A Performance Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional Laboratory Spectrometer for Rapid Biomass Compositional Analysis

Edward J. Wolfrum [1] · Courtney Payne [1] · Alexa Schwartz [1] · Joshua Jacobs [2] · Robert W. Kressin [3]

Abstract

The performance of a conventional laboratory near-infrared (NIR) spectrometer and two NIR spectrometer prototypes (a Texas Instruments NIRSCAN Nano evaluation model (EVM) and an InnoSpectra NIR-M-R2 spectrometer) are compared by collecting reflectance spectra of 270 well-characterized herbaceous biomass samples, building calibration models using the partial least squares (PLS-2) algorithm to predict five constituents of the samples from the reflectance spectra, and comparing the resulting model statistics. The prediction models developed using spectra from the Foss XDS spectrometer were slightly better than the prediction models developed using spectra from either the TI NIRSCAN Nano EVM and the InnoSpectra NIR-M-R2 as measured by the root mean square error (RMSECV) and the correlation coefficient (R^2_cv) for "leave-one-out" cross-validation (CV). The models built from the two prototype units were not statistically significantly different from each other ($p = 0.05$). The Foss spectrometer has a larger wavelength range (400–2500 nm) compared with the two prototypes (900–1700 nm). When the spectra from the Foss XDS spectrometer were truncated so their wavelength range matched the wavelength range of the two prototype units, the resulting model was not statistically significantly different from the models from either prototype.

Keywords Biomass feedstocks · Multivariate analysis · Instrumentation · Chemometrics

Introduction

Rapid analysis using spectroscopy and chemometrics is considered a secondary analytical technique because it requires extensive calibration with a representative set of samples of known composition to develop a robust predictive model. For rapid analysis using near-infrared (NIR) spectroscopy, NIR spectra are collected from a set of well-characterized samples, and a calibration model is developed using a variety of multivariate statistical techniques. NIR spectra of new samples are then collected, and the calibration model is used to predict the composition of these new samples. The calibration set must be carefully chosen to reflect the concentration range of the analytes to be measured as well as the nature of the samples to be predicted. This "calibrate-collect-predict" cycle is common across all applications where spectroscopy is used for rapid analysis [1, 2].

Small, portable, low-cost instruments demonstrate the potential of ubiquitous NIR spectroscopy applications. As the portability of a NIR spectrometer increases and its unit cost decreases, spectrometers could be used at multiple points along a given conversion chain from raw materials to finished products—a network of NIR spectrometers could track the composition of specific lots of raw materials as they move through a conversion chain. The nature of the raw material and the conversion chain could vary widely—crude oil conversion to fuels and petrochemicals; natural product conversion to consumer products or pharmaceuticals; corn grain conversion to ethanol and byproducts; and agricultural product conversion to foods, or biomass harvest, transportation, storage, and conversion to biofuels and biochemicals. There have been substantial advances in the development of small,

✉ Edward J. Wolfrum
ed.wolfrum@nrel.gov

1 National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA

2 Texas Instruments Incorporated, 13532 North Central Expressway, MS 3810, Dallas, TX 75243, USA

3 KS Technologies, 11580 Black Forest Road Suite no. 60, Colorado Springs, CO 80908, USA

portable NIR spectrometers in recent years. A review by Pasquini included a comprehensive presentation of such instruments [3]. These units operate using different optical diffraction grating properties, such as Fourier-transform (Si-Ware NeoSpectra), linear variable filter (Viavi microNIR), and diffraction (Texas Instruments TI NIRScan and NIRScan Nano). Yan and Siesler [4] demonstrate the use of these low-cost FT-NIR, LVF, and diffraction NIR systems for both classification and measurement.

There have been several recent works using NIR spectroscopy in biomass-relevant areas. Tao et al. [5] used NIR spectroscopy to examine the aflatoxin content of corn kernels. The group used both partial least squares discriminant analysis (PLS-DA) to classify corn kernels as contaminated or uncontaminated, and partial least squares regression to predict the aflatoxin content. Gao et al. [6] discuss NIR among other spectroscopic approaches (Raman, MIR, GC-MS, NMR) to do nontargeted analysis of food fraud. The nontargeted analysis attempts to identify samples that are "atypical" (i.e., not likely to be part of a population of known unadulterated samples) without having to identify the specific adulterant or contaminant. Essentially, the spectroscopic signatures of a set of samples serve as a fingerprint, and then, multivariate statistical or machine learning approaches can be used to determine whether the fingerprint of a given sample is "typical" or "atypical." Curzon et al. [7] used NIR spectroscopy in the "calibrate-collect-predict" manner mentioned above to nondestructively predict the protein content of wheat and spelt samples, but also used the raw NIR spectra directly as an additional phenotype dataset along with genetic data to develop classification models via logistic regression to differentiate among multiple cultivars of wheat and spelt grain grown in different locations under different conditions.

Equally important for ubiquitous biomass characterization across the biomass value-chain will be the opportunities presented by the combination of very low-cost, ubiquitous NIR spectroscopy and robust networking to enable advanced analytics on the spectroscopy data itself. It is likely that (in addition to predicted chemical information via calibration models) advanced machine learning algorithms will be used to derive useful information from the dataset provided by a collection of NIR spectrometers each generating a large number of spectra, particularly when combined with operating data from the process being monitored. Such "big data" approaches are already being investigated in the chemical process industries [8, 9]. In this sense, low-cost NIR spectrometers represent a new class of data-rich process sensors.

To realize these opportunities, it is first necessary to have robust, well-characterized, and functional spectrometers that reliably generate reproducible spectra. Thus, the objective of this work was to compare the performance of a Foss XDS NIR laboratory spectrometer to two next-generation NIR spectrometer prototypes.

Materials and Methods

The experimental work consisted of collecting near-infrared spectra from a well-characterized set of 270 herbaceous biomass samples and developing prediction models from these spectra using standard multivariate statistical methods. In this section, the biomass samples used, the spectrometer hardware, the spectra collection techniques, and the chemometric modeling performed are described.

Biomass Samples

A total of 270 mixed herbaceous feedstock samples were used for this work. These samples have been reported on previously [10, 11]. The feedstock types in the set include corn stover, miscanthus, switchgrass, sorghum, cool season grasses, and rice straw. All samples were characterized using standard biomass compositional analysis methods [12, 13]. For this work, data comprising the structural carbohydrates glucan and xylan, total extractives, ash, and lignin content were used.

Spectrometer Hardware

The conventional laboratory unit was a Foss XDS near-infrared (NIR) spectrometer (Foss North America, Eden Prairie MN, USA). The XDS instrument uses a pre-dispersive moving grating as monochromator, a spectral range of 400–2,500 nanometers (nm) with a resolution of 0.5 nm, and silicon (Si, 400–1100 nm) and lead sulfide (PbS, 1100–2500 nm) detectors. Figure S1 in the Electronic Supplementary Material shows a photograph of the three spectrometers used in this work: a conventional laboratory spectrometer and two low-cost and potentially portable spectrometers.

The second and third units were a NIRSCAN Nano NIR spectrometer evaluation module (EVM, Texas Instruments Incorporated, Dallas, TX, USA) and an NIR-M-R2 NIR spectrometer (InnoSpectra Corporation, Hsinchu, Taiwan). Both spectrometers share the same design, which consists of a pair of broadband tungsten filament lamps, a sapphire window, collimating and focusing lenses, a post-dispersive fixed reflective diffraction grating, a digital micromirror device (Texas Instruments DLP2010NIR DMD), and a single-point InGaAs detector [14]. They have a spectral range of 900–1700 nm. The DMD in both units has approximately 400,000 individual micromirrors arranged in an 854 × 480 array. In these units, the DMD acts as a monochromator; the fixed grating disperses the reflected light across the DMD. Unless the mirrors are actuated, the light is directed away from the single-point detector. Individual mirrors (or small groups of these mirrors) can be actuated to direct the light into the single-point detector. The resolution of the spectrometers is a

function of the number of mirrors in the DMD array and how they are selectively actuated.

NIR Spectra Collection

All 270 samples were milled to a 2-mm particle size and packed into Foss "quarter cup" cells with an optical glass window for scanning NIR spectra of the biomass samples in reflectance mode. The samples were scanned on the Foss XDS from 400 to 2500 nm averaging 32 scans with a resolution of 0.5 nm with a total scan time of approximately 1 min. The samples were scanned on the InnoSpectra NIR-M-2 and the TI NIRSCAN Nano EVM from 900 to 1700 nm averaging 32 scans with a resolution of approximately 10.5 nm and a total scan time of 55 s. The Foss XDS spectrometer uses an internal white reference. For the other two spectrometers, an external white reference (Calibrated Diffuse Reflectance Target, p/n AA-00823-000, Labsphere, N Sutton, NH) was used. The external reference was scanned before scanning samples. Additionally, the reference was re-scanned every 120 min during sample scanning. All samples were scanned in duplicate; the cell holding the biomass samples was repositioned before rescanning.

Spectra from the Foss XDS NIR spectrometer had a range of 400–2500 nm and a spacing of 0.5 nm, resulting in 4200 points for each spectrum. Spectra from the TI NIRSCAN Nano EVM and the InnoSpectra NIR-M-2 had ranges of 900–1700 nm and a spacing of approximately 3 nm, resulting in 289 points for each spectrum. All spectrometer lamps were enabled 30 min prior to collecting spectra, and the lamps remained illuminated for the duration of the sample scanning.

The repeatability and reproducibility were measured according to the method of Sirisomboon et al. [15]. In brief, a single biomass sample (corn stover milled to pass through a 2-mm screen) in a quartz cell (Foss quarter-cup cell, P/N NIR-65-039) was used for this work. Repeatability was determined by scanning the sample 10 consecutive times without moving the cell, and reproducibility was determined by scanning the sample 10 consecutive times while repositioning the cell between scans. The relative standard deviation (RSD, standard deviation divided by the mean) of the derivatized spectra was calculated at 1352 and 1416 nm for the spectra from the Foss XDS instrument and 1353 nm and 1401 nm for the spectra from the NIRSCAN Nano EVM and InnoSpectra NIR-M-2 instruments. These wavelengths correspond to peaks in the derivatized spectra from all three instruments.

Chemometric Modeling

Spectral data from the Foss NIR spectrometer were exported from the instrument operating software (Win-ISI) and, along with the biomass compositional analysis data, were stored in Excel spreadsheets. Spectral data from the TI and InnoSpectra units were stored in comma-separated value (CSV) text files. All data manipulation and chemometric modeling were performed using the R language (version 3.5.1) [16] using the following packages: pls (version 2.7-2) and prospectr (version 0.1.3).

Partial least squares (PLS-2) models to simultaneously predict the glucan, xylan, lignin, extractives, and ash content of the biomass samples were built—one model each using the spectra from the TI EVM and the InnoSpectra unit. Two models were built using the spectra from the Foss XDS spectrometer; one using the spectral range 1100–2500 nm to match previous work with corn stover feedstock samples [17] and mixed herbaceous feedstocks [11], and one using the spectral range of 900–1700 nm to match the spectral range of the TI EVM and the InnoSpectra unit.

The spectra were mathematically pretreated prior to the PLS-2 modeling using standard normal variate (SNV) normalization, detrending, and a Savitzky-Golay smoothing algorithm (1st derivative, 2nd order polynomial smoothing). For the spectra from the Foss unit, a window of 25 points was used for smoothing, while for the spectra from the TI and InnoSpectra units, a smaller window of 9 points was used. All samples from the initial calibration model whose predicted value was greater than 2.5 times the model's RMSEC value for any of the constituents were removed as an outlier. The purpose of using a standard algorithm for outlier removal was to minimize any subjectivity in the development of the model. Unlike outlier removal, the selection of the optimal number of principal components (#PCs) was done qualitatively; there was not a clear minimum in a plot of RMSECV value for all components vs. the number of principal components in the model. The optimal number of principal components was chosen where the slope of the RMSECV value vs. #PCs line was either a minimum or decreasing slowly for all constituents. Twelve (12) principal components were chosen for each of the four models presented here.

The models were validated using the "leave one out" (LOO) cross-validation algorithm. In LOO cross-validation, each sample is removed from the calibration set and a new PLS-2 calibration model is developed and used to predict the sample not in that model. The model quality was determined by the coefficient of determination (square of correlation coefficient R) between the predicted and measured values for both the calibration and validation models (R^2-calibration and R^2-LOO CV) and the root mean square error of the calibration and cross-validation models (RMSEC and RMSECV). Quantitative comparisons of RMSE and R^2 values from different models (both calibration and LOO cross-validation) are made using a modified F-test for RMSE values [18] and of correlation coefficients using a t test after applying the Fisher z-transform. Since the square of RMSE values is essentially a variance measure, the square of the ratio of two RMSE values (either RMSEC or

RMSECV) of different models can be compared with a critical F ratio given the number of degrees of freedom in each model. For the 270-sample calibration set used in this work, ratios of RMSEC or RMSECV values greater than approximately 1.1 are statistically significantly different ($p = 0.05$). Similarly, correlation coefficients (R values) can be compared after applying the Fisher z-transform to convert them to continuous variables. The "Tukey Honest Significant Difference" test was used to identify differences in R^2 values among spectrometers and constituents.

Results and Discussion

In Fig. 1, the near-infrared (NIR) spectra (plotted as the pseudo-absorbance, or negative base-10 logarithm of the reflectance, $-\log_{10}(R)$) collected with the three spectrometers used in this work are compared. Each spectrum represents the mean of the 270 mixed herbaceous feedstock samples used to build models. The two spectra labeled "Foss" indicate the wavelength ranges used for the two models built with Foss spectra (1100–2500 nm and 900–1700 nm). The shaded areas represent ± 1 standard deviation around the mean, illustrating the variability of the mixed herbaceous feedstock spectra.

These pseudo-absorbance spectra are quite similar among all three spectrometers, although the absolute values differ slightly among the different instruments (the spectra are offset slightly for ease of comparison). There is a minor absorption band around 1200 nm, and a much larger band around 1450 nm present in the spectra from all three spectrometers, and substantial absorption features in the Foss spectra above 1700 nm. The spectra from the Foss instrument show slightly less variation across the 270 samples in the calibration set, as indicated by the smaller shaded area round the Foss spectra. Figure 2 shows a qualitative difference between the reflectance spectra from TI EVM and the InnoSpectra at wavelengths above approximately 1650 nm—the reflectance spectra from the TI EVM are decreasing while the spectra from the InnoSpectra are increasing. The root causes of this difference are not clear, but it should be noted that while both spectrometers shared a common reference design, they were designed and manufactured by different companies.

In Table 1, the repeatability and reproducibility of the three spectrometers used in this work are presented, expressed as the relative standard deviation at two key wavelengths of 10 repeated scans with the sample kept in place (repeatability) and re-positioned between scans (reproducibility). The data show that the Foss XDS has the smallest repeatability at either wavelength, with the repeatability of the two portable units 2–3 times larger. The reproducibility of the Foss XDS is substantially better than the NIRSCAN Nano and the InnoSpectra NIR-M-R2 instruments, which again are quite similar. It is

Fig. 1 Near-infrared (NIR) spectra collected with the three spectrometers used in this work. Each spectrum represents the mean of the 270 mixed herbaceous feedstock samples used to build models. The two spectra labeled "Foss" indicate the wavelength ranges used for the two models built with Foss spectra (1100–2500 nm, 900–1700 nm). The spectra are offset slightly for ease of comparison. The shaded areas represent ± 1 standard deviation around the mean, illustrating the variability of the mixed herbaceous feedstock spectra.
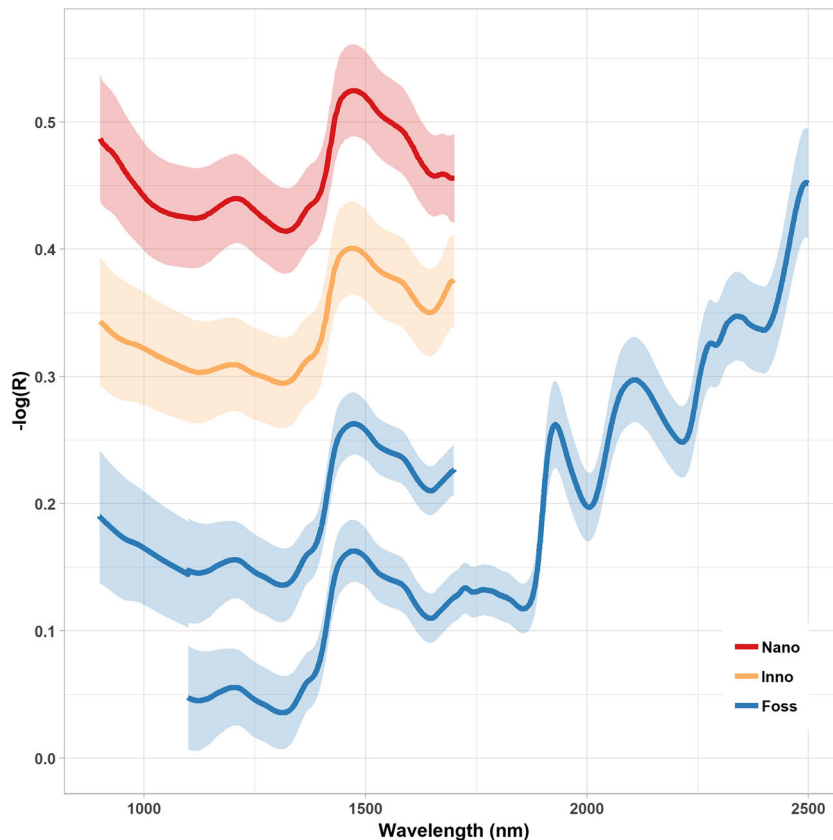
Fig. 2 Predicted vs. measured glucan content for PLS-2 models using derivatized spectra from **a** the Foss XDS (1100–250 nm), **b** the Foss XDS (900–1700 nm), **c** the TI NIRSCAN Nano, and **d** the InnoSpectra NIR-M-R2. All models used 12 principal components.
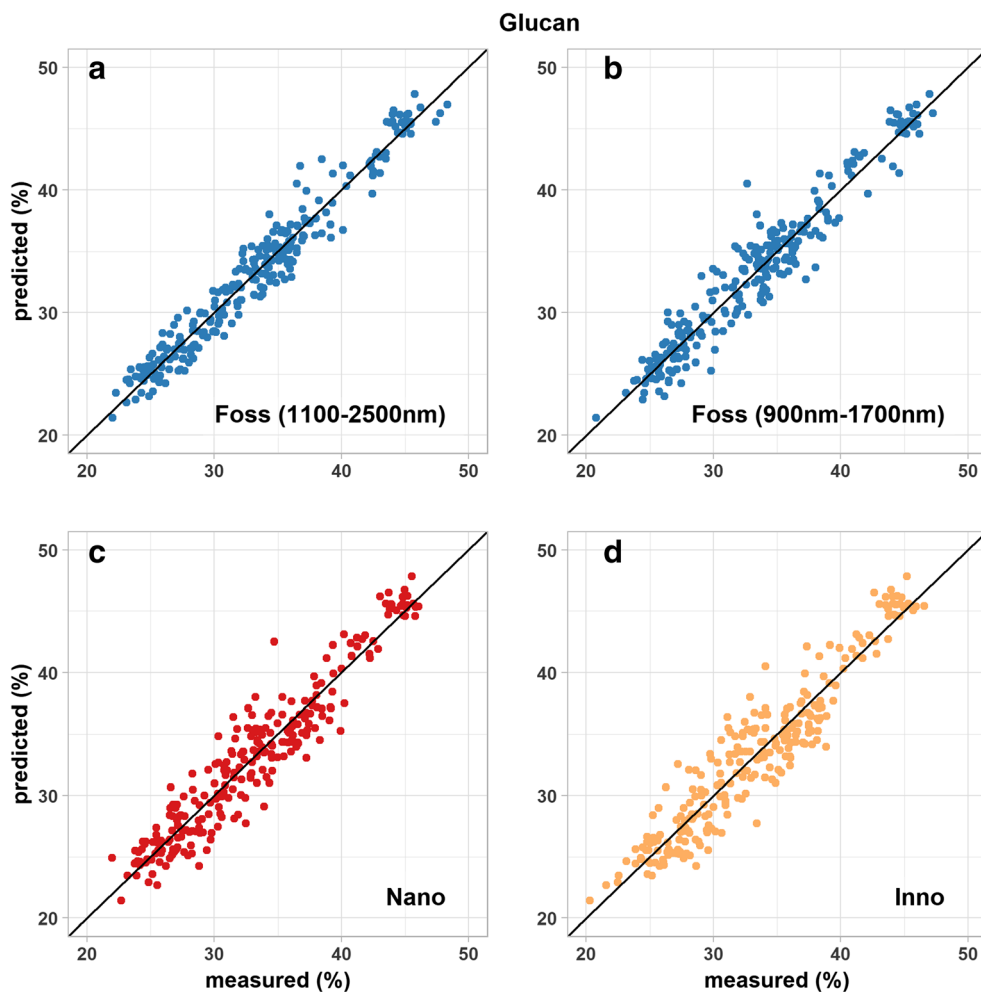
likely that this is due to the more robust sample presentation capabilities of the Foss XDS instrument compared with the two portable units.

Table 1 Results of repeatability and reproducibility tests using a sample of corn stover. Repeatability was determined as the relative standard deviation (RSD) of 10 spectra with the sample kept in place. Reproducibility was determined as the RSD of 10 with the sample repositioned between collection of the spectra. NIRSCAN Nano EVM and InnoSpectra NIR-M-R2 instruments had peaks at 1353 nm and 1401 nm, while the Foss XDS instrument had peaks at 1352 and 1415 nm

	~ 1350 nm	~ 1400 nm
Repeatability		
InnoSpectra NIR-M-R2	7.6×10^{-3}	7.7×10^{-3}
NIRSCAN Nano EMV	1.0×10^{-2}	1.2×10^{-2}
Foss XDS	3.0×10^{-3}	1.8×10^{-3}
Reproducibility		
InnoSpectra NIR-M-R2	6.7×10^{-2}	5.6×10^{-2}
NIRSCAN Nano EMV	7.3×10^{-2}	4.8×10^{-2}
Foss XDS	3.5×10^{-3}	5.4×10^{-4}

Comparison of Models

As mentioned above, one model using the spectra from the two prototypes was built, both of which have spectral ranges 900–1700 nm. Two models were built using spectra from the Foss instrument, one using the spectral range 1100–2500 nm, which has been used previously for corn stover feedstocks [17] and mixed herbaceous feedstocks [11], and one using only the 900–1700-nm range. In this way, the effects on model performance of the spectral range could be separated from other instrumental performance issues among the three spectrometers. The modeling results for four PLS-2 models developed using spectra collected using the three spectrometers in this study are summarized in Tables 1 and 2 and Figs. 2, 3, and 4.

In Fig. 2, the calibration model predictions of glucan content to the glucan content measured using primary analytical methods are compared. There is more scatter in the "prediction vs. measured" calibration plots for the models using spectra from the TI and InnoSpectra prototypes compared with the models using spectra from the Foss XDS spectrometer. The results from the two Foss models are different from models

Table 2 Summary statistics of the PLS-2 calibration models predicting the content of glucan, xylan, lignin, extractives, and ash in a mixed herbaceous feedstock population built using near-infrared (NIR) spectra from the Foss XDS NIR spectrometer, the Texas Instruments NIRSCAN Nano spectrometer EVM, and the InnoSpectra NIR-M-R2 NIR spectrometer EVM. R^2 coefficient of determination (square of correlation coefficient R), *RMSEC* root mean square error of calibration, *RMSECV* root mean square error of leave-one-out (LOO) cross-validation, *RMESP* root mean square error of prediction, *cal* calibration, *cv* cross validation

	Parameter	Foss XDS NIR[1]	Foss XDS NIR[2]	TI NIRSCAN Nano EVM	InnoSpectra NIR
Constituent					
max	Spectral range (nm)	1100–2500	900–1700	900–1700	900–1700
mean	No. of principal components	12	12	12	12
min	No. of outliers removed	22	26	17	15
Glucan	R^2-cal	0.95	0.93	0.91	0.89
47.8	RMSEC (%)	1.4	1.7	1.9	2.2
33.2	R^2_cv	0.94	0.92	0.89	0.87
21.4	RMSECV (%)	1.6	1.8	2.1	2.2
Xylan	R^2-cal	0.91	0.89	0.79	0.82
28.7	RMSEC (%)	1.0	1.1	1.5	1.4
17.8	R^2_cv	0.90	0.87	0.77	0.80
9.5	RMSECV (%)	1.1	1.2	1.6	1.5
Lignin	R^2-cal	0.94	0.94	0.87	0.84
29.0	RMSEC (%)	0.9	1	1.3	1.5
15.2	R^2_cv	0.93	0.92	0.86	0.81
6.7	RMSECV (%)	1.1	1.1	1.4	1.6
Extractives	R^2-cal	0.95	0.94	0.92	0.92
41.5	RMSEC (%)	2.3	2.4	2.8	2.8
17.7	R^2_cv	0.94	0.93	0.91	0.90
4.2	RMSECV (%)	2.5	2.6	3	3.1
Ash	R^2-cal	0.89	0.89	0.87	0.88
16.4	RMSEC (%)	1.2	1.2	1.3	1.3
6.7	R^2_cv	0.87	0.87	0.85	0.86
0.9	RMSECV (%)	1.3	1.3	1.4	1.4

[1] Foss XDS spectra truncated to 1100–2500 nm to match previous results

[2] Foss XDS spectra truncated to 900–1700 nm to match spectral range of TI and InnoSpectra units

from the two prototypes and from each other, with the Foss model using the larger wavelength range (1100–2500 nm) having less scatter than the Foss model using the smaller wavelength range (900–1700 nm).

In Table 2, the key statistics of the PLS-2 calibration models developed from the spectra are presented, as well as the range (max, mean, min) of the primary calibration data. The calibration models developed using Foss XDS NIR spectra using either of the two different spectral ranges had statistically significantly larger R^2_cv and smaller RMSECV values for all constituents compared with the models from either prototype unit ($p = 0.05$). That is, the models developed using the Foss XDS NIR spectra were superior to the models developing using either spectra of the prototype. The two models developed using the Foss spectra with different ranges are surprisingly similar to each other. The two models have statistically significantly different RMSECV values only for the structural carbohydrates glucan and xylan and do not have statistically significantly different R^2_cv values for any of the constituents. That is, using the 1100–2500-nm range improves the prediction performance only for the structural carbohydrates, not the predictions of extractives, lignin, or ash compared with the smaller 900–1700-nm spectral range. The differences in the structural carbohydrate predictions between the two models built using Foss spectra suggest that the overtones from structural carbohydrate bond vibration above 1700 nm (C–H stretch at ~ 1780 nm, O–H, C–O combination bands at ~ 2270 nm, and C–H stretch and deformation at 2280 nm) [19, 20] provide useful modeling information.

Fig. 3 (top) Root mean square error of leave-one-out cross-validation (RMSECV) of the PLS-2 models for all five constituents (glucan, xylan, lignin, ash, extractives). For all constituents, the models using Foss XDS spectra provide the smallest RMSECV values. (bottom) Correlation coefficient of leave-one-out cross-validation (R^2_cv) of the PLS2 models for all five constituents (glucan, xylan, lignin, ash, extractives). For all constituents, the models using Foss XDS spectra provide the largest R^2_cv values
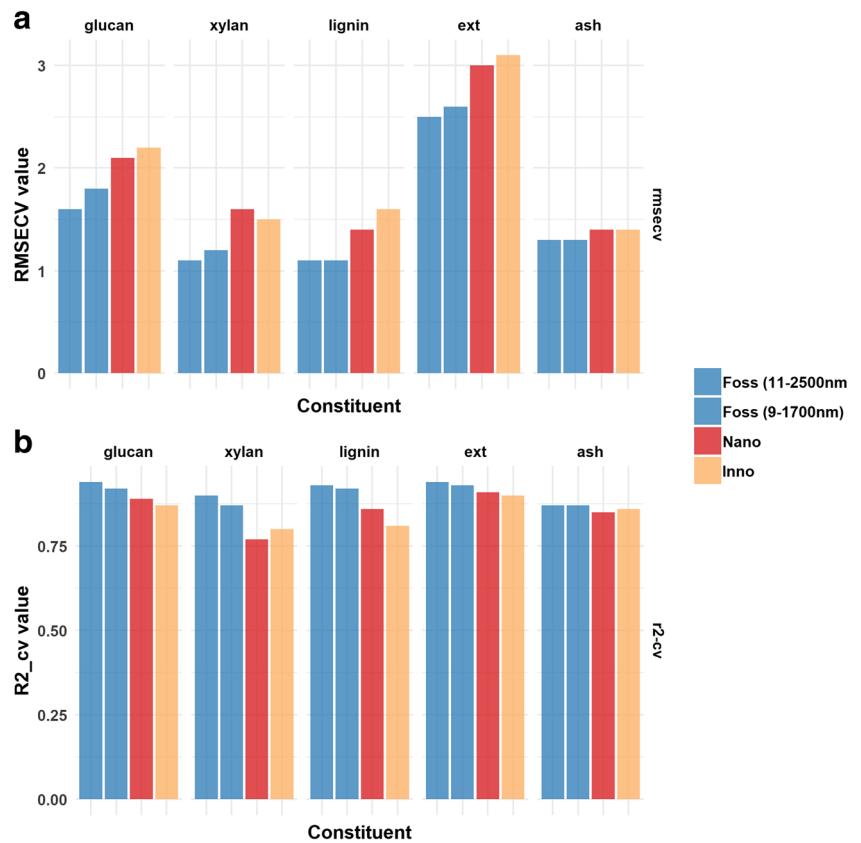
Fig. 4 Regression coefficients for glucan prediction using derivatized spectra from (top to bottom) the Foss XDS using the reduced wavelength range, the TI NIRSCAN Nano, and the InnoSpectra NIR-M-R2. The coefficient values for the three models are quite similar, suggesting the three models are using the same spectral information for prediction
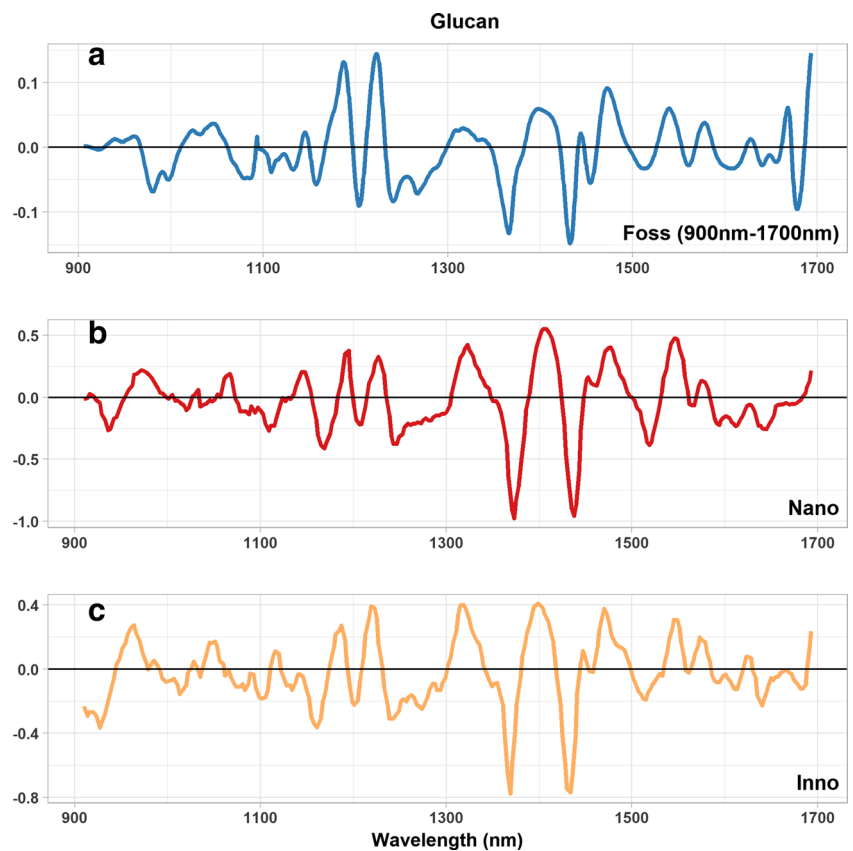
There has been previous work on the rapid spectroscopy prediction of the chemical composition determined using the same primary analytical chemistry methods as used in this work for corn stover using a FOSS XDS spectrometer [17] and for mixed herbaceous feedstocks using a Fourier-transform NIR (FT-NIR) spectrometer [11]. For the work with corn stover, the prediction uncertainties (as measured by RMSECV values) for the constituents glucan, xylan, and lignin ranged from 1.2 to 1.7, while for the work with mixed herbaceous feedstocks, RMSECV values ranged between 1.2 and 1.9 for the constituents glucan, xylan, ash, and lignin. Thus, the results presented here for all three spectrometers are generally consistent with previous work using the same primary analytical methods.

In Fig. 3, the cross-validated calibration statistics RMSECV and R^2_cv for all constituents and all models are presented. There are clear differences among the different constituents, with the xylan and lignin models having the lowest RMSECV values and the extractives prediction having the largest RMSECV values. However, the important differences are among the models themselves. The models using the Foss XDS spectra are consistently superior to the models using the spectra from either prototype (e.g., lower RMSECV values and higher R^2_cv values). As mentioned above, the predictions of structural carbohydrates from the two models using Foss spectra are different, and this is visible in Fig. 4 as well.

In Fig. 4, the regression coefficients for glucan prediction using derivatized spectra are presented from (top to bottom) the Foss XDS using the reduced wavelength range, the TI NIRSCAN Nano EVM, and the InnoSpectra NIR-M-R2. The coefficient values for the three models are quite similar, suggesting the three models are using the same spectral information for prediction. The magnitudes of the coefficients are different due to the differences in magnitude of the derivatized spectra. Nonetheless, the key features of all three plots are the same.

These results suggest that the use of low-cost spectrometers having limited spectral ranges compared with traditional laboratory spectrometers holds great promise—while the models using spectra from these prototypes are inferior by all measures compared with the models from the laboratory spectrometer, the differences are small enough to make the models quite useful. For example, the uncertainty in the lignin prediction using either model built using Foss spectra (as measured by the RMSECV value) is approximately 1.1%, while the uncertainty in the prediction using spectra from the prototype spectrometers is 1.4% and 1.5%. A prediction with an uncertainty of 1–1.5% would still be useful compared with the primary analytical method, which has a lower uncertainty [13] but can take up to a week or more to complete [12].

Opportunities and Challenges

The results presented here are quite encouraging. As mentioned above, there will be many opportunities for ubiquitous sensing across the chemical process industries in general and, as this work indicates, across the biomass-to-biofuel conversion chain in particular. However, there are significant challenges that remain to ubiquitous rapid analysis using low-cost spectrometers, including a software "ecosystem" for data collection and model application, calibration transfer among different spectrometers, and robust methods of sample presentation.

A robust software "ecosystem" permits users to collect NIR spectra, apply an existing prediction model, and immediately receive the prediction result, and is equally important as the spectrometer hardware itself. As mentioned above, typical applications of rapid analysis using NIR spectroscopy use the "calibrate-collect-predict" process using partial least squares algorithms, but there will be opportunities to use more advanced machine learning (ML) algorithms to the data generated, both the spectra from and individual unit and the aggregate spectra from units in a value chain. This will require robust data curation and will provide opportunities for both supervised and unsupervised learning approaches.

It is well-known that it is very difficult for two nominally "identical" spectrometers to generate truly interchangeable spectra. That is, small differences in manufacturing processes result in instruments that are slightly different and produce slightly different spectra even when measuring the same material. This makes using a prediction model developed for one instrument less accurate when used on another instrument, even if the second instrument is the identical make and model. This is the so-called calibration transfer problem [21, 22]. This issue will need to be thoroughly understood and addressed for the next-generation spectrometers as well, whether through existing algorithms or through novel ML approaches.

Finally, sample presentation is a critical issue in NIR spectroscopy, and different applications will require different (but reproducible) sample presentation strategies. For example, the reflectance measurements described in this work are appropriate for measuring solid materials, but the ability to perform continuous in situ process monitoring of liquid streams will require some type of immersion probe [23, 24].

Compliance with Ethical Standards

References

1. McClure W (2003) 204 years of near infrared technology: 1800-2003. J Near Infrared Spectrosc 518:487–518
2. Geladi P, Martens H (1996) A calibration tutorial for spectral data. Part 2. Partial least squares regression using Matlab and some neural network results. J Near Infrared Spectrosc 242:243–255
3. Pasquini C (2018) Near infrared spectroscopy: a mature analytical technique with new perspectives–a review. Anal Chim Acta. 1026: 8–36. https://doi.org/10.1016/j.aca.2018.04.004
4. Yan H, Siesler HW (2018) Hand-held near-infrared spectrometers: state-of-the-art instrumentation and practical applications. NIR news 29:8–12. https://doi.org/10.1177/0960336018796391
5. Tao F, Yao H, Zhu F, Hruska Z, Liu Y, Rajasekaran K, Bhatnagar D (2019) A rapid and nondestructive method for simultaneous determination of aflatoxigenic fungus and aflatoxin contamination on corn kernels. J Agric Food Chem 67:5230–5239. https://doi.org/10.1021/acs.jafc.9b01044
6. Gao B, Holroyd SE, Moore JC, Laurvick K, Gendel SM, Xie Z (2019) Opportunities and challenges using non-targeted methods for food fraud detection. J Agric Food Chem 67:8425–8430. https://doi.org/10.1021/acs.jafc.9b03085
7. Curzon AY, Chandrasekhar K, Nashef YK, Abbo S, Bonfil DJ, Reifen R, Bar-el S, Avneri A, Ben-David R (2019) Distinguishing between bread wheat and spelt grains using molecular markers and spectroscopy. J Agric Food Chem 67:3837–3841. https://doi.org/10.1021/acs.jafc.9b00131
8. Chiang L, Lu B, Castillo I (2017) Big data analytics in chemical engineering. Annu Rev Chem Biomol Eng 8:63–85. https://doi.org/10.1146/annurev-chembioeng-060816-101555
9. Stanton J (2018) Data science. National Academies Press, Washington, D.C.
10. Wolfrum EJ, Ness RM, Nagle NJ, Peterson DJ, Scarlata CJ (2013) A laboratory-scale pretreatment and hydrolysis assay for determination of reactivity in cellulosic biomass feedstocks. Biotechnol Biofuels 6:162. https://doi.org/10.1186/1754-6834-6-162
11. Payne CE, Wolfrum EJ (2015) Rapid analysis of composition and reactivity in cellulosic biomass feedstocks with near-infrared spectroscopy. Biotechnol Biofuels 8:1–14. https://doi.org/10.1186/s13068-015-0222-2
12. Sluiter JB, Ruiz RO, Scarlata CJ, Sluiter AD, Templeton DW (2010) Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. J Agric Food Chem 58: 9043–9053. https://doi.org/10.1021/jf1008023
13. Templeton DW, Scarlata CJ, Sluiter JB, Wolfrum EJ (2010) Compositional analysis of lignocellulosic feedstocks. 2. Method uncertainties. J Agric Food Chem 58:9054–9062. https://doi.org/10.1021/jf100807b
14. Gelabert P, Pruett E, Perrella G, et al (2016) DLP NIRscan Nano: an ultra-mobile DLP-based near-infrared bluetooth spectrometer. In: SPIE 9761, Emerging digital micromirror device based systems and applications VIII, 97610B. San Francisco, CA
15. Posom J, Sirisomboon P (2017) Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. J Near Infrared Spectrosc 25: 301–310. https://doi.org/10.1177/0967033517728733
16. R Core Team (2019) R: A language and environment for statistical computing. https://www.r-project.org
17. Wolfrum EJ, Sluiter AD (2009) Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover. Cellulose 16:567–576. https://doi.org/10.1007/s10570-009-9320-2
18. Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal Chem 60:1193–1202. https://doi.org/10.1021/ac00162a020
19. Xu F, Yu J, Tesso T, Dowell F, Wang D (2013) Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: a mini-review. Appl Energy 104:801–809. https://doi.org/10.1016/j.apenergy.2012.12.019
20. Shenk JS, Jerome J, Workman J, Westerhaus MO (2001) Application of NIR spectroscopy to agricultural products. In: Burns DA, Ciurczak EW (eds) Handbook of near-infrared analysis, 2nd edn. Marcel dekker, New York
21. Mark H, Workman J (2013) Calibration transfer. Spectroscopy 28: 24–37
22. Dreassi E, Ceramelli G, Perruccio PL, Corti P (1998) Transfer of calibration in near-infrared reflectance spectrometry. Analyst 123: 1259–1265
23. Hongqiang L, Hongzhang C (2008) Near-infrared spectroscopy with a fiber-optic probe for state variables determination in solid-state fermentation. Process Biochem 43:511–516. https://doi.org/10.1016/j.procbio.2008.01.012
24. Tamburini E, Vaccari G, Tosi S, Trilli A (2003) Near-infrared spectroscopy: a tool for monitoring submerged fermentation processes using an immersion optical-fiber probe. Appl Spectrosc 57:64A-85A and 117-243 (February 2003) , p. 13. https://doi.org/10.1366/000370203321535024