.71	23	0.98	18.26	69	0.72	21.47
Indispensable amino acids												
LYS	94	0.96	1.05	55	0.62	0.52	20	0.96	1.50	20	0.24	1.53
MET	94	0.94	0.42	55	0.59	0.19	20	0.95	0.61	20	0.36	0.54
CYS	94	0.94	0.44	55	0.61	0.30	20	0.94	0.75	20	0.70	0.34
THR	94	0.96	0.68	55	0.72	0.29	20	0.97	0.94	20	0.30	1.06
ILE	94	0.97	0.71	55	0.76	0.36	20	0.96	1.14	20	0.85	0.84
LEU	94	0.95	1.42	55	0.56	1.03	20	0.95	1.93	20	0.80	1.54
HIS	94	0.97	0.44	55	0.71	0.24	20	0.96	0.66	20	0.75	0.52
PHE	94	0.96	0.85	55	0.86	0.42	20	0.94	1.45	20	0.82	0.93
VAL	95	0.94	1.13	55	0.72	0.47	20	0.93	1.76	20	0.78	1.05
ARG	94	0.97	1.57	55	0.63	0.91	20	0.98	2.18	20	0.76	2.13
TRP	37	0.87	0.45	15	0.42	0.24	9	0.73	0.72	14	0.54	0.41
Dispensable amino acids												
ALA	94	0.96	0.84	55	0.75	0.40	20	0.94	1.34	20	0.78	1.04
ASP	94	0.97	1.69	55	0.82	0.60	20	0.95	2.92	20	0.87	2.09
GLU	94	0.96	3.30	55	0.91	1.94	20	0.98	4.26	20	0.72	4.84
GLY	94	0.96	0.96	55	0.68	0.43	20	0.94	1.70	20	0.80	1.01
PRO	94	0.92	1.36	55	0.90	0.98	20	0.93	1.95	20	0.80	1.30
SER	94	0.97	0.78	55	0.84	0.38	20	0.97	1.24	20	0.82	0.92
TYR	92	0.96	0.58	55	0.65	0.41	20	0.98	0.65	18	0.79	0.83

N = the number of samples; RSQ = regression coefficient; SEP = standard error of prediction; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

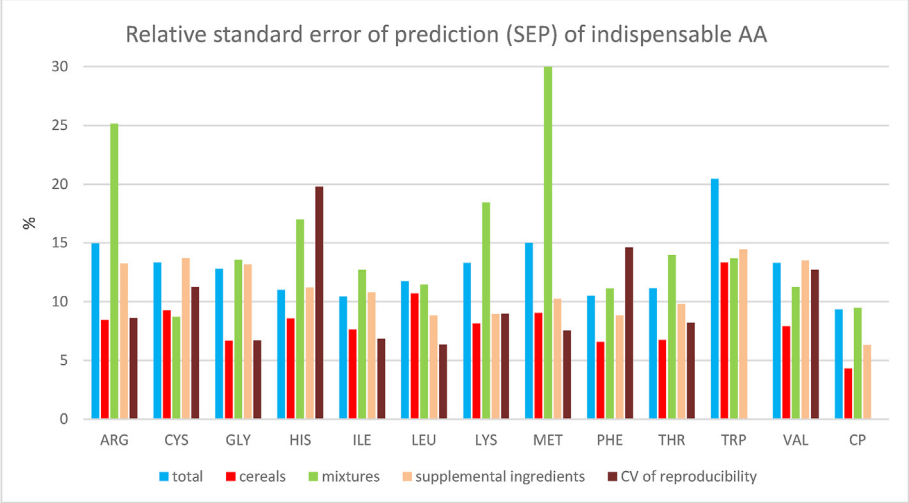

Fig. 1. Relative standard error of prediction of Near-infrared spectroscopy (NIRS) predictions of crude protein and indispensable amino acids developed on the total dataset or the 3 subgroups, cereals, mixtures and supplemental feed ingredients. CV of reproducibility is the coefficient of variation (%) for between laboratory standard deviation from 36 to 46 single determinations of amino acids in broiler finisher feed reported in the method standard ISO 13903:2005 – Determination of amino acid content. Relative standard error of prediction = SEP/mean of lab values × 100.

were higher than those reported for wheat, barley, corn, triticale, wheat bran/middling, rice bran and sorghum (Fontaine et al., 2002). Methionine, leucine and lysine had only moderate RSQ^{VAL} values of 0.59, 0.69 and 0.67 respectively, and tryptophan was poorly predicted in the independent samples (RSQ^{VAL} 0.52). However, the SECV and SEP values were low for all amino acids with relative SEP ranging from 6.6% to 13.3% (Fig. 1). Relative SEP were the lowest in the cereal group compared to all other groupings. Interestingly models built using just the cereals, estimate with roughly the same precision as models built on the whole range of samples for most amino acids except for arginine, lysine and tryptophan where the models build on just the cereal samples had an advantage for predicting cereals over the total models (Tables 2 and 5). The reason for that is unknown but could be that a relatively large proportion of indispensable amino acids, e.g., lysine, in cereals are present in the outer part of the grain tissue, which may provide specific spectral information that cannot be obtained in a general model.

The RPD values were >2 for glutamic acid, glycine, isoleucine, phenylalanine, proline, serine, threonine, tyrosine and crude protein indicating these models are good enough for adequate screening. Overall, the RSQ^{CAL} and RSQ^{VAL} values were less than those from the total models were and the RPD values were lower. However, the SEP were also lower indicating that even though there is less variability explained by the model, this model is more precise. The models developed with the cereal samples for leucine, lysine, tryptophan and methionine might not be accurate enough for all purposes.

3.5. Calibration equations with the supplemental ingredient samples

The supplemental ingredients group represents the most diversity of the subgroups and spans the range of measured values seen in the total dataset. The supplemental ingredient samples, like the total of the samples, produced equations for amino acids and crude protein with very high correlation coefficients (RSQ 0.93 to 0.99) for calibration and for validation (RSQ 0.85 to 0.99) as shown in Table 3. They had low standard errors, SECV and SEP and high RPD values (>4) indicating these models can predict with high precision. The relative SEP values were also low and compared favorably with those from the total dataset models (Fig. 1). When the total and supplemental models were tested with the same group of validation samples ($n = 20$ supplemental ingredients), the models built with the supplemental ingredients subgroup outperformed the total group models with higher RSQ and lower SEP for tryptophan, but did not perform as well as the total group models for aspartic acid and were very similar for all other amino acids and crude protein (Tables 3 and 5).

3.6. Calibration equations with the feed mixture samples

Calibration equations were developed for the diverse mixtures of pig diets. The statistics for the calibrations developed on the diet mixtures are shown in Table 4. The coefficients of determination for calibration (RSQ^{CAL} 0.75 to 0.98) were high for all amino acids and crude protein except for threonine with an RSQ^{CAL} of 0.6. The coefficient of determination for validation (RSQ^{VAL} 0.16 to 0.88) was lower than for calibration; however, the validation set only contained 20 samples, which is the recommended minimum number. The RSQ^{VAL} values for lysine and methionine were ≤0.3 whereas arginine, histidine, threonine and tyrosine were <0.7 and their RPD values were <2 indicating these models are not suitable for screening purposes. Overall, models built using the mixtures were the poorest performing of all the datasets. This is presumably because pig diets are formulated to meet minimum amino acids requirement of the indispensable amino acids and if the raw ingredients do not contain enough to meet the requirements, the diets are supplemented with free amino acids. Spectra in the near electromagnetic range measures the vibrations of molecular bonds and their overtones due to the matrix that surrounds them. Therefore, it could be expected that amino acids contained within plant material and free amino acids would not create the same spectral pattern. This could explain why it is more difficult to develop accurate models from the mixture samples. It is also possible that adding more samples and having a larger validation set will result in better calibrations. Good models were produced for predicting alanine, aspartic acid, cysteine, glutamic acid, glycine, isoleucine, leucine, phenylalanine, proline, serine, tryptophan and valine.

Compared to the other sample groupings, the mixture models also had the highest relative SEP from 8.7% to 30.0% (Fig. 1). In addition, the values for SEP were not in good agreement with the SECV for most amino acids (greater than 30% difference). This could also be due to large variation in a small validation set. Relative SEP calculated from the reported values in animal meals were generally less than our study with 7% to 13% for 9 amino acids on 40 to 50 samples (Qiao and van Kempen, 2004), and 9% to 18% for 4 amino acids on 50 samples (Gonzalez-Martin et al., 2006). The differences between these studies and ours are perhaps that we have included both conventional and experimental feed mixtures in the model, increasing the variability. When comparing relative SEP (Fig. 1) and RSQ^{val} (Tables 4 and 5), models developed using just the mixtures samples outperformed the total models on the mixture validation samples for alanine, cysteine, glutamic acid, leucine, threonine, tryptophan, valine and crude protein but the total sample models were more precise for arginine, aspartic acid, histidine, lysine, proline, serine, tyrosine and especially methionine. This may be because more accuracy was gained by having a larger range of samples in the calibration development making it possible to explain more variation in the mixtures.

3.7. Comparison with linear regression of crude protein

A cheap and quick method utilized by some feed mills to estimate amino acid content of feed ingredients is to calculate them from the protein values. Here we performed a linear regression between crude protein and amino acids for the total calibration set. Statistics for linear regression and the covariance (CV) calculated for the calibration set are shown in Table 6. The composition of protein is not similar over all the sample types as the CV for crude protein and all the amino acids vary markedly from 39% to 94%. Consequently, the RSQ_{reg} is lower than or equal to the RSQ from NIRS total equations. The mean of the difference between the measured values and the predicted values from the protein regression equation and the NIRS total equation is shown in Fig. 2.

Table 6
Linear regression of amino acids to crude protein for samples used in the total calibration (CAL).

Item	Sample statistics		Linear regression to CP		
	N^{CAL}	CV, %	Intercept	Slope	RSQ_{reg}
ALA	387	64.38	−0.51	0.44	0.94
ARG	387	83.81	−4.03	0.08	0.91
ASP	387	93.94	−7.32	0.12	0.93
CYS	395	48.48	0.76	0.01	0.83
GLU	387	48.93	8.14	0.17	0.86
GLY	387	66.67	−0.89	0.05	0.95
HIS	387	67.50	−0.56	0.03	0.96
ILE	387	72.06	−1.57	0.05	0.97
LEU	387	63.64	−0.94	0.08	0.96
LYS	395	84.81	−3.08	0.06	0.87
MET	395	57.14	0.23	0.01	0.85
PHE	387	59.26	−0.11	0.05	0.96
PRO	387	38.52	6.85	0.04	0.56
SER	387	65.00	−0.83	0.05	0.97
THR	395	67.21	−0.93	0.04	0.96
TRP	163	72.73	−0.15	0.01	0.94
TYR	369	70.91	−0.81	0.04	0.95
VAL	387	61.18	−0.38	0.05	0.98
CP	607	47.68	–	–	–

CP = crude protein; N = the number of samples; CV = (the SD of the measured values/mean of measured values) × 100; RSQ = regression coefficient; ALA = alanine; ARG = arginine; ASP = aspartic acid; CYS = cysteine; GLU = glutamic acid; GLY = glycine; HIS = histidine; ILE = isoleucine; LEU = leucine; LYS = lysine; MET = methionine; PHE = phenylalanine; PRO = proline; SER = serine; THR = threonine; TRP = tryptophan; TYR = tyrosine; VAL = valine.

The NIRS predictions are superior to protein regression for most amino acids apart from tryptophan and had a similar accuracy in cysteine, histidine, leucine, serine and valine. In other studies, NIRS predictions have been found to be more accurate than protein regression in various cereals (Fontaine et al., 2002) high protein feedstuffs (Fontaine et al., 2001) and poultry diets (van Kempen and Bodin, 1998). If the amino acids and crude protein are highly correlated as it may be within specific groups, the relationship could be predicted as well as NIRS but overall NIRS is a more accurate method as it derives more information from the spectra than protein alone can explain.

3.8. Prediction of amino acid digestibility

In a nutritional context, it is important to know the digestibility, as the capacity of protein sources to meet the demand of animals is determined not only by the absolute and relative quantities of dietary indispensable amino acids in feed but also the digestibility of the protein in the gastrointestinal tract, and the bioavailability of amino acids (FAO, 2013). Estimates from NIRS calibrations has been used to predict the digestibility of poultry feed ingredients for some amino acids, allowing for more precise formulations on different batches (van Kempen et al., 1996; van Kempen and Simmins, 1997; Hoehler et al., 2005) than is possible using crude protein as a proxy (van Kempen and Simmins, 1997). A preliminary study with 20 barley samples demonstrated NIRS to be able to predict the digestibility of lysine, methionine and cysteine in pigs (Pujol et al., 2007) but the current study is the first to apply NIRS to predict amino acid digestibility of different cereals, supplemental feed ingredients and mixed diets. In the nutrition of pigs, lysine, the sulphur-containing amino acids and threonine are in most cases the limiting amino acids and the prediction of these amino acids is therefore more important than the prediction of the dispensable amino acids. Summary statistics for the predicted apparent ileal digestibility of crude protein and amino acids are shown in Table 7 for 151 samples which have been evaluated in experiments with ileal cannulated pigs. The feedstuffs covered a large range of digestibility, especially for the dispensable amino acids glutamic acid and proline. The RSQ values are expected to be lower for the predictions of digestibility than the quantity of amino acids as the digestibility not only measures the contribution from the feedstuff but also the contribution from the variability in the pigs ability to digest food. This is determined by the structure of the protein and the presence of antinutritional factors as well as the influence of the endogenous secretion. Despite this, the RSQ for the ileal digestibility of most of the amino acids is in general high, with values mostly >0.75 and with low SECV values (<5%). It is particularly encouraging that the RSQ of lysine, methionine and threonine all are higher than 0.83. The developed models can therefore be considered good enough to give useable estimates on the ileal digestibility and although the modeling have been performed on apparent ileal digestibility data we expect the results to be transferable to standardized ileal digestibility (SID) of amino acids as the conversion from apparent to SID values are done by factors (Stein et al., 2007). The predictions for total tract digestibility (Table 8) have similar ranges as of the apparent ileal digestibility but the RSQ values are overall better for the total tract than the apparent ileal digestibility parameters, however, the SECV are similar. The most likely reason for that is that the total tract digestibility is a closer reflection of the influence of the feed whereas with the ileal digestibility the endogenous contribution would be larger. Moreover, sampling of ileal digesta is more variable compared to sampling of feces.

Tryptophan had lower RSQ values than the other amino acids for both the ileal and total tract digestibility. Tryptophan had less than

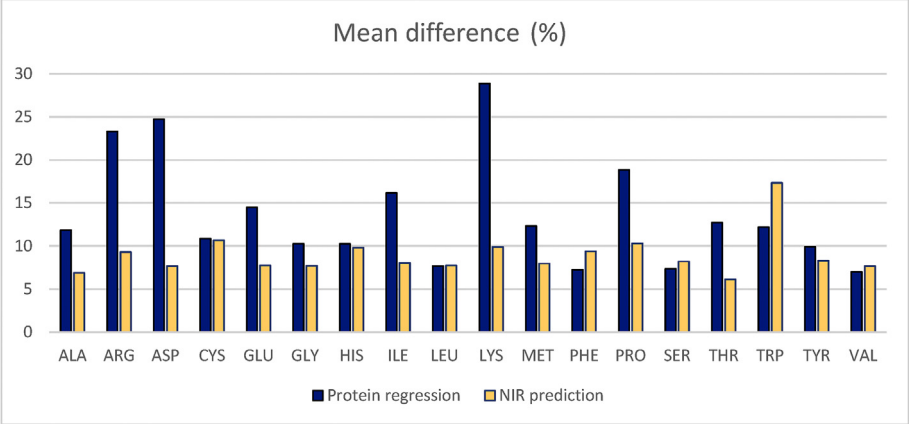

Fig. 2. Mean of the difference in amino acids predicted by protein linear regression or by NIRS calibration model compared to the measured value for the total dataset. Mean difference = Mean of [|Labn − P-Regn|/Labn × 100] or Mean of [|Labn − NIRSn|/Labn × 100], where Labn is the measured value, P-Regn is the value predicted by protein regression and NIRSn is the value predicted by the NIRS calibration.

Table 7
Summary statistics for prediction of the apparent ileal digestibility (percentage of intake) of crude protein and amino acids in pig feedstuffs.

Item	N	Min	Max	Mean	SD	Factors	SEC	RSQ	SECV
CP	144	34.9	93.7	73.4	7.2	6	3.12	0.81	4.03
Indispensable amino acids									
LYS	102	62.4	97.7	84.0	7.9	8	2.60	0.89	3.67
MET	102	55.0	97.6	85.4	6.2	9	2.31	0.86	4.17
CYS	102	15.0	93.1	74.4	10.1	4	4.99	0.76	6.04
THR	102	47.4	90.4	72.9	8.3	7	3.41	0.83	4.68
ILE	94	64.1	94.2	80.2	5.7	4	2.73	0.77	3.25
LEU	94	67.0	96.7	82.5	5.8	7	2.13	0.86	2.90
HIS	94	64.3	97.1	83.5	5.2	6	2.04	0.85	2.67
PHE	94	54.2	96.1	83.1	6.6	5	2.68	0.83	3.22
VAL	94	61.4	94.7	78.4	6.3	9	2.21	0.88	3.64
ARG	94	70.6	96.3	87.7	5.5	9	2.03	0.86	3.42
TRP	38	66.5	94.7	81.8	7.0	2	3.98	0.68	4.93
Dispensable amino acids									
ALA	94	50.7	91.4	72.4	7.0	7	3.47	0.76	4.79
ASP	94	58.9	93.2	74.8	7.4	4	3.72	0.75	4.39
GLU	94	62.9	95.6	87.6	5.4	5	2.20	0.83	3.07
GLY	94	23.5	85.2	64.7	11.1	5	6.09	0.70	7.68
PRO	94	6.5	96.6	79.3	13.1	8	4.90	0.86	8.54
SER	94	49.4	93.3	78.0	8.2	5	3.65	0.80	4.15
TYR	80	53.9	93.2	82.3	6.6	4	2.96	0.80	3.43

N = the number of samples; SEC = standard error of calibration; RSQ = regression coefficient; SECV = standard error of cross validation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

Table 8
Summary statistics for prediction of total tract digestibility (percentage of intake) of crude protein and amino acids in pig feedstuffs.

Item	N	Min	Max	Mean	SD	Factors	SEC	RSQ	SECV
CP	151	35.5	96.3	80.8	8.0	10	2.38	0.91	3.68
Indispensable amino acids									
LYS	72	24.3	97.8	79.0	11.8	8	3.16	0.93	5.36
MET	72	19.0	97.4	78.7	11.6	8	3.21	0.92	6.09
CYS	72	22.8	92.0	84.2	8.2	8	1.61	0.96	4.47
THR	72	37.3	96.1	78.3	9.8	7	3.23	0.89	4.95
ILE	64	37.5	95.9	80.4	10.0	7	2.47	0.94	3.77
LEU	64	47.0	97.5	83.7	8.4	6	2.23	0.93	3.48
HIS	64	54.5	98.0	87.8	6.8	7	1.69	0.94	3.09
PHE	64	40.6	98.1	84.2	8.7	8	1.68	0.96	2.76
VAL	64	42.4	96.8	81.4	9.0	7	2.33	0.93	3.56
ARG	64	56.4	96.6	88.0	6.5	8	1.61	0.94	2.71
TRP	23	66.2	94.5	85.3	7.2	4	2.48	0.88	6.20
Dispensable amino acids									
ALA	64	34.2	93.6	74.5	11.0	5	3.58	0.89	5.02
ASP	64	46.1	95.6	78.5	10.8	4	3.73	0.88	4.19
GLU	64	44.0	97.4	90.7	7.0	8	1.11	0.97	3.96
GLY	64	34.0	92.2	78.8	9.5	7	2.60	0.92	4.53
PRO	64	60.5	98.6	90.9	5.7	7	1.25	0.95	2.06
SER	64	47.6	95.0	84.0	8.0	7	2.01	0.94	2.80
TYR	57	27.4	95.0	80.4	10.4	7	2.37	0.95	4.96

N = the number of samples; SEC = standard error of calibration; RSQ = regression coefficient; SECV = standard error of cross validation; CP = crude protein; LYS = lysine; MET = methionine; CYS = cysteine; THR = threonine; ILE = isoleucine; LEU = leucine; HIS = histidine; PHE = phenylalanine; VAL = valine; ARG = arginine; TRP = tryptophan; ALA = alanine; ASP = aspartic acid; GLU = glutamic acid; GLY = glycine; PRO = proline; SER = serine; TYR = tyrosine.

half the number of samples than the other amino acids as this amino acid has to be analyzed separately (alkaline instead of acid hydrolysis) (Mason et al., 1980) and therefore was not included in all studies.

3.9. Sources of error

To develop accurate NIRS calibrations it is essential to have very accurate reference measurements. Measurements done in the same laboratory with a standardized protocol are therefore preferable. This is also almost what has been the case in the current project but as the samples have been collected and measured over a long period of time small deviations due to improvements in methods and instrumentations have occurred. It is also generally recommended to scan the samples within a small timeframe relative to the measurements taken by wet chemistry. This was not the case in the present study as some of the samples have been stored for a very long time and the chemical analysis completed at the time of the experiment, the earliest in 1975 whereas the NIRS scans were all performed in 2018. The high accuracy of the developed calibration equations, however, demonstrates that historical samples can be successfully used. To check if the nutritional composition of the stored samples had altered over time we used crude protein as a marker. The crude protein content was remeasured from 20 selected samples, stored from 22 to 43 years (Appendix Table 1). The new crude protein measurements were very close to the original measurement for 17 of the samples (5% difference or less) whereas 3 samples had a larger discrepancy of 7% to 15% between measurements. However, the new measurement was higher than the original indicating that crude protein had not been lost during

storage. Overall, however, it is our belief that the SEP, SECV and RSQ values could have been improved even further if the NIRS scans were undertaken at the same time as the chemical analysis.

4. Conclusions

Precise and accurate estimates of crude protein and amino acids can be made on both raw ingredients and diet mixtures of plant-based pig feeds with one calibration model. For some amino acids, however, increased precision could be obtained by using groups that are more specific but as these groups also contained a mixture of sample types a fair comparison of specific versus general could not be made. Overall, the models developed with the total sample dataset provide greater robustness, i.e. able to be used on a wider range of measured values and sample types, and good accuracy. We also found that useable predictions of the digestibility of amino acids could be made for both ileal and total digestibility, which can be used to guide the formulation of pig feed to meet nutrient requirements allowing for better production and less environmental pollution.

Author contributions

S.J. Noel is responsible for the NIR analysis, modeling and writing the draft manuscript. **H.J.H. Jørgensen** and **K.E.B. Knudsen** are responsible the construction of the sample database, performing animal feeding experiments, review and editing of the draft manuscript. **K.E.B. Knudsen** is responsible for project design and obtaining funding.

Conflict of interst

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the content of this paper.

Acknowledgements

The authors are grateful for the technical assistance of Lisbeth Marcher and Winnie Østergaard Thomsen. This project was part of the Feed-a-Gene Project and has received funding from the European Union's H2020 Program under grant agreement no 633531. The funding body had no role in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

Appendix

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aninu.2021.07.004.

References

Barnes R, Dhanoa MS, Lister SJ. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. Appl Spectrosc 1989;43:772–7.
Chen GL, Zhang B, Wu JG, Shi CH. Nondestructive assessment of amino acid composition in rapeseed meal based on intact seeds by near-infrared reflectance spectroscopy. Anim Feed Sci Technol 2011;165:111–9.
Chen L, Yang Z, Han L. A review on the use of near-infrared spectroscopy for analyzing feed protein materials. Appl Spectrosc Rev 2013;48:509–22.
Commission E. Commission Directive 98/64/EC of 3 September 1998: establishing community methods of analysis for the determination of amino-acids, crude oils and fats, and olaquindox in feedingstuffs, and amending Directive 71/393/EEC. In: Official journal of the European communities eur-lex; 1998.
Commission E. Commission Regulation (EC) No 152/2009 of 27 January 2009 laying down the methods of sampling and analysis for the official control of feed. Official Journal of the European Union 2009;54:1–130.
De La Haba J, Garrido-Varo A, Guerrero-Ginel JE, Perez-Marin DC. Near-infrared reflectance spectroscopy for predicting amino acids content in intact processed animal proteins. J Agric Food Chem 2006;54:7703–9.
Escuredo O, Martin MIG, Moncada GW, Fischer S, Hierro JMH. Amino acid profile of the quinoa (Chenopodium quinoa Willd.) using near infrared spectroscopy and chemometric techniques. J Cereal Sci 2014;60:67–74.
FAO. In: Dietary protein quality evaluation in human nutrition. Report of an FAO Expert Consultation. Rome: FAO Food and Nutrition Paper no 92; 2013. p. 67.
Fearn T. Assessing calibrations: SEP, RPD, RER and R 2. NIR News 2002;13:12–3.
Fearn T. The overuse of R 2. NIR News 2014;25. 32-32.
Fernandez JA, Jorgensen H, Just A. Comparative digestibility experiments with growing-pigs and adult sows. Anim Prod 1986;43:127–32.
Fontaine J, Horr J, Schirmer B. 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 2001;49:57–66.
Fontaine J, Schirmer B, Horr J. 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 2002;50:3902–11.
Gonzalez-Martin I, Alvarez-Garcia N, Gonzalez-Cabrera JM. Near-infrared spectroscopy (NIRS) with a fibre-optic probe for the prediction of the amino acid composition in animal feeds. Talanta 2006;69:706–10.
Han IK, Lee JH, Piao XS, Li DF. Feeding and management system to reduce environmental pollution in swine production - Review. Asian-Australas J Anim Sci 2001;14:432–44.
Hoehler D, Goodson J, Fontaine J, Jaeger A, Schirmer B. NIR spectroscopy for prediction of amino acids in feed ingredients 66th Minnesota nutrition conference and technical symposium: future of corn in animal feed. 2005. St. Paul, Minnesota.
ISO. ISO 13903 animal feeding stuffs: determination of amino acids content. International Organization for Standardization; 2005.
ISO. ISO 12099:2017 Animal feeding stuffs, cereals and milled cereal products – Guidelines for the application of near infrared spectrometry. International Organization for Standardization; 2017. www.iso.org/standard/67352.html.
Jezierny D, Mosenthin R, Bauer E. The use of grain legumes as a protein source in pig nutrition: a review. Anim Feed Sci Technol 2010;157:111–28.
Just A. The influence of crude fiber from cereals on the net energy value of diets for growth in pigs. Livest Prod Sci 1982;9:569–80.
Just A, Fernandez JA, Jorgensen H. The net energy value of diets for growth in pigs in relation to the fermentative processes in the digestive-tract and the site of absorption of the nutrients. Livest Prod Sci 1983a;10:171–86.
Just A, Jørgensen H, Fernández JA, Enggaard Hansen N. Forskellige foderstoffers kemiske sammensætning, fordøjelighed, energi- og proteinværdi til svin (The chemical composition, digestibility, energy and protein value of different feedstuffs for pigs). Denmark: National Institute of Animal Science; 1983. p. 99.
Just A, Jorgensen H, Fernandez JA. Maintenance requirement and the net energy value of different diets for growth in pigs. Livest Prod Sci 1983b;10:487–506.
Just A, Jorgensen H, Fernandez JA. Correlations of protein deposited in growing female pigs to ileal and fecal digestible crude protein and amino-acids. Livest Prod Sci 1985;12:145–59.
Kovalenko IV, Rippke GR, Hurburgh CR. Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy. J Agric Food Chem 2006;54:3485–91.
Mason VC, Rudemo M, Bechandersen S. Hydrolysate preparation for amino-acid determinations in feed constituents .6. Influence of phenol and formic-acid on the recovery of amino-acids from oxidized feed proteins. Z Tierphysiol Tierer 1980;43:35–48.
Millet S, Aluwe M, Van den Broeke A, Leen F, De Boever J, De Campeneere S. Review: pork production with maximal nitrogen efficiency. Animal : an international journal of animal bioscience 2018;12:1060–7.
Moughan PJ, Deglaire A, Wolfe RR. Amino acid metabolism – an overview. In: Moughan PJ, Hendriks WH, editors. Feed evaluation science. The Netherlands: Wageningen Academic Publishers; 2018a. p. 219–48.
Moughan PJ, Miner-Williams W, Jha R. The digestion of protein - amino acid digestibility. In: Moughan PJ, Hendriks WH, editors. Feed evaluation science. The Netherlands: Wageningen Academic Publishers; 2018b. p. 173–217.
Pazdernik DL, Killam AS, Orf JH. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. Agron J 1997;89:679–85.
Perez-Marin D, Garrido-Varo A, Guerrero JE. Non-linear regression methods in NIRS quantitative analysis. Talanta 2007;72:28–42.
Pujol S, Perez-Vendrell AM, Torrallardona D. Evaluation of prediction of barley digestible nutrient content with near-infrared reflectance spectroscopy (NIRS). Livest Sci 2007;109:189–92.
Qiao Y, van Kempen TATG. Technical note: comparison of Raman, mid, and near infrared spectroscopy for predicting the amino acid content in animal meals. J Anim Sci 2004;82:2596–600.

Rosenfelder P, Mosenthin R, Spindler HK, Jorgensen H, Knudsen KEB, Sauer N, Htoo JK, Eklund M. Standardized ileal digestibility of amino acids in eight genotypes of soft winter wheat fed to growing pigs. J Anim Sci 2015;93:1133–44.
Sapienza D, Berzaghi P, Martin N, Taysom D, Owens F, Mahanna B, Sevenich D, Allen R. NIRS white paper. Near infrared spectroscopy for forage and feed testing, NIRS consortium. US: Universtiy of Wisconsin-Extension; 2008.
Shenk J, Westerhaus M. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop Sci 1991;31:469–74.
Shenk JS, Workman JJ, Westerhaus MO. Application of NIR spectroscopy to agricultural products. In: Burns DA, Ciurczak EW, editors. Handbook of near infrared analysis, practical spectroscopy series; 2001. p. 419–74.
Spindler HK, Mosenthin R, Rosenfelder P, Jorgensen H, Bach Knudsen KE, Sauer N, Htoo JK, Eklund M. Standardized ileal digestibility of amino acids in eight genotypes of barley fed to growing pigs. Animal : an international journal of animal bioscience 2016;10:1931–40.
Stein HH, Fuller M, Moughan P, Sève B, Mosenthin R, Jansman A, Fernández J, de Lange C. Definition of apparent, true, and standardized ilealdigestibility of amino acids in pigs. Livest Sci 2007;109:282–5.
van Barneveld RJ. Chemical and physical characteristics of grains related to variability in energy and amino acid availability in pigs: a review. Aust J Agric Res 1999;50:667–87.
van Kempen T, Bodin JC. 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 1998;76:139–47.
van Kempen T, Poulenc R, France A. NIR technology: can we measure amino acid digestibility and energy values, 12th Annual Carolina Swine Nutrition Conference. 1996.
van Kempen TA, Simmins PH. Near-infrared reflectance spectroscopy in precision feed formulation. J Appl Poultry Res 1997;6:471–7.
Wang L, Wang Q, Liu HZ, Liu L, Du Y. Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. J Sci Food Agric 2013;93:118–24.
Wu JGG, Shi CH, Zhang XM. Estimating the amino acid composition in milled rice by near-infrared reflectance spectroscopy. Field Crop Res 2002;75:1–7.
Yang XS, Wang LL, Zhou XR, Shuang SM, Zhu ZH, Li N, Li Y, Liu F, Liu SC, Lu P, Ren GX, Dong C. Determination of protein, fat, starch, and amino acids in foxtail millet [setaria italica (L.) beauv.] by fourier transform near-infrared reflectance spectroscopy. Food Sci Biotechnol 2013;22:1495–500.
Zhang B, Rong ZQ, Shi Y, Wu JG, Shi CH. Prediction of the amino acid composition in brown rice using different sample status by near-infrared reflectance spectroscopy. Food Chem 2011;127:275–81.
Zhou LJ, Zhang LY, Zhang EX, Li JT, Yang WJ, Wang ZY. 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 2012;175:198–202.