EXPLORING COTTON BALE FIBER QUALITY VARIABILITY WITHIN ROUND MODULES EMPLOYING JOHN DEERE'S HID SYSTEM

Luke Fuhrer
W. M. Porter
G. Rains
J. L. Snider
S. Virk
J. He
University of Georgia
Tifton, GA
E. Barnes
Cotton Incorporated
Cary, NC

Abstract

John Deere's Harvest Identification system (HID) has allowed for the visualization of georeferenced cotton fiber quality at the module resolution. Visualization of this data is only half of the advantage to this dataset and system. Understanding why there are differences between modules and bales is the next step. The main goal of this project was to utilize round modules tracked through the HID system, and to explore the variability of the individual four bales created during the ginning process from each round module. To begin the data processing, a module's average fiber quality must be determined. This is done similarly to an early 2000's technique of module averaging. This is accomplished by taking the bale report and averaging all the fiber parameters for all bales, that are produced from a specific module. Modules are identified using the gin's unique label. The next step in understanding this data is knowing if module averaging is effectively representing the fiber quality data from the individual bales for a single module. To determine the module variability standard deviations and uncertainty analysis can be performed for the cotton bales created from each module as well as for a field total. This allows for a statistical view of the performance of averaging module fiber quality. It also shows different fiber quality parameters render at different levels, higher or lower when averaged. With current bale data and statistics, it appears that bales are relatively consistent with fiber quality and module averaging is effective in its display of fiber quality. This more statistical approach to the data begins to answer the question if averaged module fiber quality is capable of showing the variability within a cotton module.

Introduction

Through the development of a procedure to handle John Deere's Harvest Identification (HID) data (Fuhrer, 2020 & Fuhrer, 2021), a more focused understanding of the spatial variation of fiber quality can be explored. The spatial maps of the HID and fiber quality data are still only at the module level resolution. This is due to the module averaging that occurs which averages all bales from a specific module. This allows for the module's average fiber quality to be displayed at each point it represents. Module averaging, a program also offered commercially by the USDA, has been in place since 1991. This is a voluntary program that growers can be a part of, if desired. Each bale submitted to the program is compared to the running average, and outliers are identified. This program though only applies this average to certain fiber parameters. These parameters include fiber strength, micronaire, length, and uniformity. Statistical studies show that module averaging improves the accuracy of quality measurements (2019, USDA-AMS). Questions were posed to the module averaging effectiveness at displaying the variation in quality between all, typically four, bales from each module. It also prompted further questioning of which parameters performed better than others in this averaged form. Through this, the averaging of round modules on a single module basis will also gain more credibility to the program. Currently the USDA reports that in 2019 roughly 55% of cotton growers utilize module averaging. While the two largest production states, Texas and Georgia, only have around 40% of growers opting in for this program. With more statistics and knowledge on this area growers may become more confident in the program or help educate them on the opportunities available to them.

Objectives

The main objective of this study was to determine if averaged modules accurately show the true variability of fiber quality between bales from each module. The second part of the objective expands further and asks the question of how does this effect the HID fiber quality mapping. The last objective of this study is to identify parameters that do show this variability or not accurately.

Materials and Methods

The materials for this study are limited to Excel, a python coding platform, and grower bale reports. The field of interest in 2020, a grower's on farm trail, was located in Colquitt, GA. This field produced 25 modules which were used in a series of statistical analysis to evaluate the effectiveness of module averaging of round modules. This field was chosen due to an on-farm trail performed and carried out by the grower. Three seed varieties and four seeding rates were tested on the 45-acre field. Once the field was harvested, modules were scanned using a radio frequency identification (RFID) reader and labeled. These modules were then transported to the Clover Leaf Gin in Donalsonville, GA. This gin labels each module with a number identifier that includes the grower's field code and module number. This code generated by the gin, is recorded and related back to the project label. This identifier is used by the gin when collecting samples from each bale, from a specific module, for fiber classing. Once classing is complete the bale report can be generated for all the bales from a particular field. Each bale produced from a specific module are assigned the same load number. Typically, four bales per module are produced, but can vary due to module size.

The methodology does rely on the python program written previously in Fuhrer et al. 2021 for module averaging. This program allows for a quick and effective solution for averaging fiber parameters by module load number. The code utilizes the load number column and groups the bales. Each load number relates back to a single module. It then takes the mean for each specified parameter by the number of bales in that group. Finally, the program exports a new Excel file which contains these averaged values for each module. This code was adapted to perform a standard deviation function as opposed to the averaging function for this study. This allowed for the standard deviation to be examined for each parameter, for each module. The standard deviation begins to explain how much variation is present in each of the parameters. This was also graphed, and error bars used to give a visual depiction of the data range. This visual depiction gives an easy overview of the data, and allows for the identification of trends if present.
Utilizing the standard deviation values, an uncertainty analysis was also be performed. Due to a new Excel file being made with the standard deviation program, a new sheet was added to this file. This allowed for the reference to the standard deviation values much efficient. The equation (standard deviation / $\sqrt{n}$) is used for this analysis. The variable n represents the number of bales per module. The number of bales produced from each module needs to be notated to correctly input the value for n. The equation can be copied and pasted to each row with the correct referenced standard deviation inserted. This simple calculator created a much easier process to gather this data, as well as an easier and organized data view. This analysis is used to explain the variability of the output due to the variability of the inputs. In this case the module averaged values are explained by the variability of the inputted fiber quality. While standard deviation explains the single sample variation, the uncertainty analysis gives a similar view but between the samples.

Results

Utilizing a bale report from the 2020 season from our field of interest, 25 modules were tracked through the ginning process. The standard deviation portion of the study identified two parameters we need to consider not averaging. This can be seen in Figure 1. In this figure the length and loan value columns are partially or fully highlighted. A condition was put over the cells to highlight each cell that was outside of the range of two standard deviations (-2.00 to 2.00). Assuming this dataset is distributed normally, it would account for 95% of data. This range was used to identify parameters that were likely on the tail ends of the distribution or potentially outliers. The graphic depiction of length, which shows high levels of variation, can be seen in Figure 2.

	Standard Deviation of Fiber Quality Parameters by Module								
Module #	Lf	Mic	Str	Rd	b	Tr	Unif	Len	Loan Rat	Loan Value
1	0.50	0.22	1.00	0.54	0.10	0.82	1.12	1.71	0.60	10.13
2	0.50	0.06	0.74	0.53	0.17	0.50	1.05	2.89	0.48	4.29
3	0.50	0.10	0.28	0.45	0.26	0.50	0.79	1.83	0.11	4.11
4	0.50	0.17	1.20	0.41	0.23	0.82	0.93	3.42	0.66	10.01
5	0.50	0.06	1.29	0.57	0.25	0.58	0.98	2.45	0.79	11.51
6	0.50	0.05	0.98	0.22	0.10	0.50	1.01	1.26	0.13	4.60
7	0.00	0.06	0.65	0.25	0.12	0.58	0.61	0.82	0.25	3.08
8	0.00	0.06	0.38	0.50	0.15	0.00	0.67	2.63	0.02	11.63
9	0.00	0.10	1.21	0.13	0.05	0.00	0.24	2.16	0.17	6.04
10	0.00	0.06	1.56	0.50	0.22	0.00	0.76	2.22	0.78	7.06
11	0.00	0.08	1.17	0.33	0.06	0.00	0.45	0.96	0.24	9.40
12	0.00	0.06	0.63	0.17	0.13	0.50	1.20	1.71	0.51	3.82
13	0.00	0.05	1.07	0.34	0.08	0.58	0.94	1.15	0.18	8.95
14	0.00	0.08	1.79	0.10	0.08	0.58	1.09	2.36	0.73	2.90
15	0.00	0.13	0.78	0.24	0.13	0.58	1.08	0.82	0.25	7.54
16	0.00	0.00	0.90	0.25	0.14	0.50	0.77	1.50	0.14	9.68
17	0.50	0.05	0.87	0.21	0.10	0.96	0.59	0.82	0.33	5.91
18	0.00	0.05	1.01	0.55	0.41	0.50	0.76	2.22	0.24	16.64
19	0.50	0.14	0.77	0.25	0.10	0.82	0.71	2.06	0.19	7.90
20	0.00	0.08	0.26	0.17	0.17	0.50	1.06	1.00	0.04	6.09
21	0.00	0.10	0.80	0.39	0.34	0.50	0.41	4.50	0.76	2.50
22	0.00	0.06	0.74	0.83	0.17	0.00	0.93	3.21	0.33	4.03
23	0.00	0.06	0.40	0.26	0.06	0.58	0.55	1.00	0.25	3.30
24	0.00	0.05	0.51	0.41	0.08	0.58	0.97	0.96	0.36	4.57
25	0.00	0.06	0.76	0.25	0.06	0.58	0.60	2.08	0.56	8.83

If cell is highlighted, this means it is outside of range (-2.00 - 2.00)

Figure 1. A depiction of the output of the standard deviations by parameter for each module.

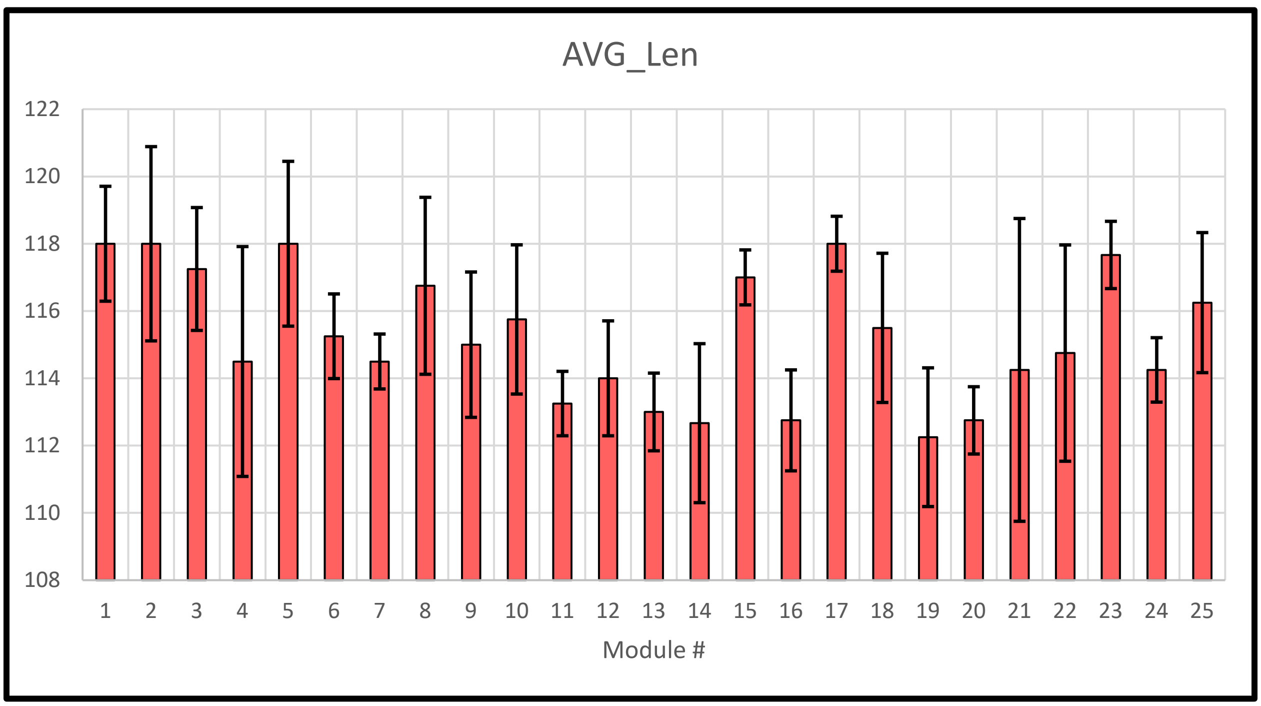

Figure 2. A graphic representation of the average standard deviation and value ranges for the length parameter.

The uncertainty analysis was used as a metric for further explanation of variability between samples. With the uncertainty analysis using the standard deviation in the equation, and having such a small sample size, it followed a similar trend as the standard deviation. The highlighted cells from Figure 1 were also cells of interest in this analysis. Figure 3 shows the uncertainty analysis. This shows that the variation between the samples of both the length and loan value are too great to confidently be explained by the averaged value. In a summarized form in Figure 4, the total average standard deviation and uncertainty analysis for all the 25 modules are displayed for each parameter. This smaller tabular output shows two major results. It shows that for our standard deviation for the length and loan value are still substantially higher than the other parameters. It also shows that with a larger sample size (n), the margin of difference for length and loan value are reduced in an uncertainty analysis.

	Uncertainty Analysis of Fiber Quality Parameters by Module									
Module #	Lf	Mic	Str	Rd	b	Tr	Unif	Len	Loan Rat	Loan Value
1	0.250	0.108	0.499	0.272	0.048	0.408	0.559	0.854	0.299	5.064
2	0.250	0.029	0.368	0.263	0.085	0.250	0.523	1.443	0.238	2.147
3	0.250	0.048	0.138	0.227	0.131	0.250	0.394	0.913	0.054	2.057
4	0.250	0.085	0.601	0.204	0.115	0.408	0.466	1.708	0.329	5.003
5	0.250	0.029	0.644	0.286	0.125	0.289	0.492	1.225	0.397	5.756
6	0.250	0.025	0.491	0.108	0.048	0.250	0.507	0.629	0.063	2.299
7	0.000	0.029	0.325	0.125	0.058	0.289	0.303	0.408	0.125	1.542
8	0.000	0.029	0.189	0.250	0.075	0.000	0.335	1.315	0.012	5.814
9	0.000	0.048	0.603	0.063	0.025	0.000	0.122	1.080	0.083	3.018
10	0.000	0.029	0.782	0.248	0.108	0.000	0.380	1.109	0.389	3.528
11	0.000	0.041	0.583	0.165	0.029	0.000	0.225	0.479	0.118	4.700
12	0.000	0.029	0.315	0.085	0.063	0.250	0.602	0.854	0.256	1.912
13	0.000	0.025	0.536	0.171	0.041	0.289	0.471	0.577	0.090	4.475
14	0.000	0.041	0.895	0.048	0.041	0.289	0.545	1.181	0.363	1.448
15	0.000	0.063	0.390	0.122	0.065	0.289	0.539	0.408	0.125	3.771
16	0.000	0.000	0.451	0.125	0.071	0.250	0.385	0.750	0.072	4.838
17	0.250	0.025	0.437	0.103	0.048	0.479	0.294	0.408	0.166	2.955
18	0.000	0.025	0.505	0.275	0.206	0.250	0.381	1.109	0.120	8.319
19	0.250	0.071	0.386	0.125	0.050	0.408	0.357	1.031	0.097	3.949
20	0.000	0.041	0.129	0.087	0.085	0.250	0.530	0.500	0.020	3.045
21	0.000	0.048	0.399	0.196	0.168	0.250	0.206	2.250	0.379	1.250
22	0.000	0.033	0.426	0.481	0.100	0.000	0.536	1.856	0.192	2.325
23	0.000	0.033	0.233	0.153	0.033	0.333	0.318	0.577	0.145	1.906
24	0.000	0.025	0.255	0.204	0.041	0.289	0.487	0.479	0.180	2.283
25	0.000	0.033	0.441	0.145	0.033	0.333	0.348	1.202	0.321	5.099

Figure 3. A depiction of the Uncertainty Analysis by parameter for each module.

Total Uncertainty by Fiber Parameter		
Parameter	Average SD	Uncertainty
Lf	0.16	0.03
Mic	0.08	0.02
Str	0.87	0.17
Rd	0.35	0.07
b	0.15	0.03
Tr	0.48	0.10
Unif	0.81	0.16
Len	1.91	0.38
Loan Rat	0.36	0.07
Loan Value	6.98	1.40

Figure 4. A tabular output of the total averaged standard deviation and uncertainty analysis for all 25 modules by parameter.

Summary and Future Research

This study answered the question of, is module average fiber quality effectively showing the true bale fiber variability. It showed that a single round module can be averaged, and the average value generated accurately accounts for the range in data for most parameters. The findings further verify that the fiber quality can be spatially displayed using these averaged values. With both a single sample analysis and between samples analysis, both views of the data provide the same variability. This further explains that the length and loan value parameters should be excluded, as there was too great of variation in the data to be explained by a single value. To validate these findings, this will be done again using bale reports from the 2021 season. Along with doing single field study's such as this, potentially a multifield scope is necessary for further understandings and confirmation. Another future potential, the ginning order, which could also be a factor in fiber variation. This project was performed at typical three gin stand commercial gin. In commercial gins there is greater potential for blending in the first and last bales of each module. Further statistical analysis is necessary to see how much blending affects this. This order was not able to be determined for the 2020 season but will monitored for analysis for the 2021 season.

Acknowledgements

The author of this paper would like to acknowledge Cotton Inc. for their funding of this project. The author would also like to thank Russ Worsley for all his help with coordination of harvest and allowing the UGA precision ag team to collect another season of data with him. Lastly, the Clover Leaf Gin in Donalsonville, GA have been very helpful and contributed greatly with the data collection of this project.

References

Fuhrer L, Porter W, Barnes E, Rains G, Snider J (2020). Mapping of In-Field Cotton Fiber Quality Utilizing John Deere's Harvest Identification System (HID). Beltwide Cotton Conference, Austin, Tx.

Fuhrer L, Porter W, Rains G, Snider J, Virk S, Barnes E (2021). The Process of Using John Deere's HID System to Develop Field Level Fiber Quality Maps. 2021 ASABE AIM, Virtual.

Fuhrer L, Porter W, Rains G, Snider J, Virk S, Smith A, Liu Y, Barnes E, Holt C (2021). Using Information From the John Deere Cotton HID System to Aid in Making Production Decisions. Beltwide Cotton Conference, Virtual.

USDA-AMS (2019). Advantages of Module Averaging. Cotton & Tobacco.