# COMPARISON OF NIR AND FT-IR SPECTRAL MODELS IN THE PREDICITON OF COTTON FIBER STRENGTH

**Yongliang Liu**
**Gary R. Gamble**
**Devron P. Thibodeaux**
**Cotton Quality Research Station, ARS, USDA, P.O. Box 792**
**Clemson, SC 29633**

## Abstract

Strength quality in cotton fibers is one of several important end-use characteristics. In routine programs, it has been mostly assessed by automation-oriented high volume instrument (HVI) system. An alternative method for cotton strength is near infrared (NIR) spectroscopy. Although previous NIR models have suggested the challenge in accurate and reliable prediction of HVI strength, in this research we have observed a much improved NIR model for HVI strength after applying the pre-screening procedure to determine appropriate calibration samples. As a validation and complementary approach, the FT-IR spectra were collected and subsequently correlated with cotton Stelometer strength. The results suggested that the capability of FT-IR model on Stelometer strength is in good agreement with that of NIR model on HVI strength, verifying the potential of NIR technique in robust, reliable and quantitative determination of cotton strength property.

## Introduction

In the latest study (Liu et al., 2011), we have not only demonstrated the consistency of strength readings between two instruments if the strength readings were divided by cotton micronaire, but also reported a much enhanced NIR model with the use of modified HVI strength as a reference. On the other hand, Stelometer strength were not correlated with NIR spectra, because the broken samples from Stelometer measurement were too small (0.4 mg x 6 replicates = 0.0024g) to be loaded into any size of NIR cells.

Mid-Infrared (IR) spectroscopy has been applied to study the structure and conformation of cotton fibers and celluloses, due to its sensitivity to local molecular rearrangement caused by inter- and intra- molecular hydrogen bonding network inside celluloses (Abidi, et al., 2008; Hsieh, 2007; Liu et al., 2010b). Though there are significant differences in spectrometer configuration and absorption origin between IR and NIR spectroscopy, IR technique provides fundamental and complementary information to NIR and also has the common ground to NIR from the standpoint of correlating spectral data to specific interest. In particular, one advantage of IR over NIR method lies in its capability of micro sampling. The objectives of this study were: (1) to correlate the HVI strength with Stelometer strength and further to propose a pre-screening procedure in determining appropriate calibration samples for NIR model development, (2) to develop NIR model on HVI strength, (3) to establish FT-IR model on Stelometer strength, and (4) to compare the modeling power between NIR and IR spectra.

## Materials and Methods

### Cotton Fibers, Visible/NIR Spectra, HVI and Stelometer Strength

A total of 405 lint cottons were collected over years to represent diverse distributions in cotton variety, growing years and locations. Among them, 104 samples were the standard cottons for maturity determination in great diversity (Hequet et al., 2006), 168 were from 2009 crop year and grown in U.S., 21 were from 2001 crop year and grown in U.S., 17 were from 2009 crop year and grown outside the U.S. (China, South Africa, Pakistan, Zambia, and Ivory Coast), 6 were from standard cottons for calibrating the HVI systems around the world, and the others were from undisclosed locations and crop years that were retained at this lab during the past year. After conditioning for 48 hrs, visible/NIR reflectance spectra (400 - 2500 nm) were acquired on a FOSS XDS rapid content analyzer (FOSS NIRSystems Inc., Laurel, MD), at 0.5 nm interval and 32 scans. Approximately 10 g of cotton fibers was pressed into a FOSS coarse granular cell, which is rectangular with internal dimensions of 3.8 cm-wide x15.2 cm-long x 4.8 cm-depth. To keep a good contact between the cotton sample and optical window, 750 g of extra weight was loaded on the top of fiber samples consistently throughout the entire experiment. Three spectra were collected for each of cotton sample by repacking and then mean spectrum was available for model development.

HVI strength and micronaire values were either measured by the standard HVI procedure at this laboratory (Liu et al., 2010a) or provided by the collaborators. Meanwhile, part of total samples was chosen randomly for Stelometer strength measurement with the established Stelometer protocol (Liu et al., 2011). All broken specimens were retained for the following FT-IR spectral acquisition.

**Broken Cotton Fibers and FT-IR Attenuated Total Reflection (ATR) Spectra**

The spectra of broken fiber specimens were collected with an FTS 3000MX Fourier transform (FT) IR spectrometer (Varian Instruments, Randolph, MA) equipped with a ceramic source, KBr beam splitter, and deuterated triglycine sulfate (DTGS) detector. The ATR sampling device utilized a DuraSamplIR single-pass diamond-coated internal reflection accessory (Smiths Detection, Danbury, CT), and a consistent contact pressure was applied by the way of a stainless steel rod and an electronic load display. Two spectra were collected at the spots closing to broken points or tips for individual breaks, over the range of 4000-600 $cm^{-1}$ at 4 $cm^{-1}$ and 32 co-added scans. This led to twelve spectra for each cotton sample and an average of them was available for model development. All spectra were given in absorbance units and no ATR correction was applied.

**Partial Least Squares (PLS) Models**

Two types of spectra were imported into PLSplus/IQ package in Grams/AI (Version 7.01, Thermo Fisher Scientific, Waltham, MA) and were smoothed with a Savitzky-Golay function (polynomial = 2 and points =13 for NIR and 11 for FT-IR), prior to respective calibration model development. In NIR model, for individual data set on the order of the smallest to the largest in HVI strength readings, the samples (numbered 2, 5, 8, 11, …) were selected to validate the calibration models that were built from the remaining samples. In the FT-IR model, on the order of the smallest to the largest in Stelometer strength readings, one hundred and sixty-eight of 252 spectra (or samples) were selected for calibration equation development, and the remaining 84 spectra (every 3rd sample) were used for model validation. To optimize the accuracy of prediction models, the spectra were subjected to different combinations of both spectral ranges (e.g., full and narrow regions) and spectral pre-treatments (e.g., mean centering (MC), multiplicative scatter correction (MSC), standard normal variate (SNV), the first and second derivatives). One-sample-out rotation cross-validation method was used, and the number of optimal factors selected for the regression equation generally corresponded to the minimum of the predicted residual error sum of squares (PRESS). Model accuracy and efficiency were assessed in the validation set on the basis of multiple coefficient of determination ($r^2$), root mean square error of prediction (RMSEP), and ratios of the standard deviation (SD) and the range to RMSEP (RPD/RER) (Williams, 2007).

## Results and Discussion

**Correlation of HVI vs. Stelometer Strength and Determination of Calibration Samples in NIR Modeling**

Figure 1 shows the plot of HVI micronaire corrected HVI against Stelometer strength for a total of 252 cotton fibers from diversified sources. Due to the nature of inhomogeneous distribution of cotton fibers in structural, chemical, and physical properties and also the concern of different sampling specimens for two independent measurements, it is reasonable to observe a number of scattered cottons (outliers) that exhibited large differences between two ratios (Pearson correlation, r, = 0.90). To this regard, the ratios (HVI strength / Stelometer strength) were calculated and then utilized to determine the outlier samples. Only those samples with the ratio range of 1.33 to 1.49 were subjectively considered to possess close (or appropriate) strength values between two physical tests. This procedure resulted in a selection of 163 samples that yielded an expected increase in r (0.98).

Back to the original strength readings (Figure 2), relatively lower correlations were observed. However, an r of 0.92 from selected cottons might suggest a good correlation between Stelometer and HVI testing, hence, either reading could be reliable and accurate for NIR model development. Meanwhile, Figure 2 indicates that Stelometer strength is smaller than the corresponding HVI value, and also reveals that Stelometer strength could be equivalent to HVI reading if multiplied by 1.4.

**Correlation of Stelometer Strength at Different Runs**

In order to unravel the variations in strength quality within one sample from different periods or runs, 104 standard cottons recommended by Hequet et al (2006) were analyzed twice by the same Stelometer instruments and operators. In Figure 3, the strength along the x-axis (horizontal) were the average of six breaks from two operators at one time, and the ones along the y-axis (vertical) were the average of three broken bundles from only one operator at differing time. Notably, an r of 0.94 in Figure 3 is good but slightly discouraging to have several scatter samples, echoing the concern of uniformity in strength distribution and subsequent reference determination for natural fibers.

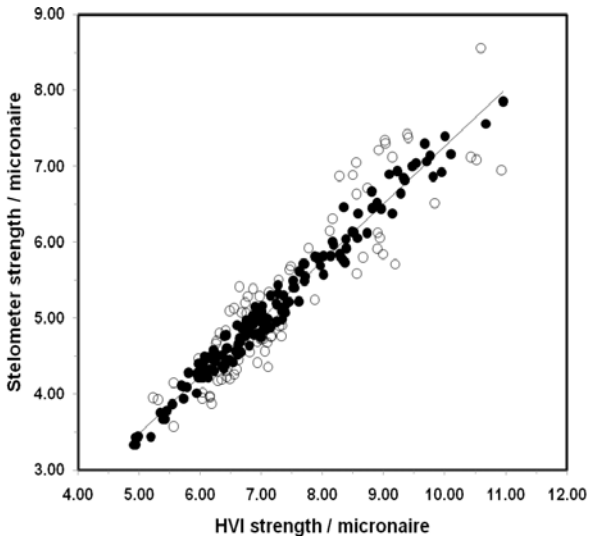


Figure1. Plot of corrected HVI vs. Stelometer strength. r = 0.90 and 0.98 for all and selected 163 samples (●).

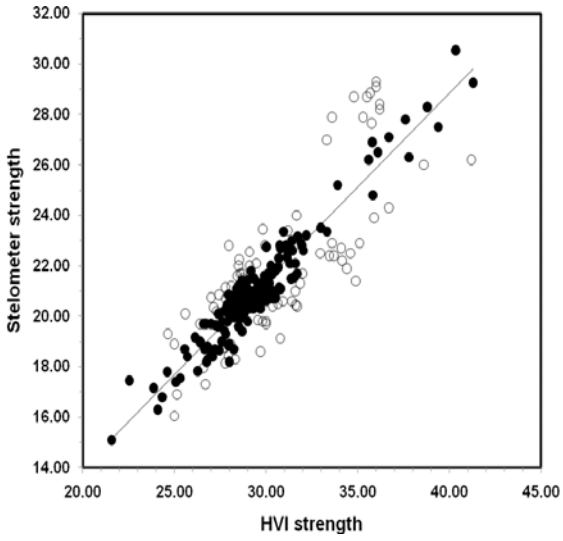


Figure 2. Plot of HVI vs. Stelometer strength. r = 0.74 and 0.92 for all and selected 163 samples (●).

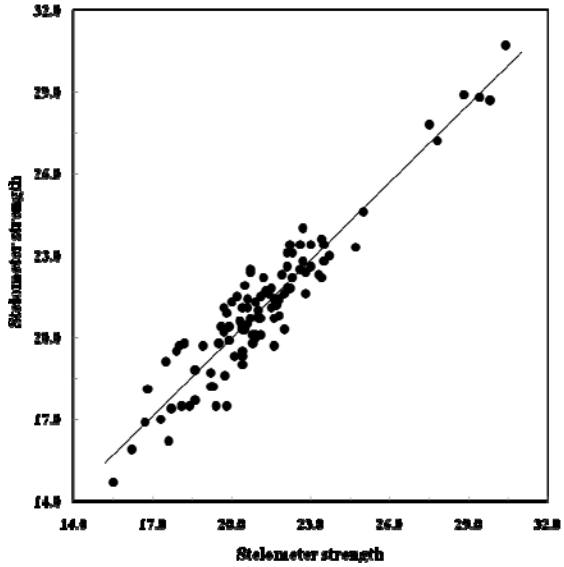

Figure 3. Plot of Stelometer strength at two differing runs. r = 0.94 for all 104 cotton standards.

**NIR Models on HVI Strength**

Representative log (1/R) spectra of 3 cotton fibers in the 1100-2500 nm NIR region are shown in Figure 4. There are at least five intense and broad bands (1490, 1935, 2105, 2270, and 2320 nm), mainly due to the (1st and 2nd) overtones and combinations of OH and CH stretching vibrations of cotton cellulose (> 90% in total mass).

To examine the effect of removing outliers on strength models, PLS model was first developed from a large sample set (SET A). This is a common practice and it produced low model characteristics (Table 1). With the use of a 90% confidence interval to exclude 15 and 10 samples that had large differences (or errors) between measured and NIR predicted HVI strength from calibration and validation sets (SET B), the recalibrated model suggested an improved $R^2$, RMSEC, $r^2$, and RMSEP, but with nearly unchanged RPD/RER.

NIR spectra of cotton samples used for Stelometer strength measurement were not collected, since the Stelometer samples were too small (0.4 mg x 6 replicates = 0.0024g) to be loaded into any size of NIR cells and also Stelometer

strength were found to have a roughly linear relationship with HVI strength (Figure 2). SET C consisted of a total 163 samples (109 in calibration set and 54 in prediction set) that were determined from Figure 2. Obviously, NIR model from SET C suggested an improvement in $R^2$, RMSEC, and $r^2$. It is of interest that both RMSEP and RPD/RER in SET C were close to those in SET B, likely indicating the presence of outlier samples in SET C. Even though samples in SET C were pre-screened on the basis of two isolated testing, it might not ensure the homogeneous and close specimens for additional NIR measurement. In addition, with respect to the amount in NIR spectral acquisition (~ 10 g), there were relatively small sample sizes (~ 0.5 mg) between 2 pairs of clamps during HVI procedure. Therefore, a 90% confidence interval was again applied to exclude the outliers (SET D). By comparing with those reported before (Liu et al., 2010a, 2011), the redeveloped model revealed a great enhancement in all model characteristics and was quite encouraging. It indicated that pre-selection of calibration samples for predicting the fiber strength or other physical properties might be necessary. Due to the distribution of sample numbers in HVI strength readings, RER might be better suited to evaluate the calibration models than RPD. An RER of > 13.0 suggested the acceptability and potential of the NIR model for the quantitative prediction of cotton fiber strength. Comparative scatter plots of measured and NIR predicted HVI strength in validation sets are shown in Figure 5. It suggests how well the NIR model predictions agree with the references from a separated measurement. For subjectively determined sample set (SET D), the solid regression line is nearly along the 45-degree direction.

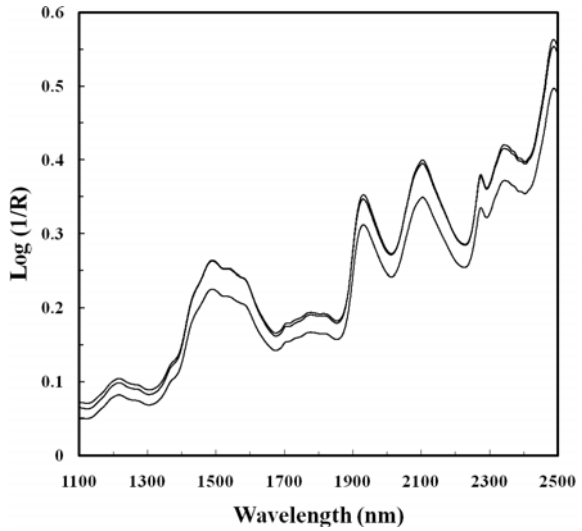


Figure 4. Typical NIR spectra of cotton fibers.

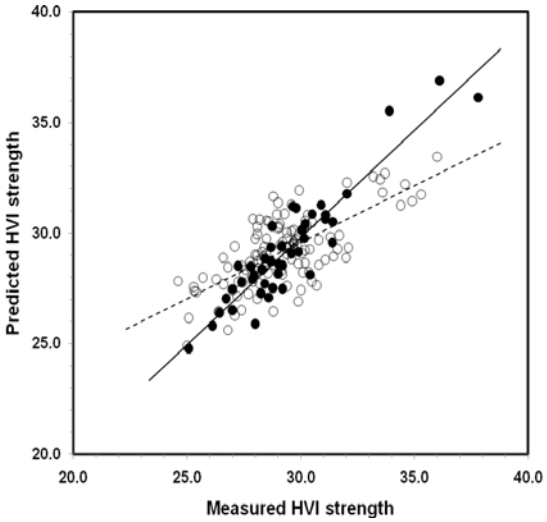


Figure 5. Plot of measured vs. NIR predicted HVI strength for SET B (○ and dashed) and SET D (● and solid).

Table 1. Optimal NIR model comparison on HVI strength in the 1105-2498 nm NIR region [a,b]

| Samples | Calibration set [c] | | | | | Validation set [c] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. | Range | SD | $R^2$ | RMSEC | No. | Range | SD | $r^2$ | RMSEP | RPD/RER |
| SET A | 273 | 21.6-41.3 | 2.79 | 0.44 | 2.10 | 132 | 24.3-39.4 | 2.58 | 0.37 | 2.06 | 1.25/7.33 |
| SET B | 258 | 22.5-36.9 | 2.26 | 0.55 | 1.52 | 122 | 24.6-36.0 | 2.13 | 0.48 | 1.54 | 1.38/7.40 |
| SET C | 109 | 21.6-41.3 | 3.00 | 0.82 | 1.29 | 54 | 24.3-37.8 | 2.45 | 0.64 | 1.69 | 1.45/7.99 |
| SET D | 92 | 21.6-38.8 | 2.64 | 0.96 | 0.54 | 44 | 25.1-37.8 | 2.36 | 0.85 | 0.97 | 2.43/13.1 |

[a] Spectral pre-treatment with MC and a Savitzky-Golay 1st function of two degrees and thirteen points for all models; Suggested 7, 8, 11, and 13 optimal factors for SET A, B, C, and D models, respectively.
[b] Samples in SET C and D were part of those and also differed from those in SET A and B. SET B and D resulted from the application of 90% confidence interval to SET A and C, respectively.
[c] SD, standard deviation; RMSEC, root mean square error of calibration; RMSEP, root mean square error of prediction; RPD, ratio of SD to RMSEP; RER, ratio of range to RMSEP.

**FT-IR ATR Spectral Characteristics of Oriented Cotton Fibers**

Figure 6 shows the scheme of broken specimen from Stelometer procedure and FT-IR ATR micro sampling positions, and Figure 7 depicts the FT-IR ATR spectra in the 3600-600 $cm^{-1}$ region of orderly oriented fibers with different Stelometer strength. These spectra were collected under similar procedure, but they exhibited large variations in relative intensity and band position. Similarity of these spectra indicated the dominant FTIR ATR bands that arise from major common chemical component in matured cottons, cellulose.

**FT-IR Models on Stelometer Strength**

PLS models 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 individual spectral region are tabulated for comparison (Table 2). In addition to the entire 3600-600 $cm^{-1}$ region, the spectra were analyzed subjectively in five narrow regions: 1800-600, 3600-800, 1800-800, 1500-800, and 1200-800 $cm^{-1}$. The reasons for choosing these spectral regions were to compare the model performance from different spectral absorptions that are indicative of unique vibration modes in celluloses, and also to facilitate the development of portable and handheld IR sensors. The optimal calibration models were obtained from the combination of MC and Savitzky-Golay 1$^{st}$ derivative (2 degrees and 13 points) spectral pre- processing in each spectral region. In comparison with those from 3600-800 $cm^{-1}$ and 1800-800 $cm^{-1}$ region, two models with an extension to 600 $cm^{-1}$ did not increase the modeling efficiency obviously, likely suggesting the insignificance of relative amount of two crystal forms ($I_\alpha$ and $I_\beta$) in strength characterization. Meanwhile, the models with reducing spectral regions of from 3600-800 $cm^{-1}$ to 1200-800 $cm^{-1}$ became worse in $R^2$ and RMSEC in calibration set, but were shown nearly unchanged $r^2$ and RMSEP in validation set. In general, the strength could be better predicted by the models from either 3600-800 $cm^{-1}$ or 1800-800 $cm^{-1}$ region than from two narrow regions of 1500-800 $cm^{-1}$ and 1200-800 $cm^{-1}$. As a compromise, the model from the 1800-800 $cm^{-1}$ region was subjectively considered as the optimal model for future applications, because this spectral region is mostly and commonly utilized in IR researches.

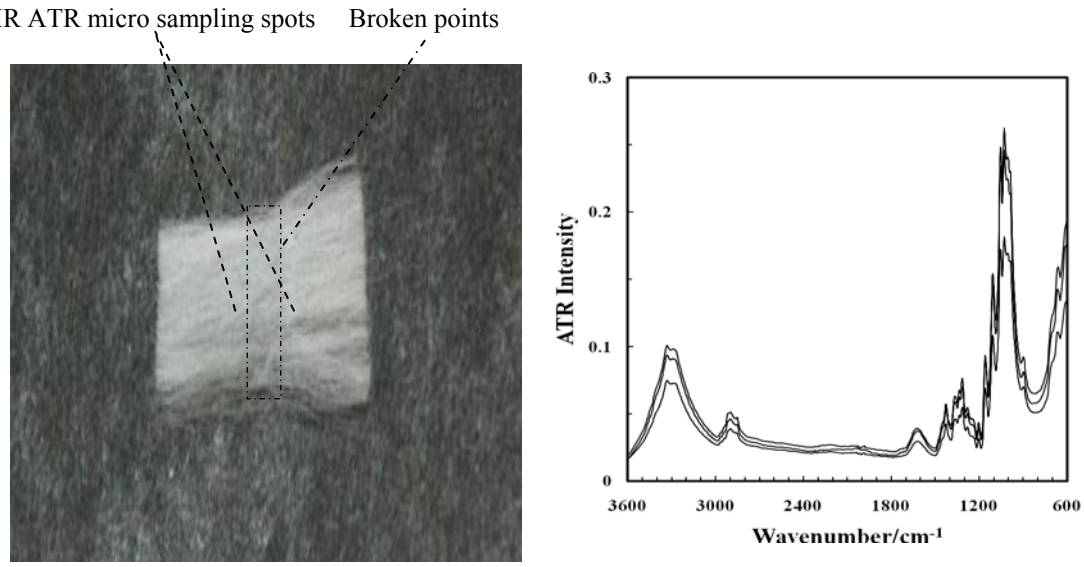


Figure 6. Scheme of FT-IR ATR micro sampling spots. Figure 7. FT-IR ATR spectra of orderly oriented fibers.

RPD or RER, quotients of SD and range of references to RMSEP in validation set, are often used as a dimensionless gauge of the ability of a spectroscopic model to predict a given property (Williams, 2007). Minimum acceptable values for the RPD and RER are suggested to be 3.0 and 10.0, respectively. With RPD of less than 2.0 and RER of less than 10.0, the Stelometer strength could not be modeled effectively by the FT-IR model created above.

NIR prediction of cotton strength using the respective HVI values as references has been performed before, in which the $r^2$ of 0.37 to 0.63 was reported (Liu et al., 2010a, 2011). Many differences existed between this FT-IR and earlier NIR study, for example, spectral region (1800-800 $cm^{-1}$ or 5555-12500 nm IR vs. 1100-2498 nm NIR), reference source (Stelometer in FT-IR vs. HVI in NIR), sampling size (0.5 mg in FT-IR vs. 0.5~10 g in NIR), and fiber status

(orderly oriented in FT-IR vs. randomly twisted in NIR). In this work, we have not only observed a strong correlation (Pearson correlation, r =0.92) between Stelometer and HVI strength, but also concluded that either readings could be reliable and accurate for NIR model development. Compared to RPD between 1.3 and 1.6 in preceding NIR studies (Liu et al., 2010a, 2011), the current FT-IR model was insignificant improved (RPD=1.8), but it still indicated the challenge in reliable and quantitative prediction of cotton fiber strength.

Table 2. Optimal Stelometer strength models in calibration and validation sets from FT-IR ATR spectroscopy [a]

| Spectral region ($cm^{-1}$) | Optimal factors | Calibration set (n = 168) | | | | Validation set (n = 84) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Range | SD | $R^2$ | RMSEC | Range | SD | $r^2$ | RMSEP | PD/RER |
| 3600 - 600 | 9 | 15.1-30.6 | 2.66 | 0.75 | 1.33 | 16.8-29.1 | 2.50 | 0.51 | 1.44 | 1.7/8.5 |
| 1800 - 600 | 10 | 15.1-30.6 | 2.66 | 0.74 | 1.37 | 16.8-29.1 | 2.50 | 0.46 | 1.56 | 1.6/7.9 |
| 3600 - 800 | 9 | 15.1-30.6 | 2.66 | 0.76 | 1.30 | 16.8-29.1 | 2.50 | 0.58 | 1.42 | 1.8/8.7 |
| 1800 - 800 | 9 | 15.1-30.6 | 2.66 | 0.70 | 1.45 | 16.8-29.1 | 2.50 | 0.59 | 1.42 | 1.8/8.7 |
| 1500 - 800 | 8 | 15.1-30.6 | 2.66 | 0.66 | 1.57 | 16.8-29.1 | 2.50 | 0.60 | 1.39 | 1.8/8.8 |
| 1200 - 800 | 8 | 15.1-30.6 | 2.66 | 0.59 | 1.70 | 16.8-29.1 | 2.50 | 0.55 | 1.44 | 1.7/8.5 |
| ***1800 - 800*** | ***9*** | ***15.1-29.3*** | ***2.35*** | ***0.84*** | ***0.95*** | ***17.4-28.3*** | ***2.18*** | ***0.74*** | ***0.84*** | ***2.6/13.0*** |

[a] Spectral pre-treatment with mean entering (MC) and a Savitzky-Golay first derivative function of two degrees and thirteen points for all models.

Even we took cautious experimental design of collecting both spectra and reference from identical broken specimens, low FT-IR modeling is not surprising. This is because of (i) highly diversification of native fibers and their heterogeneous distribution in strength quality, as the evidence from ~ 6% differences or errors between different Stelometer replicates within one individual sample, (ii) errors in Stelometer or spectral procedure from the operators, (iii) relatively small sampling size in FT-IR spectral collection (~ 2 nm in diameter) compared to that for Stelometer procedure (~ 10 x 15 nm in width x length), and (iv) displaced sampling spots between spectral and breaking force characterization.

To look into one of these concerns, a 90% confidential interval was applied to remove 26 and 14 samples from calibration and validation sets, respectively. The model was recalibrated in the 1800-800 $cm^{-1}$ and the result is also compiled in Table 2 (***Bold Italic***). It is encouraging to observe a remarkable improved $R^2$, RMSEC, $r^2$, RMSEP, and RPD/RER. Notably, an elevation of RER to 13.0 suggested the potential of FT-IR model in the quantitative determination of fiber strength. The observation is in excellent consistent with that from the NIR model (RPD/ERE = 2.43/13.1), in which calibration samples were pre-screened prior to NIR model development on HVI strength. Comparative scatter plot of measured and FT-IR predicted Stelometer strength in validation set is given in Figure 8. Undoubtedly, relatively scatter pattern in Figure 5 and 8 emphasizes the importance of both investigating the strength quality attribute and elucidating the strength mechanism by additional methods and techniques. Meanwhile, more samples were grouped in the strength range of 17-23 (Figure 8) and it suggested the need of including much diverse samples in the robust and reliable model development.

## Summary

We have proposed a pre-screening procedure to determine appropriate calibration samples on the basis of two independent measurements of identical physical property (e.g., cotton strength), prior to NIR model development. Applying a 90% confidence interval to remove outlier samples, the resultant model demonstrated the feasibility of NIR technique in the quantitative prediction of cotton HVI fiber strength for the purpose of quality control.

Though an extraordinary step was taken, the FT-IR model showed some difficulty in quantitative determination of cotton Stelometer strength. This limitation comes from an apparent lack of uniformity of strength quality distribution in native fibers, and also different sampling size between spectral (~ 2 nm in diameter) and Stelometer (~ 10 x 15 nm in width x length) measurement. With the exclusion of outliers, the recalibrated model suggested the potential of using FT-IR technique in the precise and quantitative measurement of cotton strength quality. Most importantly and interestingly, the performance of FT-IR model on Stelometer strength is consistent with that of NIR model on HVI strength, confirming the suitability of NIR in the practical determination of cotton strength quality.

**Acknowledgement**

We are grateful to J. Linda, N. Carroll, and M. Morris (ARS, Clemson) for technical assistance in collecting the data. We are also thankful to Dr. J. Knowlton (USDA-AMS, TN) for providing the cotton samples.

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

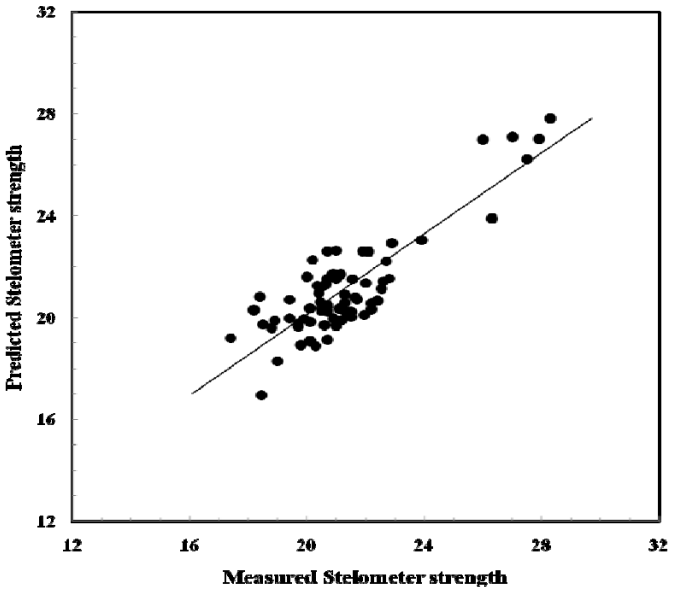


Figure 8. Plot of measured vs. FT-IR predicted Stelometer strength in validation set.

**References**

Abidi, N., E. Hequet, L. Cabrales, J. Gannaway, T. Wilkins, and L.W. Wells. 2008. Evaluating cell wall structure and composition of developing cotton fibers using Fourier transform infrared spectroscopy and thermogravimetric analysis. J. Applied Polymer Science. 107:476-486.

Hequet, E., B.Wyatt, N. Abidi, and D.P. Thibodeaux. 2006. Creation of a set of reference material for cotton fiber maturity measurements. Textile Research J. 76:576-586.

Hsieh, Y-L. 2007. Chemical structure and properties of cotton. P. 3-34. In S. Gordon and Y-L. Hsieh (Eds.) Cotton: Science and Technology. Woodhead Publish Limited, Cambridge, England.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010a. UV/visible/near-infrared reflectance models for the rapid and non-destructive prediction and classification of cotton color and physical indices. Transactions of the ASABE. 53:1341-1348.

Liu, Y., G. Gamble, and D. Thibodeaux. 2010b. Two-dimensional attenuated total reflection infrared correlation spectroscopy study of desorption process of water soaked cotton fibers. Applied Spectroscopy. 64:1355-1363.

Liu, Y., G. Gamble, and D. Thibodeaux. 2011. Correlation of HVI vs. Stelometer fiber strength and its application. Proceedings of 2011 Beltwide Cotton Conferences, National Cotton Council of America.

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, New Jersey, USA.