

RESEARCH

Open Access



Stability, variation, and application of AFIS fiber length distributions

HINDS Zachary¹, KELLY Brendan Robert^{1,2*} and HEQUET Eric Francois¹

Abstract

Background: Fiber length is one of the primary quality parameters for the cotton industry when considering the textile performance and end-use quality of cotton. Currently, many decisions regarding cotton fiber length utilize the industry standard measurement device, i.e., the High Volume Instrument (HVI). However, it is documented that complete fiber length distributions hold more information than the currently reported HVI length parameters, i.e., upper half mean length (UHML) and uniformity index (UI). An alternative measurement device, the Advanced Fiber Information System (AFIS), is able to capture additional information about fiber length distribution. What is currently not known is how much additional information the AFIS length distribution holds.

Results: The stability of differences in within-sample variation in fiber length captured by the AFIS length distribution by number characterizing differences between samples was deemed stable across the extended testing period. A diverse breeding population was evaluated and four significant sources of within sample variation in length were identified. A comparison of the ability between HVI length parameters and AFIS fiber length distribution to correctly categorize breeding lines to their family was performed. In all cases, the AFIS fiber length distribution more accurately identified germplasm families.

Conclusions: The long-term stability test of the AFIS fiber length distribution by number shows that the measurement is stable and can be used to assess differences across samples. However, more information about within-sample variation in fiber length than that can be captured by length parameters is needed to assess differences across samples in many applications. Four length parameters outperform two length parameters when trying to identify the familial background of the samples in this set. These parameters characterize distributional shape differences that are not captured by the standard AFIS length parameters, UQL and short fiber content by number (SFCn). These findings suggest that additional types of variation in cotton fiber length are not captured and are therefore not currently used in most cotton breeding programs.

Keywords: Cotton, Fiber quality, Fiber length distribution, Advanced Fiber information system (AFIS)

Background

Cotton is the most economically important natural fiber and a valuable agricultural commodity in the U.S. and even around the globe. However, the competition from other cotton-producing regions and synthetic fibers forces the U.S. cotton industry to

continually improve its product to remain competitive on the global market (Meredith 2005). The quality of the raw material is a major factor in determining the quality of the final product. In general, a bale of cotton is characterized as having good quality if its fibers are long, strong, mature, and contamination free. However, this is a rather simplified characterization of the complexity of cotton fiber quality determination. Many cotton fiber quality properties which contribute to quality yarn production are not captured by the

* Correspondence: zach.hinds@ttu.edu

¹Texas Tech University, 2500 Broadway, Lubbock, TX 79409, USA

²Texas AgriLife Research, 1102 E Farm to Market Rd 1294, Lubbock, TX 79403, USA



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

most common fiber quality evaluation system, the High Volume Instrument (HVI).

One such quality element is within-sample variation in fiber length. Variation in fiber length exists within a sample, and this variation, along with the ones in other fiber properties, may affect the quality of the finished product (Wakeham 1955; Koo and Moon 1999). Cotton fibers exhibit natural variation in length due to the environment, agronomic practices, and genetic factors (Stewart 1975; Basra 1999; Faulkner *et al.* 2011). The length of a cotton fiber will be at its maximum just before the boll opens. From this point on, the fiber is exposed to weathering, harvesting, and ginning that may lead to the shortening of its length through breakage. The impact of fiber breakage due to mechanical processes such as harvesting, cleaning, ginning, and spinning can further contribute to the variation in cotton fiber lengths (Mangialardi 1972; Hughs *et al.* 2013). These sources of variation present many unique challenges when attempting to improve fiber length.

HVI testing provides two fiber length parameters, upper half mean length (UHML) and uniformity index (UI), along with four other commonly used fiber quality parameters (micronaire, strength, reflectance, and yellow index). The HVI length measurement is based on the fibrograph principle, which measures fiber length from a beard of fibers held in a comb (Chu and Riley 1997). Length variation captured by these two length parameters are often used by spinning mills to identify quality differences and to select bales suitable for their production goals. Due to their importance in marketing, the broader cotton research community uses these parameters to help predict the type of spinning performance which they might expect from a given sample (El Mogahzy *et al.* 1990). However, these fiber quality parameters do not capture all the variations in fiber length within a sample or bale (Kelly and Hequet 2017). Standard HVI fiber length parameters cannot consistently distinguish important differences in length variation between bales that affect spinning performance. Special consideration should be given to an important fiber quality characteristic, within-sample distribution in fiber length, to improve spinning performance and yarn quality (Basra 1999).

The within-sample distribution of fiber length can be measured using the Advanced Fiber Information System (AFIS) instrument. For AFIS testing, fiber samples are formed into slivers and fed into the instrument, which mechanically separates individual cotton fibers. The individual fibers are then presented to an electro-optical sensor that measures the length of each fiber, along with several other fiber characteristics, and summarizes the length measurements into a relative frequency histogram

of 40 binned groups of similar length (Kelly *et al.* 2015). Within-sample variation in fiber length captured by the AFIS has been shown to be important in developing germplasm with the potential to produce fiber competitive on international spinning markets (Kelly *et al.* 2013).

The AFIS fiber length distribution is a complex measurement. While it is generally accepted that longer fibers will produce a stronger yarn, and fiber length distributional characteristics also have the potential to impact yarn quality. Selection work establishing the importance of the AFIS length distribution in breeding for improved yarn quality was performed by Kelly *et al.* (2013). If the differences of shape characteristics in the length distribution prove to contribute to better yarn quality, they could be quantified and targeted for improvement through breeding (Wakeham 1955; Krifa 2006; Kelly and Hequet 2017).

Differences in the fiber length distribution are important considerations for the research community. The AFIS length distribution holds more information than the HVI length parameters, UHML and UI (Kelly and Hequet 2017), but it is not known how much useful additional information the AFIS length distribution holds and how this information can be used. The objective of this research is to develop a quantitative approach to characterize between-sample differences in the fiber length distribution. The quantitative measurement is used to evaluate the stability of differences in within-sample variation in fiber length captured by the AFIS length distribution. Once the stability of these measurements is established, a diverse breeding population is developed to investigate the ways in which within-sample length variation can vary between and within families and the advantages of using the AFIS fiber length distribution in comparison with two parameters in characterizing a diverse population.

Methods

Stability of the length distribution measurement

Three cotton samples used at the Fiber and Biopolymer Research Institute at Texas Tech University for daily checks of the equipment were used to meet the objective of stability testing. The check cotton samples represent a range of variation in the length distribution by number shape and provide a measurement of this property over a long period.

Commercially grown and processed cotton bales were purchased for use as the check cottons. To achieve consistent sample quality, the bales were processed into card slivers at the FBRI. Transforming the raw cotton into card sliver involves several processing steps. First, the bales were opened and fed into the hoppers of the opening and mixing equipment (Hunter 240 BFC, Rieter B4/1, and Rieter ERM B5/5, Rieter AMH). Next, the fibers

were fed into a carding machine (Trützschler DK903) at a feeding speed of 214 m·min⁻¹ producing 40 g·yd⁻¹ card sliver. Card sliver was then fed into a drawing machine (Trützschler HSR1000) at a feeding speed of 600 m·min⁻¹ producing 35 g·yd⁻¹ D1 sliver. Finally, the samples were fed through a second drawing system (Rieter RSB 851) at a feeding speed of 400 m/min producing 35 g·yd⁻¹ D2 sliver. This D2 sliver was then placed in the FBRI Phenomics Lab for 48 h at (21 ± 1) °C and relative humidity of (65 ± 2)% to condition.

Each of the samples is evaluated on an AFIS Pro 2 (Uster Technologies AG, Memphis, TN) using a laboratory protocol for samples from commercial bales, where three slivers from each sample are evaluated with 3 000 fibers measured from each sliver. This measurement occurs once a day and is used for quality assurance management of the laboratory and produces commonly used fiber quality parameters along with length distribution measurements.

The length distribution measurement is reported by AFIS as a length-frequency histogram. The first step in developing a quantitative measure of differences in the length distribution by number is to convert this measurement into a length-response distribution (Kelly and Hequet 2017). This conversion preserves the variation captured by the AFIS length distribution by number while making length the response variable and provides a basis for stability assessment of length measurement.

Once the length distributions were converted to length response distributions, differences among the samples were characterized using a linear discriminant analysis (LDA). Significance of the discriminant axes was determined using an approximant F test ($\alpha = 0.05$).

The stability of the differences in the length distribution by number measurement among these samples was characterized by plotting the discriminant scores of these three samples over time and comparing their variation to an exponentially weighted moving average. There is no mathematical consensus on how to select a smoothing factor (Čisar and Čisar 2011). Therefore, a review of relevant literature resulted in the selection of a smoothing factor (λ) of 0.3 to give a reasonable amount of weight to both adjacent and more distant data points (Hunter 1986; Lucas and Saccucci 1990; Paudel et al. 2013). The smoothing was performed following (*current period date value* × λ) + (*previous period ewma* × (1 - λ)) = *current ewma*.

Characterizing length distribution differences in a diverse population

The check cottons provide a range of AFIS length distribution variation over time needed to develop a

quantitative measure of length differences and determine the stability of the measurement, but they are limited in the number of differences they can characterize. Any set of n sample can only vary in n-1 ways, even for a multivariate measurement like the AFIS length distribution. In practice, if a set of samples varies in n-1 ways, n-1 length parameters are needed to effectively differentiate these samples.

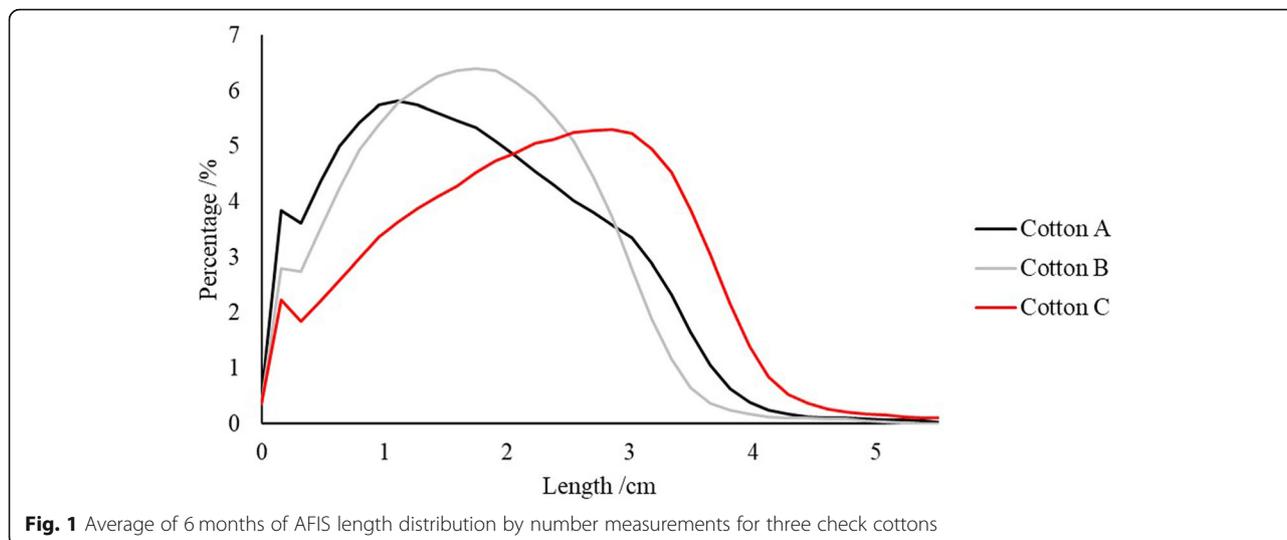
While three samples can capture two types of differences in within-sample variation in fiber length, typically two fiber length measurements are used to evaluate samples. In order to determine if this is sufficient, a larger set of diverse samples is needed.

Sixteen obsolete cotton varieties were acquired and crossed pairwise in the Texas Tech Greenhouse to produce eight F₁ populations. These plants were then selfed to produce the F₂ populations used in this experiment.

The eight F₂ populations were then planted in a completely randomized block design with three field replications at the Texas Tech Research Farm in 2017. Experimental plots were 4.6 m long with a planting density of 11 seeds per meter. The cotton was grown on

Table 1 Descriptive statistics of the three check cottons used in stability testing

	Quality parameters	Average	Min.	Max.
Cotton A	UQL /cm	3.05	2.97	3.12
	Ln /(cm)	1.8	1.7	1.93
	Ln CV /%	55.1	51.9	57.9
	SFCn /%	34.3	29.2	38.4
	L5% /cm	3.48	3.4	3.56
	Fine /mTex	162	159	166
	Mat. ratio	0.82	0.8	0.84
Cotton B	Hs	197	193	201
	UQL /cm	2.77	2.72	2.87
	Ln /cm	1.8	1.73	1.88
	Ln CV /%	48.3	45.8	51.4
	SFCn /%	29.8	26.2	33.7
	L5% /cm	3.2	3.12	3.3
	Fine /mTex	182	179	184
Cotton C	Mat. ratio	0.92	0.9	0.94
	Hs	197	195	200
	UQL /cm	3.45	3.4	3.51
	Ln /cm	2.31	2.24	2.41
	Ln CV /%	46.5	44.1	49.3
	SFCn /%	19.2	16.2	22.1
	L5% /cm	3.94	3.89	3.99
	Fine /mtex	168	166	170
	Mat. ratio	0.95	0.93	0.97
	Hs	177	173	181



a loam soil with subsurface drip irrigation. Scheduled irrigation and regional management practices were applied throughout the growing season.

Half of the mature plants from each entry were randomly selected and hand harvested for a total of 435 samples. The seed cotton was then ginned on a lab scale tabletop gin (Dennis Manufacturer, Athens, TX) to separate the seeds from the fibers. Each fiber sample was then tested on an AFIS Pro 2 (Uster Technologies AG, Memphis, TN) using a protocol of five reps testing 3 000 fiber per sample to capture the standard fiber quality parameters and to generate fiber length distributions. The goal of this section was to identify sources of variation within breeding populations and provide a potential application of the length distribution using early generation material. A 5-rep protocol was used in order to capture the increased level of within-sample variation typical of this type of sample.

The same procedure described above (Kelly et al. 2013) was then performed on the raw fiber length distributions so that within-sample variation captured by the AFIS was expressed as a length response curve. Linear discriminant analysis was performed to investigate the differences among and within the populations. Fiber length distributions from each of the populations were then averaged to generate representative distributions from each family.

Characterizing germplasm differences with the full distribution

Fiber length parameters are often used to evaluate germplasm in breeding. Therefore, the practical importance

of this variation was determined by comparing the ability of the length distribution by number to classify the familial relations of the samples. Ideally, this classification would be compared with the one based on High Volume Instrument (HVI) testing. However, the samples in this experiment were too small, i.e. individual plant selections, to support this type of fiber quality evaluation.

Instead, the length distribution by number classification was compared with a classification using AFIS upper quartile length (UQL) and AFIS short fiber content by number (SFCn). UQL was selected because it is considered similar to the HVI length parameter UHML, while SFCn was selected in order to compare the full length distribution by number against a length parameter that captures a portion of the length distribution by number that is not captured by HVI length measurements.

Linear discriminant analysis was used to classify each sample by family using each of the length parameters, the combination of the length parameters, and the AFIS length distribution by number (JMP Pro 14). The ability of the parameters and sets of parameters to differentiate samples was evaluated and compared using the total misclassification rate. A more detailed analysis of misclassification by family was used to compare a two-parameter screening approach to an approach based on the complete length distribution by number measured by the AFIS.

Table 2 Wilk's Lambda multivariate test of significance

Test	Approx. F	NumDF	DenDF	Prob > F
Wilks' Lambda	299.3	160	352	< 0.000 1*

Table 3 Significant axes of variation based on Approximant F test ($\alpha = 0.05$)

Eigenvalue	Percent	Approx. F	NumDF	DenDF	Prob > F
409.02	90.13	299.3	160	352	< 0.000 1*
44.81	9.87	100.4	79	177	< 0.000 1*

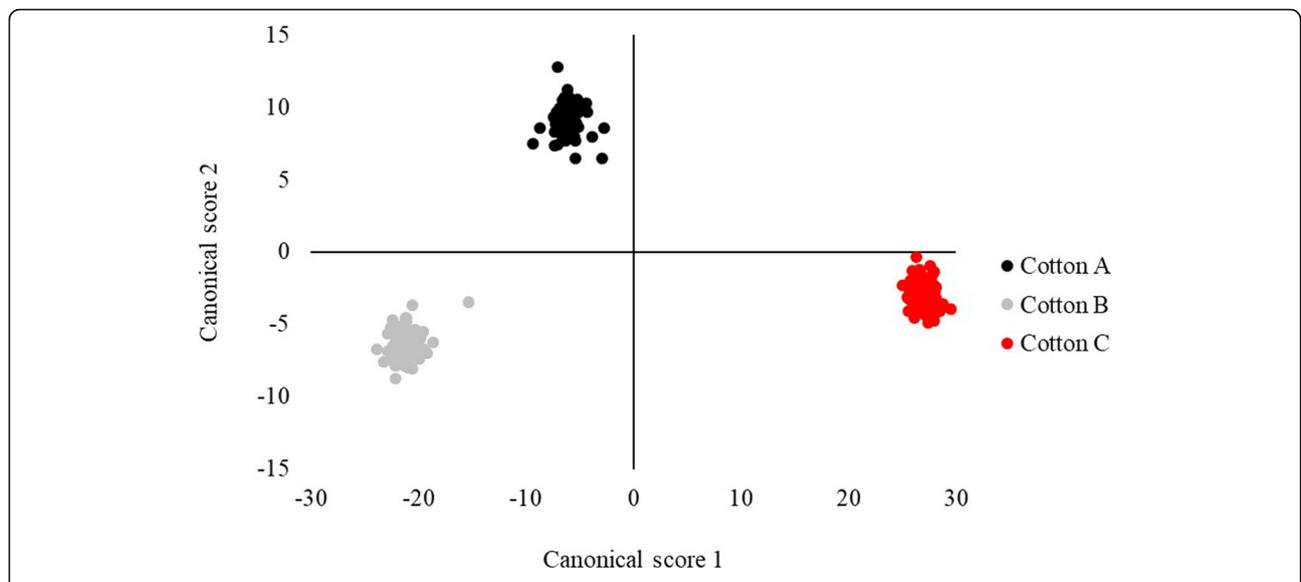


Fig. 2 Linear discriminant scores of 6 months of measurements of the length response distributions for three check cottons

Results

Stability of the length distribution measurement

The three cottons used in this section are tested every morning in the Cotton Phenomics Laboratory (CPL) to check the AFIS as part of routine laboratory management. A series of 6 months of measurements were obtained for the purposes of establishing long-term stability and determining significant differences in within-sample variation in length between the samples. A summary of the quality parameters of AFIS fiber measured over this time reveals a wide range of quality among the three samples (Table 1).

The average of the AFIS length distribution by number suggests that these samples also represent a range in within-sample length variation needed to evaluate measurement stability (Table 1). Cotton A has the largest portion of short fibers, while Cotton C has the longest staple lengths. These differences result in large shape differences amongst the three distributions (Fig. 1).

Linear discriminant analysis was then used to identify significant differences among the distributions. The number of significant axes was determined using a Wilk’s Lambda followed by an approximant F test

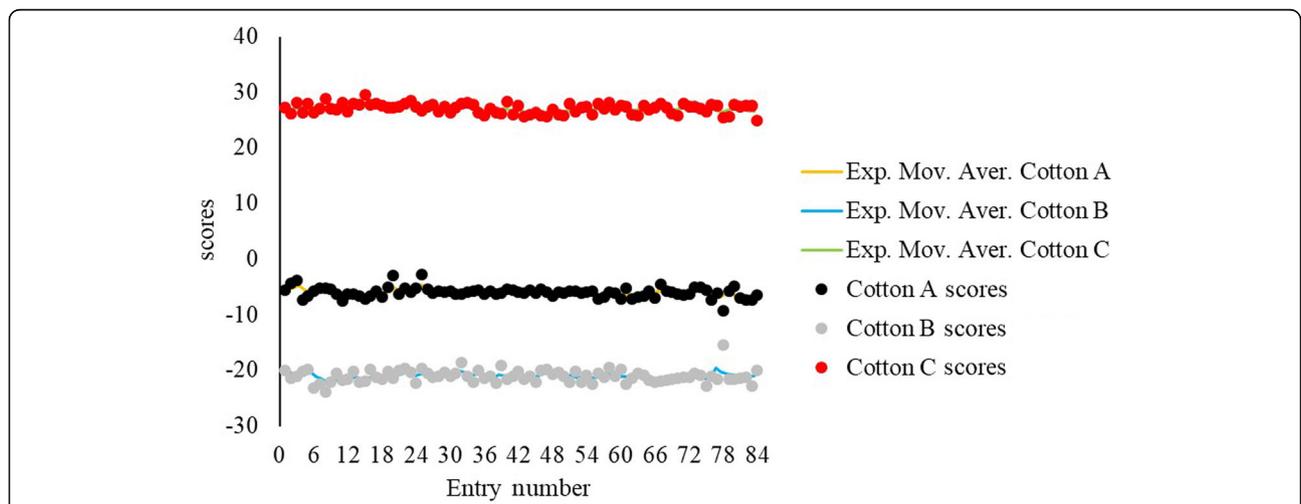
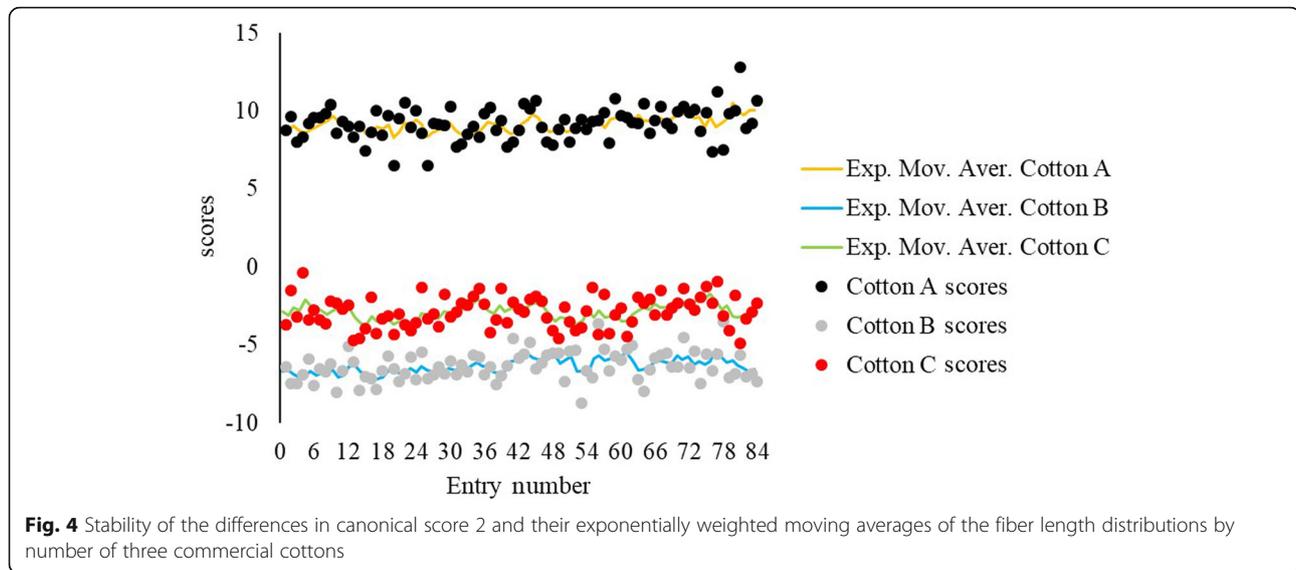


Fig. 3 Stability of the differences in canonical score 1 and their exponentially weighted moving averages of the fiber length distributions by number of three commercial cottons



(Everitt and Punnett 1991). The length distribution by number characterizes two unique ways the length varies within these three samples. It is the maximum number of ways any three samples can vary, and at least two fiber length measurements are required to identify the unique way in which these three samples contrast (Tables 2 and 3).

While the statistical test of significance suggests that two fiber length parameters are needed to characterize the differences in these samples, it does not reveal the type of difference that these two hypothetical length parameters should characterize. The nature of these differences was determined using the biplot (Fig. 2) in combination with the raw length distribution by number plot (Fig. 1).

The biplot of the canonical scores from the LDA shows that the three cottons differ in two distinct ways. Cotton C and Cotton B exhibit the largest overall difference in length (Fig. 1). Because Cotton C and Cotton B sit at the extremes of the first canonical axis and are more similar in terms of the second axis,

this suggests that the first axis is capturing an overall magnitude difference in length among the samples (Fig. 2).

Cotton A and Cotton B represent the extremes of the second canonical axis, but they also vary along the primary axis and do not well isolate the nature of the second type of variation. However, they are distinctly different in their distributional shape. The primary mode for Cotton A is shorter than Cotton B. Cotton A exhibits a distribution more closely associated with a cotton where the fibers are broken and shortened because of over-processing.

All three cotton samples separate based on the first canonical score, while Cotton A separates from the two other cottons based on the second canonical score. The clear separation in the biplot justifies using these three cottons for check samples, which represent large differences in fiber length distribution by number.

The canonical scores characterize the significant differences in length variation among the samples, and

Table 4 AFIS individual fiber quality parameters of eight cotton populations

Family	UQL (w) /cm	L(n) /cm	L(n) CV /%	SFC (n) /%	L5%(n) /cm	Fine /mTex	Mat. ratio	Hs
AB	2.57	1.88	40.73	21.61	2.95	164.3	0.85	192
CH	3.07	1.93	52.14	30.06	3.51	141.8	0.79	179
DQ	2.90	1.88	49.61	29.16	3.30	150.5	0.81	185
EL	3.30	2.11	51.03	26.50	3.76	138.5	0.82	169
FM	2.87	1.78	54.33	35.23	3.30	141.9	0.80	178
GI	2.44	1.65	47.11	34.23	2.77	159.3	0.77	207
JO	3.00	1.91	50.90	29.72	3.43	147.8	0.80	181
KN	2.74	1.91	45.38	26.12	3.12	161.1	0.81	199

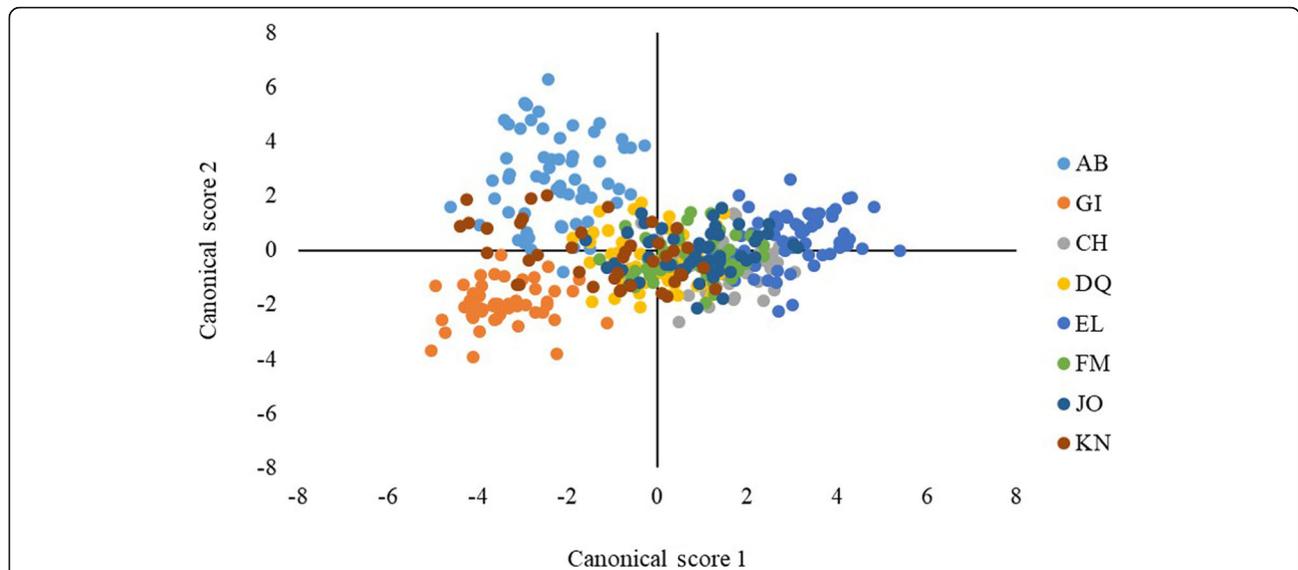


Fig. 5 Canonical plot of eight cotton populations

the stability of the length distribution by number measurement was determined by plotting these scores over time. The stability of distributional differences was based on visual comparison with the weighted moving average.

Plotting canonical score 1 for each cotton over a 6-month period along with the exponentially weighted moving average of each daily measurement shows that the measurement is stable. There are few cases in which a measurement deviates from the EWMA value, and those deviations are expected to be small when measuring variation in a naturally produced material in a laboratory environment over a long-term period (Fig. 3).

The stability plot of canonical score 2 also shows that the measurements remain stable over the 6 month testing period. The fluctuation pattern of canonical score 2 agrees with that of the score 1 (Fig. 4).

These results show that the AFIS fiber length distribution by number measurement captures significant differences among the samples, and the measurements of these differences are stable over the 6 month testing period. However, for three cotton samples we can only identify two significant sources of variation.

Characterizing length distribution differences in a diverse population

While the LDA from the previous section was able to successfully characterize two sources of variation,

Table 5 Wilk's Lambda multivariate test of significance

Test	Approx. F	NumDF	DenDF	Prob>F
Wilk's Lambda	3.5	553	2 448.4	< 0.000 1*

accessing a larger number of entries could reveal additional types of variation captured by the length distribution by number. To test this hypotheses, eight F₂ cotton families were generated from obsolete parent material. Summary statistics of the families (Table 4) show that there is a considerable range of variation for standard AFIS parameters.

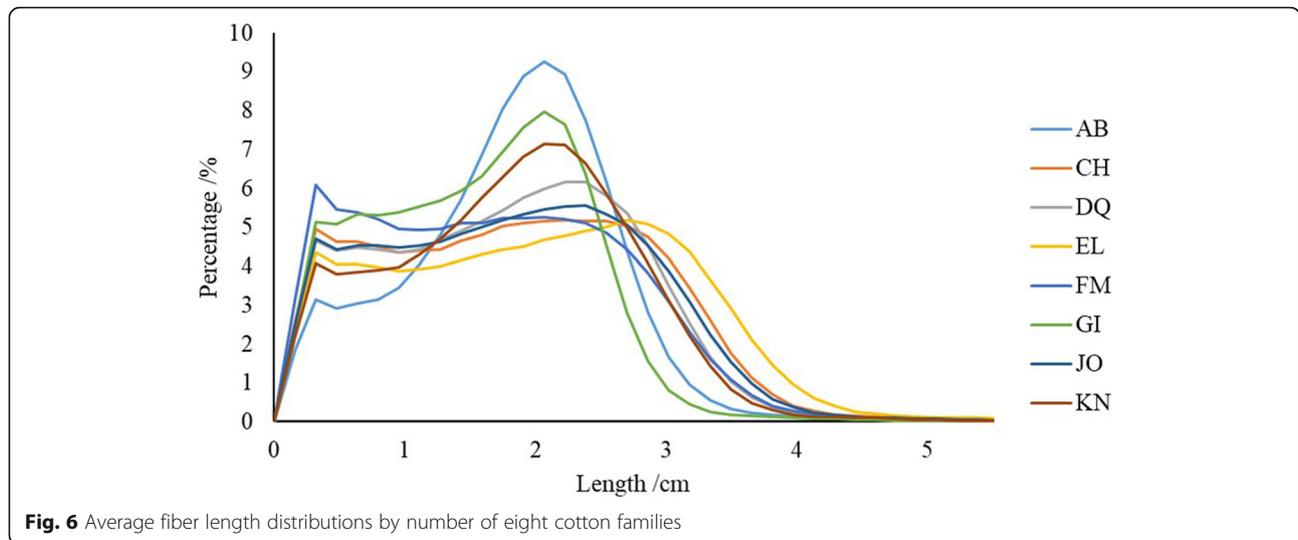
The LDA biplot illustrates the large amount of variation captured in this diverse material (Fig. 5). While it is evident that most of the samples tend to cluster among other members of the same family, there are many instances of overlap between families.

The Wilk's Lambda test shows that the differences in fiber length distributions by number are significant (Table 5). Additionally the number of significant axes in the LDA analysis reveals that there are four types of variation (Table 6). Only two length measurements, UHML and UL, are provided by HVI testing. These results suggest that two length parameters would not adequately characterize the differences in length among these samples.

The average fiber length distributions by number of the individual families illustrate the types of differences observed among the samples (Fig. 6). While

Table 6 Significant axes of variation based on Approximant F test (α = 0.05)

Eigenvalue	Percent	Approx. F	NumDF	DenDF	Prob>F
4.06	54.6	3.50	553	2 448	< 0.000 1*
1.56	20.9	2.27	468	2 104	< 0.000 1*
0.7	9.4	1.61	385	1 757	< 0.000 1*
0.48	6.5	1.28	304	1 409	0.002 0*



some of the family average fiber length distributions are relatively similar, there are also examples that exhibit extreme differences. For example, the average of family AB has a distribution pattern more closely resembling a normal distribution, albeit with an overall reduction in fiber length. Alternatively, the average of family EL is composed of fiber that are much longer but are also more variable in their lengths. Each of these average fiber length distributions could contribute both favorably and unfavorably to their performance in textile production.

To better illustrate the variation observed within and among these population, a closer examination of three entries was conducted. A separate set of three PCAs, performed on the fiber length distributions by number measured from three families, was able to characterize more than 98% of their total within-family variation using three components of variation (Table 7).

The nature of within-family variation is different than that observed among the families. While among family variation captured by the length distribution by number requires four variables to be adequately characterizes, variation within family only requires three. This is because the length within each family primarily follows a gradient. For example, PC1 explains more of the variation in populations EL and GI compared with population AB. Again, population AB has a distribution which is more peaked in comparison to the others, and this type of variation is what is captured in the second PC.

While the average fiber length distribution by number of the families clearly shows the inter-population variation, breaking the families into individual distributions shows a large amount of intra-population variation (Figs. 7, 8, and 9).

This variation included within and between family variations. The within-family variation is specific to this set of breeding families and is not necessarily represented in the commercial bales from previous sections. The within-family differences in fiber length distribution by number show a gradient of change that is capturing overall shifts in fiber length. However, between-families differences show changes in distributional shape and allude to a genetic component. The four sources of variation observed in this section show that we may need all four sources of variation if we want to better explain differences in cotton fiber length across families. These results help make the argument that using the two standard HVI fiber length parameters to characterize fiber length is inadequate to fully explain differences between cottons.

Characterizing germplasm differences with the full distribution

The additional scores required to describe the variation captured in these populations show that in some instances two length parameters are inadequate to fully characterize potential differences in germplasm. The current method of using two length parameters could lead to errors when applied to a

Table 7 Eigenvalues and explanation percentages of three components of PCA on three different populations

	PC1		PC2		PC3	
	Eigen	Percent Explained	Eigen	Percent Explained	Eigen	Percent Explained
AB	58.5	74.0	18.1	22.9	1.6	2.0
EL	68.3	86.5	8.4	10.7	1.1	1.4
GI	64.6	81.8	11.7	14.8	1.3	1.7

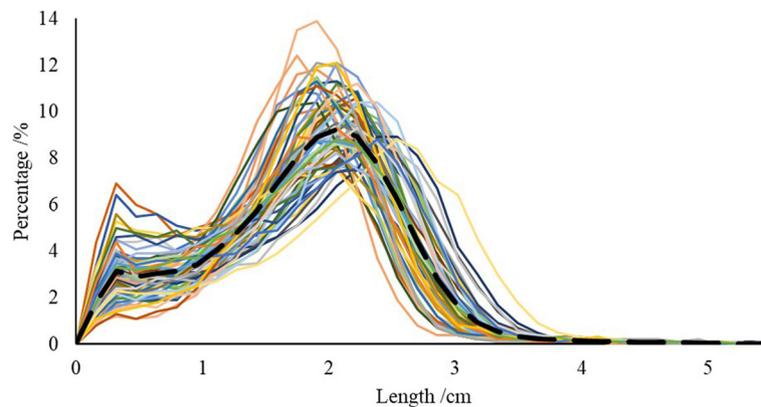


Fig. 7 Observed variation in fiber length distributions by number of individual plants within population AB

breeding program where material is often more diverse than commercial material. To test this hypothesis, LDA was applied to the populations first using the two commonly considered fiber length parameters reported with the AFIS (UQL and SFCn) then using the canonical scores generated from the AFIS fiber length distributions by number.

Table 8 summarizes the rates of misclassification of an entry back to its original family. That is, when a sample is classified as belonging to a family other than its true family, it is deemed misclassified. We would expect that misclassification rates would be low among these populations because they have no shared parentage. When the two AFIS parameters were used, individual plants were attributed to the wrong family 53% of the time. However, when the information from the AFIS length distributions were used, the percent of misclassified plants was reduced to 32%.

Examples taken from Tables 9 and 10 can highlight some of the problems with misclassification. Using

the two length parameters alone, KN is only correctly identified 5% of the time. Compare that to the alternative approach, and that value jumps to over 50%. While KN is not a high quality population which would interest a cotton breeder, a similar situation is observed in population CH in which the level of fiber quality would justify further consideration. When using the two length parameters, individual plants are correctly identified 45% of the time compared to 70% using fiber length distributions.

This could be problematic when a cotton breeder is considering which plants will be discarded and which would be carried forward in the program. For example, population CH and EL have the longest fibers among the populations tested. If forced to select one of these population to remove from their program, a breeder would eliminate population CH from their program based on average AFIS fiber length parameters. When broken down to individual plants within these two populations, entries are misidentified using

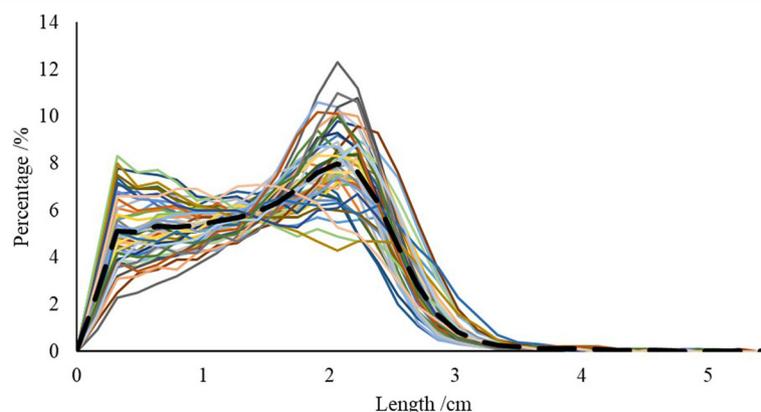
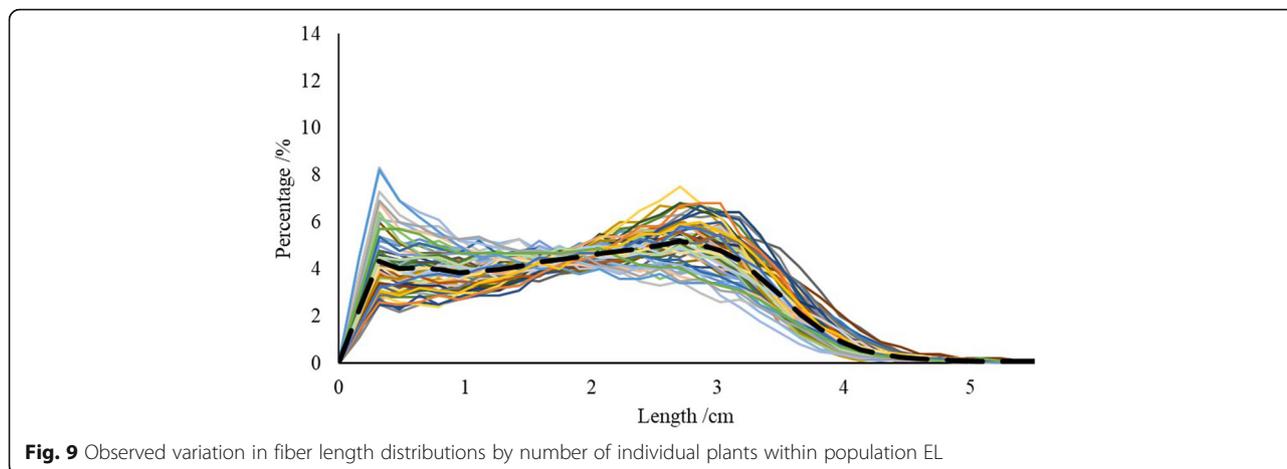


Fig. 8 Observed variation in fiber length distributions by number of individual plants within population GI



individual length parameters at more than three times the rate of using the AFIS length distribution by number. Errors in the decision-making process can result in a high cost when the long-term success of the breeding program is considered.

Conclusion

A quantitative measurement of differences in the AFIS fiber length distribution by number measured from diverse samples is stable and captures more variation than individual AFIS length parameters alone. The stability of the measurement justifies the extended application of fiber length distributions in the development of future germplasm.

HVI fiber length parameters have been used extensively to drive the development of new varieties that meet the market demands of the cotton industry. While using the industry-accepted parameters may be sufficient to develop germplasm which fits the current market, this research suggests that this is an insufficient strategy if the goal is to develop truly superior material. This analysis of the AFIS fiber length distribution shows that using only two fiber length parameters is insufficient to capture the total variation in fiber length present among the cotton samples tested.

Table 8 Summary of misclassification rates using different fiber length parameters

Parameters	Number misclassified	Percent misclassified	Entropy R-squared
UQL	255	58.6	0.30
SFC	331	76.1	0.09
UQL & SFC	231	53.1	0.37
Fiber length distribution (n)	139	32.0	0.59

The actual number of parameters needed to assess the variation in fiber length is likely population and application dependent. This would mean that depending on the types of distributions present in a population, and the objective of the experiment, the number of parameters needed to characterize the variation in fiber length distribution could vary. We showed that two parameters adequately characterized differences among the three check cottons, but four were needed to characterize the familial background of the eight breeding populations.

The application of fiber length distributions in a breeding program would lead to fewer false positive results compared with selection based on the more common method of using two length parameters. This alone would save breeders considerable time, effort, and ultimately money in their programs. If made more accessible, additional information about within-sample variation in fiber length could become a primary factor of consideration for cotton breeders aiming to produce high-quality germplasm.

Table 9 Misclassification rate by family when using UQL and SFCn measurements from AFIS

UQL(w) and SFC(n)								
Pedigree	AB	CH	DQ	EL	FM	GI	JO	KN
AB	41	0	1	0	0	7	0	11
CH	0	27	3	10	10	0	9	1
DQ	0	5	12	0	12	0	14	17
EL	0	12	0	46	1	0	1	0
FM	0	6	5	1	27	0	7	4
GI	8	0	0	0	1	38	0	3
JO	1	17	4	7	10	0	11	5
KN	11	3	7	1	3	4	9	2

Table 10 Misclassification rate by family when using the full AFIS length distribution by number

AFIS Length Distribution (n)								
Pedigree	AB	CH	DQ	EL	FM	GI	JO	KN
AB	51	0	0	0	1	1	0	7
CH	0	42	7	2	3	0	5	1
DQ	1	11	28	0	6	0	7	7
EL	0	4	0	50	3	0	3	0
FM	0	1	4	1	38	0	6	0
GI	0	0	1	0	0	47	0	2
JO	0	15	10	3	9	0	17	1
KN	6	2	4	0	0	2	3	23

While the use of AFIS fiber length distributions in cotton breeding programs will continue to be limited because of the slower testing speed compared with that of the HVI, this study shows that a more detailed measurement of within-sample variation in fiber length would be beneficial. The AFIS has been shown to capture valuable within-sample fiber length variation, but any method capable of evaluating this type of variation could be of benefit to the future development of cotton varieties.

Abbreviations

AFIS: Advanced Fiber Information System; CPL: Cotton phenomics lab; HVI: High Volume Instrument; LDA: Linear discriminant analysis; SFCn: Short fiber content by number; UI: Uniformity index; UHML: Upper half mean length; UQL: Upper quartile length

Acknowledgments

The author is grateful to Cotton Incorporated and the Cotton Incorporated Fellowship Program for funding this research.

Authors' contributions

Hinds Z, Hequet E, and Kelly B. conceived, designed, and performed the experiments. Hinds Z and Kelly B analyzed the data. Hequet E contributed analysis tools and support. Hinds Z and Kelly B wrote the paper. All authors read and approved the final manuscript.

Funding

Funds for this research were provided by Cotton Incorporated and the Cotton Incorporated Fellowship Program project number 17–539.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 16 January 2020 Accepted: 30 March 2020

Published online: 19 April 2020

References

- Basra AS. Cotton fibers: developmental biology, quality improvement, and textile processing. Binghamton: The Haworth Press; 1999.
- Chu YT, Riley CR. New interpretation of the fibrogram. *Text Res J.* 1997;67(12): 897–901. <https://doi.org/10.1177/004051759706701206>.
- Čisar P, Čisar SM. Optimization methods of EWMA statistics. *Acta Polytech Hung.* 2011;8(5):73–87.
- El Mogahzy YE, Broughton R, Lynch WK. A statistical approach for determining the technological value of cotton using HVI fiber properties. *Text Res J.* 1990; 60:495–0. <https://doi.org/10.1177/004051759006000901>.
- Everitt BS, Punn G. *Multivariate Data Analysis*. London: Wiley; 1991. p. 219–220.
- Faulkner WB, Wanjura JD, Hequet EF, et al. Evaluation of modern cotton harvest systems on irrigated cotton: yarn quality. *Appl Eng Agric.* 2011;27:523–32. <https://doi.org/10.13031/2013.38199>.
- Hughs SE, Armijo CB, Foulk JA. Upland fiber changes due to ginning and lint cleaning. *J Cotton Sci.* 2013;17(2):115–24.
- Hunter JS. The exponentially weighted moving average. *J Qual Technol.* 1986; 18(4):203–10.
- Kelly B, Abidi N, Ethridge D, Hequet EF. *Fiber to fabric*. Cotton. Madison: American Society of Agronomy Publications; 2015. .
- Kelly BR, Hequet EF. Variation in the advanced fiber information system cotton fiber length-by-number distribution captured by high volume instrument fiber length parameters. *Text Res J.* 2017;88(7):754–65. <https://doi.org/10.1177/0040517516688628>.
- Kelly CM, Hequet EF, Dever JK. Breeding for improved yarn quality: modifying fiber length distribution. *Ind Crop Prod.* 2013;42:386–96. <https://doi.org/10.1016/j.indcrop.2012.06.018>.
- Koo HJ, Suh MW. Maximizing yarn and fabric strength through variance of HVI elongation. *Text Res J.* 1999;69(6):447–56. <https://doi.org/10.1177/004051759906900609>.
- Krifa M. Fiber length distribution in cotton processing: dominant features and interaction effects. *Text Res J.* 2006;76(5):426–35. <https://doi.org/10.1177/0040517506062616>.
- Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics.* 1990;32:1–29. <https://doi.org/10.2307/1269835>.
- Mangialardi GI. Multiple lint cleaning: its effect on bale value, fiber quality, and waste composition. *Bulle: USDA-ARS Tech;* 1972. p.1456. <https://doi.org/10.22004/ag.econ.171905>.
- Meredith WR. Minimum number of genes controlling cotton fiber strength in a backcross population. *Crop Sci.* 2005;45:1114–9. <https://doi.org/10.2135/cropsci2003.0425>.
- Paudel DR, Hequet EF, Abidi N. Evaluation of cotton fiber maturity measurements. *Ind Crops Prod.* 2013;45:435–41. <https://doi.org/10.1016/j.indcrop.2012.12.041>.
- Stewart JM. Fiber initiation on the cotton ovule (*Gossypium hirsutum*). *Am J Bot.* 1975;62(7):723–30. <https://doi.org/10.1002/j.1537-2197.1975.tb14105.x>.
- Wakeham H. Cotton fibre length distribution – an important quality factor. *Text Res J.* 1955;25:422–9. <https://doi.org/10.1177/004051755502500509>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

