MODULE AVERAGING - ADVANTAGES AND BENEFITS IN CLASSIFICATION DATA

Darryl W. Earnest
USDA, AMS, Cotton Program
Washington, DC

Abstract

The USDA, AMS, Cotton Program has offered module averaging on a voluntary basis since 1991. When analyzing the historical results of module averaging for gins across the country, the Cotton Program has concluded that the reproducibility of quality measurements is significantly higher for cotton bales assigned the module average versus the bales' individual readings. This increase in reproducibility indicates less variability in the fiber measurements and more stable classification results. The crop data over the years has continued to support this without exception.

Module Averaging is advantageous to all users of the Cotton Program classification data. Accurate and reliable classification results provide the confidence needed in the marketing chain. With the majority of the U.S. cotton crop going to the export market, it is imperative that the fiber property data affiliated with each bale hold up to any retesting by receiving agents and any corresponding scrutiny. Although the individual classification data provided to each bale by the USDA is a very accurate and dependable set of measurements, the module averaging system provides even better accuracy. In its mission to provide the most accurate and reliable services possible to its customers, the USDA believes that module averaging is the best option for producers and ginners and provides the entire cotton industry with a better method for cotton classification and utilization of classing data.

Background

Introduction of Module Averaging

Module averaging is a voluntary program offered by the USDA, AMS, Cotton Program at no additional charge to its customers (Earnest, 2003). Module averaging applies only to four High Volume Instrument (HVI) quality measurements – micronaire, strength, length, and length uniformity. The Cotton Program tests all bales within a module individually and averages these individual results by quality factor. The resulting average – the "module average" – is then assigned to all of the qualifying bales within the module. Any bales qualifying as true "outliers" retain their original individual bale measurements.

The Cotton Program first implemented the module-averaging program on a voluntary basis in 1991, in response to a recommendation made by the Secretary of Agriculture's Advisory Committee on Cotton Marketing. Due to strength reproducibility being more variable than other fiber properties, it was the first fiber property offered for module averaging. The success of the 1991, pilot project resulted in increased industry participation and the expansion of module averaging to include length, length uniformity, and micronaire in addition to strength in 1992. The program expanded briefly in 1993, to include color (Rd), color (+b), and trash (percent area) but was limited to micronaire, strength, length, and length uniformity from 1994 forward because of industry recommendations.

Concept of Module Averaging

The basic concept of module averaging is that for a given module, the averages for the measurements of micronaire, strength, length, and length uniformity for all of the bales is a more representative measurement than the individual bale measurements. In the early 1990s, the problem with high variability in the strength measurement prompted the Secretary of Agriculture's Advisory Committee on Cotton Marketing to initiate a plan to address the problem. The USDA, AMS, Cotton Program (then "Division") performed several extensive studies to investigate the validity of using an average to represent the bales within a module as a possible solution to the variability problem. Studies investigated the degree of blending cotton experiences as it is picked, deposited and compressed into a module, removed from the module at the gin, ginned and baled. The studies concluded that cotton within a module undergoes significant blending throughout these processes. Therefore, it stood to reason that a bale from within a given module would be statistically representative of that module. Subsequent studies continued to support this assumption. These studies involved extensive testing of all bales from within given modules to analyze variability between and within the individual bales. The studies showed that the variability was no greater for the module than

for the bales within the module. Further, the studies concluded that any one bale from a module would be statistically representative of that module and that when retested several times, the bales' values would always approach the average of the module's values.

Additional studies followed that included cooperators from various cotton industry segments, research agencies, and educational institutions whereby cotton was HVI-tested using both the module averaging criteria and the individual bale testing criteria. The results of these studies showed that even with different instruments, testing locations, operators, etc., the module-averaging concept was valid and data obtained from module averaging was more reproducible than traditional single-bale tests.

Another set of studies used through the Cotton Program's Standardization and Engineering (S&E) Branch's procedures for establishing values on bales used for HVI calibration also supports the concept of module averaging. The bales used for making standards are very carefully selected and screened each year using classification data. Once a candidate bale is identified as having potential values conducive to a needed calibration bale, it is purchased and delivered to the S&E Branch in Memphis and prepared to for a rigorous testing program. This program includes approximately 180 tests across the entire bale conducted in multiple laboratories by different instruments and technicians. The Cotton Program has used this process to test the concept of module averaging when testing individual bales and entire modules of candidate bales for standards. The extensive amount of testing necessary for value setting has provided an excellent amount of statistical data to describe the variability present within a bale and corresponding module. Studying these results each year has provided yet another means to verify the concept that the module average is the most accurate means to quantify the individual measurements of the bales within the module.

Testing Process in Module Averaging

The Cotton Program handles and grades (classes) module-averaged bale samples in exactly the same manner as traditional samples. Each sample is received, prepared, conditioned, HVI-tested, and graded the same as all non-module averaged bales. The randomly selected checklot samples that classing offices submit to the Quality Assurance (QA) Branch for retest each day include a percentage of the module-averaged samples. The QA Branch does not differentiate between module averaged samples and traditional samples and tests all the same way. In fact, QA does not know prior to testing whether a sample represents a bale that was module averaged or retained its individual measurements. After testing, the main computer system calculates comparative statistics for evaluation.

Participation in Module Averaging Program

Participation in the module-averaging program has been consistent over the years with an average participation rate of 20-25 percent of all cotton across the country. This correlates to an average of nearly 4 million bales annually. The consistency of the participation over the years generates two assumptions. First, it shows a confidence in the program by those participating in module averaging regularly over the years. Many of the participants have remained in the program each year since its inception in 1991. In addition, the number of corresponding bales has remained consistent with insignificant increases or decreases annually (Table 1).

Table 1. Participation in Module Average Program.

Crop Year	Number of Module Averaged Bales	Percentage of Crop
1992	2.3 million	15%
1993	3.1 million	20%
1994	4.4 million	24%
1995	3.7 million	22%
1996	3.8 million	21%
1997	3.6 million	20%
1998	2.4 million	18%
1999	3.0 million	19%
2000	3.4 million	21%
2001	3.8 million	19%
2002	3.3 million	20%
2003	3.8 million	23%
2004	5.4 million	24%
2005	5.5 million	24%
2006	5.0 million	24%
2007	3.7 million *	25%
Average	3.8 million	21%

* classing totals through January 4, 2008

Second, the consistent numbers also reflect a lack of significant increase in participation despite ongoing efforts to educate and promote the program. A possible reason for this is insufficient awareness and full understanding of the program from all of the thousands of producers and ginners across the country. Even though the Cotton Program tries to make all of its customers aware of the module averaging program and its benefits, it is very difficult to get the entire message out to everyone. We rely on publications, presentations, and word of mouth to help in the education process, some of which may not reach all of the intended recipients effectively.

One important point to make on participation is regarding the minimum allowable amount of bales required to qualify for the program. Producers can opt to module average as little or as much of their cotton as they wish. The bale range for each module to be averaged is submitted to the Cotton Program and its mainframe computer system performs all of the arithmetic calculations once all individual bales have been instrument and manually tested. Although we encourage producers to module average as much as possible, there is not a set minimum to participate. The Cotton Program often encourages producers and gins to try the module averaging program using a portion of their cotton for a season and evaluate the results to help in making their long term decisions on participation.

Reproducibility of Single Bale vs. Module Bale Testing

Since 1992, the Cotton Program has recorded and analyzed the reproducibility for all quality factors within module-averaged bales using two methods: classing office single test and classing office module average. Both of these are compared to the checklot retest performed each day for each classing office by the Quality Assurance Branch located in Memphis, Tennessee. This checklot system acquires approximately one percent of all cotton samples tested in each office, selected at random, and shipped overnight into the Quality Assurance Branch. The Branch retests each sample on two separate instruments and averages the readings to generate a single measurement to compare to both the individual and the module averaged tests from the field classing offices (Gibson, 2003).

In every case since 1992, the module-average reproducibility for each of the four quality factors (micronaire, strength, length, and length uniformity) has been significantly higher than that of the individual single tests for the bales. Figures 1-4 show the annual overall reproducibility for each of the quality factors. Although some of the testing tolerances have been tightened over the years to reflect improvements in testing methods and equipment resulting in slight fluctuations in overall reproducibility, the trend on the comparison of single bale testing versus module averaging testing has remained consistent. Higher reproducibility equates to higher accuracy, less variability between instruments and classing offices, and thus, a much more stable and reliable measurement for all data users. In Figure 5, the overall average for each of the four measurements' reproducibility since 1992 is shown.

For the 16-year span, the average reproducibility for micronaire when comparing the classing laboratories' module averaging results to Quality Assurance checklot results is 85% versus 80% for individual testing. For strength, the comparison is 87% compared to 75% for individual testing. For length, the comparison is 91% versus 79% for individual testing. For length uniformity, the comparison is 94% compared to 83% for individual testing.

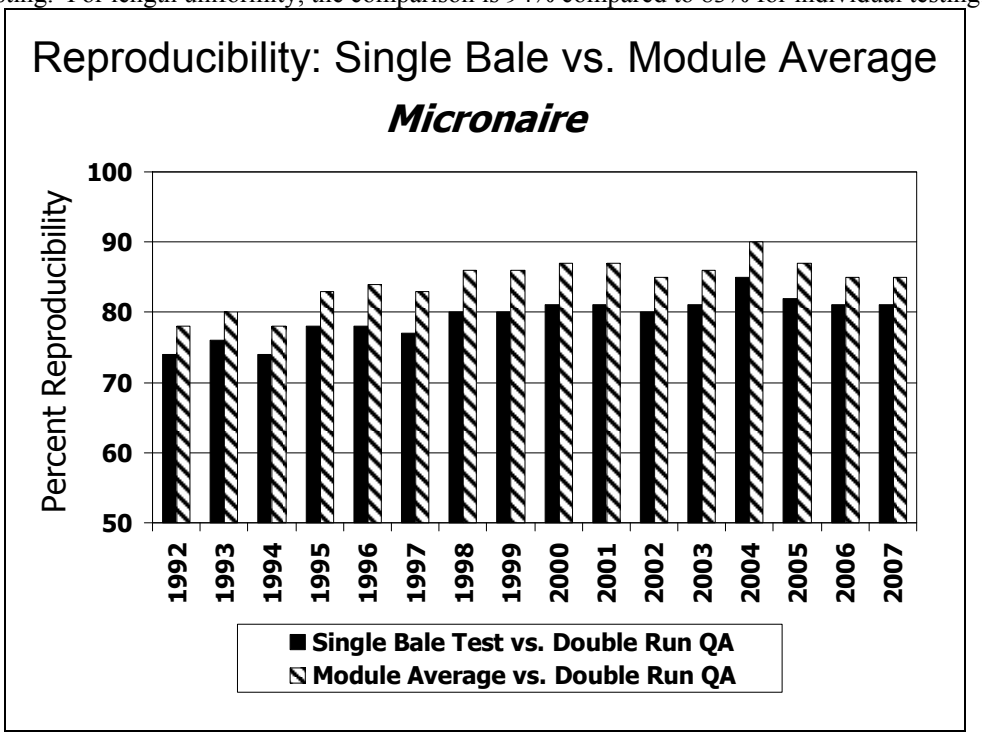

Figure 1. Reproducibility of Micronaire 1992-2007.

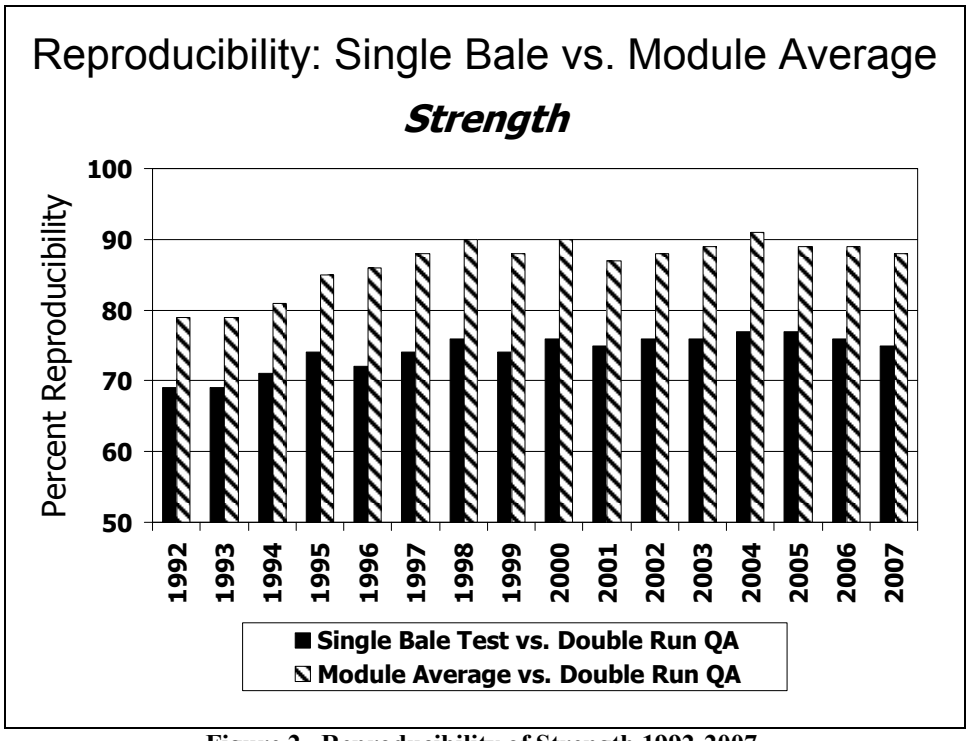

Figure 2. Reproducibility of Strength 1992-2007.

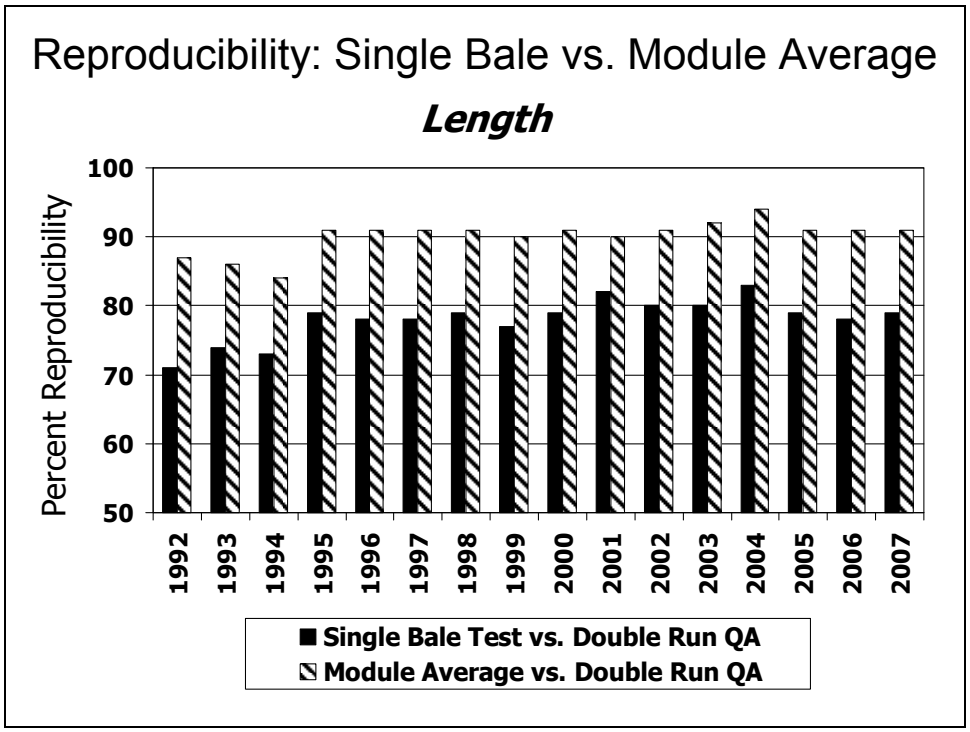

Figure 3. Reproducibility of Length 1992-2007.

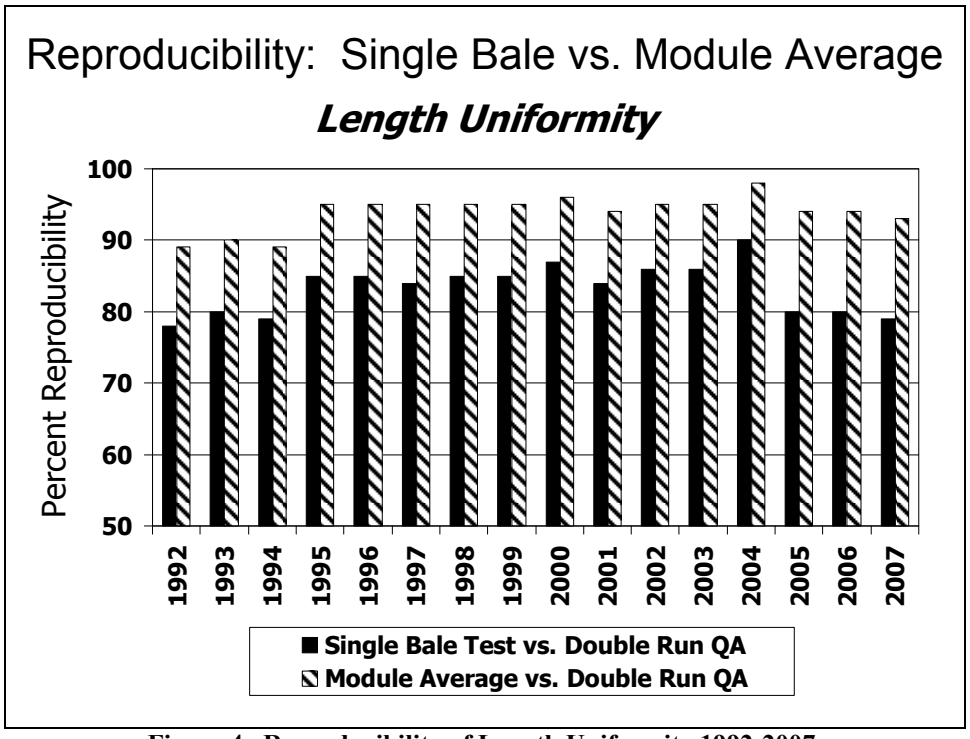

Figure 4. Reproducibility of Length Uniformity 1992-2007.

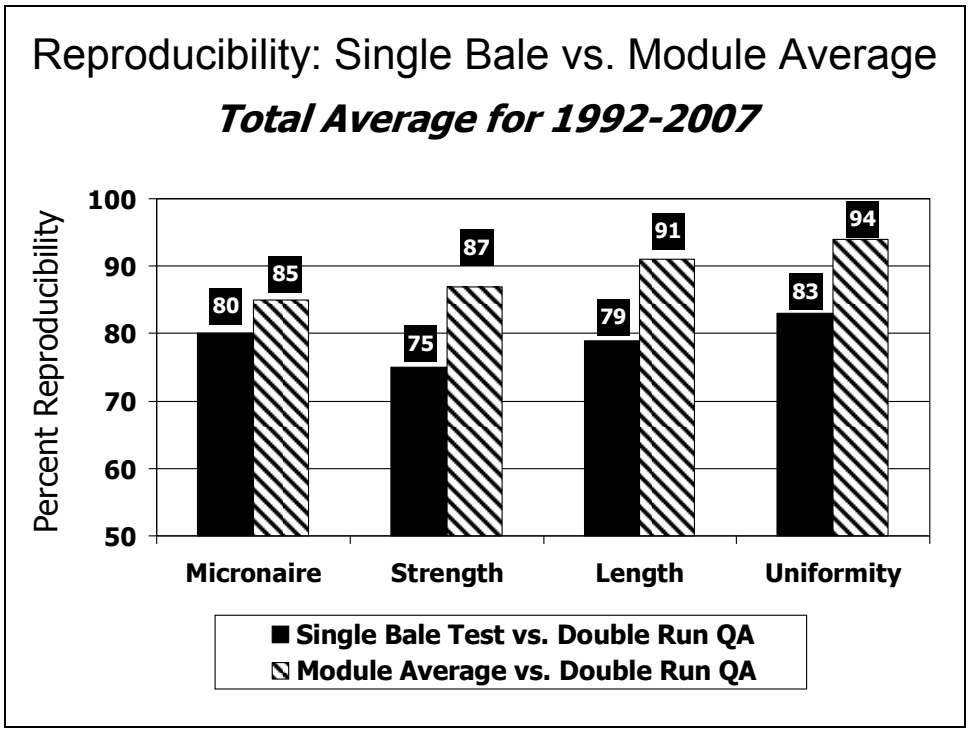

Figure 5. Total Average Reproducibility of Four Factors 1992-2007.

Looking at the same statistics for the period of 2000-07, the results are even more impressive. Since 2000, the reproducibility for micronaire averaged 87% compared to 82% for individual testing. For strength, the comparison is 89% compared to 76%. For length, the comparison is 92% compared to 80%. For length uniformity, the comparison is 95% compared to 84% (Figure 6).

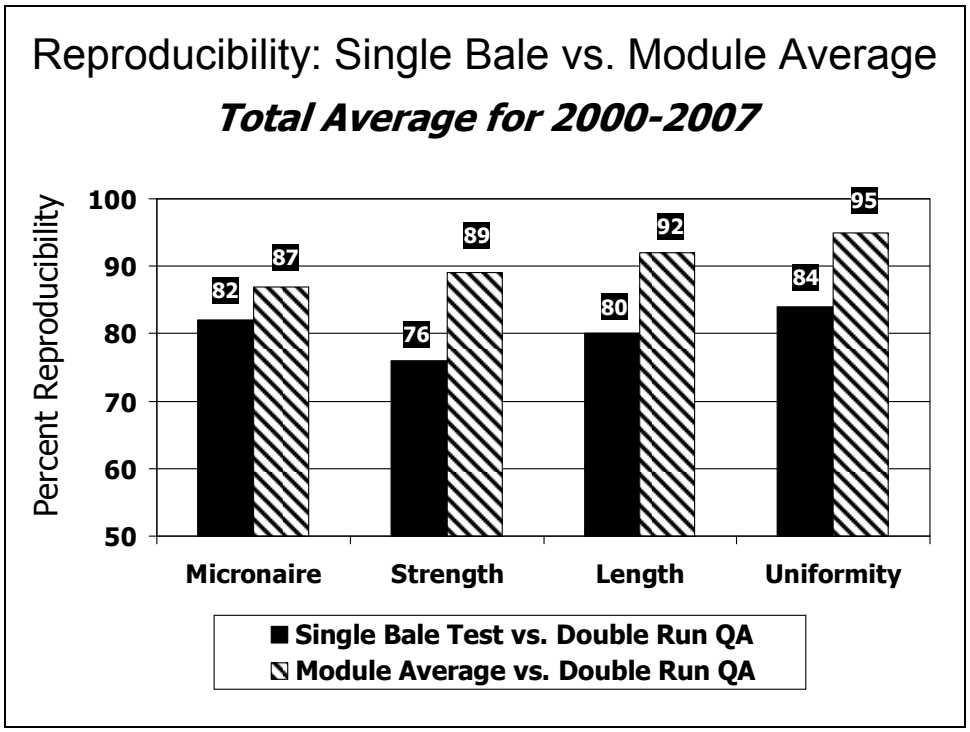

Figure 6. Total Average Reproducibility of Four Factors 2000-2007.

Further, when reviewing the 2007 classing data through December 28, 2007, the results compare favorably to these averages. The reproducibility for micronaire averaged 85% compared to 81% for individual testing. For strength, the comparison is 88% compared to 75%. For length, the comparison is 91% compared to 79%. For length uniformity, the comparison is 93% compared to 79%.

Outliers

An "outlier" bale is any bale in a module that falls outside of acceptable tolerances for testing as determined by statistical methods. Table 2 shows the current tolerances to trigger an outlier. After testing concludes for all bales within a module, the central computer calculates the average for the factors. The computer then compares each of the bale's individual measurements to the average. If the difference for a given bale exceeds the allowable tolerance for a particular quality factor, the computer removes that bale from the module-average calculation. The fiber properties for the remaining bales are re-averaged and that value is assigned to all of the bales within the module, with some notable exceptions. The two primary exceptions are 1) outliers that occur as a first or last bale of a module; and 2) any outliers in the module if the total number of outliers present exceeds 20% of the total number of bales in the module.

Table 2. 2007 Outlier Tolerances.

Outlier Tolerances			
Micronaire (units)	Strength (grams/tex)	Length (inches)	Length Uniformity (percent)
± 0.3	± 3.0	± 0.04	± 2.0

Outlier bales represent only a very small percentage of the total amount of bales module-averaged each year. For example, in the 2006 crop, the overall number of outliers flagged for all four measurements combined represented approximately 4 percent of the number of bales module averaged. Similarly for 2005, the number of identified outliers represented just 3 percent. The Cotton Program has thoroughly analyzed these outliers over the years to determine their primary cause and frequency. There are two types of outliers – first-and-last-bale outliers and middle-bale outliers. In studying outliers that occur either as first or last bales of a module, data supports that a more significant percentage of these bales are likely to be true outliers than those coming from the middle bale group. Further, in ongoing study of middle bale outliers, the Cotton Program has consistently found that a very large majority of these outliers, when retested, closely approach the module average within acceptable testing tolerances. This indicates that the measurement that created the outlier was somehow flawed, an incorrect sample was tested or submitted, or some other anomaly occurred. Since 2002, all middle-bale outliers (except noted exceptions) receive the module average without requiring a review classification. Prior to that time, the Cotton Program required a new sample to be drawn and sent to the classing office for retest. Since such a large percentage of the retested bales closely reproduced the module average, the Cotton Program discontinued the retesting process in 2002, and began assigning the module average to the middle bale outliers. The exception to this rule occurs if a module contains more than 20% outliers of any kind. In this case, all outliers retain their original values. The Cotton Program implemented the "20% rule" to provide an additional safety net in rare occasions when atypically high numbers of outliers occur in larger modules. For example, in a 20-bale module, all outliers would retain their original values if more than four outliers of any kind exist. As previously mentioned, all first-and-last-bale outliers retain their original values in all cases. As before, customers can re-submit any outliers to the Cotton Program for review classification at no additional charge.

Outlier Statistics

The Cotton Program has evaluated statistics for outliers since the inception of the module averaging program. Over the years, the outlier statistics support the applied procedures of handling first-and-last-bale outliers and middle-bale outliers. The results show that when retested as checklot bales in the Quality Assurance Branch, the middle-bale outliers reproduced the module average at a much higher rate than the original individual single-run measurements regardless of the outlier tolerances used. In looking specifically at the 2006 and 2007 crops to illustrate, it is very clear that the reproducibility results associated with comparing the QA double-run tests to both the outliers' individual measurements and that of the module average yields solid verification that the original individual tests are significantly lower for all factors than that of the module average (Tables 3 and 5). This indicates that the individual

measurements tended to reflect testing or other error. As stated earlier, outliers can occur for a variety of reasons but ongoing Cotton Program studies and analyses indicate that most middle-bale outliers result from testing flaws or correctable error and not typically from cotton variability. The 2006 and 2007 statistics for middle bale outliers fall in line with past years' data analyses and continue to support the Program's position of automatically assigning the module average to middle-bale outliers except where procedural exceptions apply (see previous section).

In studying first-and-last-bale outliers (Tables 4 and 6), the numbers indicate that there is some additional volatility in the reproducibility between the QA retests and both the module averaging results and the individual results. The differences are not as extreme with the module averages, especially with strength and length uniformity. Although the reproducibility affiliated with the module averaged values for each measurement is improving over past years, the Cotton Program maintains its position that first-and-last-bale outliers should retain their original testing values.

Table 3. 2006-07 Middle-Bale Outliers
(QA Checklot Tests vs. Module Average Values and Individual Test Values from Classing Offices)

Middle Bale Outliers (QA Check Bales)
2006-07 Season

QA Double Run vs. Office Single Bale compared to QA Double Run vs. Office Module Average

Quality Factor	QA Double Run vs. Office Single Run	QA Double Run vs. Office Module Avg.
	Reproducibility	Reproducibility
Micronaire	27.5 %	82.3 %
Strength	32.6 %	97.8 %
Length	32.7 %	93.3 %
Length Uniformity	34.2 %	98.6 %

Table 4. 2006-07 First-and-Last-Bale Outliers
(QA Checklot Tests vs. Module Average Values and Individual Test Values from Classing Offices)

First/Last Bale Outliers (QA Check Bales)
2006-07 Season

QA Double Run vs. Office Single Bale compared to QA Double Run vs. Office Module Average

Quality Factor	QA Double Run vs. Office Single Run	QA Double Run vs. Office Module Avg.
	Reproducibility	Reproducibility
Micronaire	68.5 %	45.5 %
Strength	39.1 %	90.7 %
Length	54.9 %	69.2 %
Length Uniformity	50.0 %	95.2 %

Table 5. 2005-06 Middle-Bale Outliers
(QA Checklot Tests vs. Module Average Values and Individual Test Values from Classing Offices)

Middle Bale Outliers (QA Check Bales)
2005-06 Season

QA Double Run (checklot) vs. Office Single Bale compared to QA Double Run (checklot) vs. Office Module Average

Quality Factor	QA Double Run vs. Office Single Run	QA Double Run vs. Office Module Avg.
	Reproducibility	Reproducibility
Micronaire	21.0 %	85.6 %
Strength	31.8 %	94.8 %
Length	32.8 %	91.8 %
Length Uniformity	33.6 %	100 %

Table 6. 2005-06 First-and-Last-Bale Outliers
(QA Checklot Tests vs. Module Average Values and Individual Test Values from Classing Offices)

First/Last Bale Outliers (QA Check Bales)
2005-06 Season

QA Double Run vs. Office Single Bale compared to QA Double Run vs. Office Module Average

Quality Factor	QA Double Run vs. Office Single Run	QA Double Run vs. Office Module Avg.
	Reproducibility	Reproducibility
Micronaire	52.3 %	54.0 %
Strength	35.8 %	91.2 %
Length	47.8 %	74.5 %
Length Uniformity	46.0 %	97.0 %

Benefit and Value of Module Averaging

Data Product

Since its inception, module averaging has consistently demonstrated its benefits and value. One of the possible reasons for the current level of participation not being higher appears to be the lack of understanding of the benefits of the program to the industry. The primary value of module averaging is in the area of improved fiber testing accuracy. Since the beginning of module averaging in 1991, the data shows laboratory-to-laboratory and instrument-to-instrument reproducibilities are much higher with module averaging than with traditional single-bale testing. This correlates into much more dependable and reliable data for all users of the classing data. The inherent value is the confidence that the quality measurements of module-averaged bales will always circulate around the mean of that module if retested again and again. It is better for the producers of the cotton, agents and handlers of the cotton, and ultimately the mill or manufacturer that utilizes the fiber to know with the highest certainty available that the HVI measurements of micronaire, strength, length, and length uniformity are stable and reliable. Module averaging provides the certainty that assigning a fiber measurement based on all of the available information from an entire module is a much better representation than the individual bale measurement alone.

Benefits to the Cotton Program

Ultimately, the largest benefit of module averaging to the Cotton Program is an improved data product to the customers and the inherent value of data accuracy to the cotton industry. Providing the most accurate data possible is part of the overall mission of the Cotton Program and accurate data equates to more stable, repeatable, statistically reliable information for use by the cotton industry. Delivering this accurate data helps the Cotton Program meet its obligations for providing the best customer service possible.

Benefits to the Cotton Industry - Data Reliability

With the majority of the U.S. cotton going in the export market, accurate classification results that stand up to any and all scrutiny, challenge and/or retesting are imperative. This is especially true given the current global migration to instrument testing. Many countries, most significantly China, are not only moving to instrument testing but doing so in a manner that closely emulates the U.S. classification system. The USDA classification system and specifically the HVI testing is a very reliable and accurate system and the Cotton Program goes to great lengths to ensure the most accurate testing possible. Module averaging simply takes the already accurate testing and raises it to yet another level. Since it is proven to be more accurate overall than individual measurements, the data reliability for use in the international market is a tremendous benefit to the cotton industry and users of module averaged data.

Module averaging also provides added confidence to the end user of the raw cotton - the spinner. With high efficiency mills more prevalent worldwide and decreasing margins for error in the textile manufacturing industry, it has become essential for spinners to know exactly what they have in a laydown for a specific end product. This confidence helps to prevent operational problems that can cause fabric or yarn flaws or machinery downtime. Module averaging and its higher level of accuracy provide the level of confidence needed by the spinners especially when running high output mills with large laydowns.

Warehousing

In addition to the reliability of the data, there is also potential for improvements in the storage, staging, and shipping of cotton using module averaged data. Since entire modules carry the same quality measurements, larger lots of bales could be blocked at warehouses according to the quality ranges being ordered for shipment. Lots bound for containers at the shipping port could be block stacked and staged using the module average, making for more efficiency storage and reducing the necessity to pick out individual bales from different parts of a warehouse to fill a particular order. If cotton bales could be stored and staged according to their module averages, the potential exists for larger quantities of specified quality ranges to be segregated for shipment.

Additional Measurements

Cotton Program studies have shown that using module averaging as the basis for testing could yield a potential measurement for short fiber that would be viable for the industry. When using a normalized version of the module average for the HVI measurement of short fiber, the measurement becomes a much more repeatable and reliable test. This is because the instrument is only measuring two specimens for each sample and applying an algorithmic analysis to get the short fiber measurement. Based on just two tests, the Cotton Program has been able to reproduce testing results within allowable testing tolerances about 50% of the time. This is due primarily to the level of variability associated with a short fiber measurement. However, when utilizing all of the bales within a module, the number of short fiber tests multiplies significantly and the results are then statistically normalized to further reduce error. Studies have shown that when module averaged and normalized, these results can reproduce themselves between instruments and between testing laboratories well over 80% of the time. This scenario could possibly be applied to additional measurements in the future if the industry deems it necessary.

Economic Benefit

The other potential value with module averaging is a financial benefit to the producer or ginner. Over the past several years, the Cotton Program has performed analyses to determine the difference in loan value between module-averaged data as compared to the original individual bale data. This is an internal study and in no way guarantees economic benefit but does provide some valuable insight into the potential financial benefit of module averaging.

In looking at the 2007 crop data for illustration, a total of 3,666,989 bales were module averaged as of January 4, 2008. Applying the 2007 loan schedule as a basis for reference, we can compare single bale test data to module averaged bale data for economic benefit. For the overall crop, the module average yielded advantages of 2.9 points per pound for micronaire; 0.8 points per pound for strength; 4.2 points per pound for length; and 1.4 points per pound for length uniformity. This was a total of 9.3 points per pound (or $0.00093 cents per pound) increase in the value of the module average versus the single bale test results. Applying these figures to the crop, we conclude an economic advantage of $0.4663 per bale (500 pound) and an overall advantage for the module averaged crop of just over $1.7 million.

Further, taking the same figures and applying them to the entire 2007 crop classed to date (14,960,578 bales as of January 4, 2008) we conclude that had all of these bales been module averaged, an overall potential economic benefit of $0.5089 per bale would have resulted. This equates to a total economic benefit of $7.6 million for the entire crop. Looking at individual regions and the 12 USDA classing offices, we find a range in benefits from 1.9 to 14.8 points per pound (all four fiber properties combined) with no area showing a loss in points per bale compared to the singe bale test method.

Conclusions

Since its inception in 1991 and expansion in 1992, the module averaging program has been a successful method of reducing variability in the measurements of cotton fiber micronaire, strength, length, and length uniformity. Analysis comparing HVI results between classing laboratories and the Quality Assurance Branch over the years has shown, without exception, the reproducibility is always higher for module averaged testing over the individual testing. This in no way takes away from the reliability of the individual tests made by the HVI and the USDA classification system. However, module averaging uses statistical principles to provide an even more accurate measurement for fiber measurements. The Cotton Program has always taken a proactive approach in implementing any method that could provide more reliable classification results to its customers. For that reason, it has continually strived to promote and expand participation in the module average program.

Ultimately, our desire to see expansion in the acceptance and participation in the module averaging program equates to our mission to deliver the most accurate cotton classification services possible to our customers in order to assist in the marketing of that cotton worldwide. Actually, the module averaging program creates additional steps in our internal classification process and does not reduce workload. However, we feel that the module averaging program is important and beneficial enough to the cotton industry to warrant the additional steps necessary in its operations.

The Cotton Program will continue to analyze the crop's data each year to reaffirm the principals behind module averaging. Even though these have proven valid since the inception of module averaging, we owe it to all segments to ensure that the supporting data and statistics remain true. We remain convinced that the module averaging program is beneficial to all segments by better assuring that the data utilized in the marketing and processing chain is the most accurate possible.

References

Earnest, D. 2003. Module Averaging – A New Perspective. Proceedings of the 2003 Beltwide Cotton Conferences, pp. 677-684.

Gibson, W. 2003. Futures Classification and Quality Assurance in USDA, AMS, Cotton Program. Proceedings of the 2003 Beltwide Cotton Conferences, pp. 673-676.