# IMPROVED PREDICTION OF LEAF GRADE FROM REAL TIME GIN-BASED MEASUREMENTS

**Richard K. Byler**
**Agricultural Engineer**
**W. Stanley Anthony**
**Supervisory Agricultural Engineer**
**USDA-ARS, Cotton Ginning Research Unit**
**Stoneville, MS**

## Abstract

The economic value of cotton to the farmer greatly depends on the quality as determined by the United States Department of Agriculture (USDA), Agricultural Marketing Service (AMS), Cotton Division, Cotton Classing Office (CO). Optimization of gin processing depends on the accurate prediction of the AMS CO quality measurements in real time in the gin. Cotton samples with a wide variety of trash content were generated with a variety of gin cleaning machinery and sent to the Dumas CO for classing. Data were collected while ginning the samples with equipment similar to that used in gin process control in the Microgin facility at the U.S. Cotton Ginning Laboratory, Stoneville, MS. About 17 readings of samples of each of the 122 lots were made. Similar data from a commercial gin were also analyzed. Several models were evaluated and one model was chosen to predict the AMS CO leaf grade based on readings made while ginning. This model predicted leaf grade correctly 64.5% of the time and was within one grade 99% of the time. This was an improvement over the 40-60% correct predictions experienced in 1994-1996 based on AMS CO data from 1991-1992.

## Introduction

The gin process control system improves monetary returns to the farmer by ginning the cotton with the most advantageous set of equipment (Anthony et al., 1995). To do this reliably, the basic parameters which determine the sale price of the cotton must be measured or predicted at the gin (Anthony 1990). Because the gin has very little effect on micronaire and strength, those variables are not measured in the current control system but are considered to be constant. The impact of ginning on length is predicted because we have no way to measure length in real time at the gin. The remaining parameters include color, trash content, moisture content, weight, and pricing structure. Predicting the pricing structure at the time of product sale is an important part of the control system, but is not an engineering problem. This decision is left to the farmer or the gin manager (Byler and Anthony 1997). The remaining properties are measured by the system. The purpose of this work was to develop a mathematical model to correlate the manual leaf grade assessed at the AMS CO to trash measurements made in the gin by High Volume Instrumentation (HVI)-type equipment.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 2:1572-1575 (1998)
National Cotton Council, Memphis TN

## Discussion

In the past, the leaf grade has been predicted based on the measured trash content at the last station in the gin. The percent of the total area seen by the camera that is dark is considered the percent trash and is one of the two basic trash measurements. The other measurement is the count which is an indication of the number of trash particles in the sample. From 1993 through 1996 in the Cotton Ginning Laboratory's process control projects, the leaf grade has been determined from data as shown in Table 1. This table was based on the average percent area data of a large number of samples with the corresponding leaf grade values. Unfortunately, the table is not as good at predicting the leaf grade as desired. For example, at one gin in 1994, data from about 25,000 bales were analyzed and 59.6% had the correct leaf grade prediction (Anthony, Byler, and Howard 1995). In 1995, data from about 16,000 bales were analyzed and 45.7% were predicted correctly (Anthony, Byler, and Howard 1996). In 1996, analysis of about 30,000 bales showed that the prediction was correct 52.1% of the time (Anthony, Byler, and Howard 1997).

There are several problems in improving the prediction of leaf grade. One is that a visual judgement by a human operator is to be predicted by the machine readings. To compound problems at the gin level, the same samples are not used for the two readings, so the actual samples upon which the trash level is based are different. The main problem in analyzing data from the gin and comparing it to the leaf grade is that the range of the data is limited for cotton processed conventionally at the gin. In fact, much of the AMS CO data sufficiently far from the mean to provide real improvement in the prediction is actually in error. This is because the gin attempts to clean the cotton, and actually does clean it well, to leaf grade of 3 or 4. So, any reading above 4 is more likely to be in error than one with leaf grade of 3.

### Microgin

In the spring of 1997, a set of lots of cotton was created intentionally using fewer than normal cleaning machines to ensure a wider range of trash levels in the ginned lint. This data set included lots which were much trashier than normally seen in a gin, as well as normally ginned cotton with normal trash levels. In addition, three separate samples were sent to the classing office for separate readings of leaf grade for each lot of cotton, and there were three lots for each set of cleaning machinery.

Work began on the data from the Microgin's station 4 which corresponds to the color/trash meter immediately before the battery condenser at the press in a gin. Each lot corresponded to a specific set of cleaning equipment. The reflectance (Rd) mean was taken for each lot. Generally, there were about 18 Rd observations for each lot at station 4. Experience has shown that the best judge of whether a sample is sufficient to obtain valid readings is to compare the individual Rd reading with the mean of all Rd measurements for a lot. Poor samples will have a low Rd value. Individual readings with Rd of less than 2 below the mean were discarded. In the data set before this filtering, there were 2,396 observations and after filtering, there were 2,092 observations. As expected, the Rd mean increased slightly and the trash measurements were reduced slightly when the data was filtered. This process removed almost 13% of the data, but it is believed that the data which was removed was from small or poorly formed samples. Each filtered lot contained an average of over 17 observations. The mean of each variable was taken and used in further analysis as the best estimate of the actual variable value.

Next, the data from the Dumas CO were examined. There were three separate samples sent for classing from each lot, and the mean of the three readings was used for subsequent analyses. Table 2 shows that the processing of the data changed the distribution only slightly. It should be noted that the distribution of data used in this study was very different from the distribution seen for cotton ginned normally, Table 2. There were many more observations in this data set with leaf grade above 4 than in normally ginned cotton. The data set from the Microgin was then combined with the data from the Dumas CO by lot. Comparisons of the percent area measurements showed that those made in the Microgin were lower than those made at the Dumas CO.

The percent area measurement at the CO and from the gin were used to predict the leaf grade. Equations 1 and 2 were obtained by regression using the SAS (1996) procedure GLM.

$$DleafM = 2.515 + 5.1072*Garea \quad 1)$$

where: DleafM = the mean of the three leaf grade readings at the Dumas CO,
Garea = the percent area measured in the gin, whole percent.

$$DleafM = 2.095 + 4.703*Darea \quad 2)$$

where: Darea = the percent area measured at the Dumas CO, whole percent.

Next a model was formed to predict the leaf reading from the Dumas CO based on readings at station 4 in the Microgin. Initially, the model included percent area, (percent area)$^2$, count, count$^2$, log(percent area), (percent area)*count, Rd, and +b. The variables were removed from the model one at a time, each time removing the one that contributed least to the prediction resulting in equation 3.

$$DleafM = 3.22 + 0.06269*GCT + 0.47368*LGPA \quad 3)$$

where: GCT = the count measured in the gin, and
LGPA = the log of the percent area measured in the gin, whole percent.

In the model the log of percent area contributed the least and so an additional model was created with only the count and intercept:

$$DleafM = 1.78 + 0.0898*GCT \quad 4)$$

Some goodness of fit statistics are contained in Table 3. The statistics all show that equation 3, using both the percent area and trash count, is the best model of those listed. Neither equation using the percent area alone was as good as the model using trash count alone. In addition, using the percent area measured at the CO did not predict the leaf grade better than using the percent area measured in the Microgin. Table 4 shows the results when the four equations were used to predict the leaf grade. This data shows that the prediction based on the percent area alone was rather poor, and the model based on trash count alone was not too bad. However, equation 3 was clearly the best model. These analyses are based on evaluating the effectiveness of the models on the data with which they were developed.

**Servico**

The next step was to apply equations 3 and 4 to the data set from a commercial gin (Servico Gin) obtained in the fall of 1996. This data set had one value of leaf grade from the Birmingham CO per bale for 29614 bales. The mean for each bale of the approximately 12 readings of percent area and trash count was used for the data collected in the gin. Table 5 shows the distribution of leaf grade for this data. This distribution is typical for this variable. It also shows the predictions used for control at the gin during the fall of 1996. The predicted values were significantly higher than those at the classing office. The last column shows the predicted leaf grade based on equation 3. It is observed to be even higher than the model currently used.

The Birmingham CO leaf grade was considered the correct value and the two predictions were used to calculate prediction errors, Table 6. As expected based on the distributions in Table 5, both models predicted leaf grades that were too high. The prediction model used was correct 52% of the time and correct within one grade 94.1% of the time. Of course, 98.5% of the grades were within one grade of leaf 3.

Because the mean leaf grade predicted with equations 3 and 4 did not agree with the mean of the observed leaf grade, the intercepts of equations 3 and 4 were adjusted to yield equations 5 and 6. The means of the predicted leaf grade from these equations agreed with the mean of the observed leaf grade.

$$DleafM = 1.91 + 0.06269*GCT + 0.47368*LGPA \quad 5)$$

$$DleafM = 0.64 + 0.0898*GCT \quad 6)$$

Table 7 shows the distribution of the leaf grade at the Birmingham CO as well as the distribution of the predictions by equations 5 and 6. By comparing this data to Table 5, these predictions are considerably better than those used previously.

Table 8 shows that there is not much difference in the predicting power between the two models. Equation 5 is slightly better and should be evaluated on data from other gins, classing offices, and years. The difficulty of this approach is that the mean of the leaf grade is needed so that the intercept can be adjusted. Equation 5 predicted 64.5% of the data correctly and 99.0% to within one grade. This compares favorably with the prediction of 3 for the leaf grade which would have been correct 64.6% of the time and 98.5% of the grades within one grade of leaf 3. Based on the fit of the data for the Microgin and the Dumas CO, we can be assured that these new models will predict high leaf grades when appropriate.

## Summary

The purpose of this work was to develop a mathematical model to correlate the manual leaf grade assessed at the AMS CO to trash measurements made in the gin by HVI-type equipment. Samples of cotton with a wide range of foreign matter content were generated in the Microgin using different cleaning sequences. Online data were collected during sample generation and compared to static data determined at the AMS CO at Dumas, AR. About 17 online readings were used for each of the subsamples from 122 cotton treatments. Several mathematical relationships between leaf grade, percent area, and count measured at the gin were developed. The equation,

$$D\_leaf\ M = 1.91 + 0.06269 * GCT + 0.47368 * LGPA$$

was judged to be the most appropriate since it predicted the leaf grade correctly 64.5% of the time and was within one grade 99% of the time. This was better than the 40-60% correct predictions obtained in 1994-1996 using AMS established relationships on data collected at Servico Gin. A model using both the trash count and percent area was considerably better than one using percent area alone.

## Disclaimer

Mention of a trade name, proprietary product, or specific machinery does not constitute a guarantee or warranty by the U.S. Department of Agriculture and does not imply approval of the product to the exclusion of others that may be available.

## References


Anthony, W. S. 1990. Computerized gin process control. Applied Engineering in Agriculture. 6(1):12-18. American Society of Agricultural Engineers, St. Joseph, MO.

Anthony, W.S., R.K. Byler, L. Deavenport, and D. Scamardo. 1995. Experiences with gin process control in the Midsouth and West. Applied Engineering in Agriculture. 11(3):409-414. American Society of Agricultural Engineers, St. Joseph, MO.

Anthony, W. S., R. K. Byler, and N. L. Howard. 1995. Field Evaluation of video techniques for determining color and trash content of cotton. Unpublished Annual Report for FY 1995, pp. 273-289. U. S. Cotton Ginning Research Unit, Stoneville, MS.

Anthony, W. S., R. K. Byler, and N. L. Howard. 1996. Field Evaluation of video techniques for determining color and trash content of cotton. Unpublished Annual Report for FY 1996, pp. 123-141. U. S. Cotton Ginning Research Unit, Stoneville, MS.

Anthony, W. S., R. K. Byler, and N. L. Howard. 1997. Field Evaluation of video techniques for determining color and trash content of cotton. Unpublished Annual Report for FY 1997, pp. 176-189. U. S. Cotton Ginning Research Unit, Stoneville, MS.

Byler, R. K. and W. S. Anthony. 1997. Measurement concepts in a gin process control system, 1996. Proc. Beltwide Cotton Prod. Res. Conf., pp. 1572-1573. National Cotton Council, Memphis, TN.

SAS Institute, Inc. 1996. SAS, Version 6.12. Cary, NC.


Table 1. Conversion from measured percent area to predicted leaf grade used in process control.

| Percent area range | leaf grade |
|---|---|
| 0.00- <0.15 | 2 |
| 0.15 - <0.35 | 3 |
| 0.35 - <0.55 | 4 |
| 0.55 - <0.85 | 5 |
| 0.85 - <1.25 | 6 |
| 1.25 - <2.35 | 7 |

Table 2. Percentage of observations in each leaf category before and after averaging by lot as well as the overall mean for the entire cotton crop classed at Dumas AMS CO.

| Leaf level | Dumas leaf, before averaging | Average leaf | All Dumas classing, 1996 |
|---|---|---|---|
| 1-2 | 11.7 | 11.5 | 29.9 |
| 3 | 17.8 | 18.0 | 54.2 |
| 4 | 13.4 | 12.3 | 14.3 |
| 5 | 26.0 | 28.7 | 1.1 |
| 6 | 19.9 | 18.0 | trace |
| 7 | 7.7 | 9.8 | trace |
| 8 | 3.6 | 1.6 | trace |
| Number of values | 378 | 122 | 649,409 |

Table 3. Tests for goodness of fit for the equation to the data set with
which it was developed.

| | $R^2$ | Root mean square error |
|---|---|---|
| Equation 1 | 0.830 | 0.656 |
| Equation 2 | 0.744 | 0.806 |
| Equation 3 | 0.916 | 0.464 |
| Equation 4 | 0.896 | 0.513 |

Table 4. Percent distribution of the error in predicting Dumas CO leaf grade.

| Error in predicted leaf | Equation | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| -3 | 0.8 | 0.8 | - | - |
| -2 | 0.8 | 1.6 | 0.8 | 1.6 |
| -1 | 18.3 | 23.0 | 16.4 | 18.0 |
| 0 | 54.0 | 42.9 | 64.8 | 61.5 |
| 1 | 25.4 | 29.4 | 18.0 | 18.9 |
| 2 | 0.8 | 2.4 | - | - |

Table 5. Percentage of observations within a given leaf grade of the data for a commercial gin.

| Leaf grad e | Birmingham Classing Office | Prediction based on Table 1 | Prediction based on equation 3 |
|---|---|---|---|
| 1-2 | 2.7 | 0.3 | 0.0 |
| 3 | 64.6 | 39.8 | 0.7 |
| 4 | 31.2 | 47.8 | 43.3 |
| 5 | 1.5 | 10.5 | 49.7 |
| 6 | 0.0 | 1.2 | 6.1 |
| 7 | - | 0.5 | 0.4 |

Table 6. Percentage distribution of the predicted leaf grade subtracted from the classing office leaf grade for a commercial gin.

| Error | Prediction based on Table 1 | Prediction based on equation 3 |
|---|---|---|
| -4 | 0.3 | 0.1 |
| -3 | 0.9 | 2.0 |
| -2 | 4.7 | 35.7 |
| -1 | 35.8 | 56.9 |
| 0 | 52.0 | 7.3 |
| +1 | 6.3 | 0.0 |
| +2 | 0.1 | 0.0 |

Table 7. Percentage of total bales predicted for a given leaf grade for equations 5 and 6 for a commercial gin.

| Leaf | Birmingham Classing Office | Equation | |
|---|---|---|---|
| | | 5 | 6 |
| 1-2 | 2.7 | 4.6 | 5.0 |
| 3 | 64.6 | 62.2 | 60.5 |
| 4 | 31.2 | 30.4 | 31.0 |
| 5 | 1.5 | 2.6 | 3.1 |
| 6-7 | 0.0 | 0.2 | 0.4 |

Table 8. Percent distribution of leaf grade prediction error for equations 5 and 6 for a commercial gin for 29614 bales.

| Error in predicted leaf grade | Equation | |
|---|---|---|
| | 5 | 6 |
| -2 | 0.7 | 0.9 |
| -1 | 16.7 | 17.9 |
| 0 | 64.5 | 63.4 |
| 1 | 17.7 | 17.5 |
| 2 | 0.3 | 0.3 |