The Thesis Committee for Xiaowen Guo
Certifies that this is the approved version of the following thesis:

Assessing Cotton Maturity from Fiber Cross-section Measurements

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Bugao Xu

Mourad Krifa

Assessing Cotton Maturity from Fiber Cross-section Measurements

by

Xiaowen Guo, B.E.; M.E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Textile and Apparel Technology

The University of Texas at Austin

December 2013

Abstract

Assessing Cotton Maturity from Fiber Cross-section Measurements

Xiaowen Guo, M.S.T.A.T.

The University of Texas at Austin, 2013

Supervisor: Bugao Xu

The previous Fiber Image Analysis System (FIAS-I) is not reliable enough to detect fibers, especially for the immature fibers. It yields a systematic bias in the maturity distribution. Furthermore, the maturity distributions are often assumed to be normal without any normality tests in many previous studies, and those distributions are commonly measured by a sole parameter, e.g., the mean maturity value. In fact, those statistical inferences on cotton maturity may not be valid when cotton maturity does not follow a normal distribution. In light of the complexity of maturity distributions, the sole-parameter approach does not appear to be reliable and rational to rank the maturity among different samples.

In this thesis, modified algorithms are made in the previous Fiber Image Analysis System (FIAS-I) to improve the number and accuracy of detected cross-sections and

reduce the bias on immature fiber. The normality of cotton maturity distributions are analyzed through multiple parameters and patterns of cotton maturity distributions, and the experimental results on the cross section images selected from seven cotton varieties are displayed. Finally, several normality tests are introduced, and the Box-Cox transformation is applied to the maturity distribution, which makes the comparisons among the mean maturity feasible.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 MOTIVATION AND GOALS

Cotton cross sectional analysis provides fundamental and measurable information (such as cell wall thickness, area and perimeter, etc.) which are directly related to the maturity of cottons. Accurate and unbiased cross-section data are crucial, because it can be regarded as the reference for other indirect methods. In 1995, USDA Southern Regional Research Center (SSRC) published a quick embedding method [1] of cross-sectioning fiber samples sponsored by Cotton, Inc. In 2004, the cross-section images of over 100 cotton bales were analyzed by the first version of Fiber Image Analysis System (FIAS-I) developed at the University of Texas at Austin [2].

The accuracy of the FIAS-I analysis was challenged in a study by Krifa and Padmaraj [3] on 14 cotton samples with 3 cross-sections and approximately 2800 fibers per sample. The study showed an overestimation of maturity of nearly 10% due to 10-40% of undetected or eliminated immature fibers [3]. The FIAS-I contains several unsolved problems, which cause the inexactness of cotton maturity level. Even though some supplementary image editing tools were added to the FIAS to assist the operator to re-portray (manual-editing) or delete (manual-removal) wrongly detected fibers, their weaknesses limit the further use of those manual tools. The major goal of this thesis is to develop a research on the modifications of the FIAS-I to improve not only the number but also the accuracy of detected cross-sections and reduce the bias on immature fibers.

1.2 BACKGROUND OF COTTON CROSS-SECTIONS

1.2.1 Cross Section

The cross-sectional image of a cotton fiber is displayed in Figure 1.1. It consists of two parts: cell wall and lumen. The enclosed contour located in the middle of the whole cross-section is the lumen, and the part that envelops the lumen is the fiber's cell wall.

The measurements of a cross-section can be expressed using four parameters: the perimeter of the cross-section (P_c), the area of the cross-section (A_c), the perimeter of the lumen (P_l), and the area of the lumen (A_l). The values of the two perimeters are obtained by tracing their boundaries, and the areas are calculated by counting the pixels within their corresponding boundaries.

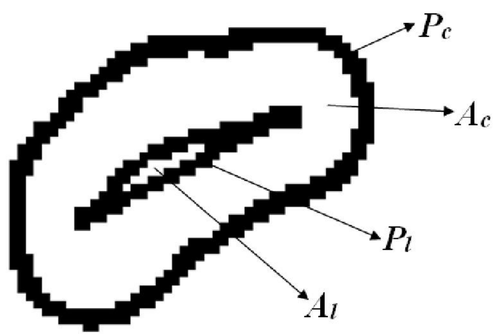


Figure 1.1 Measurements of the cross-section.

1.2.2 Cotton Maturity

In the cross-sectional analysis, the maturity of individual fibers is often represented by a shape factor called circularity or θ value. θ is the ratio of the cell wall area (A_w) over the area of a circle having the same perimeter (P_c).

$$\theta = 4\pi A_w / P_c^2 \tag{1.1}$$

in which,

$$A_w = A_c - A_l \tag{1.2}$$

Thus, increasing accuracy of locating fiber boundaries and lumens is paramount for obtaining reliable maturity distributions.

1.3 Traditional Methods for Cotton Maturity

The overall development of cotton maturity measurement can be briefly described by a few milestones. In 1956, Lord established 100 cottons to calibrate micronaire [4]. In 1980's, the ITMF in cooperation with other organizations developed a set of nine calibration cottons for use in large-scale maturity round trials [5]. In 1999, Thibodeaux et al. used a commercial image analysis system to measure the cross-sectional images of approximately 50 cotton varieties representing a wide range of genetic finenesses grown and hand-harvested in the U.S. [6]. In 2000s, Hequet and Thibodeaux's groups performed a multi-year project to create a larger-scale cotton maturity reference by analyzing images of 104 cotton bales [7] collected worldwide using a new fiber cross-sectioning protocol [1] and the customized software (FIAS) [2].

Acquiring the maturity information through the analysis of longitudinal direction of cotton fiber was commonly used several years ago. In 1986, Thibodeaux and Evans [8] measured the fiber maturity by a longitudinal width ratio value, and they found that the average maximum widths along the length direction to the average minimum widths could reflect the secondary wall of the cotton fiber. A longitudinal analysis for fibers' fineness and maturity was reported by Xu and Huang [9]. Firstly, a valid scan of longitudinal shape should be obtained, then maturity information is derived through the analysis of the fibers convolutions. The reliability of the fiber maturity information in [9] was ensured through comparison with other methods. The maturity of cotton fiber can also be measured through automated polarized light microscopy [10][11] applied onto the longitudinal direction of test fibers. Through the color difference of "yellow" and "green", the thickness of fiber wall can be obtained, and the maturity can be captured from it.

Compared with the longitudinal direction, the cross-section direction [12] could provide much more direct and accurate maturity information. However preparing the cross-section fiber slides [13] without scratches and shape distortions seems to be a challenge. Due to the consistence of longitudinal and cross-sectioning directions, Petkar *et al*. [14] and Barker *et al*. [15] investigated reference methods, from which cross-section information can be projected from the longitudinal viewer. The initial attempt of direct cross-sectioning measurement [16] used image analysis system and computer program to get the parameters from cross sections, which covers wall thickness, ribbon width, maturity factor and maturity ratio. A detailed image processing for cotton fiber's cross-sectioning analysis can be obtained in [12]. Fourier transform infrared spectroscopy [17] was utilized for the discrimination of cotton maturity due to the large difference of spectral features in cross-sectional analysis.

1.4 STRUCTURES OF THE THESIS

This thesis will cover four parts. From Chapter 2 to Chapter 5, the modified algorithms of the FIAS-I will be introduced, and data analysis between the FIAS-I and FIAS-II software will be compared to demonstrate the improvements of the performance.

Chapter 2: the major image processing routines of the FIAS-I will be introduced firstly, and then followed by several modified algorithms applied to it, which is improved as the FIAS-II software.

Chapter 3: maturity distributions obtained from the FIAS results are discussed. If the distribution is not normal, the characteristics, classifications and patterns of the distribution are analyzed.

Chapter 4: results from the previous (FIAS-I) and current (FIAS-II) software are compared through the aspects of: detected numbers of fibers, distribution parameters, distribution classifications, and distribution patterns.

Chapter 5: the normality of each distribution should be tested. In order to allow the comparisons among maturity means from non-normal distributions, the Box-Cox transformation is complemented to make such comparisons feasible among the transformed mean values.

Chapter 6: conclusions and future work for this study are presented.

Chapter 2: Modifications of the FIAS System

2.1 INTRODUCTION

The major processing routines of the FIAS-I software can be described in four steps: dynamic thresholding, background flooding, skeletonizing and lumen identification. These routines appear to be effective when image illumination is uniform and image contrast is sufficient, but not robust enough for fibers that have low contrast, broken edges, or distorted cross-sections due to self rolling and tangling with others. Modifications of the FIAS-I software are necessary.

2.2 MODIFICATIONS IN THE FIAS-II

2.2.1 Adaptive Thresholding

Illumination intensity in a captured 8-bit grayscale image can vary drastically. There is no omnipotent thresholding method that can generate consistent results across the entire image. The dynamic threshold method with adaptive parameters is a reasonable way to convert an 8-bit gray scaled image into a binary one. In the FIAS-I, a sub-window of a given size was used to calculate the threshold (T_i) based on the mean M_i and the standard deviation SD_i of pixel intensities in the window:

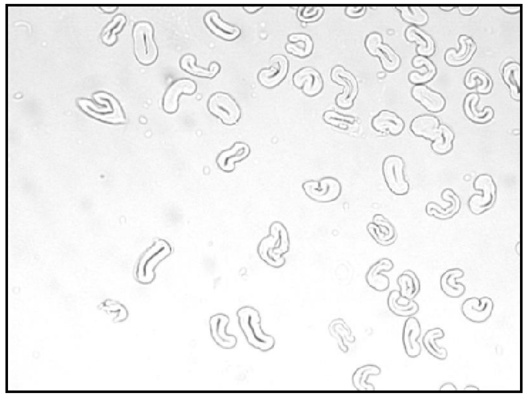
$$T_i = M_i - c \times SD_i \qquad (2.1)$$

where c was a coefficient which was set as 0.2 in the program.

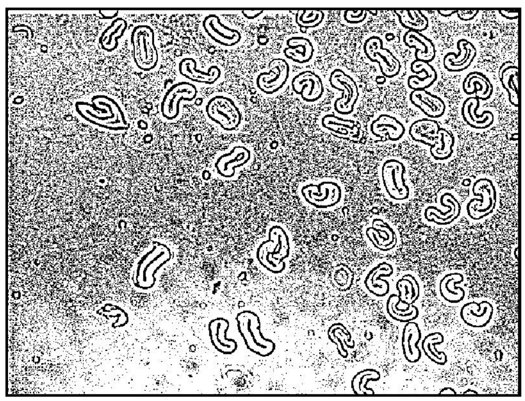
The window size was also fixed at 7×7 pixels. Since c and the window size were not changeable despite differences in intensity variability, T_i calculated using Eq. 2.1 might not always be optimal for generating binary images. Figure 2.1a displays an image with non-uniform illumination, and Figure 2.1b is the resultant image with this threshold method. The high level of noise in Figure 2.1b makes the tracing of fiber boundaries extremely difficult. It seems that in Figure 2.1a darker regions are more sensitive to background noise than brighter regions. The coefficient c should be adaptive to the mean intensity (M_i) of the current window to adjust the sensitivity of thresholding. Based on the observations on many images, it is found that a brighter region normally has a lower noise level than a darker region, and thus c for a brighter window should be increased to lower T_i. The location-dependent c is denoted as c_i. By the trial and error with a wide range of images, we established the following empirical equation for calculating c_i based on M_i:

$$c_i = 0.014M_i - 2.694\,(R^2=0.9857) \tag{2.2}$$

The binary images of Figure 2.1a with the aid of dynamic coefficient c_i is displayed in Figure 2.1c, in which the large proportion of the noisy image (Figure 2.1b) was reduced. Fiber boundaries in this image were also more complete than those in Figure 2.1b.



(a) The original input image.



(b) The previous binary output.

(c) The improved binary output.

Figure 2.1 Adaptive thresholding comparison.

2.2.2 Amending Broken Edges

After the grayscale images are converted into the binary images using the adaptive threshold, background noise, small solid objects including broken edges of fibers need to be removed to let only enclosed fibers remained. This can be done by "Background Flooding", which fills the white background with black pixels. Broken fibers are also absorbed by the "flooding" when their inner white regions (fiber walls) are occupied by black pixels. In order to keep those narrowly broken fibers from being eliminated, a step to amend almost connected edges needs to be added.

There is no generic algorithm for detecting the ends of broken boundaries. The edge amending routine had to be developed on a case-by-case basis after fiber outer boundaries were traced. It mainly involved three steps: (1) Check if a fiber boundary has a "dead-end" pixel by counting the number of its neighboring pixels. A dead-end pixel is the one that has one neighbor in the boundary chain. (2) Search for another "dead end" in the same chain. (3) Check if the open distance between the two is within an allowable limit and the boundary length between the two exceeds the threshold on fiber perimeter. The last step was to prevent a connection between two dead-ends that were not on the same boundary or too far apart. The connection distance was limited to 5 pixels in the new program, and was simply made by drawing a straight line between the two identified dead ends.

Figure 2.2a shows a few examples of typical broken boundaries, and Figure 2.2b shows the connections made. The connection lines don't alter fibers' original shapes, while preventing these fibers from being immersed with the background flooding.

(a) Cottons with broken boundaries.

(b) Amended connections.

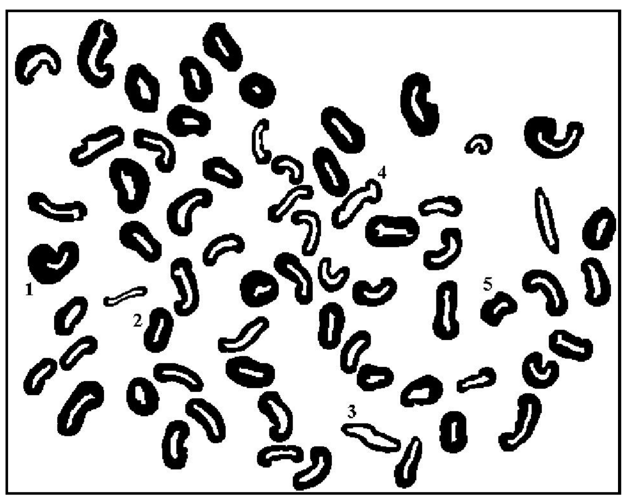
Figure 2.2 Broken boundaries and their amended connections.

2.2.3 Skeletons and Lumens Identification

In a cross-section image, some fibers often do not possess visible lumens. As seen in Figure 2.3a, fibers 1, 3 and 4 are immature fibers whose lumens are totally collapsed, fibers 2, 8 and 5 do not exhibit clear lumens because of low contrasts, and fibers 6, 9 and 7 show partial lumens. Lumen areas in these fibers can be readily overestimated or underestimated.

Once a fiber boundary is located, another threshold is calculated by using only the pixels within the boundary so that the lumen can be more precisely segmented with localized parameters (the mean and standard deviation of the pixel intensities). Since a lumen should be situated in the fiber center, the skeleton of the fiber, i.e., the middle axis of the fiber, can be used to locate the lumen if multiple black areas are present inside the boundary. For fibers whose lumens are not detected after the thresholding, their skeletons (one-pixel thick line segment) are placed as lumens. This localized thresholding and skeletonizing enhance the accuracy of lumen size and location (Figure 3b).

(a) The original input image.

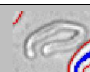
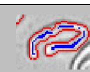
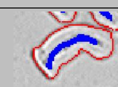
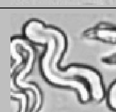
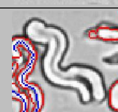
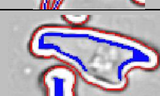
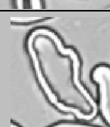
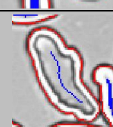
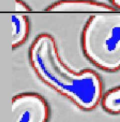
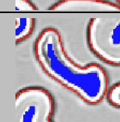
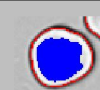


(b) The detected cross sections.

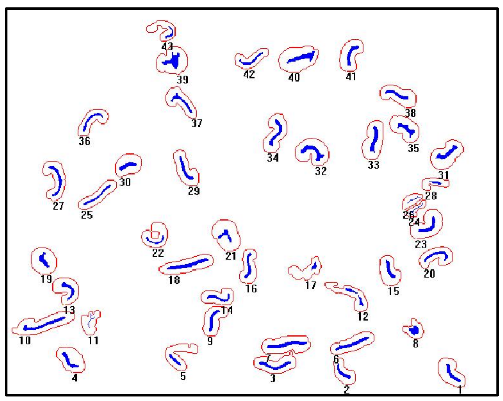
Figure 2.3 Lumen identification.

Table 2.1 displays the examples of improved results in fiber detection using the current modification methods (FIAS-II) compared to those of the previous methods (FIAS-I). Figure 2.4 gives a comparison of a whole image when being processed with these two methods.

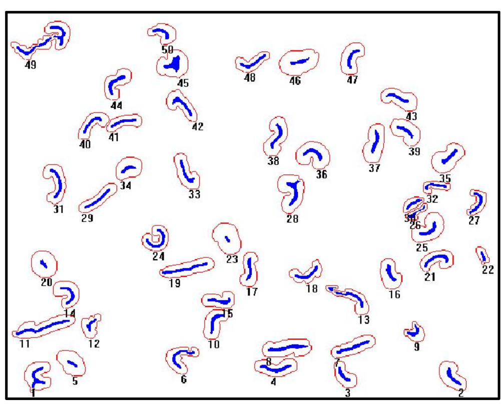
Table 2.1 Comparison of fiber detection results

	Original	Previous	Current
Missed fiber			
			
			
Underestimated lumen			
			
			
Overestimated lumen			

(a) The original input image



(b) Fiber detection of the FIAS-I.



(c) Fiber detection of the FIAS-II.

Figure 2.4 An overall comparison of fiber detection.

2.3 CONCLUSION AND DISCUSSION

The background flooding and skeletonizing steps are more standardized procedures with fewer adjustable parameters. Therefore, we focused on improving dynamic thresholding and lumen identification, which are responsible for locating accurate fiber boundaries and lumens. An edge-amending routine was also added to connect narrowly broken boundaries to prevent them from disappearing in the background flooding. The modified methods in the FIAS-II are able to detect more fibers (mainly immature fibers), and more accurately locate lumens than those in the previous system.

Chapter 3: Maturity Distributions

3.1 INTRODUCTION

Maturity distributions are often assumed to be normal without any normality tests, so the maturity level of a sample is indicated in almost all maturity testing devices (e.g., AFIS, HVI) with a single parameter, i.e., the mean maturity value of all tested fibers in the sample. In fact, many statistical inferences on cotton maturity may not be valid when cotton maturity does not follow a normal distribution.

In light of complexity of non-normal maturity distributions, the sole-parameter approach does not appear to be reliable and rational to rank the maturity among different samples. Multi-parameter and other discrimination methods should be taken into consideration.

3.2 CHARACTERISTICS OF MATURITY DISTRIBUTIONS

After a sufficient number of fiber cross sections are measured from multiple images, the descriptive statistics of fiber maturity data (θ), including mean, standard deviation, skewness and kurtosis, can be used to examine the features of maturity distribution without normality.

3.2.1 Mean and Standard Deviation

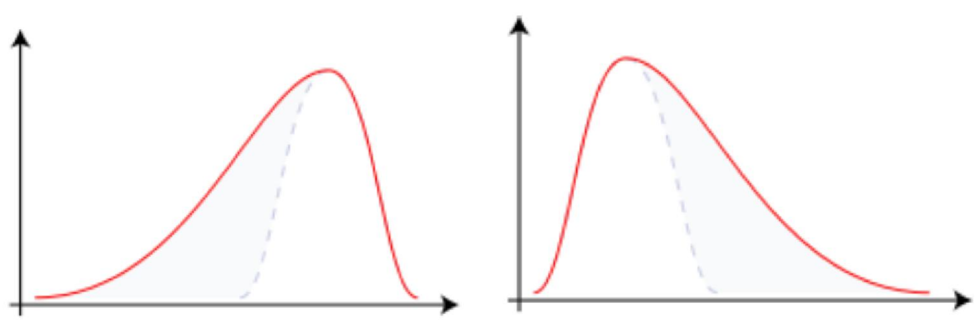
For a unimodal distribution, mean and stand deviation (*SD*) describe the central tendency and the variability of the distribution.

Mean (arithmetic average) is calculated by the sum of the observations and then divided by the number of observations. It is the most common way to describe the central tendency when compared with mode and median values.

Standard deviation (*SD*) or variance indicates how the observations deviate from the mean value. Standard deviation is the positive square root of the variance, and it is taken as "the distance of each observation from the mean, square that distance, find the average of those squares, and take its positive square root." [18]

3.2.2 Skewness and Kurtosis

Skewness indicates the "symmetry" of the distribution. According to the imbalance in a frequency distribution, skewness can be negative, positive or zero. A negative skew indicates that the distribution has a long tail on the left side (Figure 3.1a), i.e., the data concentrate more on the right side of the mean. A positive skew corresponds to the opposite case with a right-sided long tail (Figure 3.1b). A zero skew indicates that the tails on both sides of the mean balance out, which attributes to the symmetry or the even-out asymmetries.



(a) Negative skew. (b) Positive skew.

Figure 3.1 Skewed distributions [19].

Kurtosis indicates the “peakedness” of the distribution. A high kurtosis corresponds to a distribution with a sharp peak and long tails, while a low kurtosis indicates a distribution with a round peak and short, thinner tails.

When either skewness or kurtosis significantly deviates from the value corresponding to the normal distribution, the distribution loses the normality. For a non-normal distribution, many classical statistical tests, such as *t* tests, *F* tests and *chi-squared* tests may not be applied.

3.2.3 Distribution Example

Figure 3.2 displays the maturity (θ) distributions of two cotton samples (sample codes are “3044” and “3055”), in which the *x*-axis represents the fiber maturity, and the *y*-axis represents the percentage of total detected fibers.

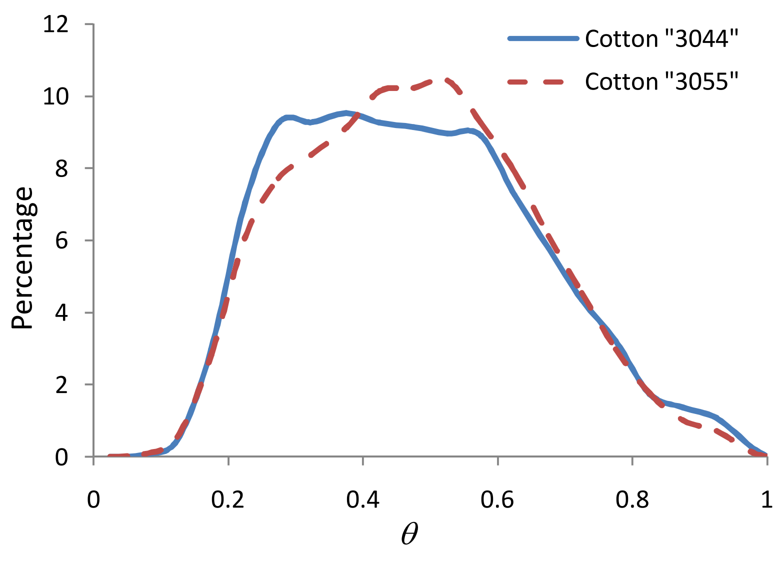


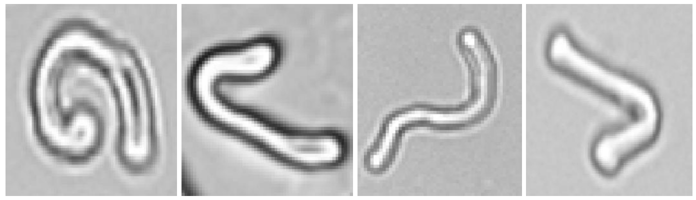
Figure 3.2 Maturity Distributions of Cottons “3044”and “3055”.

The θ of cotton “3044” and cotton “3055” have almost the same mean value (around 0.48), but different skewness values (0.37 and 0.23, respectively). Obviously, one cannot tend to use the mean θ to differentiate the maturity levels of these two cottons, which have distinctive θ distributions shown in Figure 3.2. Cotton “3044” contains more immature fibers than cotton “3055” although they have almost the same mean θ. Thus, distributional information must be taken into account when comparing or ranking maturities of different cottons.

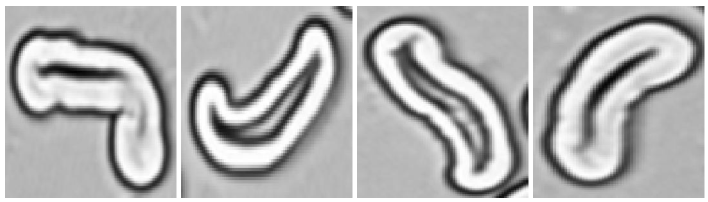
3.3 CLASSIFICATIONS OF MATURITY DISTRIBUTIONS

In order to well describe the distribution of fibers’ maturity in a comprehensive but precise manner, the maturities of fibers from each cotton variety can be classified into three representative levels according to the maturation standards. Dead fibers, as shown in Figure 3.3(a), are extremely narrow strips whose lumens are tender or even haven’t been grown out with the maturity values less than 0.3. Half-mature fibers are those growing to a certain degree and having apparent fiber walls and lumens inside. The maturity value of half-mature fiber covers the range from 0.3 to 0.6, as displayed in Figure 3.3(b). Mature fibers have the circular shapes, with fiber walls significantly greater than lumens (Figure 3.3(c)), and maturity higher than 0.6.

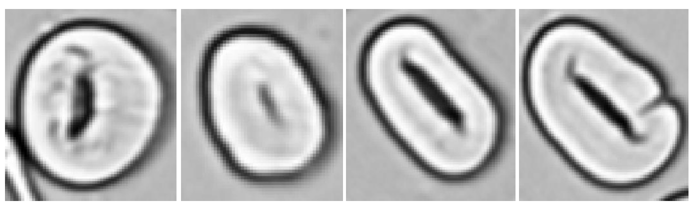
Immature fibers, especially dead fibers, have thinner walls, and thus are easier to be scratched/shredded by the cutting blade. They are also likely to be folded up transversely when the lumens collapse, increasing the difficulty of edge detection.



(a) Dead fibers.



(b) Half-mature fibers.



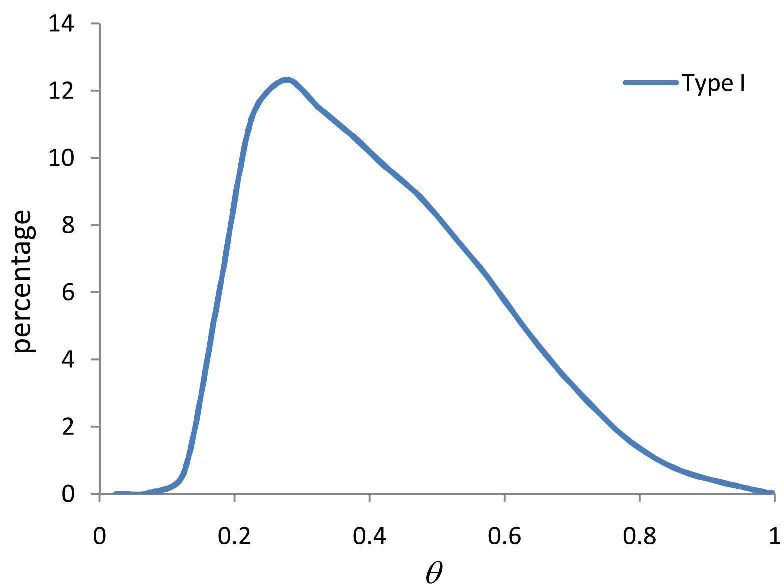
(c) Mature fibers.

Figure 3.3 Classifications of fibers.

3.4 PATTERNS OF MATURITY DISTRIBUTIONS

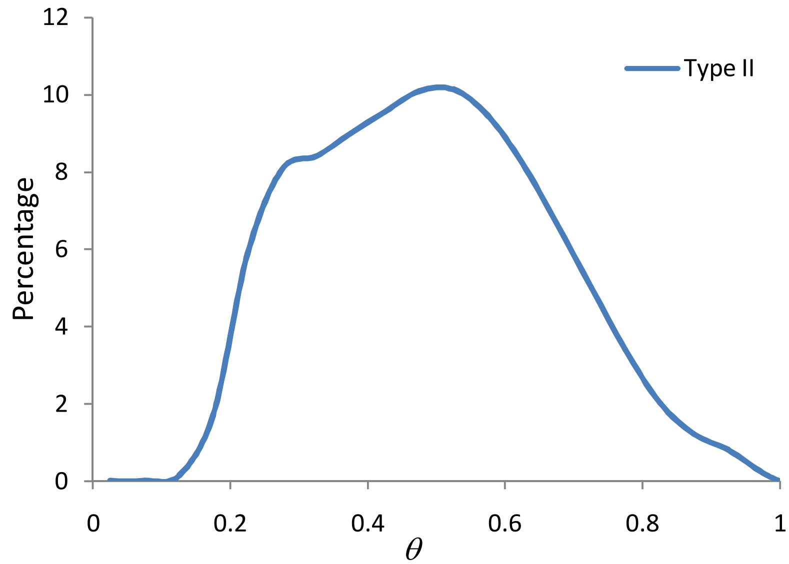
The main purpose of the Fiber Image Analysis System is to provide the most direct information to quantify the maturity extents of cottons. Except for the classification of maturity mentioned above to describe different proportions of fiber maturity within a cotton variety, maturity distribution pattern is another descriptor that clusters the maturity distributions into several major types.

Among the parameters used to depict the shape of distribution, skewness is the first principal component that can distinguish the distribution patterns notably. Four patterns can be defined to categorize maturity distributions based on skewness values. Type I: Skewness>0.3, the remarkably positive (leftward) skewed pattern reflects a large number of dead or immature fibers are detected. Type II: 0.1<Skewness<0.3, the slightly positive (leftward) skewed pattern denotes moderately more dead or immature fibers. Type III: -0.1<Skewness<0.1, the near zero skewness value indicates a symmetrical distribution. Type IV: -0.3<Skewness<-0.1, the slightly negative (rightward) skewed pattern indicates moderately more mature fibers exist. Figure 3.4 demonstrates these four pattern types. The *x*-axis represents the fiber maturity, and *y*-axis represents the percentage of total detected fibers. The detailed descriptions of these four types are listed in Table 3.1.

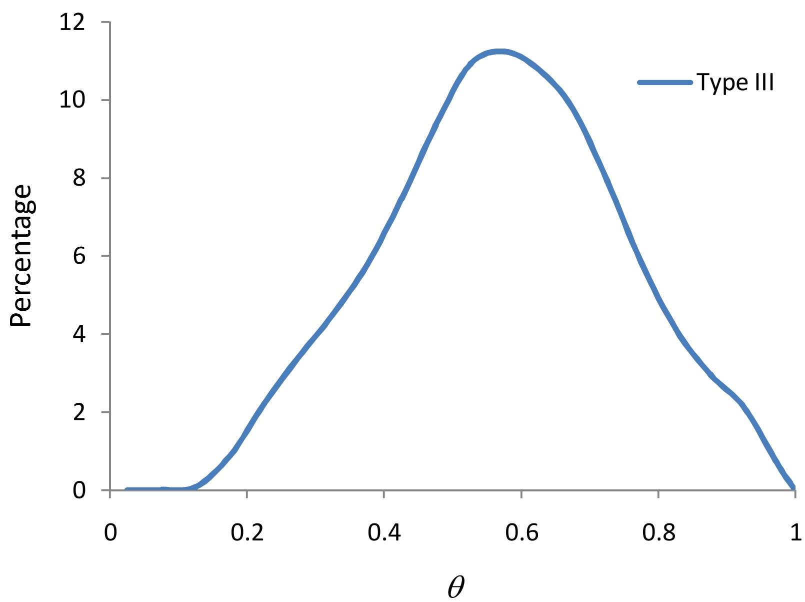


(a) Type I.

Figure 3.4 Patterns of maturity.

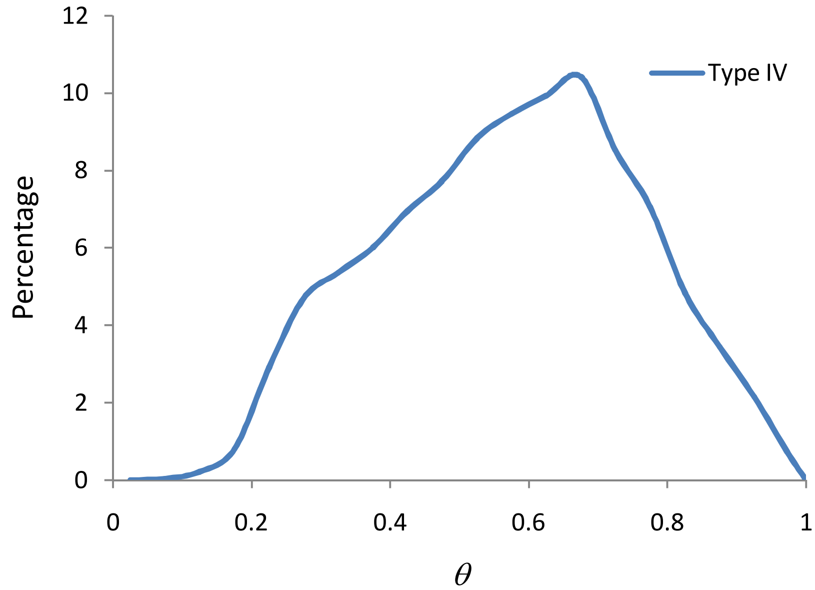


(b)Type II.

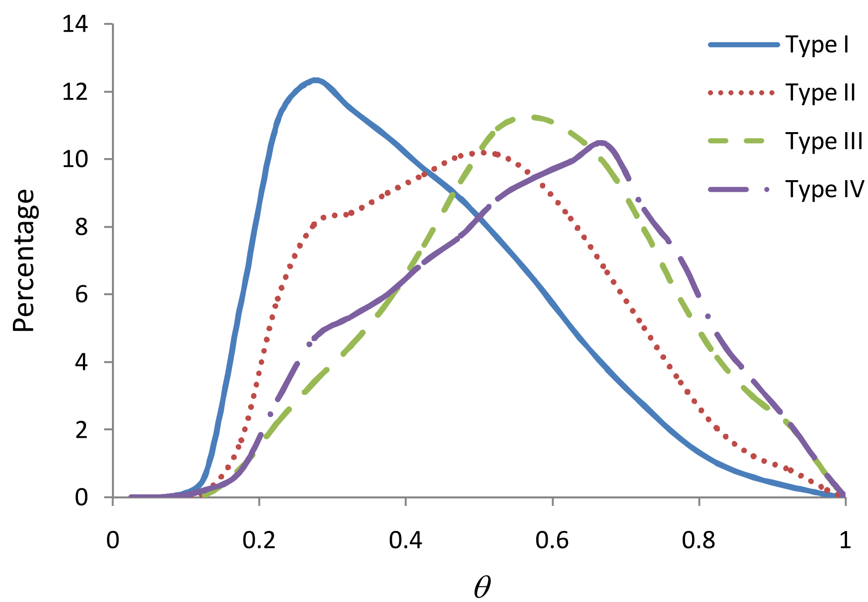


(c) Type III.

Figure 3.4, cont.



(d)Type IV.



(e) Four maturity types.

Figure 3.4, cont.

Table 3.1 Maturity Patterns

Type	Skewness (S_θ)	Characteristics	Fiber Content
I	$0.3 \leq S_\theta$	Severely leftward skewed	Many dead fiber
II	$0.1 \leq S_\theta < 0.3$	Moderately leftward skewed	Moderately leftward skewed
III	$-0.1 \leq S_\theta < 0.1$	Approximately normal	Approximately normal
IV	$-0.3 \leq S_\theta < -0.1$	Moderately rightward skewed	Moderately rightward skewed

3.5 Conclusion and Discussion

When a distribution is strongly skewed, statistic inference with a single parameter on maturity becomes invalid. Cotton maturity distributions should be described by multiple parameters, like: mean, standard deviation, skewness, and kurtosis.

Most cotton varieties exhibit positively skewed distributions, indicating more immature fiber contents than the mature ones. According to the maturation standards, the distribution can be classified into three representative levels: dead, half-mature and mature. Relying on the value of skewness, the distributions can be classified into 4 maturity patterns.

Chapter 4: Data Analysis

4.1 INTRODUCTION

The images of fiber cross sections used for the experiment were provided by the Fiber and Biopolymer Research Institute (FBRI) of Texas Tech University. A total of 15473 images from seven cotton varieties were processed with the previous and current FIAS software to examine the differences in the number of detected fibers, the maturity correlation with the data obtained from the Advanced Fiber Information System (AFIS) and High Volume Instrument (HVI), and the distributional parameters. Each of the seven varieties was assigned a unique 4-digit symbol (e.g., 2996, 2999…).

Besides the basically automated function in FIAS, it also possesses several supplementary functions. Manual editing is an optional tool for users to re-portray the unclear boundary of cross sections by their hands, and achieving the most exact results. Manual removal is another option for users to eliminate non-satisfactory fiber results. When the mouse is double clicked on a fiber, the cross section will be marked and the measurement will be deleted from the output. It saves much time with the cost of lost accuracy and number reduction.

4.2 DETECTED NUMBERS

The maturity measurements from the previous and current FIAS analyses are denoted as PA and CA, respectively. The PA data after the manual removal is symbolized as PAM. The PAM data used in this research were provided by FBRI.

Table 4.1 displays the numbers of detected fibers from CA, PA and PAM among seven cotton varieties. The “Difference Percentage” column is calculated by the number difference between CA and PA (or PAM), then over the number of PA (or PAM).

The difference percentages of Table 4.1 are varied from -0.58% to -5.61% and -145.96% to -257.77% in PA and PAM individually. Such number variations can be interpreted into two ways: 1) High accuracy in CA. Because of the automated deleting function owned by CA, far more correct cross sections are detected by CA compared with both PA and PAM. 2) High error rate in PA. A large number of wrongly detected cross sections exist in PA results, and they are eliminated manually in PAM, which leads to huge number reductions in PAM.

Table 4.1 Number Comparisons

Variety	CA	PA		PAM	
	Number	Number	Difference Percentage (%)	Number	Difference Percentage (%)
2996	135695	131630	-3.09	41844	-224.29
2999	149891	142548	-5.15	42237	-254.88
3008	155373	147121	-5.61	43428	-257.77
3009	151644	146263	-3.68	43785	-246.34
3016	134392	131319	-2.34	41165	-226.47
3074	98521	97955	-0.58	40055	-145.96
3075	141392	140240	-0.82	40312	-250.74

4.3 MEAN OF MATURITY

Maturity is a ratio number measuring the relative thickness of the cotton cell wall. Table 4.2 displays the average maturity values under PA, PAM and CA, as well as the average maturity data obtained from the Advanced Fiber Information System (AFIS) and High Volume Instrument (HVI). The correlations among PA, PAM and CA with AFIS and HVI data are also calculated in Table 4.2.

Table 4.2 Maturity Comparisons in Mean

Variety	PA	PAM	CA	AFIS	HVI
2996	0.54	0.52	0.51	0.90	4.60
2999	0.48	0.45	0.45	0.80	3.30
3008	0.50	0.47	0.46	0.84	3.20
3009	0.55	0.52	0.51	0.91	3.90
3016	0.52	0.5	0.49	0.86	4.38
3074	0.58	0.57	0.57	0.95	5.65
3075	0.44	0.42	0.42	0.80	3.09
Correlation with AFIS	0.93	0.95	0.94		
Correlation with MIC	0.74	0.85	0.88		

The correlation differences among PA, PAM and CA results are not significant because of their less distinguished mean values. It is common to understand that different populations, with different distributions, still have a large chance to get the same mean values. So the mean value is not a sensible parameter to stand for the whole population, especially when the population lacks normality.

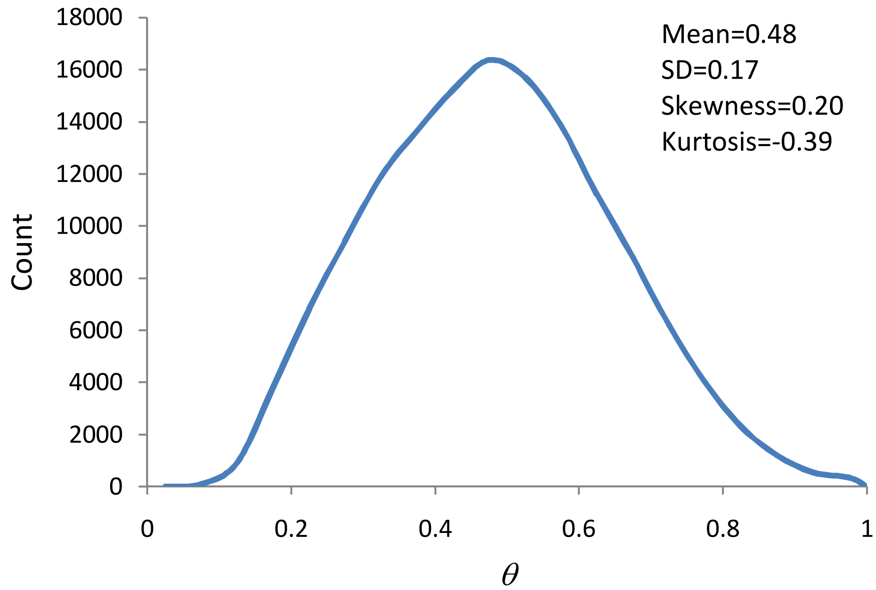
4.4 PARAMETERS OF MATURITY DISTRIBUTIONS

Table 4.3 indicates four distribution parameters (mean, standard deviation, skewness, and kurtosis) of maturity in PA, PAM and CA. The central tendency and value spread among these three methods are almost the same. While, CA shows more positive skewness, which indicates that its maturity values concentrate to the left side of the mean, so more immature fibers can be found in CA results. In most cases, the kurtosis of CA is the smallest, which demonstrates a more rounded peak with shorter and thinner tails of the distribution compared with both PA and PAM.

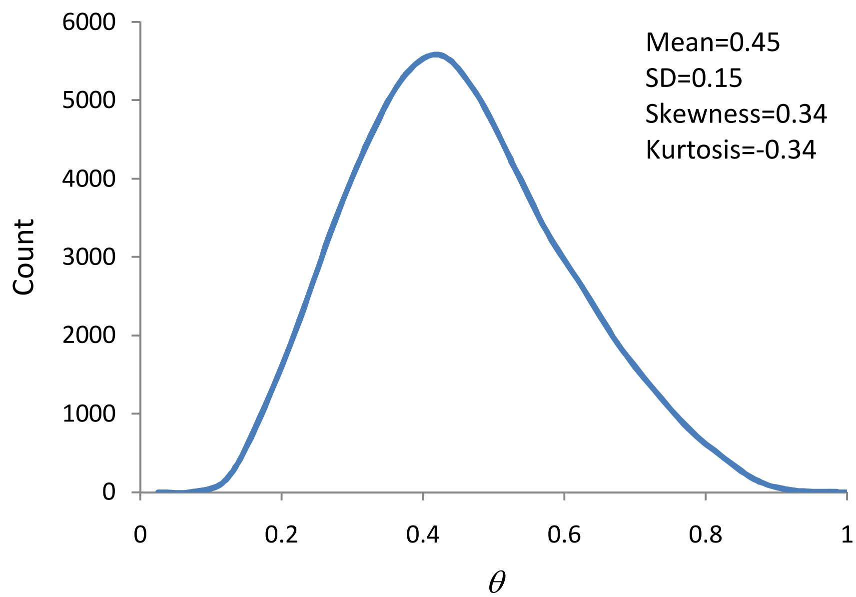
Table 4.3 Distribution Parameters of Maturity

Variety	Mean			SD		
	PA	PAM	CA	PA	PAM	CA
2996	0.54	0.52	0.51	0.16	0.15	0.17
2999	0.48	0.45	0.45	0.17	0.15	0.17
3008	0.50	0.47	0.46	0.17	0.15	0.17
3009	0.55	0.52	0.51	0.16	0.14	0.17
3016	0.52	0.5	0.49	0.17	0.15	0.17
3074	0.58	0.57	0.57	0.17	0.15	0.17
3075	0.44	0.42	0.42	0.17	0.16	0.17
Variety	Skewness			Kurtosis		
	PA	PAM	CA	PA	PAM	CA
2996	0.04	0.10	0.20	-0.28	-0.52	-0.53
2999	0.20	0.34	0.47	-0.39	-0.34	-0.42
3008	0.20	0.33	0.48	-0.40	-0.42	-0.45
3009	-0.02	0.10	0.21	-0.22	-0.42	-0.53
3016	0.05	0.15	0.25	-0.40	-0.53	-0.60
3074	-0.13	-0.11	-0.03	-0.20	-0.49	-0.48
3075	0.33	0.44	0.61	-0.38	-0.29	-0.26

Figure 4.1 shows the maturity distributions of Cotton “2999” using PA, PAM and CA. The *x*-axis represents the maturity value, and the *y*-axis represents the number of detected fibers.

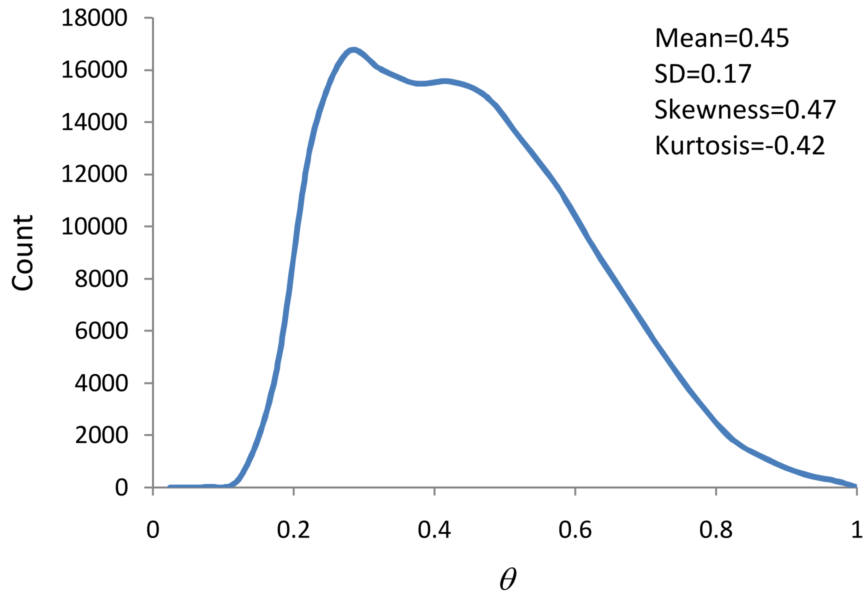


(a) Distribution of PA.

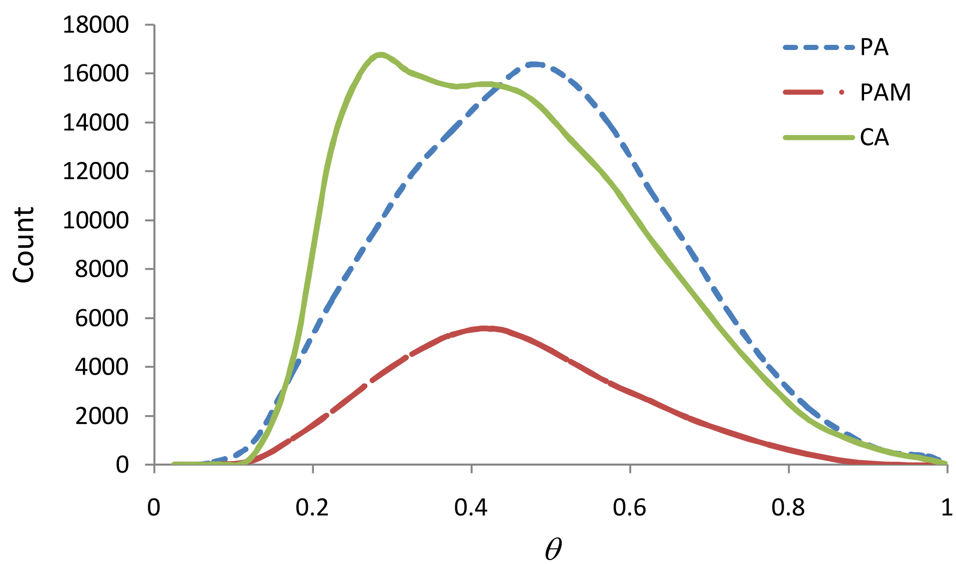


(b) Distribution of PAM.

Figure 4.1 Distributions of cotton maturity in “2999”.



(c) Distribution of CA.



(d) Distribution comparison among PA, PAM and CA.

Figure 4.1, cont.

Through the distribution comparison in Figure 4.1(d), a number of mature fibers detected by PA are eliminated manually by PAM method due to their incorrectness, and leads to the shift of skewness from type II of PA to type I of PAM. Although both of CA and PAM have incorrectness removal function, a remarkable distribution difference still can be distinguished from CA and PAM results.

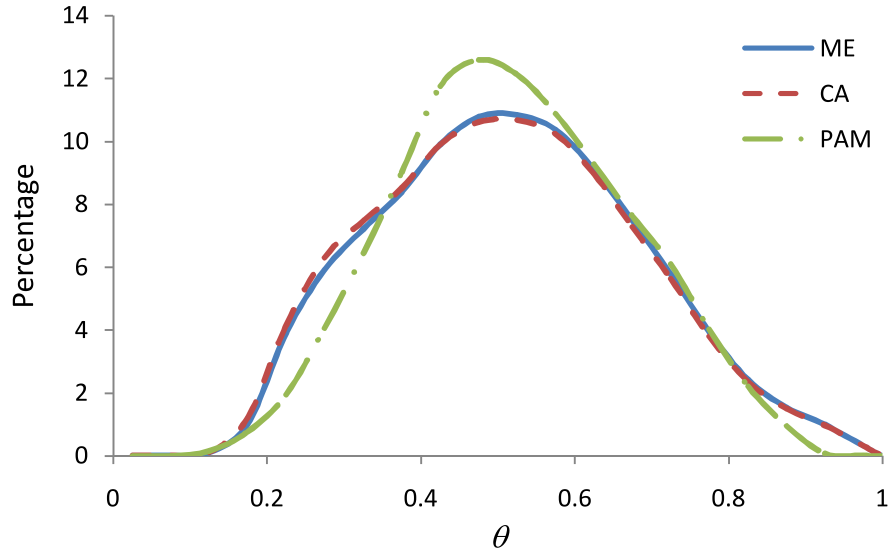
4.5 Agreement of CA method

In order to tell which result is more accurate between PAM and CA, cotton "2996" is selected to accomplish the manual editing work (ME). During the ME, users need to manually re-portray the unclear boundary of cross sections to save more fibers.

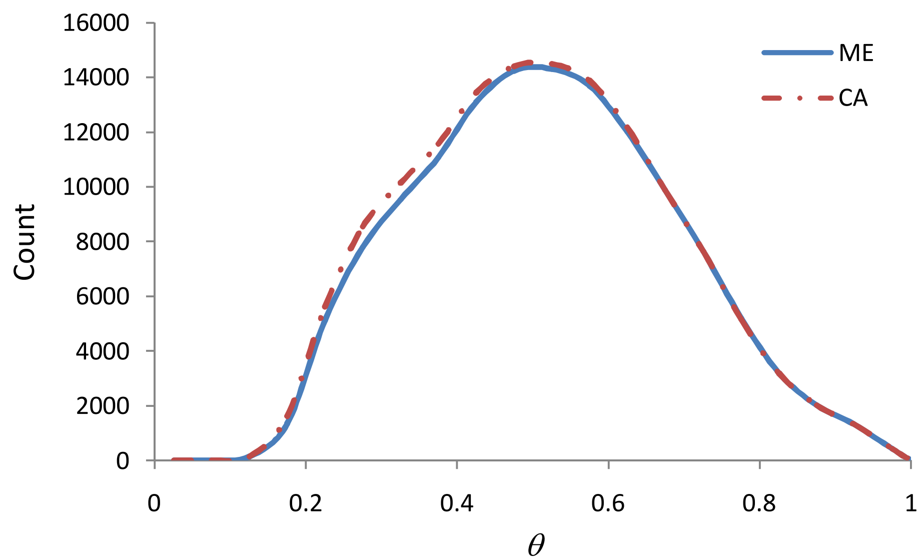
Four maturity parameters of PAM, CA and ME are listed in Table 4.4, and their maturity distributions are drawn in Figure 4.2. The x-axis represents the circularity value, and the y-axis represents the percentage of total detected fibers in Figure 4.2(a) and the number of detected fibers in Figure 4.2(b).

Table 4.4 Distribution Parameters of Maturity in "2996"

Method	Mean	SD	Skewness	Kurtosis
ME	0.52	0.17	0.19	-0.52
CA	0.51	0.17	0.20	-0.53
PAM	0.52	0.15	0.10	-0.52



(a) Distributions of PAM, CA and ME.



(b) Distributions of CA and ME.

Figure 4.2 Distributions of cotton maturity in “2996”.

Through the comparisons from Table 4.4, the parameters between CA and ME are almost the same. While, PAM and ME show a sizeable skewness difference, about 50%. The inconsistencies between PAM and ME parameters attribute to their absolutely deviated distributions in Figure 4.2(a).

As for the distributions between CA and ME in Figure 4.2(b), there is almost no discrepancy in the measurement results, as well as their total detected numbers. Although slight discrepancies in fiber frequency can be observed near θ =0.3, the agreement of the two distributions is extremely high (R^2=0.999), and the two sets of the descriptive statistics are approximately identical. This proves that the CA test is able to produce reliable fiber detections, except for irreparable immature cross sections. Because of the high agreement between the CA and ME data, there is no need to perform the manual editing after the CA test.

4.6 Distribution Classifications

The proportions of maturity classifications among dead fibers, half-mature fibers, and mature fibers for these seven cotton varieties using PA, PAM and CA are calculated in Table 4.5. Analyzing the results obtained from PA and PAM, the proportions of dead fibers are almost the same, and they are sharply less than those in CA results. It is reasonable to believe that PA has a large difficulty in detecting dead fibers because of their narrow, small or folded shapes. Relying on the modifications realized in CA algorithm, more dead fibers are detected.

Table 4.5 Proportions of Maturity Classifications

Variety	PA		
	Dead	Half- Mature	Mature
2996	7.98%	56.90%	35.12%
2999	15.45%	60.92%	23.63%
3008	13.11%	59.63%	27.26%
3009	6.71%	56.14%	37.15%
3016	10.50%	57.02%	32.48%
3074	5.72%	47.97%	46.31%
3075	23.02%	58.21%	18.77%
Variety	PAM		
	Dead	Half- Mature	Mature
2996	6.94%	62.80%	30.26%
2999	16.38%	66.44%	17.18%
3008	13.97%	65.83%	20.20%
3009	5.84%	65.15%	29.01%
3016	10.65%	63.48%	25.87%
3074	3.54%	52.42%	44.04%
3075	23.28%	62.25%	14.47%

Table 4.5 continued

Variety	CA		
	Dead	Half- Mature	Mature
2996	11.76%	57.41%	30.83%
2999	23.06%	57.92%	19.02%
3008	20.07%	57.69%	22.24%
3009	10.87%	58.76%	30.37%
3016	15.80%	56.64%	27.56%
3074	6.59%	49.23%	44.17%
3075	29.91%	54.78%	15.31%

Half-mature fibers are the major bodies in the three distributions across the seven cottons, taking more than 50% of the total identified fibers, except in cotton 3074. Mature proportions in PA are largely higher than those in PAM, which mean a number of mature fibers in PA results are eliminated from PAM only because of their inexactness. Such kind of inexactness can be caused by mistaking dead or half-mature fibers as mature fibers (fiber 5 in Figure 2.3b, due to its incomplete lumen contour).

4.7 CONCLUSION AND DISCUSSION

Owing to the algorithm modifications and the automatic incorrectness removal function applied to the FIAS-II, its latest version CA has been upgraded from the number and accuracy of detected cross-sections as well as the bias reduction on immature fibers. Consequently, the currently improved FIAS generates much fewer unreasonable fiber detections and has an automatic function of deleting wrong cross sections. No manual revision is needed after the CA.

Chapter 5: Transformation for Normality Analysis

During the dataset analysis, the normality of the distribution should be tested firstly. In this chapter, several common normality tests are introduced briefly [20]. Then, if the normality result can't be guaranteed, an effective transformation can be made to validate the assumption of normality. For example, Box-Cox Transformation is a known choice. Finally valid parameters and comparisons can be made through the interpretation after transformation.

5.1 Normality Tests

5.1.1 Moment Test

Karl Pearson [21] proposed the test of normality firstly with the hypothesis that the standardized third ($\sqrt{\beta_1}$) and fourth (β_2) moments of a distribution can be used to depict the normality of the distribution [22].

The non-zero value of $\sqrt{\beta_1}$ describes an asymmetric distribution about its mean value, and β_2 is the parameter to describe the peakness and the tail thickness of a distribution. So $\sqrt{\beta_1}$ equals to "skewness" and β_2 equals to "kurtosis" as mentioned in Chapter 3. The definitions of these two can be expressed as:

$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} \tag{5.1}$$

$$\beta_2 = \frac{\mu_4}{\sigma^4} \tag{5.2}$$

where σ^2 is the variance, and

$$\mu_i = E(X - \mu)^i \tag{5.3}$$

$$\mu = E(X) \tag{5.4}$$

5.1.2 Empirical Distribution Function Tests

Another famous group of tests used to test the normality of distributions is empirical distribution function (*EDF*) tests. The main idea of this test is to get the difference *D* between $F(x:\mu,\sigma)$, the theoretical cumulative distribution function of the normal distribution, and $F_n(x)$, the empirical distribution function.

$$F_n(x) = \frac{(X \leq x)}{n} \tag{5.5}$$

With the increase of difference *D*, the extent of the non-normality will be more and more considerable.

The most famous and fundamental *EDF* test is Kolmogorov-Smirnov Test [23] with known parameters μ and σ:

$$D = \sup_x \left| F_n(x) - F(x,\mu,\sigma) \right| \tag{5.6}$$

Based on such main idea, a group of *EDF* tests even with unknown parameters μ and σ, like: the Cramer-von Mises W^2 test [24], the Kuiper V test [25], the Watson U^2 test [26], and the Amderspm-Darling A test [27] are achieved. Their realization can be attributed to the simulations from Stephens [28].

5.1.3 Other Tests

The application of Chi-square test [29] can also discriminate the normality of a distribution. Under the null hypothesis that: the tested distribution is normal distributed, the observed values O_i and the expected values E_i are compared together:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{5.7}$$

In addition to the tests with the assistance of equation calculation as mentioned above, there is another special method called “Normal Probability Plots”. This method measures the normality situation through a direct visual effect. In Figure 5.1, the “+” symbols compose a straight line, which works as a reference to mimic a normal distribution, and the “*” symbols indicate the cumulative frequency distribution of the test sample, whose deviation from the reference line reveals the non-normality.

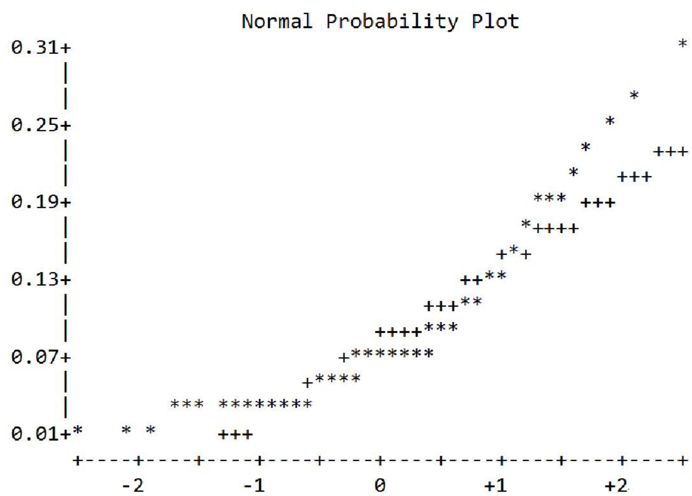


Figure 5.1 Normal probability plot.

5.2 Box-Cox Transformation

Through the Moment Test of the “skewness” and “kurtosis” data from Table 4.3, almost all the maturity distributions of cotton varieties obtained from the FIAS-II are non-normally distributed. In order to avoid the potentially misleading comparisons among maturity means from non-normal distributions, a transformation can be relied to make such comparisons feasible. Box-Cox transformation [30] is a well-known and useful transformation method to realize this function.

Box-Cox transformation can be expressed as:

$$T(x) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(x) & \lambda = 0 \end{cases} \tag{5.8}$$

where x is the input variable, and λ is the transformation parameter. Through the adjustment of parameter λ, the transformation process (Eq. 5.8) yields the output result *T(x)*.

5.3 Results Analysis

Table 5.1 gives three columns of maturity means using CA method. The 1st maturity column stands for the original mean value captured from CA results, which are distributed between "0" to "1" of each cotton variety. The 2nd column indicates the mean value of transformed results $T(x)$ when CA results of each cotton variety are used as the input variable x in Box-Cox transformation. After the transformation, the transformed results $T(x)$ are distributed normally in the negative axis range.

The 3rd maturity column traces the second maturity column values back to the state before their Box-Cox transformation. Such kind of "Re-transformation" makes the normal-distributed mean values (2nd column) come back to the range from "0" to "1" and be named as "equivalent maturity", which has the comparability among different cotton varieties with a sensible physical meaning.

Table 5.1 Mean of Maturity of CA

Variety	Before Transformation	After Transformation	Equivalent Maturity
2996	0.51	-0.55	0.50
2999	0.45	-0.75	0.42
3008	0.46	-0.73	0.44
3009	0.51	-0.56	0.50
3016	0.49	-0.60	0.48
3074	0.57	-0.43	0.57
3075	0.42	-0.87	0.39
2684	0.51	-0.56	0.50
2792	0.57	-0.43	0.57
2888	0.58	-0.41	0.59
2952	0.53	-0.52	0.52
2952rr	0.53	-0.52	0.52
3004	0.55	-0.49	0.54
3022	0.51	-0.56	0.50
3029	0.50	-0.57	0.49
3030	0.45	-0.72	0.43
3033	0.49	-0.61	0.48
3033rr	0.49	-0.60	0.48
3035	0.43	-0.81	0.41
3038	0.47	-0.65	0.46
3039	0.49	-0.62	0.47

Table 5.1 continued

3039rr	0.48	-0.63	0.47
3042	0.51	-0.56	0.50
3043	0.47	-0.69	0.45
3044	0.48	-0.68	0.45
3045	0.55	-0.47	0.55
3046	0.53	-0.51	0.52
3051	0.45	-0.75	0.42
3051rr	0.44	-0.81	0.41
3054	0.50	-0.59	0.48
3055	0.48	-0.61	0.47
3055rr	0.49	-0.61	0.47
3056	0.46	-0.71	0.44
3057	0.43	-0.82	0.40
3068	0.50	-0.61	0.48

Box-Cox transformation only plays an effect onto the distribution without the normality. Cotton "3074" is almost normally distributed due to its skewness (-0.03), so its re-transformed "equivalent maturity" value (0.57) has no difference to its original maturity mean value before the transformation (0.57).

In order to display the Box-Cox transformation effect, another 28 more cotton varieties are processed. After searching the 1st maturity column in Table 5.1, both "3044" and "3055" have almost the same original maturity mean (0.48). Through Box-Cox transformation, two distributions gain their own normal-distributed means (-0.68 for

“3044” and -0.61 for “3055”), which makes their re-transformed “equivalent maturity” in the last column be different (0.45 for “3044” and 0.47 for “3055”).

As shown in Figure 3.2, the θ distributions of Cotton “3044” and “3055” are different. Only relying on the mean maturity value cannot tell their difference. In addition to several methods mentioned in Chapter 4, like: multiple parameters (mean, standard deviation, skewness, and kurtosis), maturity classifications, and maturity patterns, the re-transformed “equivalent maturity” value is another way to discriminate the maturity distributions. The equivalent maturity of Cotton “3044” is 0.45 and Cotton “3055” is 0.47, which indicates fibers in “3044” have more immature fibers when compared with fibers in “3055”, so these two cotton varieties belong to disparate maturity distributions.

This conclusion is in agreement with the analysis in Chapter 3. According to the maturity classifications, the proportions of dead and half-mature fibers in “3044” are higher than those in “3055”. According to the maturity distribution patterns, “3044” belongs to Type I and “3055” belongs to Type II. Thus, the point that more immature fibers exist in “3044” than those in “3055” is also supported by maturity classifications and maturity patterns methods.

5.4 CONCLUSION AND DISCUSSION

In this chapter, several tests to check the normal distributions are introduced: Moment Test, Empirical Distribution Function Tests, Chi-square Test and Normal Probability Plots. In order to make the comparisons among the non-normal distributions feasible, Box-Cox transformation is applied. The re-transformed “equivalent maturity” value after the Box-Cox transformation works as another effective representation to display the “real” maturity means of the distributions lacking normality.

Chapter 6: Conclusions and Future Work

6.1 Summary of the Thesis

The algorithm changes made in the FIAS-II improved the consistency in detecting fibers of various maturities, and effectively reduced the bias on immature fibers in previous maturity distributions. The maturity distribution generated in the CA test was very similar to that of the manually traced fibers. The dramatic reduction in the number of wrong fiber detections eliminates the need for manual editing in a routine test.

Cotton maturity of a large quantity of fibers should be evaluated with multiple distributional parameters (like: mean, standard deviation, skewness, and kurtosis) since any single parameter is insufficient in characterizing the whole maturity distribution and especially invalid when the distribution is not normal. The basic shapes of cotton maturity distributions can be categorized into four patterns, each representing a major class of cotton maturity. The skewness of a maturity distribution is the principal parameter to classify the distribution pattern. Most cotton varieties exhibit positively skewed distributions, indicating more immature fiber contents than mature ones.

From a maturity distribution, the contents of dead, half-mature and mature fibers can be calculated by the cumulative probabilities in the specific θ ranges. It seems that half-mature is the major body of the maturity distributions for most cotton samples, but only the dead and the mature fibers highly correlate with the overall maturity level. Since dead fibers lack strength and dyeability, the dead fibers should be a discount factor in maturity evaluation. On the other hand, the mature fibers should be used as a premium in the ultimate rating of fiber quality.

Before the data analysis, the normality of the distribution should be tested to ensure the validation of evaluation process. The normality of a distribution can be tested by several methods, and the most traditional one is Moment Test. In order to make the comparisons among mean values of each distribution feasible when they are tested to be not normal, Box-Cox transformation is applied to correct the non-normal situation to a normal state. The equivalent maturity value works as another valid parameter to represent the maturity standard. The deviation of the equivalent maturity from its original maturity value is proportional to the extent of the distribution non-normality.

6.2 SUGGESTED FUTURE WORK

The room for current FIAS software to improve can be developed in the aspect of human-computer interaction. The software is expected to provide much more intelligent reaction to the operator.

When using the manual editing tool, the operators no large need to re-portray the boundary pixel by pixel manually. The program could re-trace the wrongly detected period or self recover the missing part of the boundary relying on the “editing range” provided by the operator.

As for the adherent fibers, there is still no accurate method to restore the individual ones due to the missing part hidden under the overlap. Using 3D microscopy is an effective way to disclose the hidden part when capturing the image of fibers through the depth adjustment. The operator only needs to match each adherent fiber with its hidden boundary, and then the restoring process would be fulfilled by the program automatically. After the restoring, operator could manually do some slight revisions to perfect the results.

Bibliography

[1] E. K. Boylston, J. P. Evans, and D. P. Thibodeaux, "A Quick Embedding Method for Light Microscopy and Image Analysis of Cotton Fibers," *Biotechnic & histochemistry*, vol. 70, no. 1, pp. 24-27, 1995.

[2] B. Xu, and Y. Huang, "Image Analysis for Cotton Fibers Part II: Cross-sectional Measurements," *Textile Research Journal*, vol. 74, no. 5, pp. 409-416, 2004.

[3] L. Padmaraj, M. Krifa, and B. Xu, "Evaluating Immature Fiber Bias in Fiber Cross-section Analysis," *Proceedings Beltwide Cotton Conferences*, pp. 1476, 2011.

[4] E. Lord, "Airflow through Plugs of Textile Fibers. Part II. The Micronaire Test of Cotton," *Journal of the Textile Institute Transactions*, vol.47, no.1, pp. T16–47, 1956.

[5] E. K. Boylston, D. P. Thibodeaux, and J. P. Evans, "Applying Microscopy to the Development of a Reference Method for Cotton Fiber Maturity," *Textile Research Journal*, vol. 63, no.2, pp. 80-87, 1993.

[6] D. P. Thibodeaux, and K. Rajasekaran, "Development of New Reference Standards for Cotton Fiber Maturity," *Journal of Cotton Science*, vol. 3, pp. 188–193, 1999.

[7] E. Hequet, B. Wyatt, N. Abidi, and D. P. Thibodeaux, "Creation of a Set of Reference Material for Cotton Fiber Maturity Measurements," *Textile Research Journal*, vol. 76, no.7, pp. 576–586, 2006.

[8] D. P. Thibodeaux, and J. P. Evans, "Cotton Fiber Maturity by Image Analysis," *Textile Research Journal*, vol. 56, pp. 130-139, 1986.

[9] Y. Huang, and B. Xu, "Image Analysis for Cotton Fibers Part I: Longitudinal Measurements," *Textile Research Journal*, vol. 72, no. 8, pp: 713-720, 2002.

[10] R. L. Long, M. P. Bange, S. G. Gordon, and G. A. Constable, "Measuring the Maturity of Developing Cotton Fibers Using an Automated Polarized Light Microscopy Technique," *Textile Research Journal*, vol. 80, no. 5, pp: 463-471, 2010.

[11] J. Rodgers, C. Delhom, C. Fortier, and D. Thibodeaux, "Rapid Measurement of Cotton Fiber Maturity and Fineness by Image Analysis Microscopy Using the Cottonscope," *Textile Research Journal*, vol. 82, no. 3, pp: 259-271, 2012.

[12] B. Xu, B. Pourdeyhimi, and J. Sobus, "Fiber Cross-Sectional Shape Analysis Using Imaging Techniques," *Textile Research Journal*, vol. 63, no. 12, pp. 717-730, 1993.

[13] P. A. Annis, T. W. Quigley, Jr., and K. E. Kyllo, "Useful Techniques in Textile Microscopy," *Textile Chemist and Colorist*, vol. 24, no. 8, pp. 19-22, 1992.

[14] B. M. Petkar, P. G. Oka, and V. Sundaram. "The Cross-Sectional Shapes of a Cotton Fiber Along its Length," *Textile Research Journal*, vol. 50, no. 9, pp: 541-543, 1980.

[15] R. L. Barker, and D. W. Lyons. "Determination of Fiber Cross-Sectional Circularity from Measurements Made in A Longitudinal View," *Journal of Engineering for Industry*, vol. 101, pp: 59-64, 1979.

[16] R. Matic-Leigh, and D. A. Cauthen. "Determining Cotton Fiber Maturity by Image Analysis Part I: Direct Measurement of Cotton Fiber Characteristics," *Textile Research Journal*, vol. 64, no. 9, pp: 534-544, 1994.

[17] Y. Liu, D. Thibodeaux, and G. Gamble, "Development of Fourier Transform Infrared Spectroscopy in Direct, Non-destructive, and Rapid Determination of Cotton Fiber Maturity," *Textile Research Journal*, vol. 81, no. 15, pp: 1559-1567, 2011.

[18] D. R. Krathwohl, "Chapter 17: The Numeric Description of Data: Descriptive Statistics," *Methods of Educational and Social Science Research*, 2nd Edition, pp. 401, 2004.

[19] [Online] http://en.wikipedia.org/wiki/Skewness.

[20] P. Armitage, and T. Colton, "Normality, Tests of," *Encyclopedia of Biostatistics*, vol. 5, pp. 3744-3748, 2005.

[21] K. Pearson, "Contributions to the Mathematical Theory of Evolution," *Philosophical Transactions of the Royal Society of London*, vol. 91, pp. 343, 1895.

[22] K. O. Bowman, and B. R. Shenton, "Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2," *Biometrika*, vol. 62, pp. 243-250, 1975.

[23] A. Kolmogorov, "Sulla Determinazione Empiricadi una Legge di Distribuzione," *Giornale dell' Institute Italiano degli Attuari*, vol. 4, pp. 83-91, 1933.

[24] H.Cramer, "On the Composition of Elementary Errors, Second paper: Statistical Applications," *Skandinavisk Aktuarietidskrift*, vol. 11, pp. 141-180, 1928.

[25] N. H. Kuiper, "Tests Concerning Random Points on a Circle," *Proceedings, Koninklijke Nederlandse Akademievan Wetenschappen, Series A*, vol. 63, pp. 38-47, 1960.

[26] G. S. Watson, "Goodness-of-fit Tests on a circle," *Biometrika*, vol. 48, pp. 109-114, 1961.

[27] T. W. Anderson, and D. A. Darling, "A Test of Goodness-of-fit," *Journal of the American Statistical Association*, vol. 49, pp. 765-769, 1954.

[28] M. A. Stephens, "EDF Statistics for Goodness-of-fit and Some Comparisons," *Journal of the American Statistical Association*, vol. 65, pp. 1597-1600, 1974.

[29] K. Pearson, "On a Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that It Can Be Reasonably Supposed to Have Arisen in a Random Sampling," *Philosophical Magazine, 5th Series*, vol. 50, pp. 157-175, 1900.

[30] G. E. p. Box, and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, vol. 26, pp. 211-252, 1964.