TEXTILE TECHNOLOGY

Obtaining Cotton Fiber Length Distributions from the Beard Test Method Part 2 – A New Approach through PLS Regression

Rachid Belmasrour, Linxiong Li, Xiaoliang "Leon" Cui*, Yiyun Cai, and James Rodgers

ABSTRACT

Cotton fiber length can be measured by the rapid method of testing fiber beards instead of individual fibers. In this method, only the fiber portion projecting from the fiber clamp can be measured. The length distribution of the projecting portion is very different from that of the original sample. Part 1 of this research reported that the original fiber length distribution of cotton can be modeled by a five-parameter mixed Weibull distribution, and the length distribution of the projecting portion can also be modeled by a five-parameter mixed Weibull distribution with parameters different from the original sample. Based on the results reported in Part 1, this work provides a new approach to estimation of the fiber length distribution of the sample based on the length distribution of the observed projecting portion. The proposed approach is the Partial Least Squares (PLS) regression. The probability density function (PDF) curves from the PLS regression method show a good match with the PDF curves obtained from the experimental data except in the short fiber region. Comparisons of some commonly used length quality parameters between experimental data and PLS regression showed good agreement for UHML and ML. The results indicate the proposed PLS approach for obtaining fiber length distribution from the beard test method is very promising, but additional work is needed to improve the estimation accuracy of short fibers.

Fiber length is a key property of cotton for marketing and yarn processing. As a natural product, cotton fiber length has a skewed, bell-shaped distribution with a high variation. There have been various studies detailing fiber length distributions (Prier and Sasser, 1971; Krowichi et al., 1996). In general, thousands of fibers need to be measured to generate a representative and statistically meaningful length distribution. Therefore, quantifying cotton fiber length distribution is time consuming and costly. Obtaining the entire fiber length distribution instead of a limited number of length parameters will enable a more complete evaluation of the cotton sample's quality. For example, the change of a length distribution curve may indicate impacts from cotton processing (Krifa, 2008). Also if the entire length distribution can be rapidly obtained from the beard testing method, any length parameters can be determined to suit customers' needs globally.

The beard testing method used in current testing devices such as High Volume Instrumentation (HVI) is based on the fibrogram theory developed in the 1940s (Hertel, 1936, 1940). In its industrial implementations, the engineering complexity of selecting fibers and forming a beard alters the original fiber length distribution observed by the instrument. This causes challenges in obtaining the entire original length distribution of a cotton sample. In other words, the fiber length distribution observed by the instrument is not the original fiber length distribution in the beard testing method. This could cause errors in length parameter measurements (Suh et al., 2006). New approaches to inferring fiber length distribution from the beard test results have been explored. This report details studies on the relationship between the observed length distribution and the original length distribution, thus providing a new approach for obtaining the original length distribution from the observed length distribution.

When using an HVI fiber clamp to make a fiber beard, a portion of each fiber is held inside the teeth of the clamp. This portion cannot be scanned (Cui et al., 2007; Cai et al., 2010). Another portion extrudes out of the clamp, this portion is defined as the projecting portion. The length distribution of the projecting portion is very different from that of the original fibers (Cui et al., 2009), therefore a model is

R. Belmasrour and L. Li, University of New Orleans, New Orleans, LA 70148; X. Cui* and J. Rodgers, USDA, ARS, SRRC, 1100 Robert E. Lee Blvd., New Orleans, LA 70124; Y. Cai, Louisiana State University, Baton Rouge, LA 70803
*Corresponding author: leon.cui@ars.usda.gov

needed to infer the actual, original length distribution of a cotton sample from the length distribution of the projecting portion.

In Part 1 of this research, the Gauss-Newton algorithm and nonparametric least squares principle was used to investigate the length distributions in connection with HVI tests (Cui et al., 2009). Specifically, a mixture of two Weibull distributions was used to fit the experimental data accurately. In the mixed Weibull distribution model, each distribution function has five parameters. Therefore, finding the fiber length distribution is equivalent to determining the five parameters. This paper reports work in developing a Partial Least Squares (PLS) regression model that estimates the five parameters of the actual original fiber length distribution by number from the corresponding parameters of the projecting portion length distribution by number.

MATERIALS AND METHODS

As discussed in a previous report of this research, eight varieties with different length characteristics were tested by using both Advanced Fiber Instrumentation System (AFIS) and HVI. The mean lengths by number of these cottons ranged from 1.65 cm to 2.29 cm, and Short Fiber Content by number (SFCn) from 17.7% to 28.0%. For each cotton, the length distributions discussed are the length distribution of the original sample and the length distribution of fibers projecting from the HVI clamp.

The original fiber length distributions were directly tested using AFIS, with samples randomly selected by hand in small pinches from the sample population. The HVI Fiber Sampler was used to prepare samples for length distribution of the projecting portion. Beards were made by the Fiber Sampler, combed, and brushed to remove loose fibers as in HVI length testing, with strength testing disabled. Then the projecting fibers were cut off along the baseline of the HVI clamp. The projecting fibers cut off from the beards were collected, gently and thoroughly opened to form thin fiber slivers, and tested on AFIS. Each distribution of each sample contained at least 35,000 individual fibers. The measured individual fiber length data from AFIS were used to construct the probability density functions (PDF) for fiber length distributions by number and by weight. The PDF by weight can be computed from the PDF by number (Cui et al.,

2008). Mixed Weibull function parameters of each length distribution were obtained from them (Cui et al., 2009).

It needs to be pointed out that in order for AFIS to measure each individual fiber, a sample sliver passes through a fast rotating fiber individualizer. The mechanical force to separate fibers introduces a degree of fiber breakage. In addition, fibers that are not individualized are not included in the computation of the fiber properties. The bias is reduced by the AFIS calibration algorithm, but may not be eliminated completely. In this paper, "original" refers to fiber length distributions measured by AFIS.

The length distributions (by number) of the projecting portions were from AFIS measurement of the fibers cut off along the HVI fiber sampler combs. An HVI instrument actually measures the same projecting portions, obtains the fibrograms, and calculates length parameters. In theory, the PDF by number can be obtained from the fibrogram to calculate the length parameters (Zeidman et al., 1991). Therefore, the method proposed here to obtain the fiber length distribution has the potential to be applied to the beard testing method as used by HVI.

The Weibull distribution function can be written as:

$$f(x; \lambda, \theta) = \lambda \theta x^{\lambda-1} e^{-\theta x^{\lambda}}, \ x > 0, \ \lambda > 0, \ \theta > 0 \quad (1)$$

where x is the variable, λ is the shape parameter, and θ is the scale parameter.

The mixed Weibull function has the following format:

$$f(x; \alpha, \lambda_1, \theta_1, \lambda_2, \theta_2) = \alpha f_1(x; \lambda_1, \theta_1) + (1-\alpha) f_2(x; \lambda_2, \theta_2) \quad (2)$$

where α is a proportion parameter and $0 < \alpha < 1$, and f_1 and f_2 are two Weibull distributions with different parameters λ and θ.

From equation 2, in the mixed Weibull models, five parameters are needed to describe the length distributions of cotton fibers: α, λ_1, θ_1, λ_2, θ_2. Therefore, if these five parameters of a cotton sample are known the fiber length distribution of this sample can be obtained.

As discussed earlier, the fiber portion actually scanned by a beard method is the projecting portion. The length distribution of this portion is very different from the original fiber. Again, a mixed Weibull distribution can be used to describe the projecting portion:

$$f(x; \alpha', \lambda_1', \theta_1', \lambda_2', \theta_2') = \alpha f_1(x; \lambda_1', \theta_1') + (1-\alpha') f_2(x; \lambda_2', \theta_2') \quad (3)$$

The five parameters α', λ'_1, θ'_1, λ'_2, θ'_2 describe the length distribution of the projecting portion. Therefore, if a regression can be established that computes the five parameters of the original distribution α, λ_1, θ_1, λ_2, θ_2 from the five parameters of the observed projecting portion distribution α', λ'_1, θ'_1, λ'_2, θ'_2, the entire fiber length distribution can be obtained.

The mixed Weibull function parameters of the original length distributions by number for the eight varieties were obtained. The response matrix **Y** can then be constructed:

$$
\begin{matrix}
\alpha^{(1)} & \lambda_1^{(1)} & \theta_1^{(1)} & \lambda_2^{(1)} & \theta_2^{(1)} \\
\alpha^{(2)} & \lambda_1^{(2)} & \theta_1^{(2)} & \lambda_2^{(2)} & \theta_2^{(2)} \\
& & \vdots & & \\
\alpha^{(8)} & \lambda_1^{(8)} & \theta_1^{(8)} & \lambda_2^{(8)} & \theta_2^{(8)}
\end{matrix}
$$

The mixed Weibull function parameters of the projecting portion length distributions by number were obtained for the eight varieties respectively. The predictor matrix **X** can then be constructed:

$$
\begin{matrix}
1 & \alpha'^{(1)} & \lambda_1'^{(1)} & \theta_1'^{(1)} & \lambda_2'^{(1)} & \theta_2'^{(1)} \\
1 & \alpha'^{(2)} & \lambda_1'^{(2)} & \theta_1'^{(2)} & \lambda_2'^{(2)} & \theta_2'^{(2)} \\
& & & \vdots & & \\
1 & \alpha'^{(8)} & \lambda_1'^{(8)} & \theta_1'^{(8)} & \lambda_2'^{(8)} & \theta_2'^{(8)}
\end{matrix}
$$

By utilizing the response matrix **Y** and predictor matrix **X**, a regression model can be established, so that with a given set of input α', λ'_1, θ'_1, λ'_2, θ'_2, this model can be used to compute the corresponding α, λ_1, θ_1, λ_2, θ_2.

Initial efforts included using the ordinary least squares (OLS) regression approach to infer the parameters of the original length distribution from those of the projecting portion. The results indicated that it was very difficult to yield satisfactory results by using OLS because of the relatively small sample size and multicollinearity.

Partial Least Squares (PLS) regression was used to carry out the above-mentioned inference. PLS is a statistical technique that generalizes and combines features from principal component analysis and multiple regression (Geladi and Kowalski, 1986; Garthwaite, 1994). PLS has some similarity with the principal component regression (PCR) (Wold, 1994), but the PLS method is less restrictive. It can better handle situations such as small sample size and multicollinearity.

Instead of searching for principal components (latent variables) that only depend on independent variables from the **X′X** matrix, the PLS method utilizes both independent variables and response variables to find latent variables from the **Y′XX′Y** matrix. Specifically, PLS regression searches for a set of factors that performs a simultaneous decomposition of **X** and **Y**. A constraint is that these factors explain the maximum covariance between **X** and **Y**. PLS regressions can be established to 1) find all factors, 2) determine how many factors are desired, and 3) use the selected factors to form the desired PLS regression coefficient matrix **B**:

$$ (\hat{a}, \hat{\lambda}_1, \hat{\theta}_1, \hat{\lambda}_2, \hat{\theta}_2) = (1, a', \lambda_1', \theta_1', \lambda_2', \theta_2') \mathbf{B} \quad (4) $$

where $\hat{a}$, $\hat{\lambda}_1$, $\hat{\theta}_1$, $\hat{\lambda}_2$, $\hat{\theta}_2$ are predicted values approaching the five parameters a, λ_1, θ_1, λ_2, θ_2 of the original fiber length distribution.

Garthwaite gave an excellent description on the fundamental ideas of PLS, formulas for obtaining factors, and ways to determine the number of factors desired (Garthwaite, 1994). The Nonlinear Iterative Partial Least Squares (NIPALS) algorithm was used in this research to carry out the PLS regression. In the computation, the inputs for the response matrix **Y** were from the results reported in Part 1, which were the mixed Weibull function parameters of the original length distributions (Table 3). The inputs for the predictor matrix **X** were the mixed Weibull function parameters of the projecting portion length distributions (Table 4).

RESULTS AND DISCUSSION

Cotton Fiber Length Distributions' Mixed Weibull Function Parameters. The PLS regression equation (2) was applied for all the eight cotton samples. The resulting regression coefficient matrix **B** is:

$$
\begin{bmatrix}
1.242 & -0.143 & 6.910 & 8.815 & -1.088 \\
-11.489 & 24.765 & -50.167 & -47.097 & -11.400 \\
-0.028 & 0.053 & -0.269 & 0.014 & -0.044 \\
0.002 & -0.007 & 0.008 & 0.013 & 0.006 \\
0.033 & -0.095 & -0.006 & 0.218 & 0.261 \\
-0.068 & 0.265 & -0.398 & -0.815 & 0.720
\end{bmatrix}
$$

Table 1 lists the PLS estimated mixed Weibull function parameters by using the above regression coefficient matrix. To illustrate the fitness of the curves, the graphs of PDF's from the experimental data and from the PLS prediction (with parameters given in Table 1) are shown in Figures 1-8. As shown from the figures, the length distributions by number

of the original fiber sample matched well with the predicted distributions, except in the region of fiber length less than about 0.5 inch. As it is well known, the lengths of shorter fibers are more difficult to handle and measure. Even though shorter fibers are high in number, they account for a smaller amount by weight of a sample. Therefore, the discrepancy in the shorter fiber region does not significantly affect the accuracy of the calculated length parameters such as UHML as show in the following paragraph.

Table 1. PLS Predicted mixed Weibull function parameters of the original length distributions.

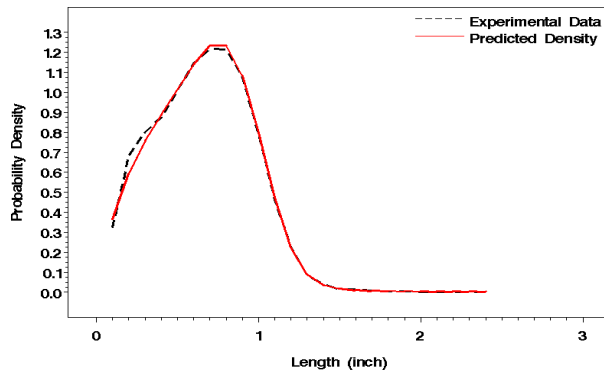| Sample	$\hat{a}$	$\hat{\lambda}_1$	$\hat{\theta}_1$	$\hat{\lambda}_2$	$\hat{\theta}_2$
30	0.506	1.815	2.673	4.187	1.754
31	0.468	1.730	2.349	4.493	1.774
33	0.423	1.890	2.092	4.556	0.947
34	0.457	1.765	1.945	5.106	0.784
35	0.388	1.843	2.137	4.597	0.786
36	0.423	1.847	2.002	4.841	0.663
37	0.428	1.758	2.410	4.829	0.498
38	0.408	1.685	2.048	5.312	0.384

Figure 1 Probability density functions (by number) of ID 30 original fibers.
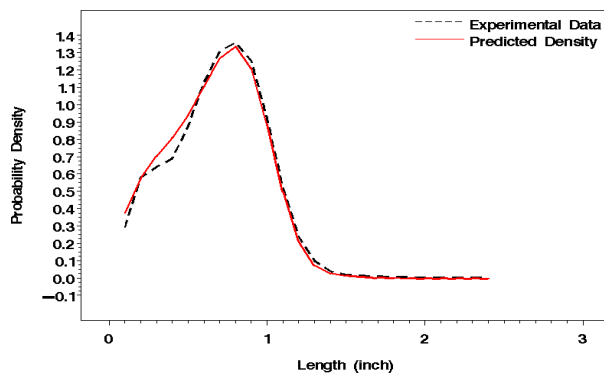

Figure 2 Probability density functions (by number) of ID 31 original fibers.
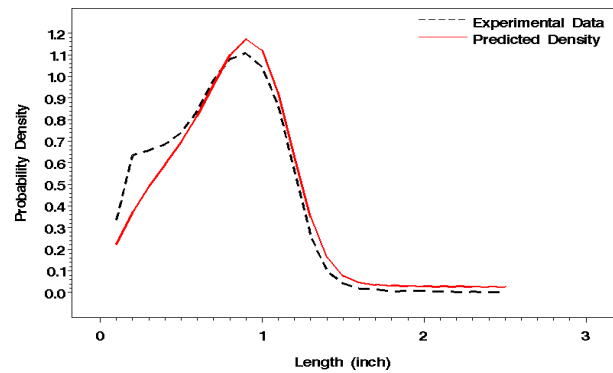

Figure 3 Probability density functions (by number) of ID 33 original fibers.
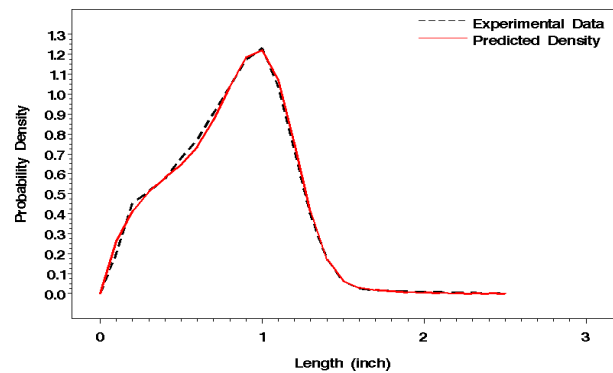

Figure 4 Probability density functions (by number) of ID 34 original fibers.

Figure 5 Probability density functions (by number) of ID 35 original fibers.

Figure 6 Probability density functions (by number) of ID 36 original fibers.

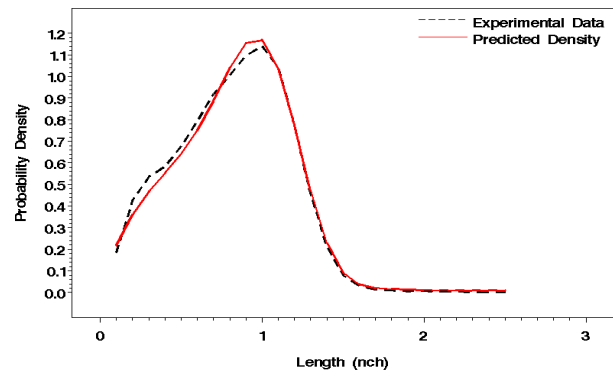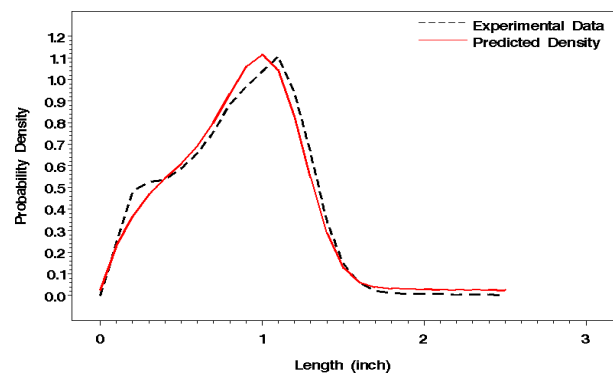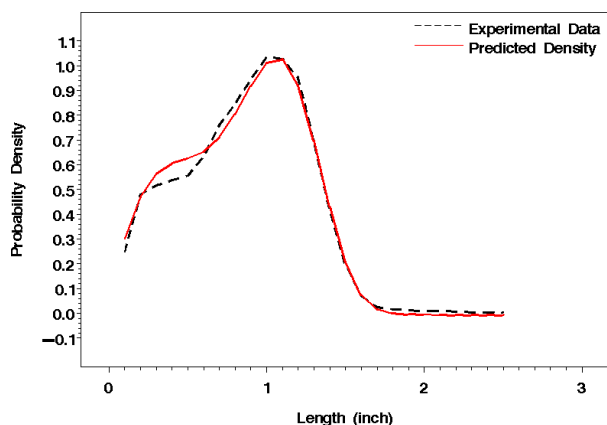**Figure 7 Probability density functions (by number) of ID 37 original fibers.**
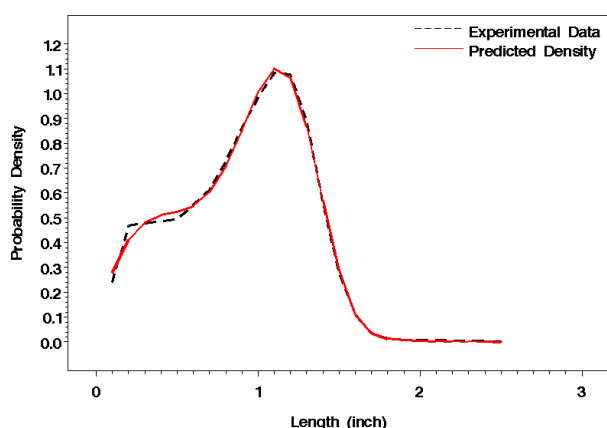

Figure 8 Probability density functions (by number) of ID 38 original fibers.

Quality Parameters Calculated from the Predicted Length Distributions. In addition to comparing the agreement of the predicted and the actual length distribution curves, length parameters from the predicted length distributions were calculated and compared with those of the AFIS measured data of the original samples. The length parameters include Mean Length by number (MLn), Upper Half Mean Length (UHML), Lower Half Mean Length (LHML), and Short Fiber Content by number (SFCn). In Part 1, the equations for computing these parameters from the mixed Weibull distribution function (Cui et al., 2009) were discussed. Table 2 lists the prediction results of some length parameters of the eight calibration varieties. The R^2 values are 0.98 for UHML, 0.94 for MLn, 0.89 for LHML, and 0.68 for SFCn. The length parameters from the PLS predicted distributions match those from the actual original distributions very well for UHML, but not as well for short fiber content. This may be related to the

bias introduced during the sampling process, shorter fibers have less possibility of being selected by the needles of the sampler comb (Cai et al., 2010). An effort is being made to improve the estimation accuracy of the shorter fiber length.

Table 2. Comparison of measured and predicted cotton fiber length parameters.

Sample		MLn	UHML	LHML	SFCn(%)
30	Original	0.658	0.983	0.494	31.7
	Predicted	0.654	0.977	0.492	31.5
31	Original	0.691	0.997	0.528	26.1
	Predicted	0.682	0.996	0.518	27.8
33	Original	0.737	1.105	0.553	26.7
	Predicted	0.787	1.122	0.607	20.5
34	Original	0.804	1.153	0.617	20.8
	Predicted	0.803	1.154	0.615	21.1
35	Original	0.810	1.168	0.620	20.5
	Predicted	0.818	1.163	0.630	19.4
36	Original	0.845	1.233	0.642	20.8
	Predicted	0.834	1.192	0.641	19.4
37	Original	0.857	1.257	0.650	20.5
	Predicted	0.837	1.248	0.629	22.8
38	Original	0.899	1.299	0.687	19.2
	Predicted	0.891	1.293	0.680	19.8
R^2		0.94	0.98	0.89	0.68
Standard deviation of residual		0.022	0.016	0.023	2.63

Table 3. Mixed Weibull function parameters of the original length distributions.

Sample	α	λ_1	θ_1	λ_2	θ_2
30	0.560	1.844	2.595	4.389	1.706
31	0.424	1.683	2.419	4.350	1.796
33	0.301	1.881	2.254	4.042	1.102
34	0.492	1.841	1.880	5.178	0.799
35	0.485	1.921	1.990	4.931	0.720
36	0.458	1.743	1.983	5.096	0.534
37	0.377	1.738	2.482	4.632	0.549
38	0.405	1.682	2.052	5.304	0.385

Table 4. Mixed Weibull function parameters of the projecting portion length distributions.

Sample	α	λ'_1	θ'_1	λ'_2	θ'_2
30	0.044	2.003	7.758	2.210	3.898
31	0.054	2.904	52.543	2.408	3.708
33	0.050	4.152	2.544	2.196	3.051
34	0.043	6.300	5.847	2.194	2.830
35	0.064	2.795	31.340	2.361	2.671
36	0.049	5.043	1.871	2.153	2.723
37	0.060	2.566	20.736	2.165	2.360
38	0.073	2.179	2.291	5.029	1.494

CONCLUSION

In fiber length measurement by a beard method, only the projecting fibers can be measured. The length distribution of the projecting portion is very different from that of the original sample. Part 1 of this research showed that the fiber length distribution of a cotton sample can be modeled by a five-parameter mixed Weibull distribution, and further, the length distribution of the projecting portion can also be modeled by a mixed Weibull distribution with different parameters to the original cotton. This work provides a new approach to infer the fiber length distribution by number of the original sample from the observed length distribution by number of the projecting fibers of a beard. The method proposed in this paper is Partial Least Squares regression, where distribution parameters of projecting length are independent variables and those of original length are response variables. The probability density function (PDF) curves from the PLS regression method show a good match with the PDF curves obtained from the experimental data except in the short fiber region. Comparisons of some commonly used length quality parameters between experimental data and PLS regression showed good agreement for UHML and ML. The preliminary results from the limited sample size indicate that the proposed PLS approach for obtaining fiber length distribution from the beard test method is very promising, but additional work is needed to improve the estimation accuracy especially in short fiber region.

DISCLAIMER

Names of companies or commercial products are given solely for the purpose of providing specific information; their mention does not imply recommendation or endorsement by the U.S. Department of Agriculture over others not mentioned.

GLOSSARY OF TERMS

OLS: Ordinary Least Squares

PLS: Partial Least Squares

PCR: Principal Component Regression

Mixed Weibull distribution model: A distribution model that consists of two Weibull distributions by using a proportional parameter

α: The proportional parameter of the original cotton fiber length distribution

λ, θ: Weibull function parameters of the original cotton fiber length distribution

α': The proportional parameter of the projecting portion fiber length distribution

λ', θ': Weibull function parameters of the projecting portion fiber length distribution

Subscripts 1 and 2: The first and second Weibull functions in the mixed Weibull distribution model

REFERENCES

Cai Y., X. Cui, J. Rodgers, V. Martin, and M. Watson, 2010, An Investigation on the Sampling Bias of the Beard Method as Used in HVI, J. Textile Inst. 101(11): 958-966.

Cui, X., D. Thibodeaux, K. Robert, J. Rodgers, V. Martin, and M. Watson, 2007. A Study to Improve the Measurement of Cotton Length Distribution from a Fiber Beard. p. 1296-1301. *In* Proc. Beltwide Cotton Conf., New Orleans, LA. 9-12 Jan. 2007. Natl. Cotton Counc. Am., Memphis, TN.

Cui, X., J. Rodgers, Y. Cai, L. Li, R. Belmasrour, and S. Pang, 2009. Obtaining Cotton Fiber Length Distributions from the Beard Test Method, Part 1 - Theoretical Distributions Related to the Beard Method. J. Cotton Sci. 13:265–273.

Cui, X., T.A. Calamari Jr., and M.W. Suh, 2008. Theoretical and Practical Aspects of Fiber Length Comparisions of Various Cottons. Textile Res. J. 68:467-472.

Garthwaite, P.H. 1994. An Interpretation of Partial Least Squares. J. Amer. Stat. Assoc. 89:123-127.

Geladi, P., and B.R. Kowalski. 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta. 185:1-17

Hertel, K.L. 1936. An optical method for the length analysis of cotton fibers. Text. Res. J. 6:331-339.

Hertel, K.L. 1940. A method of fiber-length analysis using the fibrograph. Text. Res. J. 10:510-525.

Krifa, M. 2008. Fiber Length Distribution in Cotton Processing: A Finite Mixture Distribution Model. Text. Res. J. 78:688-698.

Krowicki, R.S., D.P. Thibodaux, and K.E. Duckett. 1996. Generating fiber length distribution from the fibrogram. Text. Res. J. 66:306-310.

Prier, H.W., and P.E. Sasser, P.E. 1971. The mathematical basis of fiber-length analysis from fibrogram data. Text. Res. J. 42:410-419.

Suh, M.W., S.J. Doh, and M.D. Watson. 2006. Short Fiber Content Estimation – New Findings and A New Approach. pp. 1963-1969. *In* Proc. Beltwide Cotton Conf., San Antonio, TX. 3-6 Jan. 2006. Natl. Cotton Counc. Am., Memphis, TN.

Wold, S. 1994. PLS for Multivariate Linear Modeling. p.195-218. *In*: van de Water-beemd H (Editor) Chemometric Methods in Molecular Design. Verlag- Chemie, Weinheim, Deutschland.

Zeidman, M.I., S.K. Batra, and P.E. Sasser. 1991. Determining Short Fiber Content in Cotton Part I: Some Theoretical Fundamentals. Text. Res. J. 61: 21-30.