# TEXTILE TECHNOLOGY

## Feasibility of Visible and Near Infrared Spectroscopy in the Determination of Cotton Trash Components

Yongliang Liu* and Devron Thibodeaux

### ABSTRACT

The co-occurrence of different types of trash in commercial cotton bales compromises the value of cotton, requires more cleaning, and influences the quality of yarn and fabric. To meet the challenge of determining the trash content, two testing methods (i.e., High Volume Instrument [HVI™] and Shirley analyzer [SA]) have been utilized by trade and regulatory offices and laboratories in the cotton industry. However, these methods only report the trash amounts in total, instead of the content for individual trash components. Likely, the complexity of the co-existence of various trash types, including leaves (leaf and bract), seed coats, hulls, and stems, contributes to this limitation. To address this problem, a set of mixtures with known amounts of both clean lint fibers and individual trash components (leaves, seed coats, hulls, stems, and sand/soil) was prepared and the visible/NIR spectral response was related to corresponding trash contents. Comparison of model performances revealed the feasibility of visible/NIR technique in the precise and quantitative determination of total trash, leaf trash, and non-leaf trash components.

The co-occurrence of different types of trash in commercial cotton bales compromises the trading value of cotton, requires more cleaning, and influences the quality of yarn and fabric. To determine the amount of cotton trash, the Shirley Analyzer (SA) (ASTM, 2012a), High Volume Instrument (HVI) (ASTM, 2012b), Advanced Fiber Information System (AFIS) (ASTM, 2012c), Micro Dust and Trash Analyzer (MDTA) (Boykin et al., 2009), and FibroLab (Matusiak

Y. Liu*, USDA-ARS, Southern Regional Research Center, 1100 Robert E. Lee Blvd., New Orleans, LA 70124 and D. Thibodeaux, 103 Long View Ct., Pickens, SC.
*Corresponding author: yongliang.liu@ars.usda.gov

and Walawska, 2010) have been developed in the U.S. and other countries. Among them, the results from HVI and SA measurements have been utilized routinely at trade and regulatory offices and laboratories in the cotton industry. Briefly, the automation-based HVI procedure identifies the number of non-lint particles ($HVI_{Count}$) and measures the surface area covered by non-lint (or trash) particles ($HVI_{Area}$) on a sample's surface; whereas SA yields the weight of trash in terms of visible and invisible trash content (%) within a bulk sample, by separating the trash components through mechanical means and then weighting the respective visible trash and cleaned fiber portions. These methods yield the trash amount only in total or general terms, instead of the content for individual or targeted trash components, such as leaves (leaf and bract), seed coats, hulls, and stems. The co-existence of various trash components makes it a challenge to determine the amount of each trash type in commercial cotton bales by these instrumental systems.

Near infrared (NIR) spectroscopy, a rapid and low-cost technique that can be used in places such as ginning sites, has been used for the quantitative prediction of trash contents by HVI and SA indices (Liu et al., 2010a, 2010b; Thomasson and Shearer, 1995). Thomasson and Shearer (1995) reported the development of optimal NIR models for eight HVI cotton quality characteristics and observed the lowest $R^2$ value (0.60) for HVI trash content. In recent studies, NIR trash models built from $HVI_{Count}$, $HVI_{Area}$, and SA visible trash indices were observed to be promising in their quantitative determination (Liu et al., 2010a, 2010b). Considering the different sampling species between trash reference and spectral measurement, a 90% confidence interval rule was applied to remove outlier samples that exhibited larger differences between NIR-predicted and -measured references in these approaches (Liu et al., 2010a, 2010b).

To look into the potential of NIR technique to predict specific and unique cotton trash categories,

a set of mixtures was prepared consisting of known trash components (leaves, seed coats, hulls, stems, and sand/soil) and clean lint. By correlating the spectral response with individual trash content, this study examined and further compared the feasibility of visible/NIR spectroscopy to determine targeted trash components in the gravimetric version of weight mass (%).

## MATERIALS AND METHODS

**Clean Fibers and Cotton Trash**. Clean fibers were obtained from routine SA (Shirley Developments, Ltd., Stockport, UK) processing of 10 lint cotton bales at the USDA ARS's Cotton Quality Research Station (Clemson, SC). Each of the five cotton trash components, namely leaves (including bracts), seed coats, hulls, stems, and sand/soil, was collected either from the trash remains (or wastes) of the SA processing of commercial cottons or from the seed cotton manually. Prior to cutting, weighing, mixing, and acquiring visible/NIR spectra, the samples were conditioned at 65% relative humidity and 21 ± 2°C temperature.

**Ground Samples and Mixtures.** Both clean cotton fibers and five trash components were ground once in a Wiley mill and passed through a 20-mesh screen. This resulted in target weights of approximately 10, 9, 9, 4, and 4 g for leaves, stems, hulls, seed coats, and sand/soil trash categories, respectively. The sieve hole size for the 20-mesh screen was 0.841 mm, thus the size of the resulting particles should be smaller than that. The distribution of particle size for each trash type is unclear, but in general, leaf trash contained more small or fine particles than other cotton plant trash. Next, 100 mixtures were prepared to be as homogeneous as possible by rescreening each mixture. Each mixture weighed 5.0 g in total and consisted of cut fibers and five trash components at varying amounts in weight (g). The range, mean, and standard deviation (SD) for each of seven trash indices within these 100 samples are summarized in Table 1. Non-leaf trash component is a total of seed coat trash, hull trash, stem trash, and sand/soil trash in one mixture. Previous research has shown that SA visible trash was approximately 2.8% in average among 406 commercial cottons (Liu et al., 2012), and this information was used for this experimental setup.

**Table 1. Range, mean, and SD values for each of seven trash indices among 100 samples.**

| Components | Range | Mean | SD |
|---|---|---|---|
| Total trash (%) | 0 - 15.0 | 4.92 | 2.77 |
| Leaf trash (%) | 0 - 5.0 | 1.44 | 1.04 |
| Non-leaf trash (%) | 0 - 12.0 | 3.48 | 2.38 |
| Stem trash (%) | 0 - 5.0 | 1.28 | 1.09 |
| Hull trash (%) | 0 - 5.0 | 1.16 | 1.18 |
| Seed coat trash (%) | 0 - 3.0 | 0.55 | 0.72 |
| Sand/soil trash (%) | 0 - 3.0 | 0.50 | 0.69 |

**Visible/NIR Reflectance Spectral Acquisition.** Approximately 1.3 g of each mixture was loaded into a sample cell (1 cm in depth and 3.8 cm in diameter) and scanned on a FOSS XDS rapid content analyzer (FOSS NIRSystems Inc., Laurel, MD), with the use of a built-in internal reference background. The log (1/Reflectance) (log [1/R]) readings were obtained in the 400 to 2500 nm visible/NIR range at 0.5 nm intervals. Three spectra, 32 scans per spectrum, were obtained for each sample by repacking and their mean spectrum was utilized in the following model development.

**PLS Model Development.** Both spectral and trash reference data were loaded into PLSplus/IQ package in grams/AI (Version 7.01, Galactic Industrious Corp.[Thermo Fisher Scientific], Salem, NH) for partial least-squares (PLS) regression model development. For individual trash components, 33 (every 3rd sample in order of smallest to largest in trash content) spectra were chosen to validate the model built from the remaining 67 spectra. To compare the prediction models, different combinations of both spectral ranges (e.g., full and narrow regions) and spectral pretreatments (e.g., mean centering [MC], multiplicative scatter correction [MSC], and the first and second derivatives) were used. During the protocol, leaving-one-sample-out cross-validation method was used, and the suggested optimal factor that generally corresponded to the minimum of the predicted residual error sum of squares (PRESS) was recorded. The models were then applied to the validation and calibration samples, and their performances were evaluated in the validation set by the parameters of the coefficient of determination ($r^2$), root mean square error of validation (RMSEV), and residual predictive deviation (RPD) (Williams, 2007). In general, an acceptable model should have a lower RMSEV and higher $r^2$ and RPD.

## RESULTS AND DISCUSSION

**Cotton Trash Contents and Visible/NIR Spectral Response.** Representative visible/NIR log (1/R) spectra of four samples with total trash contents of 0.0, 4.0, 8.0, and 13.0% are shown in Fig. 1. As a comparison, the spectrum of a pure (or 100% total) trash sample was also included. There are at least five intense and broad bands with one (< 700 nm) in the visible region (400-750 nm) and four (1490, 1935, 2105, and 2340 nm) in the NIR region (750-2500 nm). Although SA-cleaned fiber was assumed to contain 0.0% total trash, it showed the broad absorptions in the visible region of 400 to 750 nm. This was expected because this visible range reflects the color information and represents the contributions from the pigmentation compounds present in natural fibers, for example, flavonoids and/or degraded products between a reducing sugar and an amino acid (Hua et al., 2007; Gamble, 2008). Chlorophyll and its degradation derivatives in cotton plant parts significantly contribute to the bands in this region, as indicated by the intense band below 900 nm in the spectrum of pure trash.

The NIR bands (1100-2500 nm) are mainly due to the first (1$^{st}$) and second (2$^{nd}$) overtones and combinations of OH and CH stretching vibrations of both cotton cellulose and trash cellulose-related compounds. Expected NIR spectral differences exist between pure trash and cotton fiber because of the difference in chemical composition. One difference is the reduced intensity at 2105 nm in pure trash, which has been assigned to the combination band of the first overtone of the O–H deformation vibration and the first overtone of the C–O stretching mode in cotton cellulose (Liu et al., 1998).

Although Fig. 1 suggests that mixtures with low total trash have visible/NIR bands in common with high total trash, there are intensity variations induced by the total trash concentration. For example, the spectra of samples with high total trash indicate log (1/R) intensity increase in the visible/short-wavelength NIR region (< 1100 nm) and also relatively weak intensity reduction in the 1100 to 2500 nm NIR range. Thus, the sensitivity of visible spectroscopy to reflect the trash content is dependent on spectral intensity change among these samples.
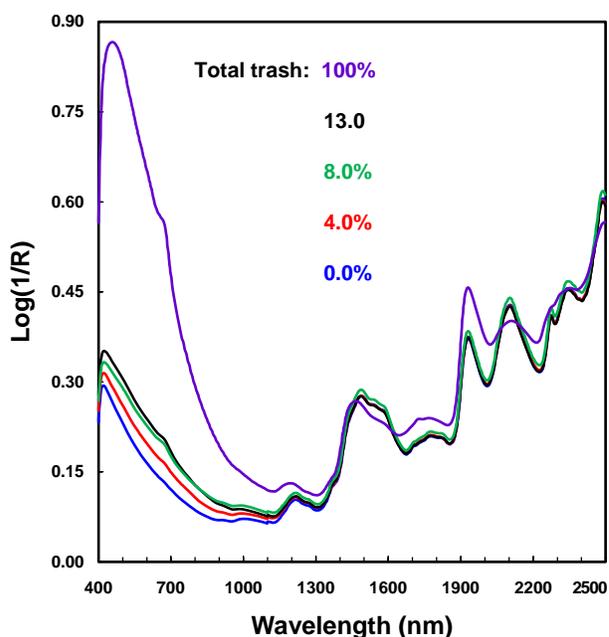


**Figure 1. Representative visible/NIR log (1/R) spectra of cotton-trash mixtures at total trash amount of 0.0, 4.0, 8.0, 13.0, and 100%.**

**Reference Values.** Range, mean, and SD values of reference values are tabulated in Table 2 for seven trash representations in calibration and validation sets, including total trash, leaf trash, non-leaf trash, as well as the components of non-leaf trash such as stem trash, hull trash, seed coat trash, and sand/soil trash. The variations of reference values included most of the variability in commercial cotton bales. For each component, the range, mean, and SD values in the validation set were comparable to those in the calibration set, meaning that the selection of samples for individual constitutes was appropriate.

**Table 2. Range, mean, and SD values for each of seven trash readings in calibration and validation sets.**

| Components | 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 |

**PLS Models.** Combinations of full/narrow spectral regions and different spectral preprocessing were utilized to optimize the PLS models for all constituents. Table 3 compares the model statistics in calibration and validation sets from various spectral regions. These prediction models were developed from such spectral pretreatments as MC and 1st derivative. The use of 2nd derivative generated poorer results for all variables (not shown). This observation is in good agreement with NIR prediction of other cotton fiber properties reported by Montalvo et al. (1994). Further investigation is necessary to determine why the additional smoothing effect and loss of information from another derivation generated poor results in this study.

RPD has been used to assess the capability of a spectroscopic model to predict a targeted property (Williams, 2007). RPD is defined as the ratio of SD of reference values against RMSEV. In general, an RPD value of 1.0 or less means a lack of modeling power, a value of between 2.5 and 3.0 suggests the suitability of the model for a screening program, and a value of greater than 3.0 suggests the acceptability of the model for a quantitative application.

**Total Trash.** For total trash constituent, the models from three narrower ranges (405-1095 nm, 1105-2495 nm, and 900-1700 nm) were slightly better than that from the 405 to 2495 nm full region, with a smaller RMSEV and the greater $r^2$ and RPD. Notably, the model from the 900 to 1700 nm region could be applied for the prediction of total trash in cottons quantitatively due to the greatest RPD (= 3.6). Figure 2 shows a comparative scatter plot of referenced and spectral predicted total trash in both calibration and validation sets, and indicates how well the visible/NIR model predictions agree with the references.

**Table 3. Comparison of $R^2$ and RMSEC in calibration set and $r^2$, RMSEV, and RPD in validation set.**

| Component | Optimal factor | Calibration Set | | Validation Set | | |
|---|---|---|---|---|---|---|
| | | $R^2$ | RMSEC | $r^2$ | RMSEV | RPD |
| **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 |
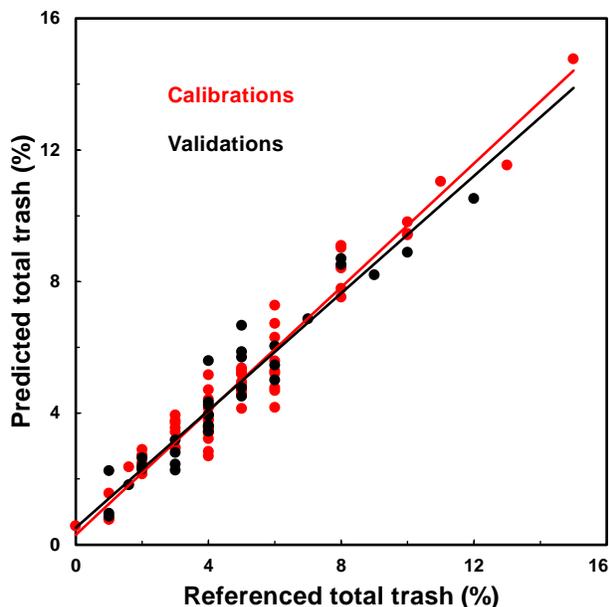
**Figure. 2. Plot of referenced vs. visible/NIR model predicted total trash content in calibration (•) and validation (•) sets.**

We (Liu et al., 2010a, 2010b) reported the NIR prediction of SA visible trash content (%) with observed RPD values of 3.0 and 2.4. There are a number of differences between this study and previous ones, and these are reflected by mainly four factors: (1) sample type (subjectively mixed and cut trash vs. cotton waste from SA cleaning process [Liu et al., 2010b] and commercial lint cottons [Liu et al., 2010a]), (2) trash content (0.0-15.0% vs. 0.0-65.2% [Liu et al., 2010b] and 1.2-7.4% [Liu et al., 2010a]), and (3) spectral range (400-2500 nm vs. 220-2500 nm [Liu et al., 2010b] and 400-2500 nm [Liu et al., 2010a]).

In these studies (Liu et al., 2010a, 2010b), we applied a 90% confidence interval rule to remove outlier samples that had large differences (or errors) between measured and NIR-predicted values from calibration and validation sets. The rationale behind this operation was due to the concerns of (1) highly diversified trash types and their heterogeneous distribution, (2) relatively small sampling size in NIR spectral collection (0.5 g x 4 replicates or 10 g x 3 replicates) compared to that for SA procedure (100 g x 2 replicates), and (3) different sampling species between spectral and reference measurement. As anticipated, the recalibrated models were improved (RPD = 3.7), indicating the potential of the NIR model for the quantitative determination of visible trash in cotton waste (Liu et al., 2010b).

The total trash model in this study was much improved (RPD = 3.6), compared to previously reported RPDs of 3.0 and 2.4. This result might highlight the concern of trash uniformity in accurate and reliable

model development, because the distribution of trash in cotton bales is heterogeneous.

**Leaf Trash and Non-leaf Trash.** PLS models were developed for leaf trash and non-leaf trash components by the same procedure. Based on the RPD scale, it could be concluded that leaf trash could be better predicted in either the 405 to 1095 nm or 405 to 2495 nm region than in other two ranges (1105-2495 nm and 900-1700 nm), whereas non-leaf trash might be modeled effectively by one of four models examined here. Hence, this observation demonstrates the potential of visible/NIR technique in the precise and quantitative measurement of cotton trash within the categories of leaf and non-leaf class.

Table 3 reveals the best model in the 405 to 1095 nm region for leaf trash (RPD = 3.9), and one in the 900 to 1700 nm region for non-leaf trash (RPD = 3.5). This is expected because of significant distinctions in color and compositions between leaf and non-leaf trash. Meanwhile, it suggests the importance of different spectral wavelengths to optimize modeling efficiency.

Typical spectra of clean cotton fibers, leaves, and non-leaf trash samples are compared in Fig. 3. Owing to large variations in physical, chemical, composition, and color among them, their spectral responses are different. It is of interest to understand the spectral feature of individual or pure plant parts and for their identification; this concept was reported by Fortier et al. (2011). Relative to that of clean cotton fibers, the spectra of leaf and non-leaf trashes exhibited the larger log (1/R) readings in the 400 to 1000 nm region and a weaker intensity in the 2020 to 2200 nm region. Probably, more chlorophyll (represented by the 675 nm band) and its degradation derivatives in leaf and non-leaf trash than in cotton fiber are responsible for intense bands in the 400 to 1000 nm region. As anticipated, there is more cellulose (indicative of the 2105 nm band) in cotton fiber than in leaf and non-leaf trash.

**Individual Non-leaf Trash.** Non-leaf trash components, such as stem trash, hull trash, seed coat trash, and sand/soil trash, were also examined. Unfortunately, the results in Table 3 suggest some hindrance in the prediction of these individual trash components, because their RPDs are much less (< 3.0). That is, non-leaf trash could not be modeled as effectively as leaf trash. One factor might be due to particle size and density of these trash components and their uniform distribution. Additional research is needed to fully understand the effect of trash size, type, and weight distribution on the NIR performance.
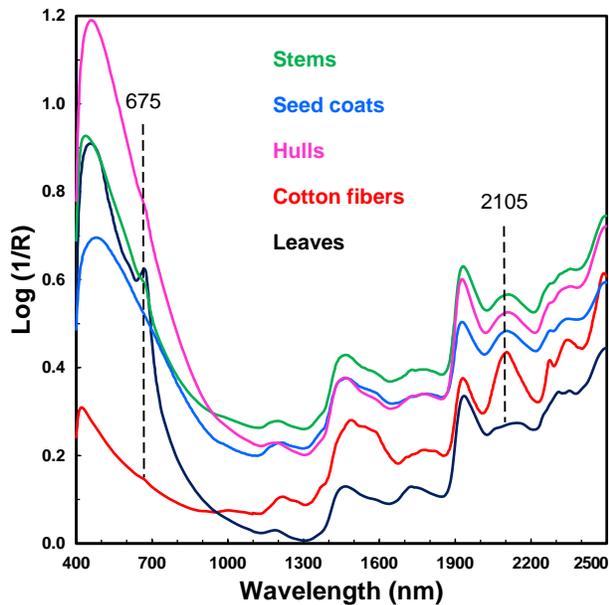
**Figure. 3. Typical visible/NIR log (1/R) spectra of clean cotton fibers, leaf trash, and non-leaf trashes that included stems, seed coats and hulls. Spectra were shifted vertically for clarification.**

## CONCLUSION

This study explores the potential of visible/NIR technique to determine individual trash components within regular cotton by analyzing a number of mixtures consisting of known trash components (e.g., leaves, seed coats, hulls, stems, and sand/soil) and clean lint fibers. The results showed the feasibility of using visible/NIR technique for the precise and quantitative determination of total trash, leaf trash, and non-leaf trash components. In addition, it indicated the difficulty in the prediction of non-leaf trash such as stem, hull, seed coat, and sand/soil. A probable rational for this limitation was the particle size and density of these trash components and their uniform distribution. Further study is necessary to understand the relationship between spectral response and non-leaf trash components.

## ACKNOWLEDGMENT

## DISCLAIMER

Mention of a product or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

## REFERENCES

American Society for Testing and Materials (ASTM). 2012a. Standard test method for non-lint content of cotton (D2812). *In* Annual Book of ASTM Standards. ASTM International, West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2012b. Standard test method for measurement of physical properties of cotton fibers by high volume instruments (D5867). *In* Annual Book of ASTM Standards. ASTM International, West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2012c. Standard test method for neps in cotton fibers (D5866). *In* Annual Book of ASTM Standards. ASTM International, West Conshohocken, PA.

Boykin, J., C. Armijo, D. Whitelock, M. Buser, G. Holt, T. Valco, D. Findley, M. Barnes, and M. Watson. 2009. Fractionation of foreign matter in ginned lint before and after lint cleaning. Trans. ASABE, 52:419–426.

Fortier, C., J. Rodgers, M. Cintron, X. Cui, and J. Foulk. 2011. Identification of cotton and cotton trash components by Fourier transform near-infrared spectroscopy. Textile Res. J. 81:230–238.

Gamble, G. 2008. Method for the prediction of the rate of +b color change in upland cotton (*Gossypium hirsutum* L.) as a function of storage temperatures. 2008. J. Cotton Sci. 12:171–177.

Hua, S., X. Wang, S. Yuan, M. Shao, X. Zhao, S. Zhu, and L. Jiang. 2007. Characterization of pigmentation and cellulose synthesis in colored cotton fibers. Crop Sci. 47:1540–1546.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010a. Evaluation of 3 cotton trash measurements methods by visible/near-infrared reflectance spectroscopy. ASABE paper no. 1008708. St. Joseph, MI.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010b. 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., S. Kokot, and T.A. Sambi. 1998. Vibrational spectroscopy investigation of Australian cotton cellulose fibers. Part 2. A Fourier transform near-infrared preliminary study. Analyst. 123:1725–1728.

Liu, Y., D. Thibodeaux, and G. Gamble. 2012. Relationship between three cotton trash measurements and near infrared spectral response. J. Cotton Sci. 16:272–281.

Matusiak, M. and A. Walawska. 2010. Important aspect of cotton color measurement. Fibers Textiles East. Eur. 18:17–23.

Montalvo Jr., J.G., S.E. Buco, and H.H. Ramey Jr. 1994. Studies to measure cotton fiber length, strength, micronaire and color by vis/NIR reflectance spectroscopy. Part II: Principal components regression. J. Near Infrared Spectrosc. 2:185–198.

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.