United States
Department of
Agriculture

National
Agricultural
Statistics
Service

NASS

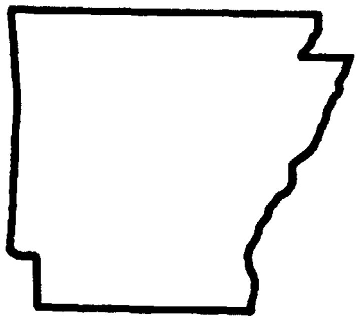
Research Division

RD Research Report
Number RD-97-03

November 1997

Fiber Properties As Components Of Cotton Yield: The 1996 Arkansas Study

Timothy Keller



FIBER PROPERTIES AS COMPONENTS OF COTTON YIELD: THE 1996 ARKANSAS STUDY, by Timothy P. Keller, Sampling and Estimation Research Section, Research Division, National Agricultural Statistics Service, United States Department of Agriculture, Washington, D.C. 20250-2000, November 1997. Research Report No. RD-97-03

ABSTRACT

Fiber properties are routinely used by the cotton industry as a measure of quality, crop maturity, and, to a lesser extent, yield. The National Agricultural Statistics Service is always searching for methods of improving its yield estimates; hence the motivation for studying the potential use of fiber properties in yield estimation.

The 1996 Arkansas study, a refinement and continuation of the 1995 Arkansas study, seeks to understand the variability of fiber properties, the relationships between fiber properties, and the relationship of fiber properties with yield. Of particular interest is the fiber property known as micronaire.

The association of first position micronaire and overall micronaire may be useful in improving early season predictions of cotton yield. Although the available data are inadequate to build an operational model, some indication of how a model of cotton yield using fiber properties might be constructed is presented.

KEY WORDS

Micronaire; fiber length; upper half mean; length uniformity index; first position; fruiting zone; lint; seedcotton, lint-percent, gin-turnout.

ACKNOWLEDGMENTS

The author wishes to express thanks for the help and advice of Dr. Hal Lewis; the staff of the Arkansas SSO, particularly NASDA enumerator Barbara Roach; Kevin White of the AMS-Classing Office, Hayti, Missouri; Roger Latham of the NASS Estimates Division; and the continued support of the NASS management.

TABLE OF CONTENTS

SUMMARY

Basic to yield estimation is the axiom: "The information necessary to make accurate yield forecasts is contained in the plants themselves." This axiom is the basic precept behind the entire objective yield program.

Early season forecasts of cotton yield are not always as accurate as one might desire; and, following the basic axiom of yield estimation, one looks to the cotton plant itself for new information. Both cotton researchers and the cotton industry in general make a significant effort to measure, understand and evaluate the *fiber properties* of cotton; but the NASS objective yield forecast does not make use of fiber properties. Before models of cotton yield that do make use of fiber properties can be developed and subsequently evaluated, it is necessary to have a basic understanding of fiber properties and their interrelationships. Achieving such a basic understanding is the goal of the ongoing research work in Arkansas. This research began with the 1995 cotton objective yield season. After working out some research design issues the first season, the research design for 1996 and succeeding years was resolved into the form outlined in the section entitled "Description of the 1996 Data Collection Process."

The fiber property that seems most promising for eventual use in modeling yield is micronaire, and the relationship between first position micronaire and overall micronaire may prove to be useful in improving the early season yield forecast for cotton. A definitive judgement of the utility of measuring fiber properties can not be made on the basis of a single season.

Some general discussion of how fiber properties might be used in yield forecasting is presented in the section entitled "Modeling Cotton Yield."

In addition to describing the 1995-96 Arkansas research, this paper also has a frankly educational intent: to provide a *very* brief introduction to the specialized vocabulary and knowledge proper to the cotton industry.

INTRODUCTION AND STUDY RATIONALE

The components of cotton yield under investigation are:

- micronaire
- fiber length
- fibers/seed
- seeds/boll
- gin turn-out

Some technical background on these variables is presented in the appendices.

The current objective yield estimates for cotton do not make use of fiber properties, e.g. micronaire, fiber length and fibers/seed. It may prove useful to incorporate these components of yield into the cotton objective yield models; however, only after sufficient data have been gathered for a sufficient number of years can the potential value of such additional information be adequately assessed.

One should not expect that any one component will serve as the basis for an improved yield estimate: this would be as if one were to attempt an estimate of the weight of a truck load of lumber, consisting of boards of various lengths, widths, breadths, and densities, by taking a random sample of boards and measuring width alone. To further illustrate this point, contrast the fiber from *Gossypium arborensis* (tree cotton) with the fiber from *Gossypium hirsutum* (upland cotton). The former species produces cotton with a very high micronaire-- so high that the fiber feels coarse to the touch-- and relatively low fiber length. The latter species produces cotton with relatively lower micronaire and relatively longer fibers. These differences are observed to effectively cancel each other to produce almost identical fiber weights, and, in consequence, almost identical yields. That fiber weight alone is also inadequate for yield estimation is demonstrated by the following data in Table 1.

Note the relatively high yield in 1994. The low micronaire and fiber weight was offset by a relatively high number of bolls per plant.

Given a suitable set of yield component variables, one may hope to obtain better yield estimates *earlier* in the season. The development of cotton bolls generally progresses from the bottom most branches upward, and from boll positions closest to the stalk to those further away.

Table 1. Missouri Cotton

Year	mean micronaire* (micrograms/in.)	mean fiber length* (inches)	fiber weight* (micrograms)	yield ** (lbs./acre)
1993	4.497	0.912	4.101	539
1994	4.145	0.906	3.755	856
1995	4.433	0.890	3.945	559

* source: USDA classing office report,Hayti, Mo. / ** source: NASS.

The branches of the cotton plant are often numbered according to their temporal development, and the bolls on a particular branch are similarly termed *first position bolls, second position bolls*, and so on, according to their proximity to the stalk. It is a reasonable conjecture, supported by some research findings, that the fiber properties of first position lint should be positively correlated with the fiber properties of the total lint harvested. Indeed, there must be a certain degree of positive correlation simply because of the typically high proportion of lint which comes from first position cotton. Table 2 shows how percent of total yield and relative micronaire varies by fruiting zone. Note that almost two thirds of the yield for this cotton came from first position bolls.

Table 2. Cotton: North Mississippi Delta-1992

Fruiting Zone			**% of Total Yield**	**Relative Micronaire (Percent of Zone 1 micronaire)**
	Boll Position	**Branches**		
1	1	1-4	23.6	100
2	1	5-8	27.7	104
3	1	9-up	15.8	80.6
4	2	1-up	16.7	84.4
5	3	all	4.7	75.8
6	4-up	all	11.5	87.3

* Source: Hal Lewis, Arkansas Experiment Station Special Report #162

DESCRIPTION OF THE 1996 DATA COLLECTION PROCESS

The 1996 research was based on about 60 cotton research plots associated with the same number of regular objective yield samples in the northeastern corner of Arkansas. Just prior to harvest survey enumerators working on behalf of NASS collected cotton from each of the two units comprising a sample. The plants in each unit were counted; then the first position bolls on the first four fruiting branches ('fruiting zone 1') and the bolls from the remaining fruiting zones were separately picked, counted and tagged. Dr. Hal Lewis --- scientist, businessman, cotton producer, and scholar, graciously gave survey enumerator the use of his micro-gin and laboratory facilities. (These facilities are located near Dell, Arkansas.) In particular, for each sample of cotton the following laboratory measurements were made:

1) Seed cotton weight
2) Weight of lint.
3) Weight of 25 seeds.
4) A set of three micronaire measurements.

After these measurements were performed each sample of cotton was tagged and mailed to the USDA cotton classing office in Hayti, Missouri. At the Hayti classing office the following measurements were made:

1) Micronaire
2) Upper half mean length
3) Length uniformity index

as well as routine measurements of other fiber properties with no immediately clear relevance to yield estimation.

In 1995 study measurements similar to those of the 1996 study were made. While the first position cotton measurements came from the research plots in both 1995 and 1996, in 1995 the 'remaining' cotton measurements were made on cotton from the adjoining objective yield unit. Given the potential variability of fiber properties with location, it was felt that a more valid analysis of the relationship between first position micronaire and overall micronaire could be made if both measurements were made on cotton from the same set of plants. Hence in 1996 all measurements were made on cotton from the research plots. Summary statistics from the 1996 study are presented in the following section. Where meaningful, comparable results from the 1995 study are presented for purposes of comparison.

SUMMARY STATISTICS

Micronaire

Table 3 shows the mean of the three NASS micronaire measurements on each sample averaged about 0.3 microgram/inch higher than the micronaire measurement made at the Hayti classing office. Two different enumerators made the NASS micronaire measurements; but enumerator identity was not recorded.

A detailed discussion of the factors that lead to such differences is given in Appendix 1, Part C. In the remainder of the discussion the micronaire measurements used are the average of the four independent measurements (3 NASS, 1 Hayti).

Table 3. *Micronaire and Differences*: *NASS versus Classing Office Measurements*

	# of Obs.	Mean	Std. Dev.	Min.	Max.
Hayti/First Pos.	87	4.613	0.653	3.200	5.900
Lab./First. Pos.	88	4.954	0.728	3.133	6.467
Hayti./Rem.	86	4.636	0.670	2.600	6.000
Lab./Rem.	89	4.957	0.693	2.400	6.533
Diff/First	86	-0.325	0.317	-1.033	0.400
Diff/Rem.	86	-0.324	0.330	-1.333	0.533

Fiber Length And Fibers Per Seed

A discussion of fiber length, the upper half mean and the length uniformity index is presented in Appendix 2.

The weight of a fiber is micronaire × fiber length; hence, given the total weight of lint in the sample, the average fiber length, and the average micronaire, it is a matter of simple division to estimate the number of fibers in a sample. (Of course, the fiber length and micronaire are different for every fiber in the tested sample of lint, so using the measured values for micronaire and length in this computation really only *approximates* average fiber weight, and hence also only approximates the true number of fibers.) Subtracting lint weight from seed cotton weight gives the weight of seeds in a sample. After determining the weight of a given number of seeds one may estimate the total number of seeds by using the proportionality of number of seeds and weight of seeds. Thus one derives fibers/seed.

To give some idea of the variability of fibers/seed from year to year, consider the following data in Table 4 for the region served by the Hayti, Missouri classing office.

Table 4.

YEAR	1993	1994	1995	1996
fibers/seed	13,900	14,500	11,000	15,000
Source: Hayti, Mo. USDA cotton classing office. Figures are rounded to the nearest 100.				

Seeds Per Boll

A cotton boll is divided into four segments, or *locks*; a 'perfect' boll has 8 seeds/lock, for a total of 32 seeds/boll.

If there are fewer than about 12 viable seeds in a boll, the plant usually aborts the boll early on.

The Research Plot Numbers

The 1995 statistics in Table 5 are based on about 205 observations; the 1996 statistics are based on about 86 observations. Recall that reported micronaires are the average of the three NASS measurements and the micronaire measured by the Hayti classing office.

Table 5.

Year	*Position*	*Statistic*	*Fiber Length*	*Micronaire*	*Fibers per Seed*	*Seeds per Boll*
1995	First	Mean	0.9325	4.88	13025	26.84
1995	Other	Mean	0.9275	4.60	12477	25.92
1996	First	Mean	0.9284	4.61	14974	27.26
1996	Other	Mean	0.9153	4.64	15367	26.41
1995	First	Std. Error	0.0028	0.0405	122	0.371
1995	Other	Std. Error	0.0026	0.0371	132	0.353
1996	First	Std. Error	0.0053	0.0704	242	0.688
1996	Other	Std. Error	0.0058	0.0723	316	0.964

Other Yield Components

About 38% of the total lint weight was from first position bolls. While it is possible that some bolls from positions 2 and higher were misidentified as first position bolls, this percentage is by no means extraordinary. The gin turn-out, as calculated by the ratio of total lint weight to total seed cotton weight was 39%, consistent with gin turn-outs of 32% to 36% typically cited for commercial cotton gins.

Although the data set is fairly small, the component measurements obtained in the 1996 study seem to be both reliable and representative.

In Table 6 the sample correlations for the fiber properties under investigation are presented. The product of unbiased estimators for a set of random variables is not necessarily an unbiased estimator for the product of the random variables. The size of the bias is a function of the covariances between variables; so even small correlations may result in large biases if the variances are large. Hence if one is considering a multiplicative components model to estimate yield, a very careful analysis of the correlation structure of the components is necessary.

<u>Table 6.</u> 1996 Data

Cell Legend:
Pearson Correlation Coefficients/Prob > |R| under Ho: Rho=0/Number of Observations

	FIRST POS. MICRONAIRE	*FIRST POS. FBR/SEED*	*FIRST POS. SEEDS/BOLL*	*REMAINING LENGTH*	*REMAINING MICRONAIRE*	*REMAINING FBR/SEED*	*REMAINING SEEDS/BOLL*
FIRST POS. LENGTH	0.13081 0.2299 86	-0.11847 0.2773 83	0.078253 0.4582 83	0.43488 0.0001 85	0.14027 0.2004 85	0.5333 0.6278 85	0.01030 0.9264 83
FIRST POS. MICRONAIRE		-0.42888 0.0001 86	0.19577 0.0761 83	0.09395 0.3924 85	0.54924 0.0001 85	-0.07769 0.4797 85	-0.08835 0.4270 83
FIRST POS. FBR/SEED			-0.46277 0.001 83	-0.01744 0.8741 85	-0.17621 0.1067 85	0.16321 0.1356 85	0.10816 0.3304 83
FIRST POS. SEED/BOLL				-0.04576 0.6812 83	0.13622 0.2195 83	0.03015 0.7867 83	0.09050 0.4101 85
REMAINING LENGTH					0.13107 0.2290 86	-0.22740 0.0352 86	0.03325 0.7640 84
REMAINING MICRONAIRE						-0.33182 0.0018 86	-0.00490 0.9647 84
REMAINING FBR/SEED							-0.22450 0.0401 84

CORRELATION ANALYSIS

Some of the statistically significant correlations in this table are mathematical artifacts. For example, the negative correlation of first position fibers/seed with first position seeds/boll is nothing more than a reflection of the numerator of one quotient being the denominator of the other.

Other studies have found a weak, but statistically significant, correlation between fiber length and fiber micronaire. Drought stress in midsummer can significantly reduce average fiber length. If the latter part of the growing season is more favorable, the cotton plant deposits an average or above average *mass* of cellulose over a shorter than average fiber length, which translates into a higher micronaire. This sort of phenomena illustrates the importance of maintaining component studies over a number of years.

Note that the correlation of first position micronaire with the micronaire from the remaining cotton was 0.5492, which was highly significant. The correlation of first position micronaire with the micronaire *averaged over all fibers* in a sample was 0.7779. Also using the Hayti micronaire as the 'true' micronaire gives a correlation of 0.8140. These results suggest that further study of the relationship between first position micronaire and the micronaire of the remaining cotton is worth investigating.

MODELING COTTON YIELD

A. The Traditional Model -- General Description

Two different calculations are used to obtain estimates of final cotton yield : one from a linear regression model and the other from a so-called traditional model. The latter model for cotton yield is but one instance of an entire class of yield models. The generic type is represented by the equation:

yield = (plants/acre) $\times$ (fruiting forms/plant) $\times$ (fruit weight/fruiting form)

In the case of cotton, the 'fruiting form' is a boll, and 'fruit weight per fruiting form' is lint weight per boll. Although plant counts are made as part of the objective yield survey, there is no formal use made of the number of plants per acre or bolls per plant; hence the yield equation for cotton may be condensed to:

yield = (bolls/acre) $\times$ (lint weight/boll)

It is only a matter of elementary algebra to interpose other variables and regroup the resulting expression into a wide variety of equivalent products. One might call the yield calculated by estimating each of the terms in such a product a *multiplicative components model*. There are a number of general statistical issues involved in using such a model, but these issues deserve a separate treatment. The purpose of the present discussion is to indicate how fiber properties can be incorporated into such a model for the purpose of obtaining a better estimate of lint weight per boll.

B. Current Estimation Procedures General Description

i) Expanding the Sampled Area

Since yield is production *per acre*, every model must expand the measurements for the sampled area up to an acre. The objective yield survey makes measurements of the distance between two rows and the distance between five rows, so average row width may be estimated and the expansion factor calculated. The objective yield sample in a field consists of two units, each of which is a 10 foot by 2 row plot. The objective yield summary generates estimates of yield at both the state level and the sample level, although the sample level estimates are not published. To see how the calculations work, it suffices to consider one unit.

Example:
Let N be the number of objective yield units per acre. (So the units of N are 1/acre.)

There are 43,560 ft^2/acre; so a 10 foot by 2 row objective yield unit and a row to row distance of 3.15 feet(a typical figure) gives $N = 43{,}560/\ (10\times3.15\times2) = 629$/acre.

ii) Counting/Estimating Number of Bolls

The word 'boll' in the heading subsumes what is actually a number of carefully delineated biological structures. What is actually being counted early in the season may be mostly blooms, or squares. Later in the season when there are actual bolls, the bolls are counted separately as small bolls (under an inch in diameter), large bolls (an inch or more in diameter), open bolls, partially open bolls and unopened bolls.

Whatever counts are available at a given point in the season are used in conjunction with historic boll counts to forecast the number of bolls that will survive to be harvested.

Burrs, which are damaged bolls, are also included in gross weight computations. The post-harvest gleaning visit produces a burr count and an estimate of the lint remaining in the field after harvest, so that yield may be adjusted for harvest loss. During the season, harvest loss is estimated by historic averages.

iii) Weight per Boll

The estimate of weight per boll is based on a combination of the season's accumulated dry weight per boll and a 5 year moving average of the end of season values for weight per boll. Early in the season the historic data predominates---and that is the key yield estimation problem!

C. Fiber Properties And Weight Per Boll

To help see the relationship between fiber properties and the estimation of lint weight per boll, it's useful to run through a yield calculation using all of the key fiber properties. For purposes of illustration the unit level will suffice.

Let W be the pounds of lint that will be harvested from the unit. Then the yield for the unit (in pounds/acre) is simply NW, where N is the expansion factor as described in section B (i).

Computing W:
If w denotes the weight of an individual cotton fiber, then

$$w = \text{micronaire} \times \text{fiber length} \equiv M \times L$$

The typical units of M and L are micrograms per inch, and inches, respectively; hence w has units of micrograms.

The total number of fibers F in the unit is:
$F = \text{fibers/seed} \times \text{total seeds in the unit} \equiv f \times S$.

So $W = w \times F$.

The weight per boll is, of course, W divided by the estimated number of bolls.

Example: To take typical values for our sample calculation, let $M = 4.6$ and $L = 0.92$, $f = 15{,}400$, and $S = 8354$.

$$W = w \times F = (M \times L) \times (f \times S) =$$

$$544{,}453{,}572 \text{ micrograms} = 544 \text{ grams},$$

which, using the conversion factor of 454 gm/lb., gives $W = 1.200$ lb.

Then, using the value of N from the example in B(ii), the calculated yield for the unit is $Y = WN = 1.200 \times 629$ or about 829 pounds per acre.

D. Using First Position Micronaire To Estimate Yield

i) The Basic Equations

The subscript 'F' will indicate that the corresponding quantity refers to first position cotton, and the subscript 'R' will indicate that the corresponding quantity refers to the remaining (i.e. other than first position) cotton.

If M is micronaire and L is lint weight per boll, then

$$L_R = M_R X_R$$

and, (1) $L_F = M_F X_F$

where X denotes the product of all the yield components besides micronaire. Note that these equations are essentially true by definition.

At some point in the season $X_F \approx X_R$. If one assumes a simple linear relationship between micronaires:

$$(2)\quad M_R = \alpha M_F + \beta$$

it follows that

$$(3)\quad L_R = \frac{(\alpha M_F + \beta)}{M_F} L_F .$$

Hence if B_R and B_F are the estimated boll counts, and L_T is total lint weight then

$$(4)\quad L_T \approx L_F B_F + \frac{(\alpha M_F + \beta)}{M_F} L_F B_R .$$

ii) Research Unit Yield Estimates--

There were 45 usable samples in the 1996 Arkansas study. The yields for the 3 foot by 1 row research units associated with the corresponding objective yield units are summarized as:

	Number of Obs.	*Net Mean Yield (lbs./acre)*
Unit 1	45	777
Unit 2	44	804
'96 ARK Final Yield	--	776

The 1996 Arkansas project was confined to the northeastern corner of the state, so some deviation from the state level yield is to be expected. The sample correlation between unit 1 and unit 2 yields was only 0.5252, illustrating the wide variation in yield even within a single field.

Estimating the value of α and β in equation (2) for unit 1 and unit 2 observations separately, and letting the first position lint represent itself, then using equation (3) one obtains the estimates of the remaining lint weight given in table 7:

Table 7:

MEAN YIELD (lbs./acre)	*Modeled*	*Actual*	*First Position*
Unit 1	788	777	300
Unit 2	805	804	306

The modeled yield and the actual yield are fairly close, but this result should be taken with some caution. The first position yield, common to both actual and modeled yield, is a sizable fraction of total yield. Also, the estimates of the parameters α and β used in the model were based on the yield to be estimated; hence, they were, in some sense, the best possible estimates. Whether estimates of the parameters based on historic data would give similar results is an open question. (On the other hand, it is certainly a hopeful sign that the estimates obtained using the unit 1 and unit 2 plots as replicates give estimates for the parameters that are both numerically close and not statistically different.)

At present the *only* data available in August are historic data. There is the possibility, however, that if the microwave drying of 'green' cotton currently under investigation in a separate research project is implemented as part of the regular objective yield program, a sufficient quantity of lint would be available in August.

RECOMMENDATIONS/CONCLUSIONS

The summarized data from the 1996 study represents the beginning of the process of accumulating knowledge of 'typical' values for fiber components, the variability of those components, and the relationship between those components in the context of the objective yield survey. Of particular interest is the potentially useful relationship between first position micronaire and overall micronaire.

Replication of the 1996 Arkansas study over several years will take this work from a promising start to a fruitful conclusion.

APPENDIX 1. UNDERSTANDING MICRONAIRE

A. Agricultural Background

Picture an individual cotton fiber as a long hollow tube. The length of the fiber is set relatively early in the growing season. In the latter part of the growing season the cotton plant fills out the inside of the fiber by cellulose deposition. The rate at which the fiber fills out is quite variable with weather conditions (particularly accumulated heat units), soil type, and variety. Given the right conditions, the rate of deposition can be quite rapid. The extent to which the fiber is filled out is expressed in terms of weight per unit of fiber length. It is common practice to present this measurement in units of micrograms per inch, termed the *micronaire* of the cotton fiber. There is an obvious connection between micronaire and yield: everything else being equal, higher micronaire means higher yield. In addition, micronaire values relate to the quality of the cotton: high micronaire lint does not spin well, low micronaire lint does not dye well. Cotton with micronaire values outside the desirable range of values is subject to a schedule of discounts.

In many cotton growing regions it has become a common practice to hasten the end of the growing season through the use of chemical defoliants; hence, the final micronaire value is capable of indirect manipulation by the cotton producer. This problem of regulating cotton fiber maturity via the timing of defoliant application has led Dr. Lewis to investigate the manner in which the average micronaire varies with fruiting zone, and the relation between those micronaires. The same considerations may prove valuable for earlier and better yield estimates.

B. Measuring Micronaire

Few scientific measurements are as direct as laying a tape measure across a board, and micronaire measurement is not one of the exceptions. The usual measurement procedure takes a fixed weight of lint, compresses the lint to a fixed volume and measures the rate of air flow through the sample at a fixed air pressure. The linear density of the cotton fibers is calculated from the rate of air flow via the Lord Equation, named for the Englishman Peter Lord who derived the equation. The basic idea is easily understood. Let V be the fixed volume to which the lint is compressed; let V_i be the volume of the empty space *inside* the cotton fibers; let V_b be the volume of the empty space between the cotton fibers; let V_f be the solid volume of the cotton fibers.

Then: $V = V_f + V_i + V_b$ or

$$V - V_f = V_i + V_b$$

Since the weight of the cotton sample is fixed and the density of cellulose is essentially a constant, it follows that V_f is a constant, and hence the quantity $V - V_f$ is a constant. By the last equation it follows $V_i + V_b$ is also a constant. High values of micronaire correspond to low values of V_i and hence high values of V_b. At a fixed pressure the rate of air flow is an increasing function of V_b (more space between fibers means less obstruction to air movement); so V_b is some function of the rate of air flow. The precise determination of this functional relationship allows one to infer micronaire.

In recent years it has become possible to measure the micronaire of individual cotton fibers in a more or less direct manner.

However the associated cost of these measurements in both time and money is high relative to the gains in precision: the accumulated empirical evidence and the evidence provided by the new measurement technology shows that the indirect measurement of micronaire via the Lord Equation is quite reliable.

All but a very small fraction of the US cotton production is classed at a USDA classing facility. With two samples per bale to analyze, speed as well as precision is an important consideration. It is mandated that the entire sequence of measurements be performed in under a minute. The mathematics underlying the determination of fiber properties dictate that micronaire be the first measurement of the sequence. For the sake of speed, the sample of lint used for the micronaire determination is taken mechanically. The weight of the sample is therefore not precisely the fixed weight required by the Lord Equation. If the weight deviates too far from the required weight the system gives the operator a red light (literally).

On the spot a computer program adjusts the observed value of micronaire to correct for the variation in weight. Within a range of weight deviations the corrected micronaire, although not perfect, is deemed acceptable.

The automated system is internally recalibrated every 15 minutes of operation, and externally recalibrated daily. Ambient temperature and humidity are carefully controlled, and incoming cotton is 'conditioned', i.e. allowed to sit in open trays until its measured moisture content is at the prescribed level.

C. Comparability of Micronaire Measurements by NASS with Classing Office Measurements

The device used to measure micronaire at the Lewis' laboratory is fundamentally no different and no less accurate than that used at the Hayti classing office; and the modest skills needed to perform the measurements are quickly acquired. Since speed was not as pressing a concern for the personnel taking the NASS measurements, the deviation of sample weight from the prescribed value is not as great as it was at the classing office. On the other hand, the ambient temperature and humidity are not controlled at the Lewis' laboratory; so systematic differences in micronaire measurements were not easily accounted for.

On basic statistical principles, in the absence of a systematic bias one would expect the average of the three micronaire measurements performed by NASS enumerators to vary less than the single classing office measurement. The classing office estimates one standard error in repeated micronaire measurements to be about 0.3 micronaire units. This figure is based on experience with samples from baled cotton---a highly blended source, the standard error in repeated micronaire measurements on a sample of lint from a *particular fruiting zone* is probably a bit higher.

Given that the average difference in classing office measurements and average of the three NASS measurements was about 0.3 micronaire units, it seems wisest to accept the average of the four *individual* measurements as the best available value for micronaire.

Appendix 2. Fiber Length, The Upper Half Mean and the Length Uniformity Index

A. Notes on the Upper Half Mean

The upper half mean and length uniformity index are summary statistics commonly used in the cotton industry. The relationship of the upper half mean and length uniformity index with the mean and variance is capable of an elementary derivation. Although the derivation is elementary, and the relationship is of some practical significance, these results are not well known; hence, a short presentation of the basic definitions and results seems to be in order.

I. Some Mathematics

Definition. For a random variable X with finite expected value μ, and for which the median is also μ, the upper half mean of X is defined to be $E[X|\, X \geq \mu]$, and is denoted UHM(X), or simply UHM when the random variable is understood.

If the distribution of X is normal with mean μ and standard deviation σ, then UHM(X) is calculated very simply in terms of μ and σ.

Proposition. If $X \sim N(\mu, \sigma)$ then UHM(X)

$= \mu + \sqrt{\frac{2}{\pi}}\,\sigma$ Proof.

Since $P[X \geq \mu] = 1/2$, the conditional expectation $E[X|X \geq \mu]$ is

$$\frac{2}{\sqrt{2\pi}\,\sigma} \int_{\mu}^{\infty} x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

Let $w = (x - \mu)/\sigma$, then this integral is

$$\frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} (\mu + \sigma w)\, e^{-\frac{1}{2}w^2} dw$$

which may be written as

$$2\mu \left[\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}w^2} dw\right] +$$

$$\frac{2}{\sqrt{2\pi}}\,\sigma \left[\int_{0}^{\infty} w e^{-\frac{1}{2}w^2} dw\right]$$

The first integral in square brackets is just the probability a standard normal random variable is nonnegative, which is 1/2; using the substitution $z = 1/2\ w^2$, it is easy to compute the value of the second integral in square brackets, and that value is 1.

Hence UHM (X) =

$$(2)\quad (2\mu)\,1/2 + \left(\frac{2}{\sqrt{2\pi}}\,\sigma\right) 1 = \mu + \sqrt{\frac{2}{\pi}}\,\sigma$$

An Obvious Generalization

(For convenience $\alpha \equiv \sqrt{\frac{2}{\pi}} \approx 0.7979$

in the rest of the discussion.)

It is of course possible to define, and (I'm told) to measure, the upper quartile mean, the upper decile mean, etc. In particular, if the lower half mean is defined by

$$LHM = E[X|\, X < \mu],$$

then since

$$E[X] = E[X|\, X < \mu]\, P[X < \mu] + E[X|\, X \geq \mu]\, P[X \geq \mu]$$

for a random variable X ~ N (μ, σ) one has μ = LHM(1/2) + UHM(1/2) = LHM(1/2) + (μ + $\alpha\sigma$)(1/2).

Hence LHM = μ - $\alpha\sigma$.

B. Some Cotton

A typical cotton boll contains 12,000 to 16,000 cotton fibers *per seed*. The length of those fibers varies greatly. It is the longer fibers that can be successfully woven into cloth; the shorter fibers can be only be used for less profitable manufactured goods, e.g. felt. Hence the mean of the distribution of fiber length does not adequately characterize the quality of the lint cotton: a population of cotton fibers distributed N(1.0, 0.5) is less desirable than a population of cotton fibers distributed N(1.0, 0.25). The UHM of fiber length is given as an attempt to summarize the quality of the lint cotton with a single number. 'Docking' for fiber quality is often based on another quantity, the length uniformity index (LUI) , which is defined by LUI = μ/UHM.

Obviously, since $\mu > 0$, $0 < \text{LUI} \leq 1$, in any case. In the case for which the fiber length is normally distributed one has

$$\text{LUI} = \frac{\mu}{\mu + \alpha\sigma} .$$

In this case the LUI really *is* a measure of length uniformity: as σ decreases the LUI increases, and LUI = 1 if and only if σ = 0, i.e. all the cotton fibers have the same length.

Given the UHM and LUI it is possible to compute the proportion of fibers shorter than a given length, since μ = LUI (UHM) and, as is easily derived

$$\sigma = \frac{(1 - \text{LUI})\text{UHM}}{\alpha} .$$

For example, consider a sample of Arkansas cotton with UHM = 1.09 inches and a LUI = 0.84, then one computes μ = 0.84(1.09) = 0.9156 and

$$\sigma = \frac{(1 - 0.84)1.09}{0.7979} = 0.2602 .$$

A commonly quoted lower bound for cotton fibers intended for textile manufacture is 0.5 inches, so the proportion of 'unsuitable' fibers in this sample is:

$P[\, Z < (0.5 - 0.9156)/0.2602 \,] =$

$P[\, Z < -2.594 \,] = 0.0048$

One should keep in mind that 0.48% of the *number* of fibers being unsuitable means *less than* 0.48% of the lint *weight* is unsuitable: shorter fibers weigh less than longer fibers, so a given number of short fibers weigh less than the same number of longer fibers, everything else being equal. As to the assumption that 'everything else is equal', there is some evidence that fiber length and micronaire have a positive correlation---so shorter fibers tend to weigh less not just because they are shorter, but because shorter fibers tend to be thinner.

Micronaire is another characteristic of cotton fiber used to determine quality. The UHM for micronaire measurements certainly exists, but is not of interest since the range of micronaire values corresponding to acceptable quality is bounded below *and* above: micronaire values too high mean poor spinning characteristics, micronaire values too low mean poor dyeing characteristics.