

ORIGINAL RESEARCH ARTICLE

Forage & Grazinglands

Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value

J. J. Acosta¹  | M. S. Castillo²  | G. R. Hodge¹

¹ Camcore, Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA

² Crop and Soil Sciences Department, North Carolina State University, Raleigh, NC 27695, USA

Correspondence

Miguel S. Castillo, Crop and Soil Sciences Department, North Carolina State University, 2413 Williams Hall, Raleigh, NC 27695.

Email: [mcastil@ncsu.edu](mailto:mscastil@ncsu.edu)

Assigned to Associate Editor Thomas Griggs.

Abstract

The quality of predicted plant-, soil-, and animal-response values from near-infrared (NIR) reflectance spectra depends on the ability to generate appropriate NIR models. The first step in the development of NIR models is collection of spectral data. Limited work, however, has been reported that compares NIR models for prediction of forage nutritive value when the spectra are obtained from devices with different spectral ranges and resolutions. The objectives of this study were to (a) develop and evaluate NIR spectroscopy models using a benchtop-type (FOSS) and two handheld NIR devices (microPHAZIR and DLP NIRscan Nano EVM) to predict crude protein (CP), acid detergent fiber (ADF), amylase and sodium sulfite-treated neutral detergent fiber (aNDF), and in vitro true dry matter digestibility (IVTD) of dried ground forage grass samples and (b) compare predictions among the three NIR devices. Switchgrass (*Panicum virgatum* L.) and bermudagrass [*Cynodon dactylon* (L.) Pers] hay samples were scanned with the NIR devices and analyzed with wet chemistry for development of NIR prediction models. Among devices, the r^2 of validation values for aNDF models ranged from .81 to .87; all other r^2 values were $>.86$ and as high as .98 with standard error of prediction (SEP; g kg^{-1}) ranging from 8.1 to 11.5 for CP, 19.1 to 23.8 for aNDF, 14.2 to 20.0 for ADF, and 26.8 to 49.9 for IVTD. The FOSS benchtop NIR prediction models consistently had the highest r^2 and lowest SEP values; however, the predictive power for both handheld devices was similar to the benchtop-type device.

Abbreviations: ADF, acid detergent fiber; aNDF, amylase and sodium sulfite-treated neutral detergent fiber; CP, crude protein; DT, detrend; IVTD, in vitro true dry matter digestibility; LOF, local outlier factor; MSC, multiplicative scatter correction; NIR, near-infrared; R^2_{cv} , coefficient of determination of cross-validation; RPD, ratio of performance to deviation; SECV, standard error of cross-validation; SEP, standard error of prediction; SNV, standard normal variate.

© 2020 The Authors. Crop Science © 2020 Crop Science Society of America

1 | INTRODUCTION

Near-infrared reflectance spectroscopy is a nondestructive analysis technology that can be used for rapid determination of constituents of various materials. The foundational relationship between NIR spectra and biology is that NIR-irradiated objects absorb and reflect light based on their molecular composition (Shenk, Workman, & Westerhaus, 2008). The work by Norris, Barnes, Moore, and Shenk (1976) marked a milestone for the use of NIR

spectra to predict forage nutritive value. We direct the reader to Roberts, Stuth, and Flinn (2004) for basic aspects of NIR spectroscopy applications in forages and feedstuffs, to Batten (1998) for a review of the contributions of NIR technology to agriculture, and to Burns (2011) for a historical perspective of its use in the animal–plant interface. Noteworthy extensions of NIR spectroscopy applications are determination of nutrient digestion and dry matter intake by animals from dried ground forage samples (Norris et al., 1976) and direct estimation of animal responses (i.e., dry matter intake, digestibility, and diet selection) via spectral scans of feces (Burns, 2011; Dixon & Coates, 2009).

The assumption is that NIR spectroscopy applications are as good as the prediction models developed out of meaningful biochemical entities, hence they are considered a spectrochemical model. Therefore, the quality of predictions for plant-, soil-, and animal-response values from NIR spectra depend on the ability to generate appropriate NIR models. The first step in development of NIR models is collection of spectral data from a sample population. There are several devices available in the market to collect spectral data with a wide range of specifications (e.g., spectral range and resolution, benchtop vs. portable) and market price (Workman & Burns, 2007). Traditional NIR devices have a benchtop-type configuration, high resolution and associated high cost, and they have been mainly used by trained personnel working in commercial and research laboratories. Recent technological advances brought to market a variety of portable spectroscopy devices (Crocombe, 2018) that are more affordable and could potentially be a tool in the hands of farmers or consultants for in situ analysis (Pérez-Marín, Paz, Guerrero, Garrido-Varo, & Sánchez, 2010; Starks, Brown, Turner, & Venuto, 2015; Warburton, Brawner, & Meder, 2014). A great deal of literature has reported successful development of NIR models to predict many traits of several forage species including prediction of chemical constituents (Saha et al., 2018), nutritive value (Burns, Fisher, & Rottinghaus, 2006), ethanol yield (Vogel et al., 2011), ergot alkaloid concentration (Roberts, Benedict, Hill, Kallenbach, & Rottinghaus, 2005), and botanical composition (Coleman, Christiansen, & Shenk, 1990; Karayilanli, Cherney, Sirois, Kubinec, & Cherney, 2016).

The methodology developed by Shenk and Westerhaus (1991) for population definition, sample selection, and calibration has been widely adopted for NIR model development in the realm of forage work. Limited work, however, has been reported that compares predictions of basic estimates of forage nutritive value when the sample dataset is scanned with different NIR devices. The specific objectives of this study were to (a) develop and evaluate NIR spectroscopic models using reflectance acquired from one benchtop- and two handheld-type NIR devices to predict

Core Ideas

- Forage nutritive value estimates were successfully predicted using NIR models.
- The predictive power of the handheld devices was very similar to that of the benchtop device.
- Different spectral transformations optimized the NIR models per analyte and per device.
- Further investigation of handheld NIR devices is warranted.

CP, ADF, aNDF, and IVTD of dried ground forage grass samples and (b) compare predictions among the three NIR devices.

2 | MATERIALS AND METHODS

2.1 | Database description and sample preparation

Samples from two warm-season perennial grass species, switchgrass and bermudagrass, were randomly selected from three independent trials. Switchgrass has dual potential as a bioenergy and forage crop and to complement and improve the overall productivity of traditional tall fescue [*Lolium arundinaceum* (Schreb.) Darbysh.] pasture-based livestock systems in the transition region of the United States (Burns, Mochrie, & Timothy, 1984; USDA, 2018). Bermudagrass is the most important warm-season grass species for livestock production in the southeast United States (Hill, Gates, & West, 2001).

The complete database for CP, aNDF, and ADF had a total of 210 samples (138 switchgrass and 72 bermudagrass). For IVTD, the database had a total of 100 samples (from switchgrass only). The switchgrass samples originated from two experiments conducted for 2 yr. The first experiment evaluated the effects of harvest frequency (clipped every 3, 6, 9, and 12 wk) and intensity (10, 20, 30, and 40 cm stubble height) of cv. Performer (Bekewe, Castillo, & Rivera, 2018); the second experiment evaluated harvest timing (before and after frost) of cultivar BoMaster when clipped once or twice per year. The bermudagrass samples originated from a 3-yr experiment that evaluated five cultivars (Tifton 85, Tifton 44, Ozark, Midland 99, and Coastal) fertigated with swine (*Sus scrofa domestica*) lagoon effluent. The bermudagrass samples were collected from plots harvested to 8 cm stubble height every time the canopy height was approximately 35–40 cm. In the field, the sampling protocol for both forages consisted

TABLE 1 Near infrared devices used to predict crude protein, acid detergent fiber, amylase and sodium sulfite-treated neutral detergent fiber, and in vitro true digestibility of switchgrass (*Panicum virgatum* L.) and bermudagrass [*Cynodon dactylon* (L.) Pers] samples

Device	ID	Type	Spectral range	Wavelength interval	No. of Wavelengths
			—nm—		
FOSS 6500 NIRSystems	FOSS	Benchtop	1100–2498	2	700
Thermo Scientific microPHAZIR	microPHAZIR	Handheld	1600–2400	8	100
Texas Instruments DLP NIRscan Nano EVM	Nano	Handheld	900–1700	5	160

of clipping within an area ≥ 3 m²; then, a subsample of ~ 1 kg was dried in a forced-air oven at 60 °C until the samples achieved constant weight. The dried samples were ground using a Wiley mill with four rotating and six stationary knives that produce a shearing action (A. H. Thomas Co.) to pass through a 1-mm screen and subsequently stored in Whirl-Pak bags. Each sample in the database represents a field experimental unit (hence, a unique replicate-sampling event-treatment-year combination).

2.2 | Near-infrared devices and wet chemistry analyses

In the laboratory, 4 g of ground material of each sample were placed in spinning-sample module cups with quartz bottoms used in the FOSS 6500 device and then scanned using one benchtop and two handheld NIR spectroscopic devices. Information about spectral range, wavelength interval, and number of wavelengths for the three devices is provided in Table 1. The FOSS 6500 is a benchtop-type device and has the capability to obtain spectra in the visible-NIR spectral range (400–2498 nm); however, for the purpose of this study we used the spectra in the NIR region (1100–2498 nm) only. A total of 32 scans per sample were performed by the FOSS device and the software provided the mean spectrum for each sample. The two handheld devices were the microPHAZIR (ThermoFisher Scientific, 2020) and DLP NIRscan Nano EVM (Texas Instruments, 2020). For the handheld devices, samples were scanned in four positions by rotating the sample cups 90° and taking measures while in static position. Custom R scripts (R Core Team, 2016) were developed to calculate the mean reflectance spectrum of each sample for the handheld devices. The handheld devices were not in direct contact with the ground sample (although they have capability for direct contact with the sample); instead, scans were taken through the cover glass of the FOSS module cups that contained the samples. In this way, the exact same sample packed in the same cup was scanned with all three devices.

Wet chemistry analyses were performed by the Dairy One Forage Laboratory (Ithaca, NY). In summary from the laboratory analytical procedures of Dairy One (2015), concentration of CP was calculated by multiplying the concentration of total N (determined by dry combustion using a LECO CN628) by 6.25. The aNDF and ADF concentrations were determined using Methods 13 and 12, respectively, of the ANKOM Fiber Analyzer (ANKOM Technology). For IVTD, samples were analyzed using a 48-h in vitro digestion procedure (Daisy II incubator; Method 3) (Dairy One, 2015). Standard deviation (and associated coefficients of variation, %) for replicate wet chemistry analyses of a standard sample from the Dairy One Forage Laboratory are 2.6 (1.15%), 10.2 (3.94%), 11.2 (2.95%), and 26.2 (1.81%) g kg⁻¹ for CP, ADF, aNDF, and IVTD, respectively. Descriptive statistics of laboratory results are presented in Figure 1.

2.3 | Near infrared model development

Model development was performed using a data analysis pipeline written in R environment (R Core Team, 2016). The pipeline was previously used in the successful development of NIR models to determine chemical properties of wood (Hodge, Acosta, Unda, Woodbridge, & Mansfield, 2018) and nutritive value of switchgrass (Bekewe, Castillo, Acosta, & Rivera, 2019) and a mixture of native warm-season grasses (Castillo, Tiezzi, & Franzluebbbers, 2020). The pipeline has two separate phases: (a) transformations and outlier detection and (b) model training, cross-validation, and prediction of new observations. In summary for the NIR pipeline: first, mathematical transformations of the spectra were applied to raw ($\log R^{-1}$) NIR spectra to remove the scattering of diffuse reflections associated with sample particle size and to improve subsequent regression analyses. Scatter-correction methods included multiplicative scatter correction (MSC), standard normal variate (SNV), and detrend (DT). Spectral derivative methods included second-order polynomial and second derivative of Savitzky-Golay smoothing with two different window sizes of five and seven points (SG5 and

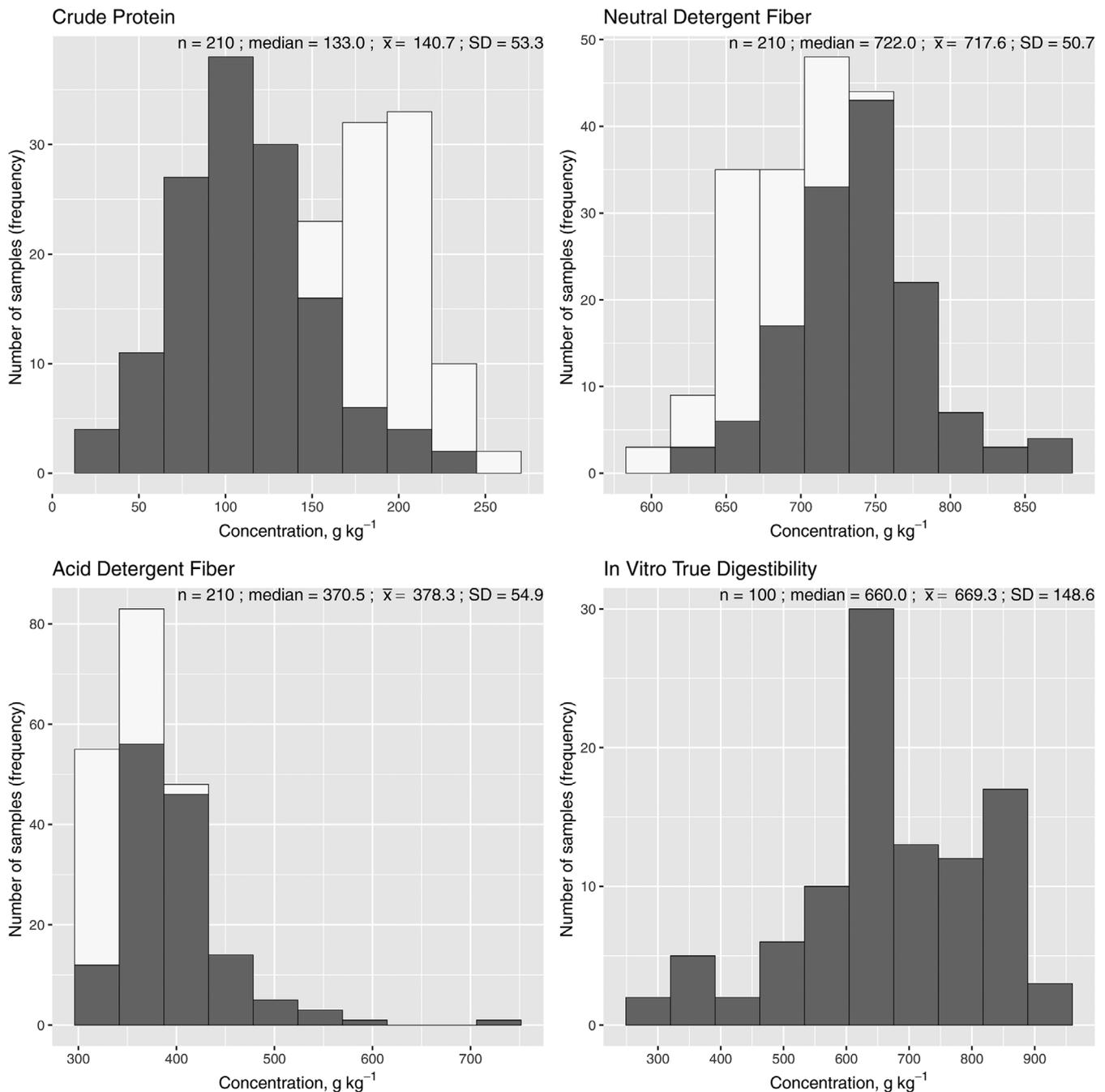


FIGURE 1 Histograms and descriptive statistics of nutritive value estimates for switchgrass (*Panicum virgatum* L.) and bermudagrass [*Cynodon dactylon* (L.) Pers] samples. Filled bars correspond to switchgrass samples and white-filled bars correspond to bermudagrass samples. Crude protein, amylase and sodium sulfite-treated neutral detergent fiber, and acid detergent fiber samples in the database included switchgrass and bermudagrass samples; for the in vitro true digestibility database, only switchgrass samples were included

SG7). We also combined pairs of transformations (SNV + DT, MSC + DT, SNV + SG, MSC + SG, and DT + SG). Local outlier factors (LOFs) were used to filter out atypical spectral data (Breunig, Hans-Peter, Raymond, & Sander, 2000). Traditional distance-based multivariate methods for outlier detection identify outlier observations if their deviation from a global distribution centroid is greater than a defined threshold. Measurements like Mahalanobis

distance typically take a global view of the data set and identify outliers as observations that are extremely distant from the global centroid. This may not be ideal if the dataset exhibits a complex structure (e.g., with several clustered populations). In contrast, the LOF is a density-based outlier detection method that measures the degree to which an observation is isolated from its nearest surrounding neighbors, and it scores each observation according

TABLE 2 Number of samples identified as outliers and number of samples in the resulting outlier-free databases used for development of near-infrared models for crude protein (CP), amylase and sodium sulfite-treated neutral detergent fiber (aNDF), acid detergent fiber (ADF), and in vitro true digestibility (IVTD). Outliers were defined as samples with local outlier factor values >2

Model ^a	Variable	Device	Spectral trans-formation ^b	No. of outliers	No. of samples in outlier-free database
Full-dataset	CP	FOSS	SNV	1	209
		microPHAZIR	MSC	0	210
		Nano	SNV	0	210
	aNDF	FOSS	DT_SG7	0	210
		microPHAZIR	MSC	0	210
		Nano	DT	0	210
	ADF	FOSS	SG7	0	210
		microPHAZIR	NIR	0	210
		Nano	SG7	0	210
	IVTD	FOSS	DT	0	100
		microPHAZIR	SNV	0	100
		Nano	DT_SG7	0	100
Split-dataset	CP	FOSS	MSC	1	159
		microPHAZIR	NIR	0	160
		Nano	NIR	0	160
	aNDF	FOSS	SNV	1	159
		microPHAZIR	MSC	0	160
		Nano	MSC_SG7	0	160
	ADF	FOSS	SG7	0	160
		microPHAZIR	NIR	0	160
		Nano	SNV_SG7	0	160
	IVTD	FOSS	SNV_SG7	0	75
		microPHAZIR	DT	0	75
		Nano	SNV	0	75

^a Full-dataset models used all samples in the database for calibration and cross-validation; split-dataset models used $\approx 75\%$ of samples in the database for calibration and the remaining 25% of samples as test-set.

^b SNV, standard normal variate; MSC, multiplicative scatter correction; DT_SG7, Detrend plus Savitzky–Golay smoothed spectra using seven point; NIR, raw spectra ($\log R^{-1}$; R = reflectance).

to the ratio between the average densities of its neighbors and itself. If the distance from a given observation to its nearest neighbors is always large, that observation is considered a local outlier. Breunig et al. (2000) recommended using at least 10 neighboring points for the calculation of the LOF, and they demonstrated that the distribution of the LOF scores for a population is centered around one, with a very narrow standard deviation (ranging from 0.1 to 0.4). For our analysis, we set the algorithm to calculate an LOF score for each observation based on its 20 nearest neighbors. Samples with LOF values greater than two were excluded from the analysis (since $\text{LOF} > 2$ will be roughly equivalent to the mean + 3 SD). The number of samples identified as outliers in the database as well as the resulting number of samples in the outlier-free datasets for each device and variable are presented in Table 2.

Second, outlier-free datasets for the raw NIR spectra and all transformations were used to fit NIR models between spectral data and wet chemistry laboratory values. Partial least squares regression was implemented using the R-package pls (Mevik & Wehrens, 2016) and model performance was evaluated using leave-one-out cross-validation (Browne, 2000; Diana & Tommasi, 2002). The leave-one-out method uses as many cross-validation points (groups) as there are number of observations present in the model (n). For example, if $n = 100$, then a total of 100 models will be trained and tested using this type of cross-validation. Desirable models are those that maximize the coefficient of determination of cross-validation (R^2_{cv}), minimize the standard error of cross-validation (SECV), and have a small number of latent variables (projection factors). Since the objective of the study was to compare different NIR devices, we followed a specific two-step decision

TABLE 3 Fit statistics of near-infrared models for the full- and split-datasets for prediction of crude protein (CP), amylase and sodium sulfite-treated neutral detergent fiber (aNDF), acid detergent fiber (ADF), and in vitro true digestibility (IVTD) of switchgrass (*Panicum virgatum* L.) and bermudagrass [*Cynodon dactylon* (L.) Pers]. Model performance was evaluated using leave-one-out cross-validation

Model ^a	Variable	Device	Spectral transformation ^b	Factors ^c	R ² _c ^d	SEC ^e	R ² _cv ^f	SECV ^g
						g kg ⁻¹		g kg ⁻¹
Full-dataset	CP	FOSS	SNV	18	.99	5.0	.99	6.0
		microPHAZIR	MSC	9	.98	7.5	.97	8.7
		Nano	SNV	12	.97	8.6	.96	10.9
	aNDF	FOSS	DT_SG7	5	.90	16.2	.89	16.6
		microPHAZIR	MSC	8	.89	17.1	.86	19.2
		Nano	DT	12	.91	15.3	.85	19.5
	ADF	FOSS	SG7	6	.84	21.8	.82	23.6
		microPHAZIR	NIR	8	.84	21.6	.80	24.7
		Nano	SG7	5	.84	22.1	.79	25.3
	IVTD	FOSS	DT	10	.99	16.9	.98	20.7
		microPHAZIR	SNV	4	.97	26.3	.96	28.7
		Nano	DT_SG7	3	.96	30.8	.95	33.4
Split-dataset	CP	FOSS	MSC	16	.99	4.7	.99	5.9
		microPHAZIR	NIR	7	.98	7.1	.98	7.8
		Nano	NIR	12	.98	8.0	.96	10.1
	aNDF	FOSS	SNV	9	.90	15.8	.88	17.8
		microPHAZIR	MSC	8	.90	15.7	.87	18.5
		Nano	MSC_SG7	6	.91	15.2	.86	18.8
	ADF	FOSS	SG7	6	.83	23.5	.79	26.0
		microPHAZIR	NIR	8	.84	22.6	.78	26.9
		Nano	SNV_SG7	6	.84	22.4	.76	27.6
	IVTD	FOSS	SNV_SG7	6	.99	13.8	.98	19.3
		microPHAZIR	DT	3	.97	26.7	.96	29.0
		Nano	SNV	13	.99	13.0	.96	30.7

^a Full-dataset models used all samples in the database for calibration and cross-validation; split-dataset models used $\approx 75\%$ of samples in the database for calibration and the remaining 25% of samples as test-set.

^b SNV, standard normal variate; MSC, multiplicative scatter correction; DT_SG7, detrend plus Savitzky–Golay smoothed spectra using seven points; NIR, raw spectra ($\log R^{-1}$).

^c Number of loading factors (latent variables) in the partial least squares regression models.

^d R²_c, coefficient of determination, calibration.

^e SEC, standard error of calibration.

^f R²_cv, coefficient of determination, cross-validation.

^g SECV, standard error of cross-validation.

algorithm to identify the best NIR model for each analyte–device combination without any subjective input into the process. First, one model for each spectral transformation (total of 14) was selected based on the first local minimum SECV as a strategy to prevent overfitting (Osten, 1988). Second, the resulting 14 models were then ranked according to their R²_cv and the best selected model for each analyte–device combination was the model with the highest R²_cv value. Hence, the number of latent variables of each model is a consequence of the previous selection approach. The resulting models are presented in Table 3.

Two types of models are presented in Table 3 based on the number of samples used in the training set. The models that were fitted using all observations in the outlier-free dataset are hereafter referred to as full-dataset; models developed using $\approx 75\%$ of the observations in the outlier-free database are hereafter referred to as split-dataset. The leave-one-out cross-validation was used to develop the calibration equation for both types of models. Then for the split-dataset models, the $\approx 25\%$ of the observations not used in the calibration (hereafter referred to as test set) were used to evaluate prediction performance. Sample selection for the split-dataset models and

the methodology for comparison of NIR predicted values vs. wet-chemistry laboratory values is described in the following section.

2.4 | Near-infrared model comparison

Performance evaluation of the predictions comparing the three NIR devices was done using the split-dataset models (Table 3). The full outlier-free database was randomly split in two sets: training set and test set. The training set consisting of $\approx 75\%$ of observations was used for development of the split-database models. The test set (remaining $\approx 25\%$ of observations) was used for performance evaluation of predictions. The samples in both datasets, training and test set, are unique and do not overlap between sets. For CP, ADF, and aNDF, a total of 160 observations were randomly selected to form the training set and the 50 remaining samples were used for the test set, with the exception of the CP and aNDF models in the FOSS device, which had 159 samples (Table 2). To verify that the distribution of samples in the dataset followed the species composition of the full database, we calculated the proportion of bermudagrass samples. The full database had 72 bermudagrass samples out of 210 (proportion = .34); the training set had 54 bermudagrass samples out of 160 (proportion = .34); and the test set had 18 bermudagrass samples out of 50 (proportion = .36). For IVTD (switchgrass samples only), the training set was created by randomly selecting 75 observations and the remaining 25 observations were then used for the test set. We assessed the predictive power of the NIR models by plotting scatterplots for CP, aNDF, ADF, and IVTD of NIR predicted values on the x -axis and wet-chemistry laboratory values on the y -axis (Figure 2); in addition, we calculated the coefficient of determination (r^2), SEP, ratio of performance to deviation (RPD) (Williams, Darnenne, & Flinn, 2017), bias, intercept, and slope (Table 4).

3 | RESULTS AND DISCUSSION

The data analysis pipeline generates summary tables with fit statistics and figures for all the possible NIR models for each response variable per device. An example table with 14 models generated to predict IVTD from samples in the split-dataset scanned with the Nano device is presented in Table 5. Additionally, Figure 3 shows an example output of SECV and (R^2_{cv}) as a function of the number of factors (latent variables) for the SNV model presented in Table 5. It is apparent from Table 5 that there are several models developed with different spectral transformations that all have excellent fit statistics and that

could be used to predict IVTD. For the variable IVTD, the model selection algorithm previously described identified the SNV model with 13 factors as the best calibration model for the Nano device because it had the highest $R^2_{cv} = .96$ and lowest SECV = 30.7 among all spectral transformations (Table 5); it was then further evaluated with the test set as presented in Table 4. If choosing a model for operational use, the final model selection would be a decision to be made by the researcher and could consider a compromise among the fit statistics of each model. Some analysts might consider the number of latent factors to be a very important model selection criterion, and in that case, might prefer a different spectral transformation. For example, the DT_SG7 model with six factors had only slightly lower values for $R^2_{cv} = .95$, and slightly higher values for SECV = 33.4 than the 13-factor SNV model. Or some analysts might instead choose the DT model with nine factors, and $R^2_{cv} = .96$, and SECV = 31.0 (Table 5). The lower number of latent factors might mean that these models would be more precise and accurate for future operational use than the SNV model. However, these very slight differences in R^2_{cv} and SECV would not change any conclusions in this study regarding the relative utility of the three NIR devices.

The fit statistics for models generated with the full-dataset (using 100% of samples the database) and the split-dataset (using 75% of the database) were very similar (Table 3). However, it is worth noting that different mathematical transformations and with different number of latent variables optimized the NIR models for each analyte and device, highlighting the advantages of a flexible data analysis pipeline for development of models. In general, R^2_c and R^2_{cv} values were $>.95$ for CP and IVTD, $>.85$ for aNDF, and $>.79$ for ADF (Table 3). Models with greater R^2_c and R^2_{cv} values had lower SEC and SECV values. The values of SEC and SECV (g kg^{-1}) in the split-dataset models (Table 3) were lowest for CP (≤ 10.1), ranged from 15.2 to 18.8 for aNDF, from 22.4 to 27.6 for ADF, and from 13.0 to 30.7 for IVTD. The aforementioned values are comparable to, and in some instances better than, the fit statistics of the NIRS Forage and Feed Testing Consortium level 2 equations as disclosed by Pittman et al. (2016) and Saha et al. (2018) for grass hay ('13GH50-2.eqa') and mixed hay ('16mh50-2.eqa'), respectively.

Prediction performance of the NIR models was assessed using the split-dataset models and the test-set (Table 4; Figure 2). With the exception of the models for aNDF where r^2 values ranged from .81 to .87, all other r^2 values were $\geq .86$ and as high as .98, with SEP (g kg^{-1}) ranging from 8.1 to 11.5 for CP, from 19.1 to 23.8 for aNDF, 14.2 to 20.0 for ADF, and from 26.8 to 49.9 for IVTD (Table 4). In addition to r^2 and SEP estimates, Table 4 presents the results of RPD, bias, and upper and lower limit of 95% confidence

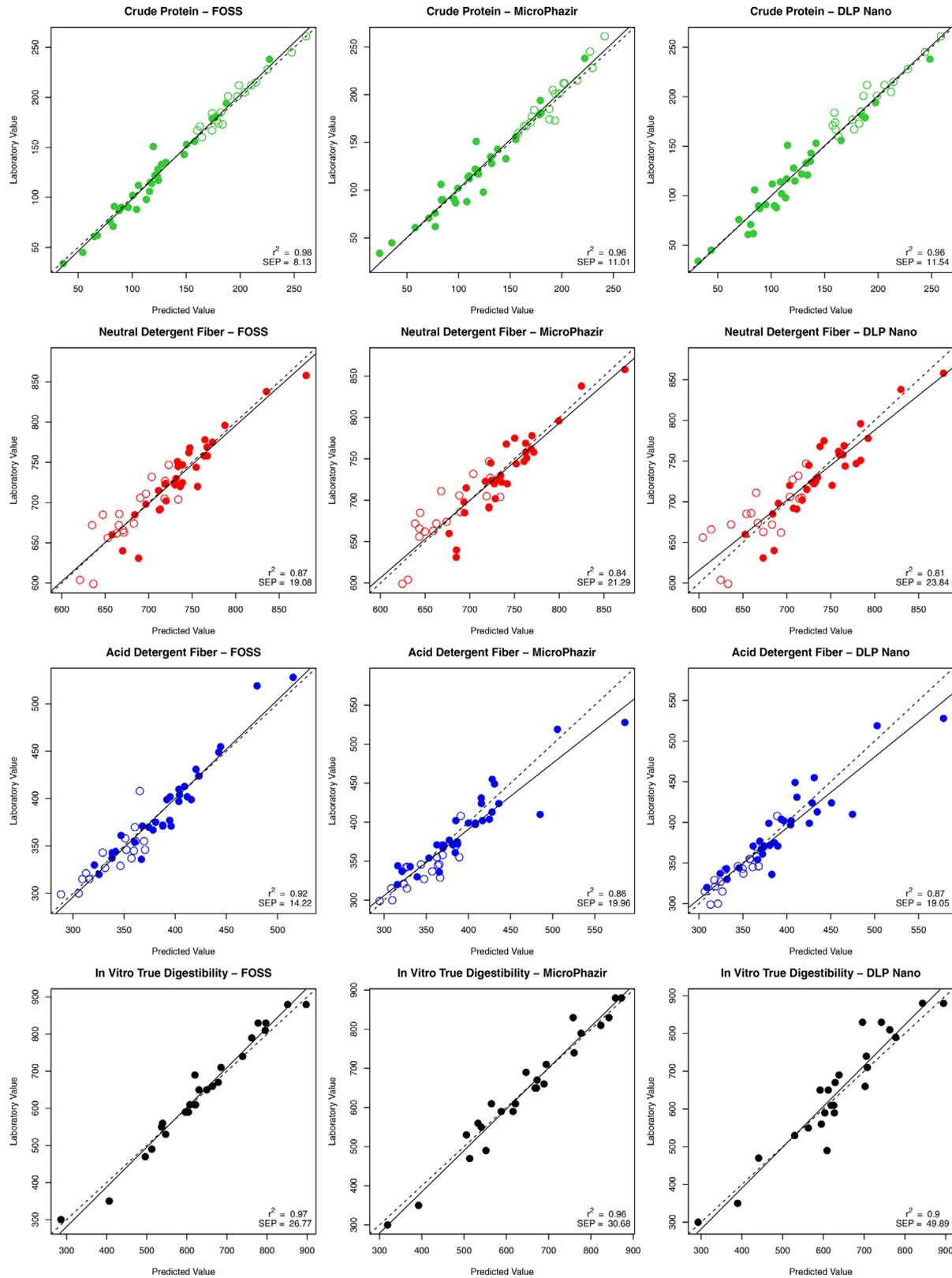


FIGURE 2 Validation scatterplots of wet-chemistry values (ordinate) and near infrared (NIR) predictions (abscissa) for crude protein, amylase and sodium sulfite-treated neutral detergent fiber, acid detergent fiber, and in vitro true digestibility of bermudagrass and switchgrass samples. Filled circles correspond to switchgrass and white-filled circles to bermudagrass samples. The dotted line in each figure represents a line with slope = 1 and the solid line represents the linear-regression line for the data

TABLE 4 Fit statistics for performance evaluation of predictions of near infrared models developed using 75% of the observations in the database (split dataset) and evaluated on the remaining 25% of the observations (test set). The fit statistics correspond to the scatter plots in Figure 2

Variable ^a	Device	Spectral transformation ^b	r^2	SEP g kg ⁻¹	RPD	BIAS	g kg ⁻¹			
							β_0 LL	β_0 UL	β_1 LL	β_1 UL
CP	FOSS	MSC	.98	8.1	6.8	0.4	-12.6	0.8	1.0	1.1
	microPHAZIR	NIR	.96	11.0	5.1	2.6	-9.8	8.6	1.0	1.1
	Nano	NIR	.96	11.5	4.8	0.5	-11.1	8.4	0.9	1.1
aNDF	FOSS	SNV	.87	19.1	2.7	-1.7	-58.7	101.2	0.9	1.1
	microPHAZIR	MSC	.84	21.3	2.4	-1.3	-40.9	132.6	0.8	1.1
	Nano	MSC_SG7	.81	23.8	2.2	-0.3	11.2	185.9	0.7	1.0
ADF	FOSS	SG7	.92	14.2	3.4	-0.9	-50.4	18.7	0.9	1.1
	microPHAZIR	NIR	.86	20.0	2.5	-6.0	13.1	89.5	0.7	0.9
	Nano	SNV_SG7	.87	19.1	2.6	-4.0	5.3	81.7	0.8	1.0
IVTD	FOSS	SNV_SG7	.97	26.8	5.6	5.2	-91.3	9.5	1.0	1.1
	microPHAZIR	DT	.96	30.7	4.9	-1.8	-98.1	22.5	1.0	1.1
	Nano	SNV	.90	49.9	3.0	9.3	-146.2	60.1	0.9	1.2

Note. r^2 , coefficient of determination, prediction; SEP, standard error of prediction; RPD, ratio of performance to deviation (SD/SEP; SD = standard deviation of reference samples in validation test-set); β_0 , intercept (of the regression line in Figure 2; LL and UL represent lower and upper limits, respectively, of 95% confidence interval); β_1 , slope of the regression line in Figure 2; LL and UL represent lower and upper limits, respectively, of 95% confidence interval.

^a CP, crude protein; aNDF, amylose and sodium sulfite-treated neutral detergent fiber; ADF, acid detergent fiber; IVTD, in vitro true digestibility.

^b MSC, multiplicative scatter correction; NIR, raw spectra (log R^{-1} ; R = reflectance); SNV, standard normal variate; DT, detrend; SG7, Savitzky-Golay smoothed spectra using seven points.

TABLE 5 Example summary table of fit statistics of several near-infrared models ($n = 75$; split-dataset) developed to predict in vitro true digestibility (IVTD) from spectra collected using the Nano handheld near-infrared spectroscopic device described in Table 1. The model in the first row was selected as the best model and further evaluated using a test set in Table 4 and Figure 2

Spectral transformation ^a	Factors ^b	R^2_c ^c	SEC ^d g kg ⁻¹	R^2_{cv} ^e	SECV ^f g kg ⁻¹
SNV	13	.99	13.0	.96	30.7
MSC	10	.98	18.6	.95	32.0
DT	9	.98	19.3	.96	31.0
SG5	5	.97	25.5	.93	38.6
SG7	8	.99	15.0	.94	34.4
SNV_DT	9	.98	19.3	.96	31.0
MSC_DT	9	.98	19.3	.96	31.0
SNV_SG5	6	.98	20.6	.94	35.5
SNV_SG7	6	.98	19.5	.95	32.9
MSC_SG5	6	.98	21.6	.93	38.4
MSC_SG7	8	.99	15.5	.93	37.4
DT_SG5	6	.98	20.8	.94	35.7
DT_SG7	6	.98	19.7	.95	33.4
NIR	12	.98	19.1	.95	33.1

^aSNV, standard normal variate; MSC, multiplicative scatter correction; DT, detrend; SG5, Savitzky-Golay smoothed spectra using five points; NIR, raw spectra (log R^{-1}).

^bNumber of loading factors (latent variables) in the partial least squares regression models.

^c R^2_c , coefficient of determination, calibration.

^dSEC, standard error of calibration.

^e R^2_{cv} , coefficient of determination, cross-validation.

^fSECV, standard error of cross-validation.

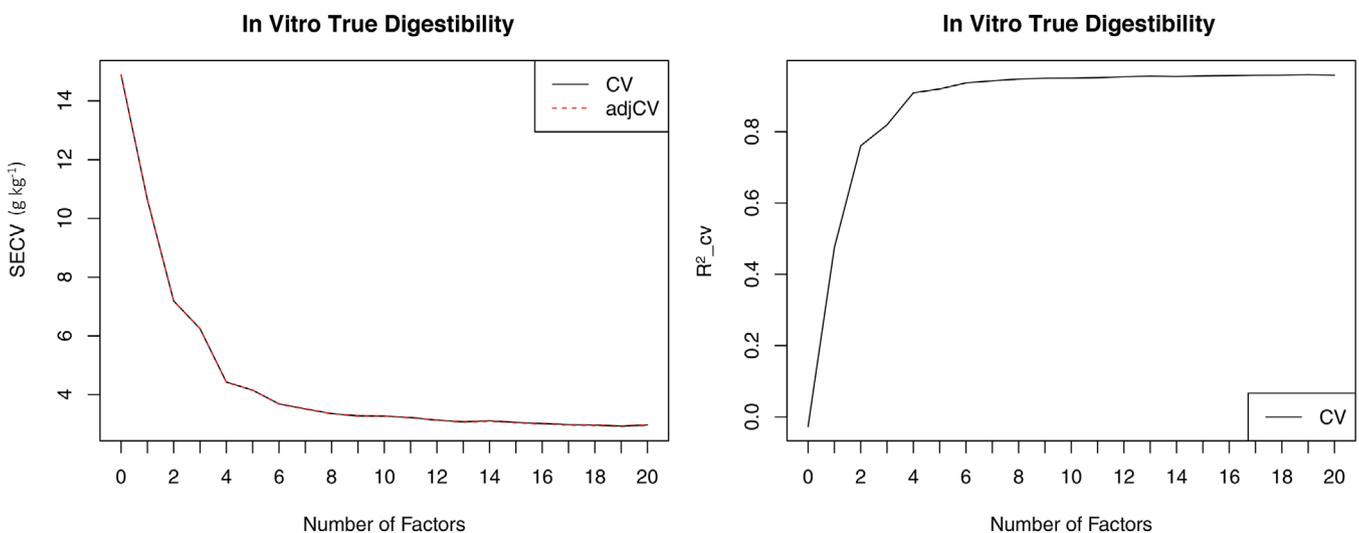


FIGURE 3 Example of fit statistics as a function of number of latent variables (factors) for the first model presented in Table 5

intervals for the intercept (β_0) and slope (β_1) of each of the regression models presented in Figure 2. Acceptable models are those with bias values closer to zero, slope values closer to 1, and intercept values closer to zero. The RPD values ranged from 2.2 to 6.8. Chang, Laird, Maus-

bach, and Hurburgh (2001) suggested that RPD values >2 indicate a model with good prediction ability; however, Williams (2014) presented different and higher threshold values for RPD. Minasny and McBratney (2013) indicated that reporting RPD values may be redundant when already

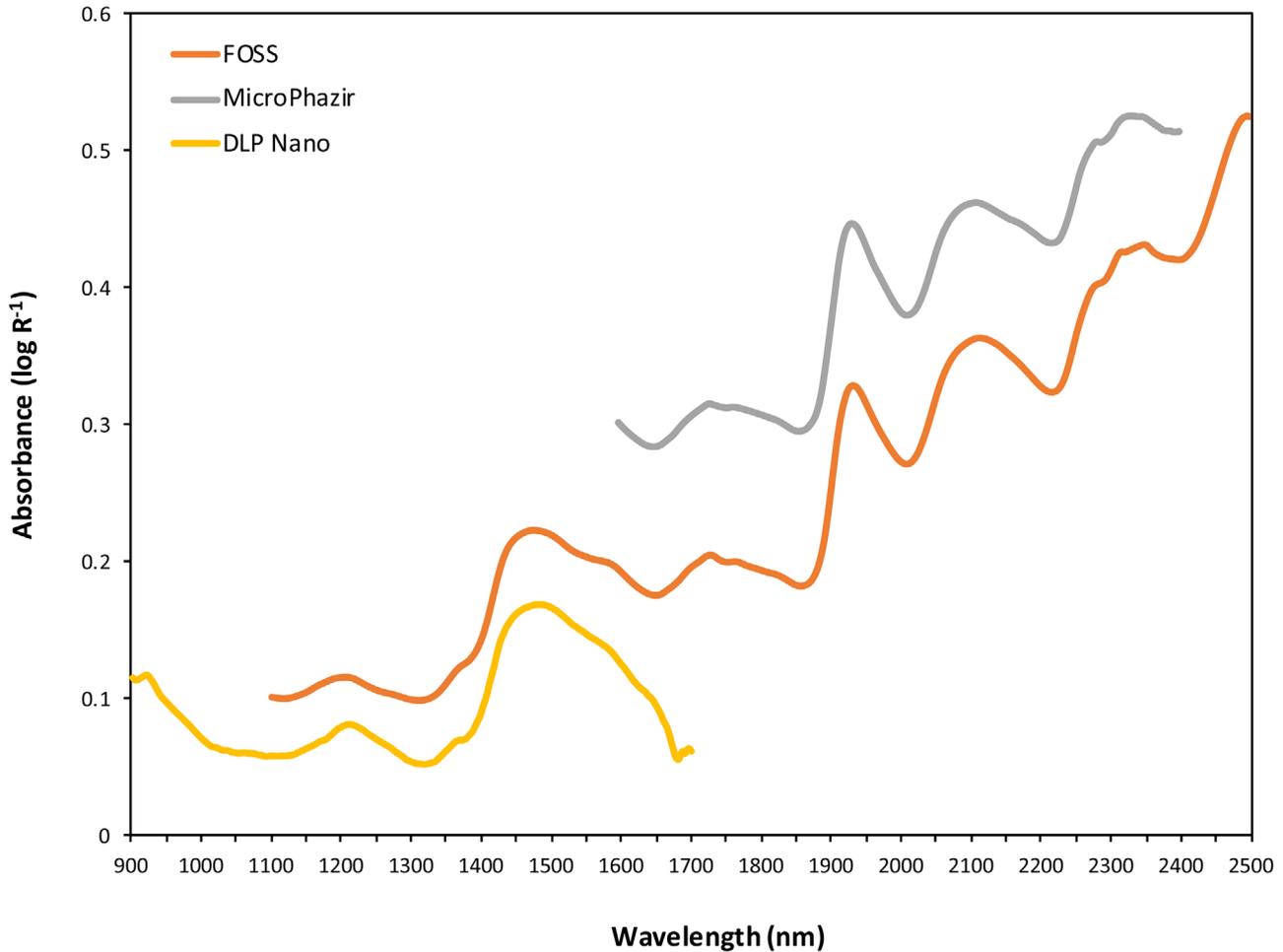


FIGURE 4 Raw spectra measured by three near infrared spectroscopy devices. The plotted spectra correspond to the mean values of all scanned samples across instrument resolutions

reporting r^2 values for normally distributed variables and large datasets.

The average raw spectrum was different among devices (Figure 4). The three devices have different light sources, generate different intensities of light at different wavelengths, and use different reflectance detection mechanisms, so it is expected that the spectra look somewhat different. However, it is clear that the peaks and valleys of spectra from the different devices are generally aligned in the overlapping wavelength ranges (Figure 4). In general, the models obtained with the FOSS 6500 NIRSystems device had better fit statistics, followed closely by the Thermo Scientific microPHAZIR, and the Texas Instruments DLP NIRscan Nano EVM. The FOSS device covered a wider spectral range (1100–2498 nm) at a 2-nm wavelength interval vs. the microPHAZIR (1600–2400 nm; every 8 nm) and the Nano (900–1700 nm; every 5 nm). In a study with pharmaceuticals, Kolomiets and Siesler (2004) reported that spectral resolution did not have a systematic influence on the accuracy of the chemomet-

ric evaluation. Similarly, in a study that evaluated three NIR devices (covering 1000–2500 nm, 1140–2200 nm, and 1580–2400 nm) to determine the concentration of rosmarinic acid in *Rosmarini folium*, Kirchler et al. (2017) concluded that resolution of the devices seemed to be of secondary importance relative to spectral range. Rosmarinic acid concentration was ~3% of the leaf mass (Kirchler et al., 2017), and this study showed that rosmarinic acid and the surrounding molecules produced from 6 to 11 important absorption bands in the spectral range, depending on the device. In our study, the FOSS machine covered a very large range, while the microPHAZIR covered about one-half of that range, and the Nano covered the other half of the range. The concentration of analytes in this study ranged from 140.7 to 717.6 g kg⁻¹ of dry matter (CP and aNDF, respectively). All three devices produced models that predicted all forage nutritive values with very similar precision (Table 4). This implies that there are a number of informative bands (sufficient for very accurate model development) across the entire NIR range from 1100

to 2498 nm, and that no specific portion of the range is absolutely critical to good NIR models for the analytes in this study. It is clear that for all nutritive value estimates, the NIR-predicted values are very close to the laboratory-measured values, indicating a strong goodness of fit for the selected models (r^2 between .81 and .98; SEP of 8.1–11.5 for CP, 19.1–23.8 for aNDF, 14.2–20.0 for ADF, and 26.8–49.9 for IVTD) (Figure 2). Additionally, indications of heteroskedasticity or under–over estimation are not observed. However, dispersion of points around reference lines (linear model and $Y = X$) varied across device and specific nutritive value estimate. For instance, among the nutritive values, the smallest dispersions are observed in the CP models; among devices, the dispersion of the points is always lower for the FOSS model than for the microPHAZIR or Nano.

4 | CONCLUSIONS AND IMPLICATIONS

Forage nutritive value estimates were successfully predicted using NIR models. Consistently, the prediction models from the benchtop-type device (FOSS) had the highest r^2 and lowest SEP values; however, the predictive power of the handheld-type devices (microPHAZIR and Nano) was very similar to the benchtop-type device, notwithstanding instrument differences and spectral range differences. Different mathematical transformations and different number of latent variables optimized the NIR models per analyte and per device, thus highlighting a benefit of flexible NIR model developing software. Our results indicate that handheld NIR spectroscopy devices have potential to generate comparable NIR models to benchtop-type devices for prediction of forage nutritive value estimates; hence, further investigation of handheld devices is warranted.

ACKNOWLEDGMENTS

The authors express their appreciation to Stephanie Sosinski, Research Technician of the Forage and Grassland Program at NC State University, and to Willi Woodbridge, Camcore's database administrator, for their excellent technical support.

ORCID

J. J. Acosta  <https://orcid.org/0000-0002-9429-5166>

M. S. Castillo  <https://orcid.org/0000-0003-1066-5906>

REFERENCES

- Batten, G. D. (1998). An appreciation of the contribution of NIR to agriculture. *Journal of Near Infrared Spectroscopy*, 6, 105–114. <https://doi.org/10.1255/2Fjnirs.127>
- Bekewe, P. E., Castillo, M. S., Acosta, J. J., & Rivera, R. (2019). Defoliation management effects on nutritive value of 'Performer' switchgrass. *Crop Science*, 60, 1682–1689. <https://doi.org/10.1002/csc.20036>
- Bekewe, P. E., Castillo, M. S., & Rivera, R. (2018). Defoliation management affects productivity, leaf/stem ratio, and tiller counts of 'Performer' switchgrass. *Agronomy Journal*, 110, 1467–1472. <https://doi.org/10.2134/agronj2018.01.0003>
- Breunig, M., Hans-Peter, K., Raymond, T., & Sander, J. (2000). LOF: Identifying density-based local outliers. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93–104). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/342009.335388>
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108–132. <https://doi.org/10.1006/jmps.1999.1279>
- Burns, J. C. (2011). Advancement in assessment and the reassessment of the nutritive value of forages. *Crop Science*, 51, 390–402. <https://doi.org/10.2135/cropsci2010.06.0334>
- Burns, J. C., Fisher, D. S., & Rottinghaus, G. E. (2006). Grazing influences on mass, nutritive value, and persistence of stockpiled Jesup tall fescue without and with novel and wild-type fungal endophytes. *Crop Science*, 46, 1898–1912. <https://doi.org/10.2135/cropsci2005.09-0327>
- Burns, J. C., Mochrie, R. D., & Timothy, D. H. (1984). Steer performance from two perennial *Pennisetum* species, switchgrass, and a fescue—'Coastal' bermudagrass system. *Agronomy Journal*, 76, 795–800. <https://doi.org/10.2134/agronj1984.00021962007600050020x>
- Castillo, M. S., Tiezzi, F., & Franzluebbers, A. J. (2020). Tree species effects on understory forage productivity and microclimate in a silvopasture of the Southeastern USA. *Agriculture, Ecosystems & Environment*, 295, 106917. <https://doi.org/10.1016/j.agee.2020.106917>
- Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65, 480–490. <https://doi.org/10.2136/sssaj2001.652480x>
- Coleman, S. W., Christiansen, S., & Shenk, J. S. (1990). Prediction of botanical composition using NIRS calibrations developed from botanically pure samples. *Crop Science*, 30, 202–207. <https://doi.org/10.2135/cropsci1990.0011183X003000010044x>
- Crocombe, R. A. (2018). Portable spectroscopy. *Applied Spectroscopy*, 72, 1701–1751. <https://doi.org/10.1177/2F0003702818809719>
- Dairy One. (2015). Analytical procedures. Retrieved from <https://dairyone.com/download/forage-forage-lab-analytical-procedures/>
- Diana, G., & Tommasi, D. (2002). Cross-validation methods in principal component analysis: A comparison. *Statistical Methods & Applications*, 11, 71–82. <https://doi.org/10.1007/BF02511446>
- Dixon, R., & Coates, D. (2009). Review: Near infrared spectroscopy of faeces to evaluate the nutrition and physiology of herbivores. *Journal of Near Infrared Spectroscopy*, 17, 1–31. <https://doi.org/10.1255/2Fjnirs.822>
- Hill, G. M., Gates, R. N., & West, J. W. (2001). Advances in bermudagrass research involving new cultivars for beef and dairy produc-

- tion. *Journal of Animal Science*, 79, E48–E58. <https://doi.org/10.2527/jas2001.79E-SupplE48x>
- Hodge, G. R., Acosta, J. J., Unda, F., Woodbridge, W. C., & Mansfield, S. D. (2018). Global near infrared spectroscopy models to predict wood chemical properties of Eucalyptus. *Journal of Near Infrared Spectroscopy*, 26, 117–132. <https://doi.org/10.1177/0967033518770211>
- Karayilanli, E., Cherney, J. H., Sirois, P., Kubinec, D., & Cherney, D. J. R. (2016). Botanical composition prediction of alfalfa–grass mixtures using NIRS: Developing a robust calibration. *Crop Science*, 56, 3361–3366. <https://doi.org/10.2135/cropsci2016.04.0232>
- Kirchler, C. G., Pezzeri, C. K., Beć, K. B., Mayr, S., Ishigaki, M., Ozaki, Y., & Huck, C. W. (2017). Critical evaluation of spectral information of benchtop vs. portable near-infrared spectrometers: Quantum chemistry and two-dimensional correlation spectroscopy for a better understanding of PLS regression models of the rosmarinic acid content in Rosmarini folium. *Analyst*, 142, 455–464. <https://doi.org/10.1039/C6AN02439D>
- Kolomiets, O., & Siesler, H. W. (2004). The influence of spectral resolution on the quantitative near infrared spectroscopic determination of an active ingredient in a solid drug formulation. *Journal of Near Infrared Spectroscopy*, 12, 271–277. <https://doi.org/10.1255/2Fjnirs.435>
- Minasny, B., & McBratney, A. (2013). Why you don't need to use RPD. *Pedometron*, 33, 14–15. Retrieved from <http://www.pedometrics.org/Pedometron/Pedometron33.pdf>
- Mevik, B., & Wehrens, R. (2016). Introduction to the pls package, R project. Retrieved from <http://cran.r-project.org>
- Norris, K. H., Barnes, R. F., Moore, J. E., & Shenk, J. S. (1976). Predicting forage quality by infrared reflectance spectroscopy. *Journal of Animal Science*, 43, 889–897. <https://doi.org/10.2527/jas1976.434889x>
- Osten, D. W. (1988). Selection of optimal regression models via cross-validation. *Journal of Chemometrics*, 2, 39–48. <https://doi.org/10.1002/cem.1180020106>
- Pérez-Marin, D., Paz, P., Guerrero, J. E., Garrido-Varo, A., & Sánchez, M. T. (2010). Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality of refrigerated storage behavior in plums. *Journal of Food Engineering*, 99, 294–302. <https://doi.org/10.1016/j.jfoodeng.2010.03.002>
- Pittman, J. J., Arnall, D. B., Interrante, S. M., Wang, N., Raun, W. R., & Butler, T. J. (2016). Bermudagrass, wheat, and tall fescue crude protein forage estimation using mobile-platform, active-spectral and canopy-height data. *Crop Science*, 56, 870–881. <https://doi.org/10.2135/cropsci2015.05.0274>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/
- Roberts, C. A., Benedict, H. R., Hill, N. S., Kallenbach, R. L., & Rottinghaus, G. E. (2005). Determination of ergot alkaloid content in tall fescue by near-infrared spectroscopy. *Crop Science*, 45, 778. <https://doi.org/10.2135/cropsci2005.0778>
- Roberts, C. A., Stuth, J., & Flinn, P. (2004). Analysis of forages and feedstuffs. In C. A. Roberts, J. Workman, & J. B. Reeves (Eds.), *Near infrared spectroscopy in agriculture* (pp. 231–267). Agronomy Monograph 44. Madison, WI: ASSA, CSSA, and SSSA.
- Saha, U., Vann, R. A., Reberg-Horton, S. C., Castillo, M. S., Mirsky, S. B., McGee, R. J., & Sonon, L. (2018). Near-infrared spectroscopic models for analysis of winter pea (*Pisum sativum* L.) quality constituents. *Journal of the Science of Food and Agriculture*, 98, 4253–4267. <https://doi.org/10.1002/jsfa.8947>
- Shenk, J. S., & Westerhaus, M. O. (1991). Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Science*, 31, 469–474. <https://doi.org/10.2135/cropsci1991.0011183x003100020049x>
- Shenk, J. S., Workman, J. J., & Westerhaus, M. O. (2008). Application of NIR spectroscopy to agricultural products. In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of near-infrared analysis* (pp. 347–382). Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Starks, P. J., Brown, M. A., Turner, K. E., & Venuto, B. C. (2015). Canopy visible and near-infrared reflectance data to estimate alfalfa nutritive value before harvest. *Crop Science*, 56, 484–494. <https://doi.org/10.2135/cropsci2015.03.0162>
- Texas Instruments (2020). DLP® NIRscanTM Nano EVM user's guide. Retrieved from <http://www.ti.com/lit/ug/dlpu030g/dlpu030g.pdf>
- ThermoFisher Scientific. (2020). MicroPHAZIR. Retrieved from <https://www.thermofisher.com/order/catalog/product/MICROPHAZIRPC?SIDsrch-srp-MICROPHAZIRPC#/MICROPHAZIRPC?SIDsrch-srp-MICROPHAZIRPC>
- USDA–NRCS. (2018). Economic and production performance of grazing native grasses in the fescue belt. Retrieved from <https://www.farmers.gov/sites/default/files/documents/ScienceToSolutionsBobwhiteForage.pdf>
- Vogel, K. P., Dien, B. S., Jung, H. G., Casler, M. D., Masterson, S. D., & Mitchell, R. B. (2011). Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses. *BioEnergy Research*, 4, 96–110. <https://doi.org/10.1007/s12155-010-9104-4>
- Warburton, P., Brawner, J., & Meder, R. (2014). Technical note: Handheld near infrared spectroscopy for the prediction of leaf physiological status in tree seedlings. *Journal of Near Infrared Spectroscopy*, 22, 433–438.
- Williams, P. (2014). The RPD statistic: A tutorial note. *NIR News*, 22. <https://doi.org/10.1255/2Fnrn.1419>
- Williams, P., Darnenne, P., & Flinn, P. (2017). Tutorial: Items to be included in a report on a near infrared spectroscopy project. *Journal of Near Infrared Spectroscopy*, 25, 85–90. <http://doi.org/10.1177/0967033517702395>
- Workman, Jr., J., & Burns, D. A. (2007). Commercial NIR instrumentation. In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of near-infrared analysis* (pp. 67–68). Boca Raton, FL: CRC Press, Taylor & Francis Group.

How to cite this article: Acosta JJ, Castillo MS, Hodge GR. Comparison of benchtop and handheld near-infrared spectroscopy devices to determine forage nutritive value. *Crop Science*. 2020;1–13. <https://doi.org/10.1002/csc2.20264>