Relationship of Fiber Properties to Vortex Yarn Quality via Partial Least Squares

Calvin Price, Herman Senter, Jonn Foulk, Gary Gamble, William Meredith

Federal Reserve Bank of Philadelphia, Philadelphia, PA; Clemson University, Clemson, SC; USDA-ARS Cotton Quality Research Station, USDA-ARS Crop Genetics and Production Research Unit,

Correspondence:
Calvin Price, email: cprice@gmail.com

ABSTRACT
The Cotton Quality Research Station (CQRS) of the USDA-ARS, recently completed a comprehensive study of the relationship of cotton fiber properties to the quality of spun yarn. The five year study, began in 2001, utilized commercial variety cotton grown, harvested and ginned in each of three major growing regions in the US (Georgia, Mississippi, and Texas). CQRS made extensive measurements of the raw cotton properties (both physical and chemical) of 154 lots of blended cotton. These lots were then spun into yarn in the CQRS laboratory by vortex spinning with several characteristics of the yarn and spinning efficiency measured for each lot. This study examines the use of a multivariate statistical method, partial least squares (PLS), to relate fiber properties to spun yarn quality for vortex spinning. Two different sets of predictors were used to forecast yarn quality response variables: one set being only HVI™ variables, and the second set consisting of both HVI™ and AFIS™ variables. The quality of predictions was not found to significantly change with the addition of AFIS™ variables.

INTRODUCTION
Cotton is a natural agricultural product whose chemical and physical properties cannot be completely controlled. Cotton quality is affected by cotton variety and growing conditions, which vary by year and harvesting location. Fiber processing and spinning can be affected by fiber properties. The following items need to be addressed in order to improve the utilization of cotton: 1.) new methods to more fully characterize cotton quality, 2.) assessment of the impact of cultivation practices and fiber varieties on fiber utilization, and 3.) relationship of fiber properties to utilization. Using our existing equipment, this study evaluated fiber properties and their relationship to processing performance and product quality. These fiber property results provide the potential to predict multiple yarn quality variables using statistical methods.

Cotton grading has progressed from subjective human classers to High Volume Instrumentation (HVI™). Practically all cotton grown in the United States is classed by United States Department of Agriculture (USDA), Agricultural Marketing Service (AMS). Fiber length, length uniformity, strength, micronaire, trash, Rd, and +b are officially classed on the HVI™ (Uster Technologies Inc., Knoxville, TN) according to established standards ([1]). HVI™ properties have been considered the most important and most readily acceptable factors in predicting cotton quality for subsequent textile processing and spinning. HVI™ data is primarily used by textile mills for bale lay downs and monitoring the spinning process ([2]).

Research relating to yarn typically involves HVI™ and AFIS™ fiber properties. The Advanced Fiber Information System (AFIS ™) (Uster Technologies Inc., Knoxville, TN) is a destructive method that aeromechanically opens fibers and separates fiber, trash, and dust for electro-optical measurements thus producing various distributions. AFIS™ analyzes fineness, upper quartile length (UQL), short fiber content (SFC), maturity ratio, neps (count/gram), and visible foreign matter (VFM). Many individuals have reported on the use of AFIS™ fiber data to study the effects of fiber properties on the quality of ring spun yarns. AFIS™ data is primarily used for monitoring the entire spinning system to maintain process quality ([3]).

Textile mills have different product end uses, management policies, and machinery types and choose their cotton based on suitability for spinning demonstrated by fewer ends down with acceptable strength levels in their individual products. Spinning quality of cotton is dependent upon a combination of fiber physical properties and other measurements. Yarn properties typically considered include strength, uniformity (evenness and defects), appearance, processing waste, and processing efficiency (ends down). Cotton properties are variable and how they influence yarn has often been an area of research. In 1921, Balls created a fiber property-based "Prediction Formula" to be used by both cotton producers and processors that would allow one to predict spinning test results from fiber quality properties ([4]). An important yarn quality measurement is strength that is determined by fiber strength and fiber interactions. Correlations of yarn strength to fiber cohesion were first presented by F. T. Peirce in 1946 ([5], [6]).

Vortex spinning is relatively new cotton spinning technology with Murata first demonstrating the economic potential of vortex spinning in 1997 by spinning a 15 tex yarn at 400 m/min ([7]). Vortex spinning drafts sliver to the desired yarn count via a four roller/apron drafting system after which fibers are sucked into a spiral orifice at the entrance of an air jet nozzle where fibers are twisted from the force of the air jet ([7]). Produced yarns are comparable to ring spun yarn and are resistant to pilling, demonstrate lower hairiness, better moisture absorption, and wash resistance ([8]).

The primary statistical method used in this paper is partial least squares (PLS). PLS is a technique that can be used to create predictive regression models for multiple response variables when there is a high degree of collinearity in the predictor variables, or when there are very many predictor variables ([9]). When the number of predictors is close to or even greater than the number of observations, the ordinary least squares (OLS) estimators of slope coefficients may be unstable or a unique solution may not exist. In order to overcome these deficiencies, the method of PLS can be used to yield stable results with superior predictive ability. Emphasis is on developing a model for prediction, with no cause and effect relationship assumed between the independent variables and dependent variables. Similarly, one is not able to judge which variables are important in "causing" any of the dependent variables. PLS is not a data reduction technique nor does it directly provide evidence or support as to which variables should be removed or as to which variables are not important in explaining the Y variables. However, PLS can be used with different sets of X variables to determine which sets lead to better predictions, and to determine how much better predictions are with different sets of X variables.

PLS replaces the original set of numerous X variables with a smaller set of orthogonal factors (also called components) that are extracted from these variables. This is similar to the more commonly used statistical procedure of principal components, where factors are extracted from a set of variables so as to maximize the variance accounted for in that set of variables. However, in PLS, the factors extracted from the X variables are chosen so as to maximize the covariance accounted for between the X and Y variables. Thus, for each number of components that could be pulled from X, there is a corresponding PLS model with its own set of predictions on all of the Y variables. Generally, these predictions are different than those that result from any other PLS model that uses a different number of components. The total number of components extracted from the X matrix ranges from 1 to the number of variables. As the number of components increases, the PLS model converges to the regular multiple regression model, where it becomes exactly the same when the number of components extracted is equal to the rank of the design matrix. It is often difficult to determine how many factors from the X matrix should be used in the optimal model. Some Y variables may be predicted best with one number of components, while other Y variables are predicted better with a different number of components.

Output after PLS regression includes information on the fit of each Y variable corresponding to each possible number of components pulled from the X matrix. So, for example, if we wanted to predict 5 Y variables from 20 X variables, the 1st Y variable would be fit with 20 different models, the 2nd Y variable would also be fit with 20 different models, as would the 3rd Y variable, and so on with the end result being a total of 100 different models and a corresponding 100 different measures of predictive ability. Consequently, we might find the 1st Y variable is predicted best with 7 components, the 2nd Y variable may be predicted best with 10 components, and so on. The measure of "best predictive ability" used in this study is called the Prediction Error Sum of Squares (PRESS) statistic.

To make our predictions we consider two main sets of fiber quality variables. Measurements on one group of 14 fiber qualities are performed on all cotton bales (HVI™), while measurements on 12 other fiber quality variables are often available (AFIS™). To evaluate cotton bale variability, HVI™ and AFIS™ measurements included the standard deviation (stdev) of each fiber quality measurement. The goal of this work was to make the best possible predictions on 12 yarn quality variables from the vortex spinning method. Benchmark determinations will determine how good predictions are with the initial set of 14 HVI™ variables; subsequently PLS will be performed with the combined 26 HVI™ and AFIS™ fiber quality variables to evaluate any changes in prediction quality.

Preliminary Inspection of the Data

A preliminary inspection of the data was performed on the 12 vortex yarn variables and 26 fiber quality variables. These fiber quality predictor variables were grouped into the following categories: HVI™ (14 variables) and AFIS™ (12 variables). The resultant database utilized in this analysis included 152 observations on 38 variables. All HVI™ (X) variable observations occurred in identical pairs, so the 152 yarn quality variable (Y) observations correspond to only 76 distinct HVI™ variable observations. Linear relationships existed between each yarn quality variable and all HVI™ variables, except for the variable micronaire, which demonstrates a quadratic relationship with many of the yarn quality variables (*Figure 1*). As a result, the square of micronaire was added as a fiber quality variable. References to "HVI™ variables" from hereon include the square of the micronaire variable.

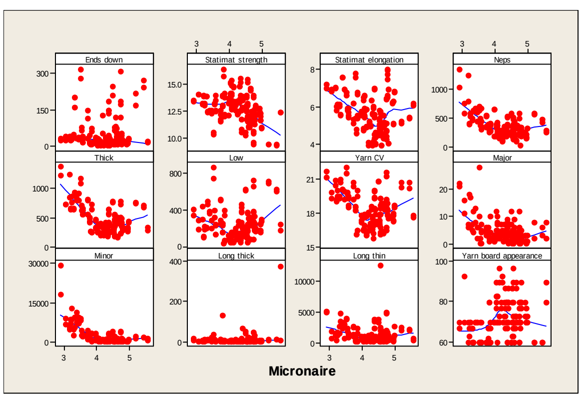

FIGURE 1. Scatter plot of ends down, Statimat strength, Statimat elongation, neps, thick, low, yarn CV, major, minor, long thick, long thin and yarn board appearance versus micronaire.

A search for outliers in the data was performed using a distance plot (*Figure 2*) which provides a scatter point corresponding to each observation. A point's distance along the horizontal axis signifies how close the observation is to the orthogonal components that were created to represent the fiber quality variables. Similarly, a point's distance along the vertical axis signifies the distance of the observation from the orthogonal components created to represent the yarn variables.

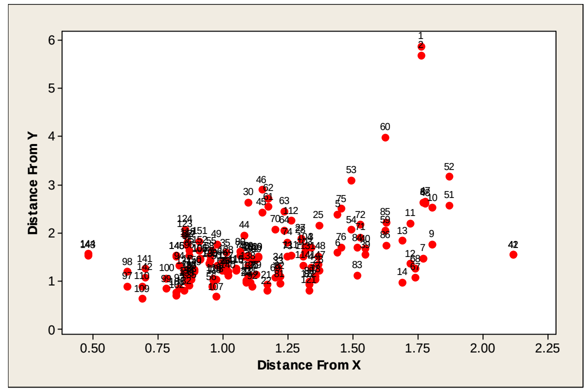

FIGURE 2. Distance plot, using 8 components.

In the distance plot, observations 41 and 42 are outliers with respect to the HVI™ variables, and observations 1, 2, and 41 are outliers with respect to the set of yarn quality variables. The analysis that follows was done in the absence of these 4 observations, and yielded essentially identical results as when they were included.

PLS With HVI™ Variables

Following a preliminary inspection of the data, PLS regression was performed using 12 yarn quality variables against 15 HVI™ variables. All yarn quality variables were modeled with 1-15 components, and measures of fit were created for every model. For each yarn quality variable, we need to choose the optimal number of components to use in making a predictive model for that variable. While this can be done in a variety of subjective ways, the approach to be used here is a combination of R-squared and predicted R-squared.

The predicted R-squared is a measure of predictive ability based on the PRESS statistic. Whereas the usual R-squared (also called fitted R-squared) uses the error sum of squares when accounting for variation not explained, predicted R-squared uses the PRESS statistic. Values for the PRESS statistic will always be larger than values for the regular error sum of squares and the predicted R-squared will always be smaller than the fitted R-squared. However, the number of components chosen will not be solely determined by which model has the highest predicted R-squared; a lower predicted R-squared will be accepted if it is accompanied by a significant increase in the fitted R-squared or a significant decrease in the number of components used. The ideal case includes an obvious ‘elbow’ where additional components begin to add only negligible increases to both predicted R-sq and fitted R-squared (see *Figure 3* for the Statimat strength response variable).

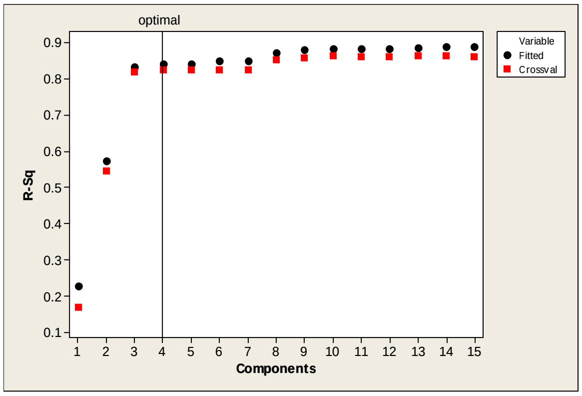

FIGURE 3. PLS model selection plot (response is Statimat strength).

Figure 3 demonstrates that additional components past 4 components yield only negligible increases to both predicted R-squared (denoted by “cross-validated”) and fitted R-squared. However, not all yarn variables were as clear as the Statimat strength response variable. For example, the long thin yarn variable (*Figure 4*) displayed an obvious elbow at 4 components for predicted R-squared, but the fitted R-squared still has considerable gains for a larger number of components; thus 9 components was selected as a compromise. It should be noted that the choice in number of components is not related to choosing which variables are important or how many variables are suitable for predicting the yarn quality. Regardless of whether 2 components or 10 components are chosen, every single component is a linear combination of all HVI™ variables.

After considering similar results for each of the 12 yarn variables, *Table I* summarizes the optimal number of components chosen for each variable.

TABLE I.
Number of components chosen in the PLS model for each variable using only HVI™ results.

Variables	Number of components	R-squared	Predicted R-squared
Ends down	10	0.60	0.51
Statimat strength	4	0.84	0.82
Statimat elongation	9	0.68	0.62
Neps	15	0.77	0.70
Thick	15	0.85	0.81
Low	10	0.77	0.72
Yarn CV	15	0.86	0.82
Major	10	0.43	0.30
Minor	15	0.70	0.60
Long thick	15	0.25	0.00
Long thin	9	0.45	0.37
Yarn board appearance	7	0.56	0.51
Mean		0.65	0.57

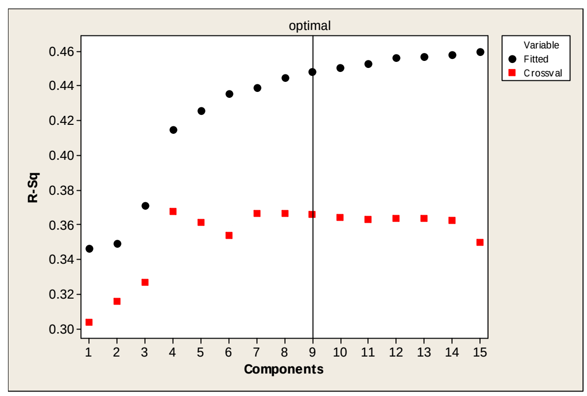

FIGURE 4. PLS model selection plot (response is long thin).

These results show that the long thick yarn variable is difficult to predict, as indicated by a 0.00 predicted R-squared. This result is not dependent at all on how many components were selected, as can be seen from *Figure 5*. Considering the low value of predicted R-squared and fitted R-squared values across all possible number of components, we conclude that this variable has essentially no relationship to the HVI™ variables and has very little of its variation explained by them. If the long thick variable is removed from the model we get better predictions on the remaining yarn quality variables because the components extracted from the data do not have to be created with the intent of trying to explain a variable that cannot be explained. In a sense, long thick is an outlier variable that skews the creation of the extracted components, which are in turn used in making predictions for all other yarn variables. As demonstrated in *Table II*, with the removal of the

TABLE II.
Number of components chosen in PLS model for variables using HVI™ results with variable long thick removed.

Variables	Number of Components	R-squared	Predicted R-squared
Ends down	10	0.60	0.51
Statimat strength	3	0.83	0.82
Statimat elongation	10	0.69	0.63
Neps 15		0.77	0.70
Thick 15		0.85	0.81
Low 11		0.78	0.73
Yarn CV	15	0.86	0.82
Major 10		0.43	0.30
Minor 15		0.70	0.60
Long thick	na	na	na
Long thin	8	0.45	0.37
Yarn board appearance	7	0.57	0.51
Mean		0.68	0.62

long thick yarn variable the average fitted R-squared improves by 3 percentage points (from 0.65 to 0.68) and the average predicted R-squared improves by 5 percentage points (from 0.57 to 0.62).

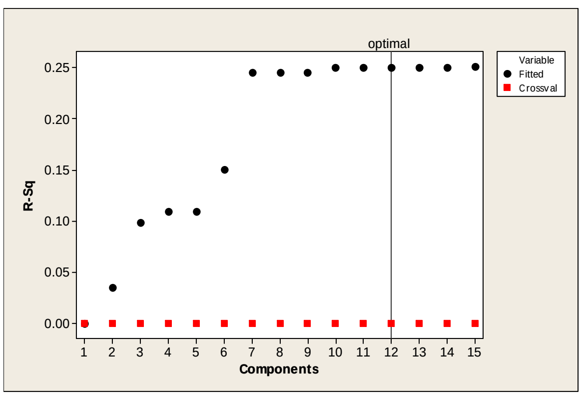

FIGURE 5. PLS Model selection plot (response is long thick).

The main results from this table are the average fitted and predicted R-squared values. These two mean R-squared values serve as the HVI™ benchmark for comparisons when we perform PLS with the additional set of fiber quality measurements (12 AFIS™ variables). If the corresponding numbers are significantly higher than these, we can consider AFIS™ variables to contribute significantly to predictions on the yarn quality response variables. Conversely, if the corresponding R-squared numbers are roughly the same (note that fitted R-squared cannot decrease when more variables are added), we will conclude that the AFIS™ variables do not contribute significantly to predictions on the yarn quality response variables when HVI™ variables are present.

In addition to assessing the fit and predictive ability of our model, we can make an indirect attempt at variable reduction by interpreting the correlations (also known as loadings) of the fiber quality variables with each of the extracted components. A component is a new variable created as a weighted sum of the original variables, where the weights are determined according to some specific criteria. The PLS components are created so that the 1st component explains the largest amount of covariance of the original variables, and the 2nd component explains the 2nd largest amount of covariance, and so on. Thus, if there is a certain group of variables that correlate highly on the 1st PLS component, we may be able to attach an interpretation based on similarities between the variables, and conclude that the underlying common feature represented by those variables is important in explaining the covariance between our fiber quality variables and our set of yarn variables. Since each of the components explains a decreasing amount of covariance, only the first two components will be of interest. *Figure 6* is a graphical representation of the loadings of each HVI™ variable on the first two components.

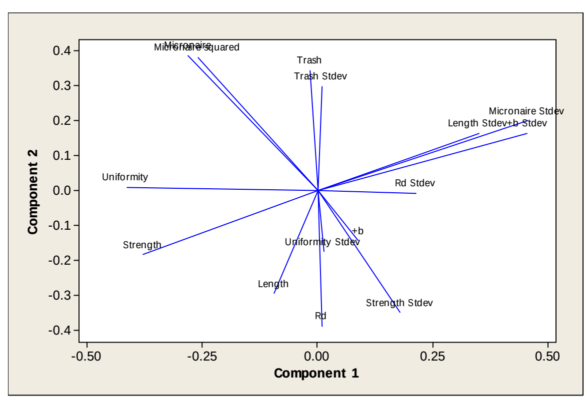

FIGURE 6. PLS loading plot using HVI™ variables.

For example, the variable uniformity has a correlation of approximately -0.40 with the 1st component, and approximately zero correlation with the 2nd component. This is the pattern we would like to see when choosing a group of variables to associate with a component: a strong correlation with a given component, and moreover, a near zero correlation with other components. This ensures that the feature represented by the variable is isolated to the given component. Looking across all of the variables, we can see none of them have particularly strong correlations with either of the first two components (all are smaller than 0.50 in magnitude), so any interpretations drawn here should be regarded as simple suggestions. Nevertheless, we can see that the fiber quality variables Uniformity, Micronaire stdev, and +b stdev have the largest isolated correlations with the first component, and Rd, Trash, and Trash stdev have the highest isolated correlations with the second component. The 1st set of variables – uniformity, Micronaire stdev, and +b stdev – appear to be capturing the common yarn quality features of strength and elongation. The 2nd set of variables – Rd, trash, and trash stdev – appear to be capturing the common yarn quality features of thick and low places. These results are generally in agreement with Deussen ([10]) who stated that key properties for vortex spinning systems are as follows: length, fineness, strength, friction, cleanliness.

PLS With HVI™ and AFIS™ Variables

PLS regression was run on yarn variables using both HVI™ and AFIS™ variables. Linear relationships between the yarn variables and the 12 AFIS™ fiber quality variables all appear acceptable. Inspection for outliers does reveal some potential outliers, but the model resulting from their exclusion had no substantial differences. In order to determine the optimal number of components to model each yarn quality variable 11 yarn variables (long thick excluded) were regressed against the 15 HVI ™ variables and the 12 AFIS™ variables (27 fiber quality predictors total).

The general pattern in the results is that a slightly larger number of components was chosen as optimal across all of the yarn variables, and further, there was no change in the average predicted R-squared and only a slight increase in the average fitted R-squared (*Table III*). Considering the lack of improvement in predictive ability, we conclude that adding AFIS™ variables is not worthwhile when making predictions for the vortex yarn quality response variables. In fact, five of the yarn variables saw a decrease in predictive ability after the AFIS™ variables were included. These included particularly large decreases in the yarn variables Major and Minor, where the predicted R-squared fell by 4 percentage points and 9 percentage points respectively. No table increases occurred in the yarn variables Statimat strength and Statimat elongation, where the predicted R-squared increased by roughly 4 percentage points and 12 percentage points respectively.

Using HVI™ and AFIS™ variables, an indirect attempt at variable reduction was made by interpreting the correlations (loadings) of each fiber quality variable on the extracted components. The corresponding loadings for the first two components is shown in *Figure 7*. As was the case before, none of these correlations are particularly large (all are less than 0.50 in magnitude) so any interpretations here are merely suggestions. The first component correlates highly with AFIS™ maturity ratio and neps, and the second component is characterized by AFIS™ UQL and HVI™ length. The 1st set of variables – maturity ratio and neps – appear to be capturing the common yarn quality feature of thick and low places. The 2nd set of variables – UQL and length – appear to be capturing the common yarn quality feature of strength and elongation. Again these results are generally in agreement with Deussen ([10]) who stated that key properties for vortex spinning systems are as follows: length, fineness, strength, friction, cleanliness.

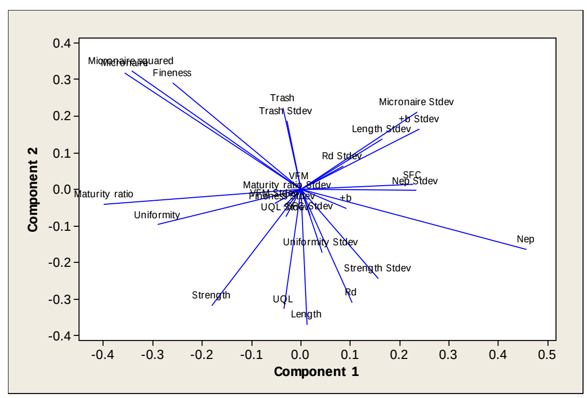

FIGURE 7. PLS loading plot using HVI™ and AFIS™ variables.

Comparison With Other Regression Methods

Having completed PLS analysis, it is interesting to compare these results with other regression alternatives such as stepwise and OLS. Since these methods require a single yarn variable, these methods will be performed on two separate yarn variables: one that PLS predicts quite well (Statimat strength) and one PLS was not able to predict very well (major). Using the basic OLS model, we will use all of the

TABLE III.
Number of components chosen in the PLS model for each variable using HVI™ and AFIS™ results.

Variables	Number of Components	R-squared	Predicted R-squared
Ends down	15	0.66	0.52
Statimat strength	8	0.89	0.86
Statimat elongation	14	0.83	0.75
Neps 18		0.78	0.69
Thick 13		0.85	0.79
Low 10		0.80	0.73
Yarn CV	10	0.87	0.83
Major 12		0.47	0.26
Minor 12		0.64	0.51
Long thick	na	na	na
Long thin	10	0.47	0.36
Yarn board appearance	10	0.60	0.51
Mean		0.72	0.62

TABLE IV.
OLS model for Statimat strength using HVI™ and AFIS™ results.

Predictor Coefficient		Standard Error of Coefficient	T	P	VIF
Constant -1	5.92	10.28	-1.55	0.124	
HVI™ Mic	-3.717	1.436	-2.59	0.011	506.6
HVI™ Mic stdev	-0.9316	0.3635	-2.56	0.012	5
HVI™ Strength	0.32413	0.05002	6.48	0	7.4
HVI™ Strength stdev	0.1773	0.1686	1.05	0.295	2.7
HVI™ Rd	0.13656	0.02397	5.7	0	4.3
HVI™ Rd stdev	0.82	0.2463	3.33	0.001	2
HVI™ +b	-0.1001	0.1073	-0.93	0.353	3.6
HVI™ +b stdev	-0.0414	0.1463	-0.28	0.778	8.1
HVI™ Trash	0.13134	0.04101	3.2	0.002	9.5
HVI™ Trash stdev	-0.2087	0.1956	-1.07	0.288	5.6
HVI™ Length	8.704	4.097	2.12	0.036	18
HVI™ Length stdev	-0.0404	0.1348	-0.3	0.765	5
HVI™ Uniformity	0.27947	0.09509	2.94	0.004	5.9
HVI™ Uniformity stdev	0.7254	0.3503 2.	07 0.	04	1.4
AFIS™ Fineness	-0.02796	0.01395	-2	0.047	10.7
AFIS™ Fineness stdev	0.00219	0.03061	0.07	0.943	1.5
AFIS™ UQL	-7.869	3.004	-2.62	0.01	14.7
AFIS™ UQL stdev	6.963	4.856	1.43	0.154	1.5
AFIS™ SFC	-0.19455	0.0443	-4.39	0	7
AFIS™ SFC stdev	0.07916	0.0799	0.99	0.324	1.4
AFIS™ Maturity ratio	2.114	2.893	0.73	0.466	6.7
AFIS™ Maturity ratio stdev	-2.276	6.499	-0.35	0.727	1.4
AFIS™ Nep	-0.0024	0.000994	-2.42	0.017	16.1
AFIS™ Nep stdev	0.000209	0.002568	0.08	0.935	1.5
AFIS™ VFM	0.08766	0.08914	0.98	0.327	2.8
AFIS™ VFM stdev	-0.0817	0.1437	-0.57	0.571	1.5
HVI™ Mic squared	0.3351	0.1588	2.11	0.037	429

HVI™ and AFIS™ variables as predictors (27 total). Results for the yarn variable Statimat strength are in *Table IV*. The foremost result to be noted here are the large Variance Inflation Factor (VIF) values, indicating a high degree of collinearity, and essentially invalidating any other conclusions that could be drawn about which fiber quality measurements are significant. Both the fitted and predicted R-squared are on the same level as that drawn from the PLS model, though the PLS model did so with a considerably smaller dimension of fiber quality measurements (only 8, versus the 27 here), further indicating excessive collinearity. OLS results for the other yarn variable are omitted, as the res ults are identical to these: that is, compared to OLS, PLS achieved the same quality of fit and predictive ability, but with th e ad ded feature of rem oving collin earity and ensuring more stable results.

TABLE V.
Stepwise regression model for Statimat strength using HVI™ and AFIS™ results.

Predictor	Coefficient	Standard Error Coefficient	T	P	VIF
Constant	-11.298	8.407	-1.34 0.	181	
HVI™ Strength	0.34589 0.	0301	11.49 0		2.8
HVI™ Rd	0.11665	0.01884	6.19	0	2.8
HVI™ Uniformity	0.27911 0.	08025	3.48	0.001	4.4
HVI™ Mic	-4.341	1.279	-3.4 0.	001 41	5.6
HVI™ Uniformity stdev	0.7529 0.	324	2.32	0.022	1.3
HVI™ Mic stdev -1	.1289	0.251	-4.5	0	2.5
HVI™ Rd stdev	0.7485 0.	2211	3.39	0.001	1.7
HVI™ Length	9.465	3.921 2.	41	0.017 17	.1
HVI™ Trash	0.07719	0.02412 3.	2	0.002 3.	4
HVI™ +b	- 0.17952	0.08859 -2	.03	0.045	2.6
HVI™ Mic squared	0.3888	0.1403	2.77 0.	006 34	6.7
AFIS™ Fineness	- 0.02141 0.	013	-1.65	0.102	9.6
AFIS™ SFC	- 0.19543	0.03986 -4	.9	0	5.9
AFIS™ Nep	- 0.00277	0.000867 -	3.2	0.002	12.7
AFIS™ UQL -8	.601	2.878	-2.99	0.003	14
AFIS™ UQL stdev	7.856	4.682	1.68	0.096	1.4

Stepwise reg ression fo r yarn varia ble Statimat strength yielded a fitted R -squared of 0.9 0 and a predicted R -squared of 0.89, com pared t o the PLS values o f 0.89 an d 0.86 ; thus we ca n see that there was a slight increase i n the pre dicted R -squared. However, this increase c omes at the e xpense of introducing collinearity into the model, as ab out one-third of the selected fib er quality measurements h ad large VIF's . Com pared to the 8 pre diction fact ors used in PLS, the stepwise procedure selected 16 fiber quality measurements (*Table V*) as significant.

Likewise, when pred icting for th e M ajor yarn variable, th e fitted an d predicted R-sq uared from a stepwise pr ocedure were 0. 43 and 0. 35, a 9% increase over the predicted R-square d found with PLS. Co llinearity was also no t as sev ere in th is model, wi th o nly t wo variables havi ng a l arge VIF. The results are shown below in *Table VI*. These

TABLE VI.
Stepwise regression model for major places using HVI™ and AFIS™ results.

Predictor	Coefficient	Standard Error Coefficient	T	P	VIF
Constant -3	8.59	16.82	- 2.29 0.	023	
AFIS™ Nep	- 0.00237	0.005395	- 0.44	0.661 8.	2
HVI™ Trash stdev -2	.642	0.9168	- 2.88	0.005 2.	1
HVI™ Mic stdev	9.367	2.628	3.56	0.001	4.5
HVI™ +b stdev -2	.107	0.9697	- 2.17 0.	031	6.2
AFIS™ SFC	0.8157 0.	2124	3.84	0	2.8
HVI™ +b	2.1904 0.	557	3.93	0	1.7
HVI™ Length 57	.81	26.28	2.2	0.029	12.8
AFIS™ UQL -3	1.59	20.34	- 1.55 0.	123	11.7
HVI™ Strength stdev 0.	96	1.121	0.86	0.394	2.1
HVI™ Uniformity stdev -5	.631	2.509	- 2.24 0.	026	1.3
HVI™ Rd stdev	3.814 1.	672	2.28	0.024	1.6
HVI™ Length stdev 1.	6594	0.9011	1.84	0.068	3.9
HVI™ Mic	-1.785 1.	28	- 1.39	0.165 7	

results suggest that a step wise procedure may be able to make more accurate predictions than PLS for some yarn variables. Th is result sh ould no t b e su rprising, since th is stepw ise pr ocedure o nly took on e sing le yarn variable into acc ount, whereas the PLS m ethod creates p redictions for 11 yarn quality response variables sim ultaneously. More over, there are m any studies that detail th e p itfalls o f using stepwise procedures ([11], [12]) where it is known that resulting models could fit well only by chance, and the final model is also likely to be unstable.

Another potentially useful comparison involves the measure of predictive ability used in this paper. This work used "leave one out" criteria related to the PRESS statistic and predicted R-squared. However, there is evidence that this method may result in overly optimistic views on prediction ability, and other alternatives (e.g., "leave *k* out") may prove more appropriate ([13]).

CONCLUSIONS

This work examines the use of PLS to predict multiple yarn quality variables based on a number of cotton quality variables. The data set indicated that OLS regression was inappropriate with the large number of fiber quality predictors compared to the number of observations, and the presence of strong collinearity in the fiber quality variables. Partial least squares was evaluated as a method that could overcome these difficulties and still provide accurate predictions. Two different sets of fiber quality measurements were used to forecast the yarn quality response variables: one set being only HVI™ variables, and the second set consisting of both HVI™ and AFIS™ variables. In forecasting yarn quality with PLS, every single component is a linear combination of all fiber quality variables so it is difficult to extract key variables. The quality of predictions was not found to significantly change with the addition of AFIS™ variables, implying that effort spent on gathering observations for these variables is not worthwhile for the sake of predicting the vortex yarn quality response variables. HVI™ data performs well for bale lay downs and predicting yarn quality while AFIS™ data allow textile mills to monitor and maintain their process quality. Relevant tasks for future work include judging predictions of the yarn quality based on other groups of fiber quality measurements beyond HVI™ and AFIS™. PLS was indirectly used to suggest which variables (HVI™ length, uniformity, Micronaire stdev, +b stdev, Rd, trash, trash stdev, and AFIS™ maturity ratio, neps, upper quartile length) are "important" for explaining vortex yarn variables (data reduction), however, other multivariate techniques are specifically designed for this.

REFERENCES

[1] ASTM International. 1999. Standard test methods for measurement of physical properties of cotton fibers by high volume instruments (D5867-95). Annual Book of Standards, ASTM International, West Conshohocken, PA. pp. 883-890.

[2] Schleth, A. and Peters, G. 2005a. Uster HVI™ Application Handbook. Uster Technologies. Knoxville, TN. 168 pp.

[3] Schleth, A. and Peters, G. 2005b. Uster AFIS™ Pro Application Handbook. Uster Technologies. Knoxville, TN. 168 pp.

[4] Balls, W.L. 1921. A Method for Measuring the Length of Cotton Hairs. MacMillan and Co., Ltd., London. 62 p.

[5] Bogdan, J. 1956. The characterization of spinning quality. Textile Res. J. 26, 720-730.

[6] Bogdan, J. 1967. The prediction of cotton yarn strength. Textile Res. J. 37(6), 536-537.

[7] Gordon, S. 2001. The effect of short fiber and nep levels on Murata vortex spinning efficiency and product quality. Commonwealth Scientific and Industrial Research Organization Textile and Fiber Technology Final Report to Cotton Research and Development Corporation. 2001. 14 pp.

[8] Anonymous. 2009. Vortex Yarn Guide Book. Cat. No. 17 1P4C2 09 -05-2(NS). Murata Machinery Ltd., Charlotte, NC. 12 pp.

[9] Tobias, R. 1997. An introduction to partial least squares regression. TS-509. SAS Institute, Cary, NC. 8 pp. Available at http://ftp.sas.com/techsup/download/technote/ts509.pdf (accessed 21 Sept. 2007; verified 21 Sept. 2007).

[10] Deussen, H. 1993. Rotor Spinning Technology. Schlafhorst Inc. Charlotte, NC. 128 pp.

[11] Judd, C., McClelland, G. 1989. Data Analysis: A Model Comparison Approach. New York: Harcourt Brace Jovanovich. 635 pp.

[12] Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B., 58(1), 267-288.

[13] Shao, J. 1993. Linear model selection by cross-validation. J. Amer. Statistical Assoc. 88(422), 486-494.

AUTHORS ADDRESSES
Calvin Price
Federal Reserve Bank of Philadelphia
Research
Ten Independence Mall
Philadelphia, PA 19106-1574
UNITED STATES

John Foulk ,William R. Meredith
USDA-ARS
Ravenel Center Rm 10
McGregor Road
Clemson, SC 29634
UNITED STATES

Herman Senter
Clemson University
Department of Mathematical Sciences
O-106 Martin Hall
Clemson, SC 29634-0975
UNITED STATES

Gary Gamble
USDA-ARS, Cotton Quality Research Station
P.O. Box 792
Clemson, SC 2963