Comparisons of commercially available NIRS-based analyte predictions of haylage quality for equid nutrition

by Le Cocq, K., Harris, P., Bell, N., Burden, F.A., Lee, M.R. and Davies, D.R.

Le Cocq, K., Harris, P., Bell, N., Burden, F.A., Lee, M.R. and Davies, D.R. (2022) 'Comparisons of commercially available NIRS-based analyte predictions of haylage quality for equid nutrition', *Animal Feed Science and Technology*, *283* (115158).

Comparisons of commercially available NIRS-based analyte predictions of haylage quality for equid nutrition

Kate Le Cocq[1 † *], Paul Harris[1], Nikki Bell[2], Faith A Burden[2], Michael R.F. Lee[1,3†], David R. Davies[1,4]

[1]Rothamsted Research, North Wyke, Okehampton, Devon, EX20 2SB, United Kingdom

[2]The Donkey Sanctuary, Sidmouth, Devon EX10 0NU, United Kingdom.

[3]University of Bristol, Bristol Veterinary School, Langford, Somerset, BS40 5DU, United Kingdom

[4]Silage Solutions Ltd. Bwlch y Blaen, Pontrhdygroes, Ystrad Meurig, Ceredigion, SY25 6DP, United Kingdom

[†] Current address: Harper Adams University, Newport, Shropshire, TF10 8NB

Declarations of interest: none

Keywords: forage, diet, donkey, equine, analysis, calibration, ration, model.

Abbreviations. ADF, acid detergent fibre expressed inclusive of residual ash; aNDF, neutral detergent fibre assayed with a heat stable amylase and expressed inclusive of residual ash; cm, centimetre; DM, dry matter; FM, fresh matter; g, gram; h, hour; kg, kilogram; ml, millilitre; n, nanometre;; NIRS, Near-Infrared Reflectance Spectroscopy; PC, principal component; PCA, principal component analysis; s, seconds; SEM, standard error of the mean; SD, standard deviation; WSC, water soluble carbohydrate;

*Correspondence to: Harper Adams University, Newport, Shropshire, TF10 8NB, United Kingdom. Tel +44 01952 820280. E-mail address klecocq@harper-adams.ac.uk

Abstract

Maintaining animal health and performance relies on the availability of an appropriate diet. For herbivores, accurate assessment of forage nutrient quality is critical for appropriate diet formulation and rationing, including potential supplementation. Near-Infrared Reflectance Spectroscopy (NIRS) is a rapid method that is used in place of traditional chemical methodologies (wet chemistry) to predict analyte contents in forage samples. The method relies on scanning a sample with near-infrared light and predicting the analyte content by comparing the reflected spectra to a model which has been developed with samples of known analyte content measured by wet chemistry. The purpose of this study was to examine the accuracy of four NIRS-based methods on haylage from seven farm holdings compared with wet chemistry (the control). We analysed 64 samples for a range of analytes (dry matter (DM), pH, ash, acid detergent fibre expressed inclusive of residual ash (ADF), neutral detergent fibre assayed with a heat stable amylase and expressed inclusive of residual ash (aNDF), crude protein and water-soluble carbohydrate (WSC)) commonly assessed for haylage quality in equid nutrition. We compared results obtained by wet chemistry to corresponding NIRS-based predictions from four commercially available NIRS services. The results revealed large discrepancies amongst all five methods. For DM, average bias (mean±SD) for three reported methods was -15.5±188.4, -10.1±50.4, 12.9±33.8 g/kg respectively and for WSC reporting positive bias from four methods of 26.9±51.3, 24.8±38.2, 26.2±50.1 and 14.5±45.2, g/Kg respectively. The extent of these discrepancies from the wet chemistry also varied by analyte where for example, predictions for DM were more reliable than those for WSC and results demonstrated that predictions obtained by NIRS could result in feeding forage outside of target nutritional values.

Introduction

Near-Infrared Reflectance Spectroscopy (NIRS) is commonly used for analysis of forages in the place of traditional chemical methodologies (wet chemistry). It is a fast, non-destructive method which allows many samples to be processed with immediate quantitative results, thus increasing the chance for data capture on a high number of samples (Shenk and Westerhaus, 1994). This can provide critical information for quality of feedstuffs, allowing quality control measures to be undertaken before samples are consumed or dispatched (Batten, 1998). The use of NIRS in agriculture has increased with the development of handheld and portable NIRS devices including the capture of live measurements to provide nutritive information during harvesting of crops, blending of total mixed rations (Evangelista et al., 2021; Montes and Paul, 2008; Piccioli-Cappelli et al., 2019) or analysis of forage quality and composition (Shenk and Westerhaus, 1994). Several commercial laboratories offer NIRS

analysis for samples received by post which enables individual livestock owners to access sample services which provide a range of forage analytes for as little as £15 Sterling per sample.

Forage analysis using NIRS methodologies have emerged as the dominant technology for routine analysis in the livestock and equine sector. Forage for equids is usually produced in bales and variation can be high between individual units even if they originate from the same batch or field (Sheaffer et al., 2000). For equids, forage fibre, water-soluble carbohydrate (WSC) and crude protein are useful nutritional analytes for diet formulation, especially if an all-forage diet is fed. In some countries feeding wrapped forages in bales has partially or totally replaced hay in equine diets (Müller, 2018). Therefore, it is advantageous to have inexpensive, fast and accurate measurement techniques which can be used to aid decision making before offering forages. Commercial laboratories offer services on fresh samples that can return results within 24 h of receipt, making it a convenient tool to aid decision making. Recently, Harris et al. (2018) reported a dataset comparing sample analysis of dried, ground NIRS to that of wet chemistry for 52 haylage samples. The study found good correlations between mean measurements made by wet chemistry and NIRS for paired subsamples, although on an individual sample basis for WSC, discrepancies of up to 20% were observed.

NIRS is a method based on the scanning of a sample with near-infrared light. The absorbance, typically from *ca*. 400- 2500 nm range of wavelengths (λ) for laboratory-based instruments is compared with a calibration developed from the known values (by wet chemistry) for samples of the same type and preparation. A calibration model is then built from the observed relationships and then used for NIRS-based analyte prediction (Reeves III, 2000). There are many factors that affect the final result obtained by NIRS for biological samples including: the state in which the sample is scanned (fresh, dried, ground), the NIRS instrument make and model and scanning methodology/vessel, the way the sample is taken and homogenised as a representative sample of the unit (e.g. bale, batch or field), and the calibration model itself. Since methods for the development of the calibration models are not standardised, there is a possibility that inaccurate results can be obtained if the sample submitted differs from the range of sample compositions included to build the calibration model (Andueza et al., 2011). As sample analysis for both scientific and animal nutrition purposes are increasingly reliant on commercially available NIRS services (with 'closed' or little detail on the calibration models), there is a need to ensure that reliable, consistent results are being obtained. Thomson et al. (2018) and Harris et al. (2018) highlighted the issue that in different forage types, NIRS predictions may differ from the results obtained by wet chemistry. Our study's aim was to build on this previous work where we analysed data for a range of analytes (dry matter (DM), pH, ash, acid detergent fibre expressed inclusive of residual ash (ADF), neutral detergent fibre assayed with a heat stable

amylase and expressed inclusive of residual ash (aNDF), crude protein and WSC) commonly assessed for haylage quality in equid nutrition on fresh scanned samples. We compared the analyte data obtained by wet chemistry (which acts as our control or 'true' measurements), to corresponding NIRS-based predictions from four commercially available NIRS services.

Materials and Methods

Experimental design and sampling

In March 2018, 77 haylage samples were obtained from seven holdings that produce haylage for Donkey consumption in Dorset and Devon, UK, where 66 samples were from individual bales. Each bale was sampled according to O'Brien et al. (2006), by taking six spatially distinct cores using a mechanical silage corer (Dairy One forage Lab, Ithaca, NY, USA) from the top to bottom. Each core sample was taken to a depth of 30 cm on the barrel side along a diameter transect of the bale. The sample was removed and the 6 samples from each bale were mixed to ensure homogeneity within the composite sample. Samples were immediately scanned with a commercial handheld NIRS device and the raw spectra used to predict analytes with two different calibration models (methods B and C) for bale and clamp silage. A sub sample was sent fresh for wet chemistry analysis (method A), and for further NIRS-based analyte predictions at commercial laboratories either on fresh material (method D) or dried and ground material (method E) for haylage nutrition analysis. The measured analytes assessed were: DM, pH, ash, ADF, aNDF, crude protein and WSC.

The commercial labs did not offer a complete dataset of NIRS-based analyte predictions, where DM, pH and ADF were absent for method E and pH and ADF were absent for method D. Two samples were also subsequently rejected as their replicate measurements for DM were outside of the recommended range indicating a sampling or measurement error, giving a final study dataset of 64 samples.

Wet chemistry (method A)

pH was determined by measurement of sample supernatant using a pH electrode at room temperature. Supernatant was obtained through adding 90 ml milli-Q water to a 10 g subsample of fresh matter (FM) which was agitated for 10 min (Merry et al., 1995). Further chemical analyses for wet chemistry were conducted by Sciantec Analytical (Cawood Scientific limited, UK) under the following standard operating procedures and methods. Samples were oven dried at 60 °C initially and then subsampled for chemical analysis outlined below. To determine the DM content of material, a further subsample was placed at 105 °C until no further loss of weight

was recorded. Dried material was then milled to 1 mm and 3 g of the resulting material was furnaced (505℃ for 12 h ramp rate 2℃/min) to determine the ash content. Forage that had been dried at 60 °C and subsequently ground was used to determine aNDF (assayed with a heat stable amylase and expressed inclusive of residual ash) (SOP S1012) and ADF (expressed inclusive of residual ash) (SOP S1129) as described in Davies et al., (1998) with the exception of using oven dried, not freeze dried material. Samples were defatted with acid detergent then starch transformed to soluble sugars by treating with α-amylase. The soluble material was removed by boiling in neutral solution and the remaining insoluble material was weighed to determine the aNDF. From the soluble material, acid detergent fibre was determined using the Ankom 220 analyser (ANKOM Technology Corp., Macedon, NY, USA). Total N content of the forage was determined by the Kjeldahl technique (FOSS Kjeltec 8400 analyser, Foss Co. Ltd, Denmark). Crude protein (SOP S1113) was calculated from total N content multiplied by 6.25 as described in Merry et al. (1995). The anthrone technique was used to measure WSC concentration of the sample (SOP S1030) (Davies et al., 1998; Merry et al., 1995). Briefly, the sample was agitated in distilled water and filtered to remove particulate material. An aliquot of the resulting supernatant was mixed with anthrone. The absorbance of the solution was measured at 625 nm by spectrophotometer and compared to a standard curve of glucose standards.

NIRS-based predictions (methods B to E)

Methods B and C used a handheld microNIR spectrometer (VIAVI Solutions Inc., Scottsdale, Arizona, USA) which is a mobile on farm device to scan the samples. Each sample was prepared following the supplier's protocol and scanned in the fresh form without any processing of the mixed cored sample, which involved a total scan time of 50 seconds split into 5 separate scan times of 10 s with mixing of the sample between each scan. The scan data collected was then fitted to one of two models (bale (method B) or clamp (method C)). Thus, exactly the same scan data was used for both models in this case. The sample scanned for methods B and C was then subsampled for the remainder of the analysis. For method D the sample was scanned by a commercial laboratory that scans using a FOSS NIR 6500 (Foss Co. Ltd, Denmark). This sample was also scanned in the fresh form with no post sample processing. Method E was scanned by a different second commercial laboratory using a bench top FOSS 5000 (Foss Co. Ltd, Denmark) NIRS instrument, however this laboratory oven dries and grinds the sample prior to scanning with NIRS and as such does not provide a NIRS prediction for DM. The laboratories are independent of each other and therefore have their own unique set of prediction models that were used to describe the NIRS data.

141

142 *Statistical analysis*

143 The analyte datasets were summarised by their mean, standard error of the mean (SEM) and data range, together
144 with the presentation of analyte boxplots conditional to each method (A to E) used. Next for each analyte set, a
145 linear correlation analysis was conducted, where coefficients (r) should tend to +1 if each NIRS-based method (B
146 to E) provided predictions that tend to those using wet chemistry (method A). An indication of which analyte
147 tended to provide the most coherent outputs across all methods was found via a principal component analysis
148 (PCA) (see Joliffe, 2002) on each analyte set, and by reporting the percentage variance explained for the first
149 principal component (PC), only. Analytes with similar outputs across all methods will provide percentages
150 approaching 100%, while those with dissimilar outputs will provide percentages approaching 0%. Given both the
151 correlation analyses and the PCAs are invariant to changes in scale in the data (i.e. constant upward or downward
152 shifts in the data would go unnoticed), a series of linear regression analyses were conducted, in a pairwise fashion
153 (i.e. Method A to B, C, D and E, in turn), similar to that done for the correlations. For each regression, the estimated
154 intercept and slope parameters should respectively equal 0 and +1 for exact 1:1 correspondence. Results (p-values)
155 from a linear hypothesis test were reported comparing this ideal model with the estimated model using a finite
156 sample F test (Fox, 2015).

157 Finally, a series of Bland-Altman plots (Bland and Altman, 2007) were found to statistically and visually
158 determine the level to which methods B to E (NIRS-based predictions) agreed with method A (wet chemistry).
159 Here the difference between the wet chemistry and an NIRS-based prediction was plotted (on the y-axis) against
160 the average of the same two outputs (on the x-axis). Bland-Altman plots were given with a series of y-axis
161 thresholds to guide interpretations, these were: (a) a threshold of 0, where exact agreement between methods
162 occurs if all points coincide with this line (i.e. zero bias); (b) the mean bias (i.e. mean of the differences) from the
163 zero line of (a), together with associated 95% confidence intervals; (c) mean-based upper and lower agreement
164 thresholds together with associated 95% confidence intervals; and (d) median-based upper and lower agreement
165 thresholds but without associated confidence intervals. Median-based thresholds provided a useful guide to the
166 influence of outliers on the Bland-Altman analysis, complementing the more usual, mean-based ones. The Bland-
167 Altman analysis was also presented with the standard deviation (SD) of the bias (not shown on the plots), as well
168 as the mean bias in (b). All statistical analyses were conducted within the R statistical computing environment

(version 3.6-3), where the Bland-Altman plots and associated diagnostics were found using and adapting functions provided in the blandr R package (Datta, 2017).

Results

Figure 1 summarises the analyte data using boxplots conditioned by the five methods (A to E), where these are reviewed below, together with the data summaries and Bland-Altman plots. Figure 2 provides the correlation coefficients for each analyte set, where each cell of each correlation matrix should reflect expected positive correlations between methods. Correlation coefficients worth noting are those > +0.5, and always including method A (the wet chemistry control). This subset consisted of method A in comparison to: methods C and D for DM; method C for ash; method D for aNDF; methods C, D and E for crude protein; and methods C and E for WSC. The correlation analysis indicated that NIRS-based predictions for pH (Figure 2b) and for ADF (Figure 2d) were of no value, where the corresponding scatterplots (not shown) for these weak positive or even negative correlations, confirmed such poor relationships. The correlation analysis also indicated that method B was of little use for predicting all seven analytes, as correlations with method A ranged from +0.29 for crude protein to -0.46 for aNDF.

The percentage of variance explained within the first PC for each analyte group was 53.8%, 35.9%, 32.6%, 38.6%, 32.4%, 35.2% and 33.1% for DM, pH, ash, ADF, aNDF, crude protein and WSC, respectively. This means that NIRS-based predictions for DM were the most likely to be consistent with each other and with the wet chemistry (as highest percentage), while this outcome was least likely for aNDF (as lowest percentage). The results (*p*-values) from the linear hypothesis tests indicated all method comparisons resulted in a strongly significant deviation (at the 99.99% level and greater) from the ideal 1:1 (45^{o}) regression line from the origin, except that for method A in comparison to method C for DM (with a *p*-value = 0.15). These test results confirmed observations from the correlation analyses, above, where NIRS-based predictions for pH and for ADF were of no value, and method B was of little use for predicting all seven analytes. Surprisingly, given the strong correlation in Figure 2a, method A in comparison to method D for DM significantly deviated from the 1:1 line. This was due to a relatively large negatively estimated intercept resulting in Method D consistently under-predicting DM.

Figures 3 to 7 provide the Bland-Altman plots for the seven analytes and five methods. For each analyte, the plots are given with a common *y*-axis scale to furnish objective comparisons. These results are discussed per analyte, and with reference to data summaries and the boxplots of Figure 1.

198 Dry matter was reported for four methods (A-D) with all methods showing similar means ±SEM (644±15.6,
199 659±17.0, 653±15.2, 631±12.6 g/kg for methods A to D, respectively), and ranges of (574.4, 578.4, 602.4, 540.0
200 g/kg), respectively (see also Figure 1a). Mean bias was calculated through Bland-Altman analysis (Figure 3) and
201 showed that on average, methods B and C over-predicted DM (to that found with method A, the wet chemistry)
202 with mean bias values of -15.5 and -10.1 g/kg and associated SDs of 188.4 and 50.4 g/kg, respectively. Method
203 D on average, under-predicted DM with a mean bias of 12.9 g/kg with an associated SD of 33.8 g/kg. The high
204 SD of the differences (bias) for method B indicated (and confirmed) an extremely poor performance of this
205 method, relative to methods C and D (and this was clearly reflected in Figure 3). Method D showed a trend of
206 over-predicting at the lower DM range and under-predicting in the higher DM range. Method C appeared to be
207 relatively more accurate at the extremes than method D, but this could be due of the lower number of samples
208 within this DM range.

209

210 Mean ±SEM of pH was reported for methods A-C as 5.97±0.053, 6.67±0.161, 6.51±0.149, with ranges of 3.39,
211 5.23 and 5.52 for each method, respectively (see also Figure 1b). From the Bland-Altman analysis (Figure 4a-b),
212 methods B and C tended to under-predict at low pH, while over-predicted at high pH values. Mean bias showed
213 that on average, methods B and C over-predicted pH with mean bias values of -0.71 and -0.54 and associated SDs
214 of 1.4 and 1.3, respectively. As already indicated, the NIRS-based predictions for pH were of little to no value.

215 Ash concentration was reported for all five methods (A-E) with all methods showing similar means ±SEM
216 (76.1±1.67, 74.2±2.12, 76.0±2.22, 69.2±1.10, 91.1±1.35 g/kg for methods A to E, respectively), and ranges of
217 64.00, 76.48, 93.10, 50.00, 48.97 g/kg, respectively (see also Figure 1c). Mean bias of the NIRS-based prediction
218 methods was calculated through the Bland-Altman analysis (Figure 4c-f) and showed that on average, methods
219 B, C and D under-predicted ash with mean bias values of 1.92, 0.76 and 6.88 g/kg. Method E over-predicted ash
220 with a mean bias value of -15.0 g/kg. The SDs associated with this bias were 23.5, 14.6, 12.2 and 14.2 g/kg for
221 methods B to E, respectively. Thus again, method B performed the poorest in this respect. There were no clear
222 incidences of consistent over- or under-prediction. Although samples analysed by method D fall into discrete
223 categories highlighted by the linear features in Figure 4e.

224 Predictions of ADF were reported for methods A, B and C with mean ±SEM of 344.8±3.37, 424.1±9.13, and
225 415.3±7.21, respectively (see also Figure 1d). Mean bias of the prediction methods was calculated through the
226 Bland-Altman analysis (Figure 5a-b) and showed that on average, methods B and C (quite severely) over-

predicted ADF with negative mean bias values of -79.3 and -70.4. The SDs associated with this bias were 84.8 and 64.4, respectively. From the Bland-Altman analysis, methods B and C tended to over-predict across the full range of the ADF values found from method A (as few of the plotted differences were positive). This over-prediction was most severe at high ADF values. Again, and as already indicated, the NIRS-based predictions for ADF were of little to no value.

The aNDF predictions were reported for all methods with mean ±SEM values of 629.9±5.26, 641.7±12.28, 657.3±12.54, 519.4±3.23, 572.6±2.38, respectively (see also Figure 1e). Bland-Altman plots (Figure 5c-f) showed a negative bias for methods B and C with respective mean bias values of -11.7 and -27.4 and a positive bias for methods D and E with respective mean bias values of 110.5 and 57.3. The SDs associated with this bias were 123.5, 103.2, 35.5 and 38.1, respectively. Methods B and C were both weak predictors of aNDF and tended to strongly over-predict for the highest aNDF values. Methods D and E tended to under-predict but with some over-prediction in the lower range of the aNDF values. Methods D and E performed relatively well for aNDF prediction.

Crude protein predictions were found with means ±SEM of 89.2±2.18, 108.6±5.47, 108.9±4.73, 111.4±0.99 and 131.7±1.60 reported for methods A to E, respectively (see also Figure 1f). All predictive methods had a negative mean bias, with values of -19.3, -19.6, -22.2 and -42.5, respectively (Figure 6); thus, over-prediction was more likely. The SDs associated with this bias were 42.2, 31.7, 13.2 and 14.7, respectively. Methods B and C were both weak predictors of crude protein while methods D and E performed relatively well. Again, samples analysed by method D fall into discrete categories highlighted by linear features, while the performance of method E suffered from two clear outlying predictions of crude protein.

The WSC predictions were found with means ±SEM of 113±6.12, 86.0±2.93, 88.0±2.27, 86.7±1.34, 98.5±0.78 for methods A to E, respectively (see also Figure 1g). All NIRS-based predictive methods had a positive mean bias, with values of 26.9, 24.8, 26.2 and 14.5, respectively (Figure 7). The SDs associated with this bias were 51.3, 38.2, 50.1 and 45.2, respectively. All methods tended to over-predict low WSC values, while all methods under-predicted high WSC values. Method E marginally appeared to be the best NIRS-based predictor in terms of the smallest bias coupled with second smallest variance (i.e. SD), but had the clearest (linear) trend in such bias/variance (i.e. moving from over- to under-prediction for low to high WSC).

Discussion

This paper reports a comparison of traditional wet chemistry methods with four commercial (two on-farm and two off-farm) NIRS-based predictions for seven analytes within a set of haylage samples obtained from seven holdings across Dorset and Devon, UK. The objective was to determine the degree to which each of the NIRS-based methods agreed with analysis by wet chemistry and examine the value of NIRS methodologies as an alternative to wet chemistry. We found that the accuracy of the results obtained by NIRS varied depending on the NIRS method and the analyte when compared with the wet chemistry. Other studies have addressed similar questions in other forage types (Davies et al., 2012; Thomson et al., 2018) and other preparation methods (Alomar et al., 2003). Harris et al. (2018) compared wet chemistry with NIRS of dried ground samples for a single commercial laboratory for a haylage sample set and found that values for sugar (WSC) concentration differed between methods by 20% in some cases but correlations were satisfactory. This demonstrates that correlations can be satisfactory but individual samples can differ greatly, with the end result being that some predicted samples may deviate so far from the real value so as to cause nutritional and health issues in the animal consuming such a forage. This is particularly true for WSC where high concentrations are known to increase the risk of laminitis in insulin resistant equids (Geor, 2008). The study described here adds substantially to the study of Harris et al. (2018) on haylage and is unique because it examines a number of current commercially available NIRS- based methodologies (whole fresh samples) that are used in current practice at commercial laboratories, alongside a dried and ground preparation method. It has also utilised the future hand-held on-farm devices which are becoming wide-spread in practice and could, for the equid owner offer the opportunity to analyse every bale immediately before feeding to check the nutritional status is correct for their particular equid's requirements. The individual analytes examined are discussed in more detail below.

Dry matter was predicted with good accuracy both by a commercial (on-farm) hand-held NIRS-based methodology (method C) and by a commercial (off-farm) lab NIRS-based methodology (method D), but the second on-farm method (method B) performed poorly in predicting DM. Method E employed a dried and ground sample and so this laboratory report the oven DM data in their final forage report (data not shown). The most interesting finding was that of the difference between method B and method C, because this was the same instrument and spectra but fitted to different models. In general, bales are higher in DM than clamp samples so we would expect that method C would be less accurate overall then method B, however, this was not the case. The unexpected finding could be due to the number of samples used to build the prediction. It was reassuring that method D was predicted with good accuracy as this method is specifically for haylage. These results highlight the

284 care needed by both analytical companies and those using the instruments in the field that the correct predictions
285 are being used for their samples.

286 pH is an important analyte as it can be used, with DM, as an indication of fermentation quality (Kung and Shaver,
287 2001). However, pH was poorly predicted by both on-farm NIRS methodologies investigated (methods B and C).
288 Ibáñez and Alomar (2008) attribute inaccurate pH predictions made by NIRS to the fact that pH measurements
289 were obtained from different sample preparation methods in the case of a dried sample being scanned. However,
290 as here we scanned a fresh sample, we were unable to corroborate this hypothesis. As there are rapid electrode-
291 based methods available that can be employed on-farm, NIRS methodologies do not appear of value for measuring
292 pH.

293 Ash is commonly used as a measure of the mineral content of a forage, where high values can be the result of soil
294 contamination within a preserved forage (McDonald et al., 1991). This can represent an increased risk of the
295 presence of undesirable bacteria especially under conditions where the pH is unusually high, where lactic acid
296 production has not inhibited growth of *Listera* or *Clostridia* species that represent a threat to animal health. The
297 range of ash content expected in silages is dependent on the herbage type ensiled. For grass dominated silages
298 ideally the ash content below 9% DM is acceptable, whereas for silages rich in clover and herbs an ash content
299 below 10 % DM is acceptable (McDonald et al., 1991). Overall, all four NIRS-based predictions for ash were
300 poor, where a general under-prediction of ash, was seen for three of the four predictive methodologies (B to D),
301 could result in forage that falls outside of an acceptable range being fed.

302 Fibre fractions, (ADF and aNDF) are important analytes for equids, especially donkeys, because they evolved to
303 thrive on rough vegetation and therefore require high fibre diets to maintain gut health (Burden, 2012; Burden and
304 Bell, 2019; Burden and Thiemann, 2015). The main factor affecting aNDF and ADF in haylage production is the
305 stage of maturity of the forage at harvest. More mature forages e.g. grasses, i.e. those reaching seed maturation,
306 having the highest level of aNDF (with a greater proportion of this being ADF), which is less digestible (Beever
307 et al., 2000). Accurate prediction of ADF / aNDF is required to help the practitioner decide the optimum time of
308 harvest to ensure adequate balance of highly digestible hemicellulose, digestible cellulose and less digestible
309 lignin. Both ADF and aNDF were poorly predicted by the hand-held methodologies (methods B and C), while
310 predictions in aNDF improved using the commercial labs (methods D and E), but these still had a tendency to
311 under-predict. As the horse market may appear to commerce to be of limited financial reward it is highly probable
312 that the NIRS prediction equations have been developed for the ruminant sector. The overall requirement of the

ruminant sector is to reduce the fibre content with lower aNDF and a lower proportion of this being ADF (Beever et al., 2000). Therefore the higher level of aNDF and ADF in haylage destined for equids was unlikely to be present in large numbers in the prediction datasets, this will in effect put an upper limit on the prediction models ability to give a higher value for these parameters and thus the current techniques were likely to underpredict these types of forage. However, the laboratories conducting these analyses for the ruminant sector should be aware that non-milking pregnant cows also have a nutritional requirement for low digestible high fibre preserved forages and if our reasoning is correct they could be underestimating fibre content not only for equid haylages but also dry cow forages. The most accurate predictions for aNDF were the dry and ground lab-based approach (method E). Harris et al. (2018) suggested that scanning samples devoid of water could be superior as it leaves peaks at other wavelengths easier to detect. In addition, drying and grinding a sample enabled a more homogenous sample to be scanned and thus improved the accuracy of the scan data making it more representative, whereas scanning fresh forage was liable to greater scanning inaccuracies as a single larger leaf of higher digestibility is likely to hide a smaller less digestible stem, making the fresh sample scanning more prone to poorer prediction due to sample heterogeneity.

Protein is an important factor in equid diets and crude protein requirements for equids will vary according to the life stage, performance and the health status of the animal in question. Knowledge of crude protein content enables keepers to ensure individual dietary protein requirements are met without under or over provision thus avoiding health related consequences in some animals. (Ringmark and Jansson, 2013). For this study, all four NIRS-based methods tended to over-predict crude protein, where the lab-based methods (methods D and E) performed more accurately than the on-farm methods (methods B and C). Relative to other analyte predictions, the prediction of crude protein was moderately accurate, at least for those lab-based.

Predictions of WSC were poor for all NIRS methodologies and there was an indication that the hand-held and lab NIRS prediction models each had a narrow sample database for this analyte. Methods D and E only predicted between a range of 4 - 10% WSC; therefore the higher the actual WSC content of the haylage by wet chemistry the greater it under-predicted. This agrees with the findings of Harris et al. (2018), suggesting that this is related to the formulation of calibrations across a number of NIRS methods rather than a specific sample set. This low range for WSC, suggested that the dataset for model calibration may have been adapted from forage samples preserved for ruminant feeding which are generally lower in DM and thus lower in WSC, as greater amounts of WSC are used during the more extensive fermentation (McDonald et al., 1991). This effect can also be seen but

342 to a lesser extent with the fibre fractions as the aNDF values were consistently under predicted, as for ruminant
343 rationing the fibre content would typically be lower to improve digestibility.

344 Harris et al. (2018) also reported a poor prediction for WSC although they suggested that the mean bias across the
345 dataset was acceptable. However, reporting only the mean bias is limited as it does not reveal situations of large
346 under-prediction and large over-prediction of equal magnitude. Samples within the dataset greater than the mean
347 bias could be over the currently recommended level for insulin resistant equids of 12% WSC on a dry matter basis
348 (Geor and Harris, 2013). The content of WSC in any forage destined for consumption by some equids is of critical
349 importance as it is an important factor to consider when selecting forage for certain equids, for example insulin
350 resistant equids (Hoffman, 2009; Morgan, 2015), where less than 12% WSC in the DM is currently recommended
351 (Geor and Harris, 2013). Thus, the gross under-prediction at the high WSC content observed in this study suggest
352 many of these haylages should not be fed to equids that are insulin resistant. The prediction range for all the NIRS
353 methods was also narrower than that found with the wet chemistry for WSC. This may explain the poor agreements
354 shown by the Bland-Altman plots as the NIRS-based methods cannot predict outside of the range of the
355 calibration. It is difficult to decipher the actual values from the results presented in Harris et al. (2018) as only
356 mean values are presented, therefore it is impossible to comment on whether the same issue existed within the lab
357 NIRS methodology that was employed in their study as was found in the current study. Rameriz et al. (2015)
358 acknowledges the requirements for different predictions depending on tissue type. In our study it is possible that
359 the sample type submitted was significantly different from that of those used to build the prediction equation due
360 to species, maturity level, climate or geographic differences.

361 NIRS-based analyte prediction models and sample preparation vary between commercial laboratories. Some offer
362 a service that consists of scanning fresh whole forages, which has the advantage of reducing sample handling time
363 resulting in information being made available about a forage before it is fed. This is recommended for livestock
364 with specific dietary requirements and could help to manage intake appropriately (Burden and Bell, 2019; Geor,
365 2008; Geor and Harris, 2013). The UK NIRS laboratories had for many years been dominated by conducting
366 preserved forage NIRS analysis using fresh unprocessed samples due to the speed and cost. This was driven by
367 the 'Feed into Milk' Research programme conducted at ARINI (Hillsborough, Northern Ireland and SAC,
368 Scotland (Park et al., 1998). There are however a number of issues with this, the first being one of homogeneity
369 as already discussed, the second being the fact that the water peak dominates the NIR spectral range and can
370 interfere or mask other peaks that are associated with other chemical and nutritional components of the forage
371 making the ability to predict these from a NIRS scan less accurate (Roberts et al., 2004). The drawbacks to this

method are described by Harris et al. (2018), and because of this, we therefore recommend the drying of samples before NIRS scanning. However, this goes against the advantageous reduction in sample processing time that development of hand-held on-farm NIRS technologies can deliver (Alomar et al., 2003). As such, in this study we included analysis for DM, pH, ash, ADF, aNDF, crude protein and WSC using three fresh forage NIRS methodologies (B-D) and one dried and ground scanned NIRS method (E), against wet chemistry (A). In addition, methods B and C both employed the scan data from a single hand-held farm NIRS instrument, but results were calculated by different prediction models.

This current study examined inter and intra laboratory variation. The differences in prediction calibrations from a given instrument were shown to influence the accuracy of prediction as did differences between sample preparation between labs. However, differences between sample preparations and labs cannot exclusively be explained by fresh versus dried analysis as study labs also use different prediction models developed in-house Our study showed that the variables in both the sample processing, and the prediction models influence the accuracy of the final prediction. As such, care needs to be taken to ensure that whatever methodology is used that the sample that is analysed is representative of the samples in the prediction data set to ensure as accurate a prediction as possible. It should also be noted that although our study showed deviation between NIRS methods and wet chemistry, one laboratory was used as the wet chemistry standard which may differ from that used to build the calibration model. The laboratory employs UKAS accredited methods that are standard for UK analyses. As stated by Harris et al. (2018), different laboratories will have different wet chemistry methodologies in their database establishment and so the analyte predictions will be biased to the background wet chemistry method. This should be considered as a limitation of the NIRS method when calibrations are developed and will contribute to the results observed in this study, where, we acknowledge that this could represent a study limitation. In addition, the sward type and maturity may influence biochemical and physical attributes of a preserved forage (Müller, 2012) and therefore affect the accuracy of NIRS result. This is addressed by the findings of Thomson et al. (2018) in their assessment of commercial lab based NIRS for forage analysis in the ruminant sector, where samples of grass/clover silages were measured commercially with grass-based prediction models. Their studies showed protein predictions to be an issue which is not surprising given the higher protein content of clover compared to grass. However, with the development of hand-held NIRS devices, improving animal health through immediate data availability is a promising prospect and has potential to extend to hygienic as well as nutritive quality (Berardo et al., 2005; Cheli et al. 2012).

Conclusion

Being able to accurately determine the nutritional content of haylage is an important factor in managing dietary intake in equids. The availability of such data through the application of NIRS technology offers an opportunity to provide fast and inexpensive information relative to wet chemistry analysis. This, however, depends on the accuracy of the analyte predictions. From our findings, it is clearly incorrect to say as a blanket statement that NIRS-based predictions are as accurate to that found with wet chemistry for haylage sampled on farm. NIRS has quickly become the go to analysis method for many in the equid and farming industry as it provides a fast analysis which can aid on-farm decision making. Most commercial labs provide services that are based on internally developed model calibrations which may have come from a limited sample set and one method of wet chemistry analysis. Therefore, the results of analysis can vary between commercial labs. In our work, we compared the results gained from sending subsamples from the same bulk sample to two commercial laboratories and also compared the results from two hand-held device calibrations predicted from the same spectra. study results highlighted that the same sample can produce vastly different results when scanned using different methods. Comparison to the wet chemistry revealed different levels of prediction accuracy dependent on analyte considered. This work highlights that NIRS is only as accurate as the calibration samples allow and currently is not reliable to advise on haylage quality on-farm.

Acknowledgements

This work was funded by The Donkey Sanctuary, Sidmouth, Devon, UK and supported as part of the Rothamsted Research, UK Institute Strategic Programme – Soil to Nutrition (BS/E/C/000I0320) funded by the UK Biotechnology and Biological Sciences Research Council.

References

Alomar, D., Fuchslocher, R., de Pablo, M., 2003. Effect of preparation method on composition and NIR spectra of forage samples. Anim. Feed Sci. Technol. 107, 191-200. https://doi.org/10.1016/S0377-8401(03)00124-X.

Andueza, D., Picard, F., Jestin, M., Andrieu, J., Baumont, R., 2011. NIRS prediction of the feed value of temperate forages: efficacy of four calibration strategies. Anim. 5, 1002-13. https://doi.org/10.1017/S1751731110002697.

429 Batten, G.D., 1998. An appreciation of the contribution of NIR to agriculture. J. Near Infrared
430 Spectrosc. 6, 105-114. https://doi.org/10.1255/jnirs.127.

431 Beever, D.E., Offer, N., Gill, E., 2000. The feeding value of grass and grass products, in:
432 Hopkins, A., (Ed.), Grass its production and utilisation, third ed. Blackwell Science, Oxford,
433 UK, pp.140-195.

434 Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A., Marocco, A., 2005. Rapid
435 detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. J.
436 Agric. Food Chem. 53(21), pp.8128-8134. https://doi.org/10.1021/jf0512297.

437 Bland, J. M., Altman, D.G., 2007. Agreement between methods of measurement with multiple
438 observations per individual. J. Biopharm. Stat. 17, 571-582.
439 https://doi.org/10.1080/10543400701329422.

440 Burden, F., 2012. Practical feeding and condition scoring for donkeys and mules. Equine Vet.
441 Educ. 24, 589-596. https://doi.org/10.1111/j.2042-3292.2011.00314.x.

442 Burden, F.A., Bell, N., 2019. Donkey nutrition and malnutrition. Vet. Clin. North Am. 35, 469-
443 479. https://doi.org/10.1016/j.cveq.2019.08.004.

444 Burden, F.A., Thiemann, A., 2015. Donkeys are different. J. Equine Vet. Sci. 35, 376-382.
445 https://doi.org/10.1016/j.jevs.2015.03.005.

446 Cheli, F., Campagnoli, A., Pinotti, L., Dell'Orto, V., 2012. Rapid methods as analytical tools
447 for food and feed contaminant evaluation: methodological implications for mycotoxin analysis
448 in cereals. Food production-approaches, challenges and task. INTECH Open. pp.185-204.
449 https://doi.org/10.5772/33254.

450 Datta, D., 2017. blandr: a Bland-Altman method comparison package for R. Zenodo. (last
451 accessed 09/11/2021). https://doi.org/10.5281/zenodo.824514.

452 Davies, D., Merry, R., Williams, A., Bakewell, E.L., Leemans, D., Tweed, J., 1998.
453 Proteolysis during ensilage of forages varying in soluble sugar content. Int. J. Dairy Sci. 81,
454 444-453. https://doi.org/10.3168/jds.S0022-0302(98)75596-1.

455 Davies, D.R., Davies, G.K., Morgan, C.T., 2012. Silage analysis-comparison of 58 Welsh farm
456 silages analysed either by traditional wet chemistry or wet NIRs, in: Kuoppala, K., Rinne, M.,
457 Vanhatalo, A., (Eds.), Proceedings of the XVI international silage conference 2012. MTT
458 Agrifood Research Finland. p. 270.

459 Evangelista, C., Basiricò, L., Bernabucci, U., 2021. An Overview on the Use of Near Infrared
460 Spectroscopy (NIRS) on Farms for the Management of Dairy Cows. Agriculture. 2021;
461 11(4):296. https://doi.org/10.3390/agriculture11040296.

462 Fox, J., 2015. Applied regression analysis and generalized linear models, third ed. SAGE
463 Publications Inc., Califonia, USA.

464 Geor, R.J., 2008. Metabolic predispositions to laminitis in horses and ponies: obesity, insulin
465 resistance and metabolic syndromes. J. Equine Vet. Sci. 28(12), pp.753-759.
466 https://doi.org/10.1016/j.jevs.2008.10.016.

467 Geor, R., Harris P.A., 2013. Laminitis, in: Geor, R, Harris, P.A., Coenen, M., (Eds.),. Equine
468 applied and clinical nutrition. Cambridge University Press, Cambridge, UK.
469 https://doi.org/10.1016/B978-0-7020-3422-0.00027-4.

470 Harris, P.A., Nelson, S., Carslake, H.B., Argo, C.M., Wolf, R., Fabri, F.B., Brolsma, K.M., van
471 Oostrum, M.J., Ellis, A.D., 2018. Comparison of NIRS and wet chemistry methods for the
472 nutritional analysis of haylages for horses. J. Equine Vet. Sci. 71, 13-20.
473 https://doi.org/10.1016/j.jevs.2018.08.013.

474 Hoffman, R.M., 2009. Carbohydrate metabolism and metabolic disorders in horses. Rev. Bras.
475 de Zootec., 38(SPE), 270-276. https://doi.org/10.1590/S1516-35982009001300027.

476 Ibáñez, L., Alomar, D., 2008. Prediction of the chemical composition and fermentation
477 parameters of pasture silage by near infrared reflectance spectroscopy (NIRS). Chilean J.
478 Agric. Res. 68(4), 352-359. http://dx.doi.org/10.4067/S0718-58392008000400005.

479 Jolliffe, I.T., 2002. Principal component analysis, second ed. Springer, New York, USA.
480 https://doi.org/10.1007/b98835.

481 Kung, L., Shaver, R., 2001. Interpretation and use of silage fermentation analysis
482 reports. Focus on forage, 3(13), pp.1-5. https://fyi.extension.wisc.edu/forage/interpretation-
483 and-use-of-silage-fermentation-analysis-reports/ (accessed 28/06/2021).

484 McDonald, P., Henderson, A., Heron, S., 1991. The biochemistry of silage second edition.
485 Chalcombe Publications, UK.

486

487 Merry, R., Dhanoa, M., Theodorou, M., 1995. Use of freshly cultured lactic acid bacteria as
488 silage inoculants. Grass Forage Sci. 50, 112-123. https://doi.org/10.1111/j.1365-
489 2494.1995.tb02304.x.

490 Montes, J.M., Paul, C., 2008. Near infrared spectroscopy on agricultural harvesters: Technical
491 aspects. NIR news, 19(1), pp.10-12. https://doi.org/10.1255/nirn.1059.

492 Morgan, R., Keen, J., McGowan, C., 2015. Equine metabolic syndrome. Vet. Rec. 177, 173-
493 179. http://dx.doi.org/10.1136/vr.103226.

494 Müller, C.E., 2012. Equine digestion of diets based on haylage harvested at different plant
495 maturities. Anim. Feed Sci. Technol. 11, 177(1-2):65-74. 10.1016/j.anifeedsci.2012.06.002.

496 Müller, C.E., 2018. Silage and haylage for horses. Grass Forage Sci. 73(4), 815-827.
497 https://doi.org/10.1111/gfs.12387.

498 O'Brien, M., O'Kiely, P., Forristal, P.D., Fuller, H.T., 2006. A note on sampling baled grass
499 silage for fungal propagules. J. Anim. Feed Sci. 15, 305-
500 311. https://doi.org/10.22358/jafs/66902/2006.

501 Park, R.S., Agnew, R.E., Gordon, F.J., Steen, R.W.J., 1998. The use of near infrared reflectance
502 spectroscopy (NIRS) on undried samples of grass silage to predict chemical composition and
503 digestibility parameters. Anim. Feed Sci. Technol. 72(1-2), 155-167.
504 https://doi.org/10.1016/S0377-8401(97)00175-2.

505 Piccioli-Cappelli, F., Calegari, F., Calamari, L., Bani, P., Minuti, A., 2019. Application of a
506 NIR device for precision feeding in dairy farms: effect on metabolic conditions and milk
507 production. Ital. J. Anim. Sci. 18, 754-765. https://doi.org/10.1080/1828051X.2019.1570829.

508 Ramirez, J.A., Posada, J.M., Handa, I.T., Hoch, G., Vohland, M., Messier, C., Reu, B., 2015.
509 Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in
510 different tissue types of a broad range of tree species. Methods Ecol. Evol. 6(9), 1018-1025.
511 https://doi.org/10.1111/2041-210X.12391.

512 Reeves III, J.B., 2000. Use of near infrared reflectance spectroscopy. In: D'Mello, J.P.F. (Ed.),
513 Farm animal metabolism and nutrition. CABI Publishing, Wallingford, UK, pp. 185–207.

514 Ringmark, S., Jansson, A., 2013. Insulin response to feeding forage with varying crude protein
515 and amino acid content in horses at rest and after exercise. Comp. Exerc. Physiol, 9(3-4), 209-
516 217. https://doi.org/10.3920/CEP13014.

517 Roberts, C.A., Workman, J., Reeves, J.B. (Eds.), 2004. Near-infrared spectroscopy in
518 agriculture (Vol. 44). American Society of Agronomy Inc. Madison, Wisconsin, USA.

519 Sheaffer, C.C., Martin, N.P., Jewett, J.G., Halgerson, J., Moon, R.D., Cuomo, G.R., 2000.
520 Sampling requirements for forage quality characterization of rectangular hay bales. J. Agron.
521 92, 64-68. https://doi.org/10.2134/agronj2000.92164x.

522 Shenk, J., Westerhaus, M., 1994. The application of near infrared reflectance spectroscopy
523 (NIRS) to forage analysis, in: Fahey G.C., Collins, M., Mertens D.R., Moser, L.E. (Eds.),
524 Forage quality, evaluation, and utilization, American society of agronomy Inc. Madison,
525 Wisconsin, USA, pp. 406-449. https://doi.org/10.2134/1994.foragequality.c10.

526 Thomson, A., Humphries, D., Rymer, C., Archer, J., Grant, N., Reynolds, C., 2018. Assessing
527 the accuracy of current near infra-red reflectance spectroscopy analysis for fresh grass-clover
528 mixture silages and development of new equations for this purpose. Anim. Feed Sci. Technol.
529 239, 94-106. https://doi.org/10.1016/j.anifeedsci.2018.03.009.

530

531

532

533

534

535

536

537

538

539

540

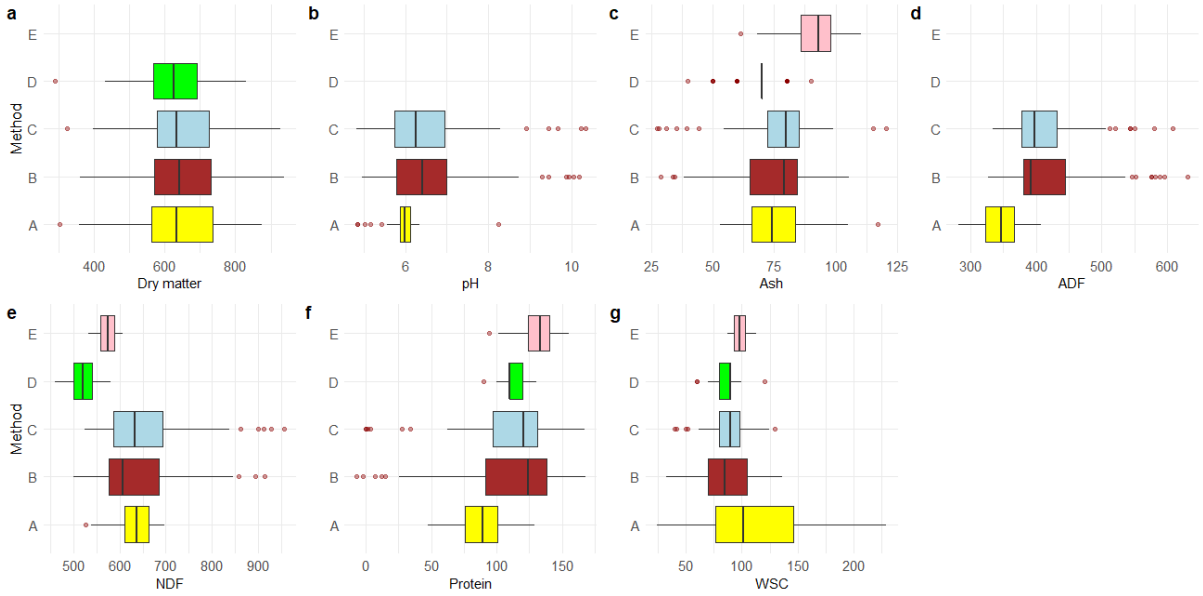
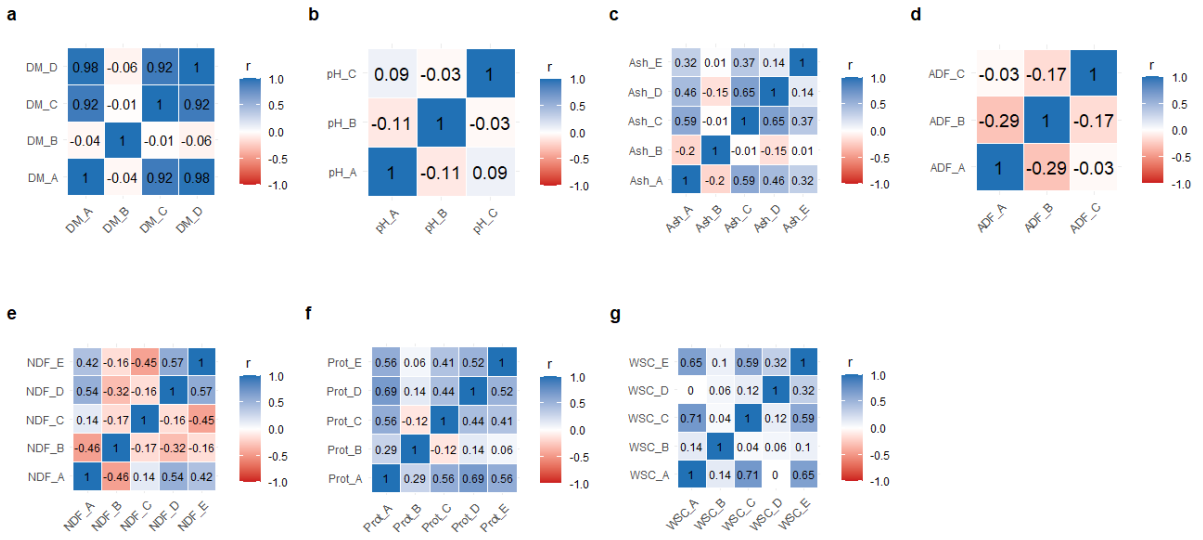


Figure 1. Conditional boxplots for the seven analytes (dry matter (DM), pH, ash, acid detergent fibre expressed inclusive of residual ash (ADF), neutral detergent fibre assayed with a heat stable amylase expressed inclusive of residual ash (aNDF), crude protein and water soluble carbohydrate (WSC)) across the five methods (A to E)


a
DM_D 0.98 -0.06 0.92 1
DM_C 0.92 -0.01 1 0.92
DM_B -0.04 1 -0.01 -0.06
DM_A 1 -0.04 0.92 0.98
DM_A DM_B DM_C DM_D
r 1.0 0.5 0.0 -0.5 -1.0
b
pH_C 0.09 -0.03 1
pH_B -0.11 1 -0.03
pH_A 1 -0.11 0.09
pH_A pH_B pH_C
r 1.0 0.5 0.0 -0.5 -1.0
c
Ash_E 0.32 0.01 0.37 0.14 1
Ash_D 0.46 -0.15 0.65 1 0.14
Ash_C 0.59 -0.01 1 0.65 0.37
Ash_B -0.2 1 -0.01 -0.15 0.01
Ash_A 1 -0.2 0.59 0.46 0.32
Ash_A Ash_B Ash_C Ash_D Ash_E
r 1.0 0.5 0.0 -0.5 -1.0
d
ADF_C -0.03 -0.17 1
ADF_B -0.29 1 -0.17
ADF_A 1 -0.29 -0.03
ADF_A ADF_B ADF_C
r 1.0 0.5 0.0 -0.5 -1.0
e
NDF_E 0.42 -0.16 -0.45 0.57 1
NDF_D 0.54 -0.32 -0.16 1 0.57
NDF_C 0.14 -0.17 1 -0.16 -0.45
NDF_B -0.46 1 -0.17 -0.32 -0.16
NDF_A 1 -0.46 0.14 0.54 0.42
NDF_A NDF_B NDF_C NDF_D NDF_E
r 1.0 0.5 0.0 -0.5 -1.0
f
Prot_E 0.56 0.06 0.41 0.52 1
Prot_D 0.69 0.14 0.44 1 0.52
Prot_C 0.56 -0.12 1 0.44 0.41
Prot_B 0.29 1 -0.12 0.14 0.06
Prot_A 1 0.29 0.56 0.69 0.56
Prot_A Prot_B Prot_C Prot_D Prot_E
r 1.0 0.5 0.0 -0.5 -1.0
g
WSC_E 0.65 0.1 0.59 0.32 1
WSC_D 0 0.06 0.12 1 0.32
WSC_C 0.71 0.04 1 0.12 0.59
WSC_B 0.14 1 0.04 0.06 0.1
WSC_A 1 0.14 0.71 0 0.65
WSC_A WSC_B WSC_C WSC_D WSC_E
r 1.0 0.5 0.0 -0.5 -1.0

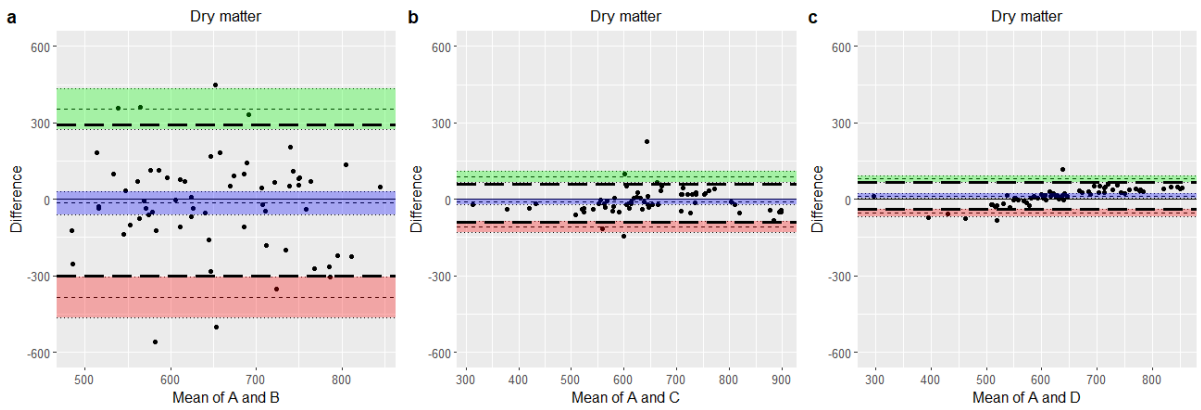


Figure 3. Bland-Altman plots for dry matter. Solid line at 0 for exact agreement between methods. Middle filled region for the 95% confidence interval around the mean bias line (short dashed line). Top filled region for the 95% confidence interval around the mean-based upper agreement threshold (short dashed line). Bottom filled region for the 95% confidence interval around the mean-based lower agreement threshold (short dashed line). Long dashed lines are for the median-based upper and lower agreement thresholds.

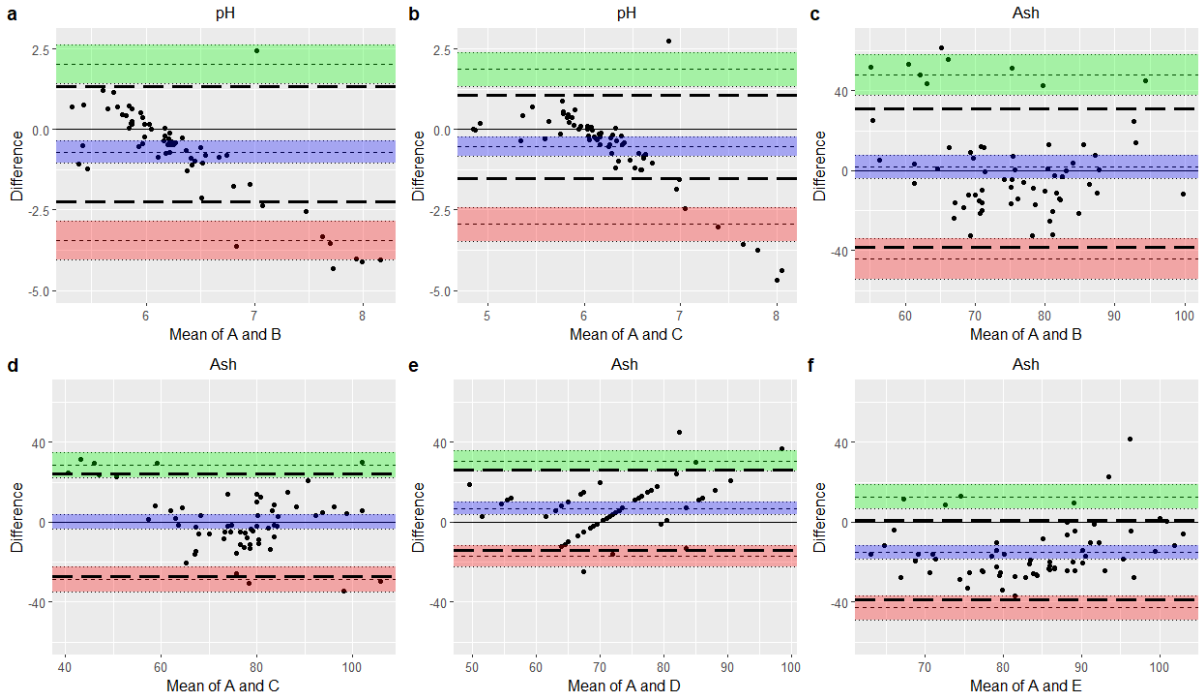


Figure 4. Bland-Altman plots for pH and ash. Solid line at 0 for exact agreement between methods. Middle filled region for the 95% confidence interval around the mean bias line (short dashed line). Top filled region for the 95% confidence interval around the mean-based upper agreement threshold (short dashed line). Bottom filled region for the 95% confidence interval around the mean-based lower agreement threshold (short dashed line). Long dashed lines are for the median-based upper and lower agreement thresholds.

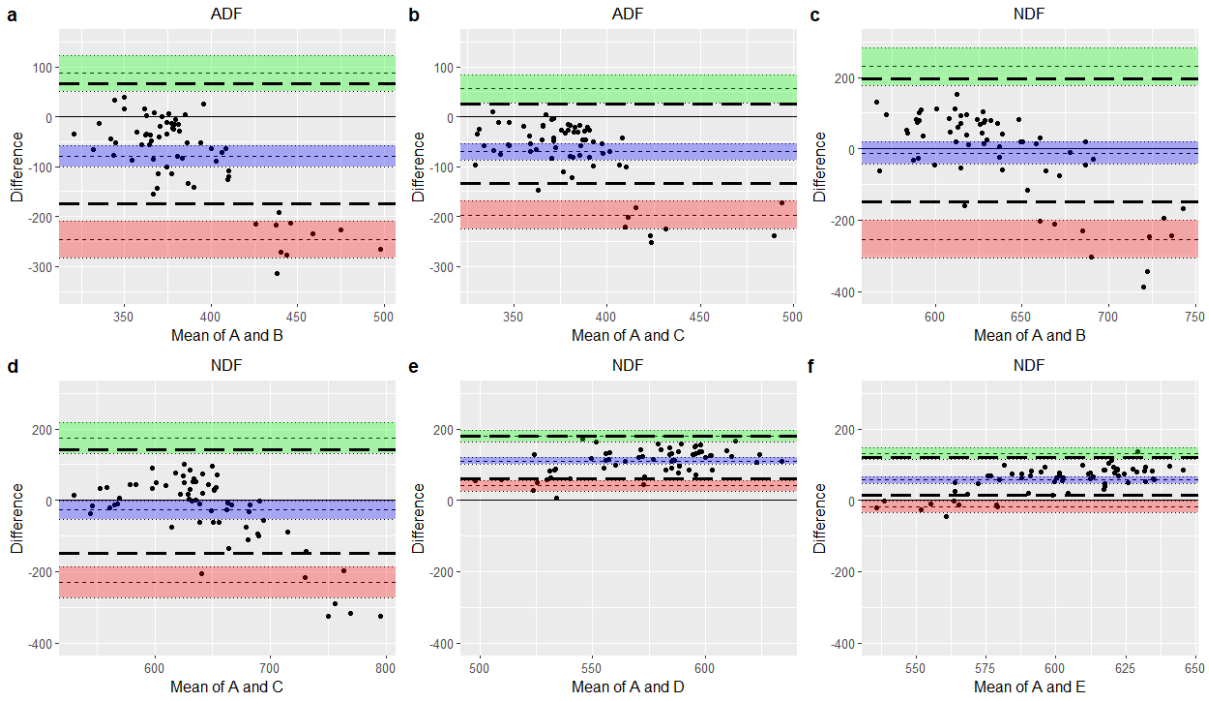


Figure 5. Bland-Altman plots for acid detergent fibre (ADF) and neutral detergent fibre (aNDF). Solid line at 0 for exact agreement between methods. Middle filled region for the 95% confidence interval around the mean bias line (short dashed line). Top filled region for the 95% confidence interval around the mean-based upper agreement threshold (short dashed line). Bottom filled region for the 95% confidence interval around the mean-based lower agreement threshold (short dashed line). Long dashed lines are for the median-based upper and lower agreement thresholds.

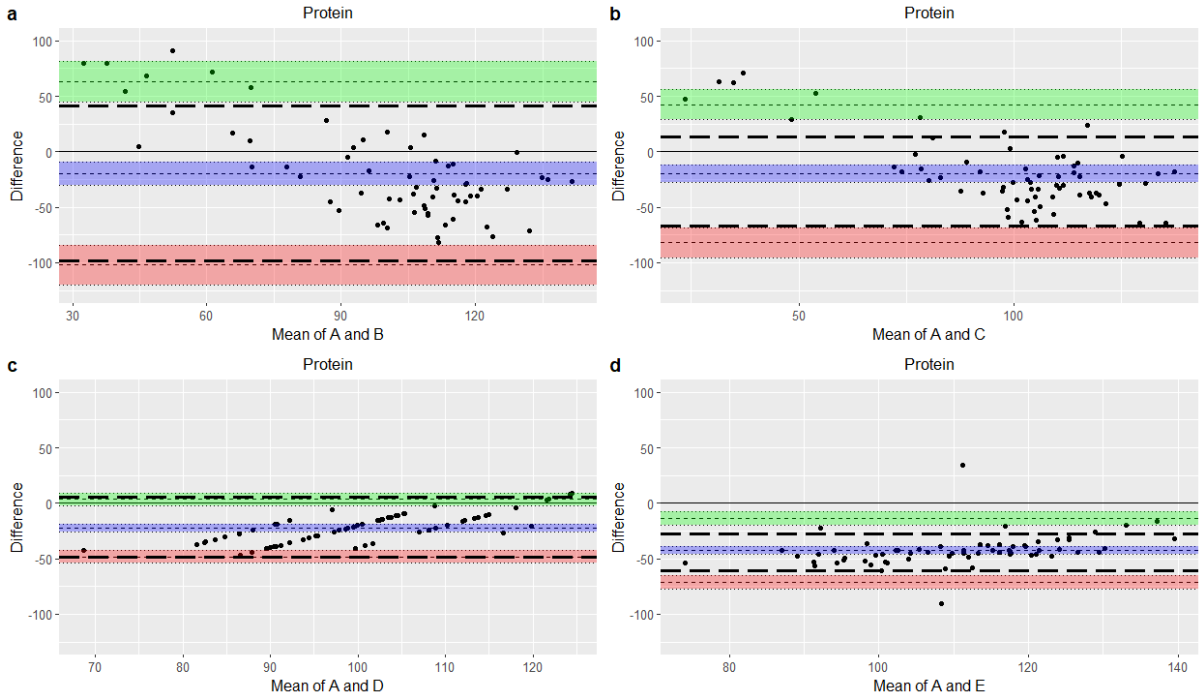


Figure 6. Bland-Altman plots for crude protein. Solid line at 0 for exact agreement between methods. Middle filled region for the 95% confidence interval around the mean bias line (short dashed line). Top filled region for the 95% confidence interval around the mean-based upper agreement threshold (short dashed line). Bottom filled region for the 95% confidence interval around the mean-based lower agreement threshold (short dashed line). Long dashed lines are for the median-based upper and lower agreement thresholds.

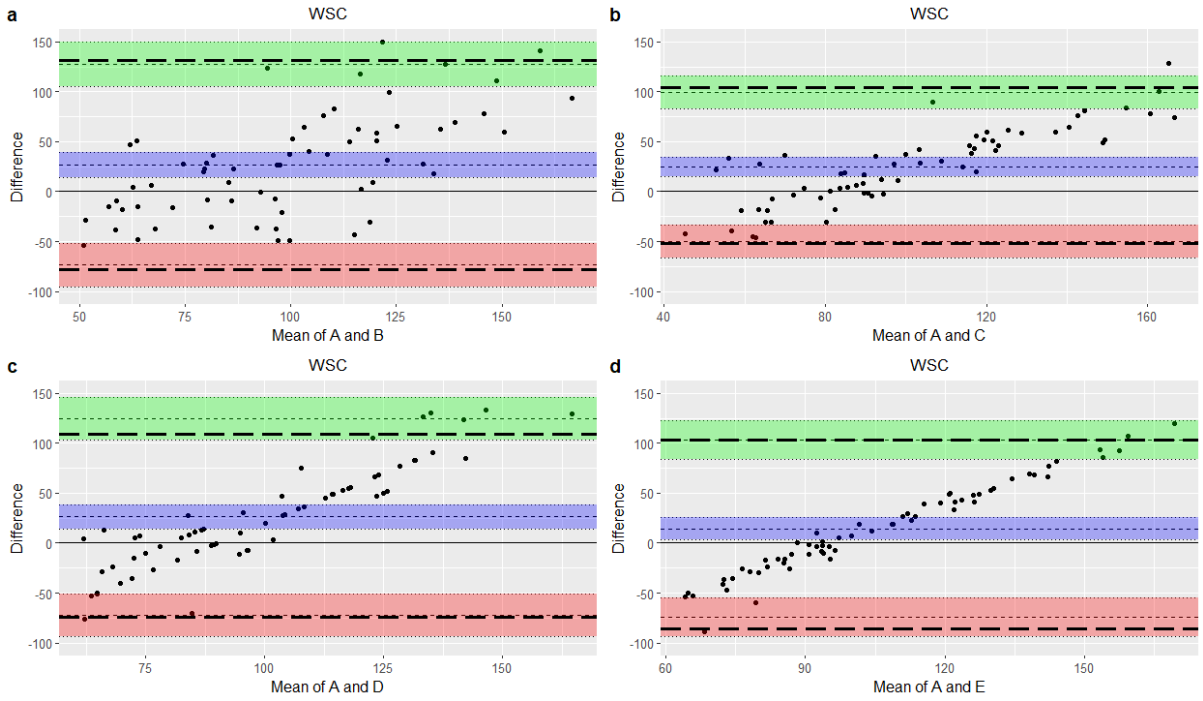


Figure 7. Bland-Altman plots for water-soluble carbohydrate (WSC). Solid line at 0 for exact agreement between methods. Middle filled region for the 95% confidence interval around the mean bias line (short dashed line). Top filled region for the 95% confidence interval around the mean-based upper agreement threshold (short dashed line). Bottom filled region for the 95% confidence interval around the mean-based lower agreement threshold (short dashed line). Long dashed lines are for the median-based upper and lower agreement thresholds.