EXPLORING NIR TECHNIQUE IN RAPID PREDICTION OF COTTON TRASH COMPONENTS

Yongliang Liu
Devron Thibodeaux
Gary Gamble
USDA, ARS, Cotton Quality Research Station
Clemson, SC

Abstract

Near infrared (NIR) spectroscopy, a useful technique due to the speed, ease of use, and adaptability to on-line or off-line implementation, has been applied to perform the qualitative classification and quantitative prediction on a number of cotton quality indices, including cotton trash from HVI, SA, and AFIS measurement. It is well-known that these current-in-use trash measuring devices only produce the trash values in some aspects, instead of the content for individual trash component. This difficulty comes from the complexity of co-existence of various trash types, for example, leaves (leaf and bract), seed coats, hulls, and stems. Regarding to this, mixtures of known trash components (e.g., leaves, seed coats, hulls, stems, and sand/soil) with cut lint fibers were prepared physically and then their NIR spectra were correlated with the respective trash contents. The results suggested the feasibility of NIR technique in the precise and quantitative determination of total trash, leaf trash and non-leaf trash components.

Introduction

Presence of non-lint materials (trashes) in commercial cotton bales at various amounts degrades the market values and further influences the end-use qualities. In order to ensure a fair-trading, the USDA Agricultural Marketing Service (AMS) has introduced the high volume instrument (HVI) measurement as a universal standard index (Knowlton, 2002). Among the indices, trash content is generated by one of three HVI modules and represents the trash portion only detectable on the surface of a sample. In addition to HVI's geometric method, gravimetric-based Shirley analyzer (SA) and advanced fiber information system (AFIS) have also been utilized to determine the trash contents. As expected, they record the trash contents in different representations from geometric based HVI to gravimetric based SA or AFIS. For example, HVI assesses the trash content in terms of particle count and percentage area on a sample's surface, while SA and AFIS yield the respective weight of trash in terms of visible trash content (%) and visible foreign matter content (%) within the bulky samples.

With the increasing acceptance of HVI readings in the domestic and international trading, there is a continued interest in the relationship between HVI trash and SA trash from domestic and foreign cotton customers as well as government regulators. Due to the complexity of not only trash type, size, and its weight distribution but also the nature of HVI and SA tests, it is understandable that there has few study available trying to correlate two types of trash readings and, apparently, this is a challenge. In a separated investigation (Liu et al., 2012), we have attempted to unravel this interest by applying new strategy of sub-grouping the samples on the basis of HVI and SA trash readings.

Notably, HVI, SA, and AFIS only yield the amount of trash in general terms, instead of the content for individual trash component. In large part, this difficulty originates from the complexity of co-existence of various trashes, such as leaves (leaf and bract), seed coats, hulls, and stems, in completely unpredicted manner within the lint cottons.

Near infrared (NIR) spectroscopy, a technique due to non-destructive, the speed, ease of use, and adaptability to on-line or off-line implementation, has been applied for the quantitative prediction of cotton trashes from three trash measurements (Thomasson and Shearer, 1995; Liu et al., 2010a & 2010b). Considering the unique nature of trash in cotton, a few NIR studies have been conducted and the obtained results have not been encouraging. For instance, Thomasson and Shearer (1995) reported the optimal NIR models for 8 cotton HVI characteristics and observed the lowest R^2 value (0.60) for trash component. Even though the UV/visible/NIR model on SA visible trash in cotton waste were much improved (R^2=0.90), it still showed the difficulty in precise and quantitative determination of visible trash portion for quality control purpose (Liu et al., 2010a). Probably, major factor leading to low trash model performance is due to highly diversification of trash types and their heterogeneous distribution. Thus, a 90% confidential interval was applied to remove outlier samples that might be related with different sampling species between reference and spectral measurement (Liu et al., 2010a & 2010b).

To address this viewpoint, mixtures of known trash components (e.g., leaves, seed coats, hulls, stems, and sand/soil) with cut lint fibers were prepared physically and then their NIR spectra were correlated with the individual trash contents. Main objectives of this study were to compare NIR models on individual trash constitute in the gravimetric version of weight mass (%), and also to examine the effect of trash uniformity on NIR model performance.

Materials and Methods

Clean Fibers and Cotton Trashes

Clean fibers were obtained from routine SA (Shirley Developments, Ltd., Stockport, UK) process of different lint cottons at the USDA ARS's Cotton Quality Research Station (Clemson, SC). While each class of five cotton trashes, namely leaves (including bracts), seed coats, hulls, stems, and sand/soil, was collected manually from 3 varieties of seed cottons harvested in 2008. All samples were well conditioned at a constant relative humidity of 65% and temperature of 68 ± 2 ºF, prior to cutting, weighing, mixing, and acquiring NIR spectra.

Ground Samples and Mixtures

Both cotton fiber and five trashes were grounded in a Wiley mill to pass through a 20-mesh screen. Then, 100 mixtures, each weighted 5.0 g in total and consisted of cut fibers and 5 trashes at varying amounts were prepared manually and simply. It led to a range of 0-15% with a mean of 4.92% for total trash, a range of 0-5% with a mean of 1.44% for leaf trash, a range of 0-5% with a mean of 1.28% for stem trash, a range of 0-5% with a mean of 1.16% for hull trash, a range of 0-3% with a mean of 0.55% for seed coat trash, and 0-3% with a mean of 0.50% for soil/sand trash. Also, it produced a range of 0-12% with a mean of 3.48% for non-leaf trash (i.e., a total of seed coat trash, hull trash, stem trash, and sand/soil trash). This experimental setup was based on the fact that SA visible trash was ~2.8% in average (Liu et al., 2012).

Reflectance Spectral Measurement

The mixtures were loaded into a sample cell (0.38 inch in depth and 2 inch in diameter) and scanned on a FOSS XDS rapid content analyzer (FOSS NIRSystems Inc., Laurel, MD). A background was recorded with a built-in internal reference before acquiring the spectra of samples. The log (1/Reflectance) readings were recorded over the visible/NIR range of 400 - 2500 nm at 0.5 nm interval and 32 scans. Three spectra were obtained for each sample by repacking; where upon the mean spectrum was utilized in model development.

Model Development

All visible/NIR spectra were imported into PLSplus/IQ package in Grams/AI (Version 7.01, Galactic Industrious Corp., Salem, NH, current part of Thermo Fisher Scientific) for partial least-squares (PLS) regression model development. On the order of the smallest to largest in total trash content, 67 spectra were selected for calibration equation development and the remaining 33 (every 3rd sample) spectra were used for model validation. To optimize the accuracy of prediction models, the spectra were subjected to different combinations of both the spectral ranges (e.g., full and narrow regions) and the spectral pretreatments (e.g., mean centering (MC), multiplicative scatter correction (MSC), and the first and second derivatives). Full (one-sample-out rotation) cross-validation method was used, and the number of optimal factors chosen for the regression equation generally corresponded to the minimum of the predicted residual error sum of squares (PRESS). The saved regression equations were subsequently applied to the validation samples. Model accuracy and efficiency were assessed in the validation set on the basis of the coefficient of determination (r^2), root mean square error of validation (RMSEV), and residual predictive deviation (RPD) (Williams, 2007). Usually, an optimal model should have lower RMSEV and higher r^2 and RPD.

Results and Discussion

Cotton Trash Contents and Visible/NIR Spectra

Figure 1 shows the typical visible/NIR log (1/R) spectra of four samples with total trash content of 0, 4.0, 8.0 and 13.0 %, respectively. As clean cotton was from cleaning effect of SA operation, its total trash amount was assumed to be 0.0%. Therefore, distinctive spectral differences should occur with trash concentrating, because of obvious difference in color and composition between the trash, a mixture of main contributions from plant parts, and cotton fiber, a majority (>94%) of cellulose species. For instance, the spectra with higher trash content showed high log (1/R) intensity in the spectral region of visible/short-wavelength (SW) NIR region (< 1100 nm). Such distinctions could form the basis for qualitative and quantitative determination of cotton trash from visible/NIR technique.

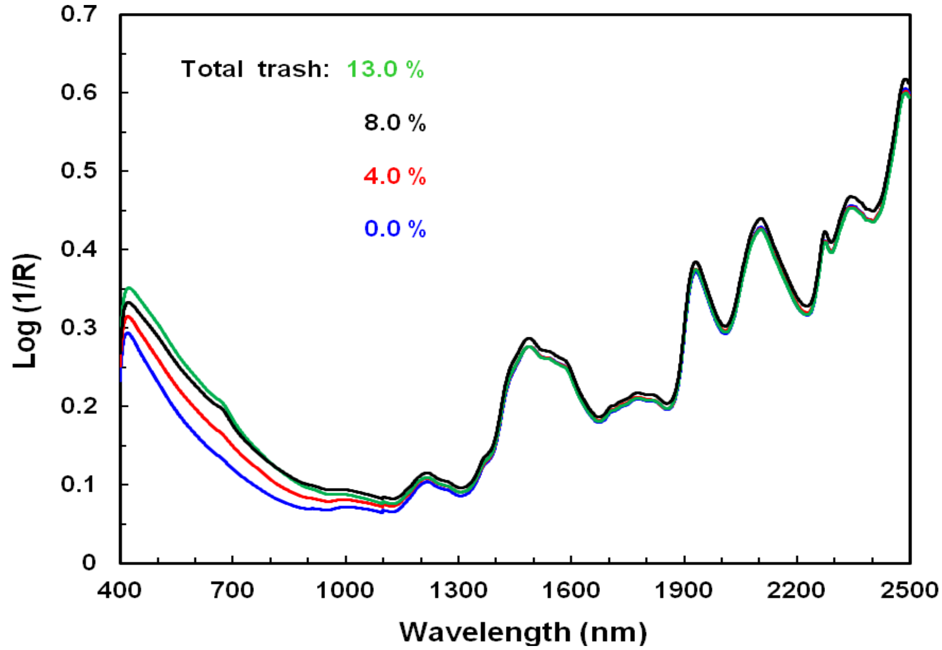

Figure 1. Typical visible/NIR log (1/R) spectra of mixtures.

Reference Values

Table 1 summarizes the range, mean, and standard deviation (SD) of reference values for seven trash representations in calibration and validation sets, including total trash, leaf trash, non-leaf trash, stem trash, hull trash, seed coat trash, and sand/soil trash. The variations of reference values covered most of the variability in commercial cotton bales. The range, mean, and SD values for each component in the validation set were comparable to those in the calibration set, indicating that the selection of samples for individual set was appropriate.

Table 1. Summary of range, mean, and SD for seven trash readings in calibration and validation sets.

Constituent	Calibration Set (n =67)			Validation Set (n = 33)		
	Range	Mean	SD	Range	Mean	SD
Total trash, %	0 - 15.0	5.08	2.83	0 - 12.0	4.59	2.58
leaf trash, %	0 - 5.0	1.40	1.05	0 - 4.0	1.52	1.01
non-leaf trash, %	0 - 12.0	3.68	2.47	0.60 - 9.6	3.07	2.10
Stem trash, %	0 - 5.0	1.34	1.13	0 - 4.0	1.15	0.97
Hull trash, %	0 - 5.0	1.24	1.33	0 - 3.0	0.99	0.77
Seed coat trash, %	0 - 3.0	0.61	0.77	0 - 2.5	0.42	0.59
Sand / soil trash, %	0 - 3.0	0.50	0.73	0 - 2.5	0.50	0.59

Prediction Model – Total Trash

PLS models for seven constituents were developed using the different combinations of full / narrow spectral regions and a variety of data pre-treatments. The statistics of optimal results in calibration and validation sets from various spectral regions are summarized in Table 2. In addition, the model performance from the 900-1700 nm NIR region was included. The best prediction models were obtained from combinations of such spectral pretreatments as MC and 1st derivative. The use of 2nd derivative, along with other data processing, yielded much poorer results for all variables (not shown).

Comparison of model performance for total trash constituent indicated that the model from the 900-1700 nm region was slighter better than those from other three regions: 405-2495 nm or 405-1095 nm or 1105-2495 nm, with the least RMSEC and RMSEV.

Table 2. Statistics in calibration and validation sets. [a]

Component	Optimal factor	Calibration Set R^2	RMSEC [b]	Validation Set r^2	RMSEV [b]	RPD [c]
Total trash						
405 - 2495 nm	6	0.93	0.76	0.89	0.88	2.9
405 - 1095 nm	6	0.93	0.75	0.92	0.75	3.4
1105 - 2495 nm	6	0.93	0.76	0.93	0.75	3.4
900 - 1700 nm	6	0.94	0.69	0.92	0.72	3.6
Leaf trash						
405 - 2495 nm	4	0.94	0.26	0.92	0.28	3.6
405 - 1095 nm	5	0.95	0.23	0.94	0.26	3.9
1105 - 2495 nm	6	0.87	0.38	0.84	0.41	2.5
900 - 1700 nm	8	0.89	0.35	0.89	0.35	2.9
Non-leaf trash						
405 - 2495 nm	7	0.93	0.67	0.92	0.66	3.2
405 - 1095 nm	6	0.91	0.73	0.90	0.68	3.1
1105 - 2495 nm	7	0.95	0.54	0.91	0.69	3.0
900 - 1700 nm	8	0.95	0.54	0.92	0.60	3.5
Stem trash						
900 - 1700 nm	7	0.78	0.53	0.65	0.59	1.6
Hull trash						
900 - 1700 nm	7	0.80	0.59	0.57	0.67	1.1
Seed coat trash						
900 - 1700 nm	10	0.79	0.36	0.24	0.57	1.0
Sand/soil trash						
900 - 1700 nm	8	0.91	0.22	0.69	0.34	1.7

[a] All spectral processing with mean centering (MC) and the first derivative (1st deri.).
[b] Root mean square error of calibration (RMSEC) and validation (RMSEV).
[c] RPD = SD / RMSEV.

RPD, ratio of SD of reference values against RMSEV, is often used as a dimensionless gauge of the ability of a spectroscopic model to predict a property (Williams, 2007). An RPD value of greater than 3.0 indicates the acceptability of the model for quantitative prediction, a value of greater than 2.5 and less than 3.0 suggests the suitability of the model for screening program, and a value of 1.0 or less means the lack of modeling power. Hence, the model for total trash from the 900-1700 nm region could be used for quantitative applications (RPD = 3.6). A comparative scatter plot of referenced and NIR predicted total trash in both calibration and validation sets is given in Figure 2. It suggests how well the NIR model predictions agree with the references.

Although there is no literature available that deals with the topic of total trash in commercial cottons, it is possible to link this concept with visible trash content (%) from traditional and gravimetric SA procedure. NIR prediction of visible trash has been reported before, in which the optimal RPD of 3.0 and 2.4 was observed (Liu et al., 2010a & 2010b). Many differences exist between this study and previous ones, for example, in sample type (subjectively mixed and cut trashes vs. cotton waste resulted from cleaning process of lint cotton (Liu et al., 2010a) and commercial lint cotton (Liu et al., 2010b)), trash content (0.0-15.0% vs. 0.0-65.2% (Liu et al., 2010a) and 1.2-7.4% (Liu et al., 2010b)), spectral range (400-2500 nm vs. 220-2500 nm(Liu et al., 2010a) and 400-2500 nm (Liu et al., 2010b)), and sampling cell dimension (0.38 inch in depth x 2 inch in diameter vs. 0.38 inch in depth x 2 inch in diameter (Liu et al., 2010a) and 1.5 inch in width x 6 inch in length (Liu et al., 2010b)).

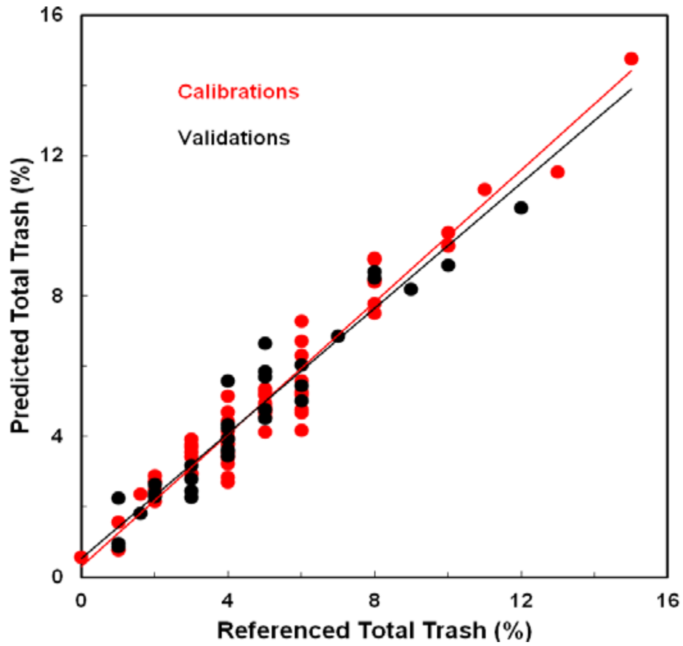

Figure 2. Plot of referenced vs. NIR model predicted total trash content in calibration (●) and validation (●) sets.

In earlier studies (Liu et al., 2010a & 2010b), a 90% confidential interval was applied to remove outlier samples that had large differences (or errors) between measured and NIR predicted values from calibration and validation sets, mostly because of such considerations as (i) highly diversification of trash types and their heterogeneous distribution, (ii) relatively small sampling size in NIR spectral collection (0.5 g x 4 replicates) compared to that for SA procedure (100 g x 2 replicates), and (iii) different sampling species between spectral and reference measurement. As a result, the recalibrated models were improved and an elevation of PRD to 3.7 suggested the potential of NIR model in the quantitative determination of visible trash in cotton waste (Liu et al., 2010a). While in another study (Liu et al., 2010b), the RPD from redeveloped visible trash model did not increase obviously, which might be due to the distribution of range and SD values in this small validation set.

Compared to the reported RPDs of 3.0 and 2.4, current total trash model was much improved (RPD = 3.6). It might address the concern of sample uniformity in accurate and reliable model development. Furthermore, this model was as effective as one after excluding the outliers in previous study (Liu et al., 2010a).

Prediction Models – Leaf Trash and Non-Leaf Trash

In the same procedure, PLS models were established for leaf trash and non-leaf trash component. Interestingly, leaf trash could be better predicted in the 405-1095 nm region than other three ranges (RPD = 3.9), and non-leaf trash might be modeled better in the 900-1700 nm region (RPD = 3.5). Likely, this reflected the distinctions in structure and compositions between leaf and non-leaf trashes, and also revealed the importance of different spectral preprocessing to optimize the modeling power.

With RPD ≥ 3.0, the PLS models for leaf trash and non-leaf trash component implied the feasibility of NIR technique in the precise and quantitative measurement of trash under the category of leaf and non-leaf class.

Prediction Models – Individual Non-Leaf Trash

It is of great interest to examine whether non-leaf components, such as stem trash, hull trash, seed coat trash, and sand/soil trash, could be modeled as effective as leaf trash. Unfortunately, the obtained results in Table 2 suggested some hindrance in the prediction of these individual trashes, because the RPDs were much less than 3.0. One of many factors might be due to particle size of these trashes and their uniform distribution. More study is needed to explore the potential of NIR model in the determination of these non-leaf trash components.

Summary

Collecting NIR spectra on cut fiber and trash mixtures, the resultant models showed the potential of NIR technique in the precise and quantitative determination of total trash, leaf trash and non-leaf trash components. However, it indicated the degree of difficulty in the prediction of such non-leaf trashes as stem, hull, seed coat and sand/soil. This limitation arises from the particle size of these trashes and their uniform distribution, and further study is necessary to understand the relationship between spectral response and non-leaf trash components.

Acknowledgements

We are grateful to Mattie Morris (ARS, Clemson) for technical assistance in collecting and cutting the cotton and trash, preparing the mixtures and colleting their spectra.

References

Knowlton, J. 2002. Development in cotton standards. Proceedings of the Beltwide Cotton Conferences, National Cotton Council of America, Memphis, TN.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010a. Assessment of cotton fiber and trash contents in lint cotton waste by UV/visible/near-infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 18: 239-246.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010b. Evaluation of 3 cotton trash measurement methods by visible / near-infrared reflectance spectroscopy (ASABE Paper No. 1008708), American Society of Agricultural and Biological Engineers, St. Joseph, MI.

Liu, Y., D. Thibodeaux, and G. Gamble. 2012. Relationship between three cotton trash measurements: HVI, SA, and AFIS. Proceedings of the Beltwide Cotton Conferences, National Cotton Council of America, Memphis, TN.

Thomasson, J. A., and S. A. Shearer. 1995. Correlation of NIR data with cotton quality characteristics. Trans. ASAE. 38: 1005-1010.

Williams, P. 2007. Grains and seeds. P. 165-217. *In* Y. Ozaki, W.F. McClure and A.A. Christy (eds.) Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, Inc., Hoboken, NJ.