# APPLYING PCA TO EXAMINE THE INFLUENCE OF FIBER PROPERTIES ON ROVING AND YARN QUALITY CHARACTERISTICS

**Wen-Bin Yu and Clarence D. Rogers**
**Clemson University**
**Clemson, SC**

## Abstract

The multi-collinearity between fiber characteristics results in unstable regression coefficients when fiber properties were used to predict/explain yarn quality in a regression analysis. Principal component analysis (PCA) and principal component regression (PCR), which were used in the previous study and generated explainable results, are techniques that can solve the interrelationship exists between fiber characteristics. In this study, PCA and PCR were used to investigate fiber-roving and fiber-yarn relationship and develop reliable models from production data. The results suggest that there might be an important factor, which has not been included in the HVI measures, that affects roving evenness.

## Introduction

Most spinning performance or yarn quality studies have been conducted in a very controlled environment or a laboratory. This has advantages and disadvantages. Some of the advantages are: controlled condition, control of input, and the ability to track inputs through the system. A primary disadvantage might be that the real world plant operates quite differently. Conditions may not be controlled, inputs may not be controlled over time, and numerous things happen in a plant environment. It is very difficult to trace fiber from the initial process to the end process in the system. Some other problems we might face and need to overcome, when developing models from plant data are:

(1) The cotton fibers purchased by different spinning plants are not similar owing to the different requirements of cotton fiber. In other words, the usage of cotton fiber has a bias natural in different plants. Thus models developed for one plant might not be realistic for another plant.

(2) Other than the fiber characteristics, there are additional factors in a plant manufacturing that can influence yarn quality. Those factors (machine setting, temperature, humidity, etc.) might be plant dependent and need to be investigated also.

(3) It is not easy to find a fiber-yarn, one-to-one relationship in the plant data. In order to obtain the proper output values for the corresponding input values, a specific pre-treatment of the plant data should be performed.

(4) The spinning of yarn includes several processes. These processes can be represented as a value chain of opening, carding, drawing, roving, and spinning. The qualities of the intermediate products contribute to the final yarn quality. Hence, the properties of intermediate products in yarn production (such as sliver, roving, etc.) and the relations between these products also need to be investigated. In a plant environment, numerous factors may be added in each process. One can not expect building a complete model by using only the fiber characteristics and yarn properties. Several intermediate models might be necessary in order to fully explain fiber-yarn relation in a spinning process.

(5) Spinning performance (such as: the production, machine stops, etc.) is another consideration in a yarn plant. It is expected that the improvements in yarn quality will not sacrifice any performance efficiency. Spinning performance and yarn quality should be examined at the same time.

Since a practical model for a yarn manufacturing plant is desired, the analysis will include all of the input/output data from the plant. That is, cotton fiber property's measurements, card sliver quality measurements, finisher drawing sliver quality measurements, breaker drawing sliver quality measurements, roving quality measurements, and yarn quality measurements from the plant will be examined. It is important that these measurements represent what is actually taking place in the plant. In essences these measures indicate how the complete plant is performing in the making of yarn.

It is found that there exists multi-collinearity between fiber characteristics.[3, 5] When severe collinearity is present among independent variables, regression coefficients in the model become unstable. In the previous study [5], PCA and PCR have been used to study the fiber-yarn relationship of sets of USDA laboratory data. Reasonable results were obtained from those models. How these techniques can be applied to the study of fiber-yarn relation was described. The same techniques were applied to examine the plant data in this study.

## Materials and Methods

### Fiber, Roving and Yarn Measurements

A 100% cotton ring spinning plant was selected for this study. Data was collected from the daily product monitoring reports from the plant. Typically, mix laydowns are established according to certain fiber property averages. For instance, if fiber length is considered to be important, the mix will be controlled according to a certain length average. Similarly, if micronaire is important, the mix will have a certain micronaire average. However,

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 2:1331-1334 (1996)
National Cotton Council, Memphis TN

establishing a mix laydown by average alone is not sufficient. Mixes must be established to control fiber property variance.

In this work, fiber property averages and variances were considered in establishing mix laydowns. For modeling purposes, the standard deviation and mean value of mix laydown fiber properties were used as input variables. The HVI fiber properties include micronaire (MIC), upper half mean length (UHM), length uniformity index (UI), strength (ST), reflectance (Rd), yellowness (+b), leaf (LEF), color (COL), and class grade (CLG). Temperature and humidity influence the processibility of fibers. These two variables are controlled in a well-conditioned laboratory. However, this might not be easy to achieve in a plant environment. It is important to establish how the variations for these environmental factors influence the conversion of fiber to yarn. Therefore, temperature and humidity are also included as input variables in the models.

The roving process output measures include roving weight (WT), Uster CV% (CV), and measure of variation for roving weight (SWT). Yarn data includes: yarn break factor (BF) and its variation (SBF), break number (B_NO), yarn weight (WT), yarn Uster CV% (CV), thick places per 1,000 yards (THICK), thin places per 1,000 yards (THIN), and neps per 1,000 yards (NEPS). These data represent the output measures for the roving and yarn manufacturing system. A study of these final output measures is the primary objective of this work. More specifically, the investigation of what input/output measures from prior process influence yarn quality and spinning performance measure is desired.

## Principal Components Analysis and Principal Components Regression

The techniques that utilized to analysis the fiber-roving and fiber-yarn relationship were principal components analysis (PCA) and principal components regression (PCR). PCA is one of the techniques to study data with interrelationship. PCA is a data re-constructing technique which involves a minimum of statistical presumptions. There is no model hypothesis required for this technique and it is much easier to use than most of the statistical procedures[6]. The basic concept of principal components analysis is every linear regression model can be reconstructed in terms of a set of independent variables. Constructed models with low correlation variables which are obtained by the linear combination of the original variables will meet the assumptions of a regression model. PCR model, applying principal component analysis prior to the linear regression model, was used to explain and show the fiber-yarn relationships. Since multi-collinearity exists between fiber characteristics, PCA should be applied prior to regression analysis. However, it should be noted that if there exist no multi-collinearity between the fiber characteristics, the PCR will generate the same result as normal regression analysis.

Many standard statistical packages provide the procedures of PCA (such as SAS, SPSS, etc.). Some individual software concentrates only on this technique, which might be more convenient in performing certain analysis. The software used in this study is UNSCRAMBLER®. One of the important features of this software is its graphics presentation capability. The variations in most of the textile properties are pronounced. Variation can be studied using mathematical or graphical techniques. Graphical presentation may be a preferred approach when there exists a large amount of variation. The reason is that a specific number of the coefficients of regression equation may not be stable if multi-collinearity exists between the input variables. However, the correct number is in the neighborhood. So rather than building a prediction equation, a graphics presentation of the results might provide or show the tendency of the relation between the input and output variables. This graphical presentation will be discussed later.

PCA uses two primary tools to interpret the behaviors of the model [6]. These are score plots and loading plots. Scores express the relation between observations and components. Loadings express the relation between variables and components. The relations can be plotted in one (against only one principal component), two (against two principal components) or three (against three components) dimensions. The loadings (or weights) of variables to the principal components are plotted in loading plots. The original variables will have their individual weights (or loadings) for each component. Higher loading of the original variable indicates a stronger influence of the corresponding component. The number of the principal components that can be reconstructed in a model is equal to the total number of the original input variables. Since the first and second principal components explains most information from the data, loading plots of the first two principal components are usually used to examine the relationship between input and output variables. Through PCR, both the fiber and yarn variables can be plotted in the loading plots. Variables located on the positive side of the principal component have a positive influence on the component and variables located on the negative side of the principal component have a negative influence on the component. The farther the distance a variable is away from the origin, the more contribution of this variable to the component.

Variables which are independent to those variables with higher influences on the component will be located near the origin. If variables (either input or output variables) are located within one cluster in the loading plot, then those variables are highly positive correlated.

An important rule to identify negative correlation is: If a straight line can be drawn through the origin to connect two different groups of variables, the variables in these two groups are negatively correlated.

Score plots explain the relation between each individual observation and the principal components. The "score" of every data point is plotted against the principal components in one or two dimensions. If the first two components describe most of the information in the data set, the observations should be successfully separated by the two components. In other words, score-plots indicate those fiber (or yarn) characteristics with a higher or lower quality rating.

Since there are several input variables, the analysis is conducted to reduce the number of input of variables. Only the variables that affect the output variables are of interest. Therefore, the analysis will consist of several iterations. Loading plots can be used to eliminate variables from the model. If a fiber variable is not located within a reasonable distance to any roving property in the loading plots, it will be eliminated from the model. In this case, the corresponding fiber properties can not be used to predict any dependent variable. If all the dependent variables are located at the intersection of the PC1 and PC2, the most varied input variables located away from the origin will be eliminated. On the other hand, if all the independent variables are located near the origin, the dependent variables that are far away from the origin will be removed from the model. In this case, those roving variables vary dramatically and fiber variables are not good predictors of the roving characteristics. The analysis will be continued until an explainable loading plot is obtained.

The same logic is applied for the output yarn variables. If the output variables vary in such a way that fiber measures or any measure from previous processes are not good predictors, the most variant output yarn variable will be eliminated. In the corresponding loading plot, those fiber variables are all located near the origin. The input variables are not able to predict or explain the most varied output variable in this case, hence the output variables will be dropped from the analysis. The analysis will be continued until the input variables in the middle of the loading plot are spread out or separated.

## Results and Discussion

Shown in Figure 1 is the loading-plot for roving. It can be seen that the roving weight (WT) and standard deviation of weight (SWT) are located in the middle of the intersection of PC1 and PC2. This indicates these two output variables are independent to most of the fiber variables. The model is not able to predict or explain the behavior of weight of roving and its variation with the fiber characteristics. However, the loading plot shows that the roving CV% has a negative correlation with Rd, UI and possibly +b. That is to say, as the value of fiber Rd, UI or +b increase, the roving CV% tends to decrease. It might be hard to explain the statistical relationship between fiber Rd, +b and roving CV%. However, an increase in fiber UI indicates more uniformity in fiber UHM, and a decrease in roving CV% indicates a more even or consistent roving. This relationship is reasonable and expected.

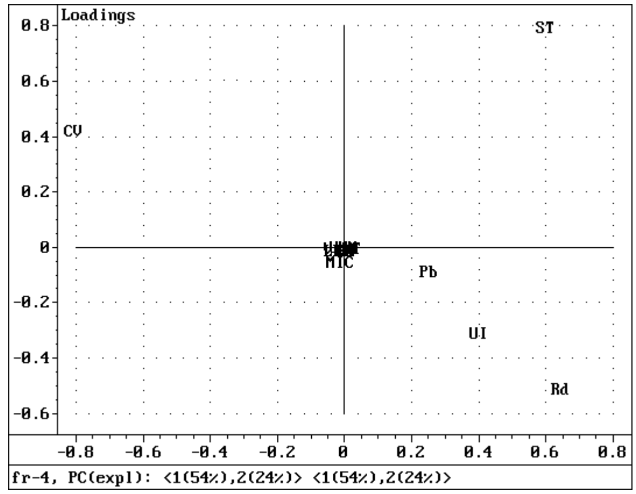


Figure 1 Principal Components PC1 and PC2 Loading Plot of Fiber-Roving Model, Comparison of Fiber Properties (Micronaire: MIC, Strength: ST, Uniformity Index: UI, Length: UHM, Reflectance: Rd, and Yellowness: +b) and Roving Quality Measures (Uster Roving CV%: CV, Weight: WT, and Standard Deviation of Weight: SWT)

There is no evidence to suggest and it might be inappropriate to conclude that there is a direct relationship between UI, Rd, and +b, although the three variables are located in the same cluster. The result that both Rd and +b have similar influence on roving CV% as UI, might suggest that there exists another important factor that really affects roving CV%. Such a factor was not included in the HVI measures, and the fiber Rd or +b somehow might represent this factor. A possible reason for this factor might be fiber maturity. However, this cannot be addressed until data shows that there is a correlation between HVI measurements for Rd, +b and fiber maturity.

As shown in Figure 1, fiber ST has a negative influence on roving CV%. Thus, stronger fibers produce a more even roving. This negative influence is different from those caused by Rd, UI or +b. It comes from PC1 only. ST and CV% are on the opposite side of PC1, but they are on the same side (positive direction) of PC2. PC1 explains 54%, and PC2 explains 24% of the variation in the two PC models. This indicates that 54% of the time ST and roving CV% has a negative correlation and 24% of the time it might have a positive correlation. Clearly, ST is not as good a predictor as Rd, +b and UI. However, ST still presents information where some of the other input variables, such as MIC, give no information at all. That is, if the MIC needs to be increased (or decreased) for some reason, it will not affect roving CV%, but a change in ST will affect the roving CV%.

From the conclusion of the above analysis, Rd, +b and UI are the most influential factors that affect CV% of roving. ST has an uncertain influence on roving CV%, 54% of the time the correlation is negative; 24% of the time the correlation is positive. Other fiber properties have little or no influence.

The loading-plot for yarn is shown in Figure 2. There might not be good interpretation for PC1 and PC2 in this case. Since most of the yarn properties except yarn CV% (CV), weight (WT) and break number (B_NO) are eliminated from this plot. This data shows that HVI measures of fiber properties are not good predictors for the other yarn properties. However, one might consider an explanation for yarn CV% as was given for roving CV%. Figure 2 shows that fiber Rd, UI, and +b have a negative effects on yarn CV%.

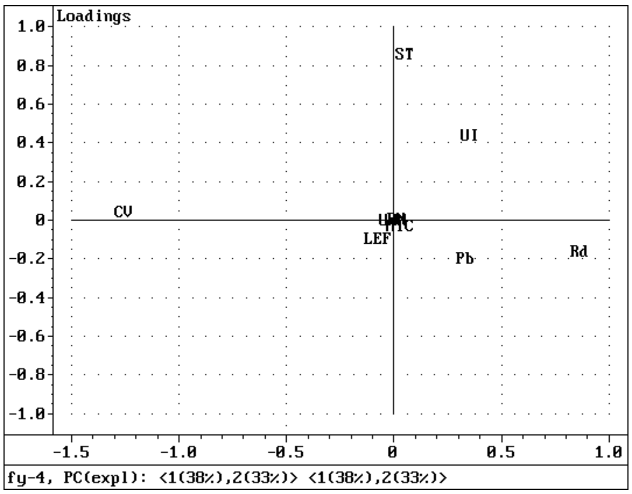


Figure 2 Principal Components PC1 and PC2 Loading Plot(b) of Fiber-Yarn Model, Comparison of Fiber Properties (Micronaire: MIC, Strength: ST, Uniformity Index: UI, Length: UHM, Reflectance: Rd, and Yellowness: +B) and Yarn Quality Measure (Uster Yarn CV%: CV)

## Conclusions

In this study, the expectations were to establish different models that include all the activities (opening, carding, drawing, roving, and spinning) in the spinning process of the plant. The data for this study is collected from the daily production monitoring reports of the plant. The purpose of these reports is to monitor and identify any abnormal operation, and it is not for investigation purpose. Data sets of several processes, such as card sliver, breaker and finisher drawing sliver, were not sufficient to perform statistical analysis. Also, the frequencies of measurements of the spinning performance were not sufficient to establish reliable models. Time limitation and insufficient information were the reasons that only fiber-roving model and fiber-yarn model were investigated. The analysis of spinning performance, such as machine stops, was not achieved, due to insufficient data. However, the systematic method has been applied to the plant data and accomplished significant results.

The relation between fiber and roving has been drawn based on the PCA and PCR analysis. In this study, the roving CV% is expected to decrease as the characteristic of reflectance (Rd), length uniformity index (UI), and yellowness (+b) in fibers are increased. That is, fibers with higher Rd value, higher UI, or higher +b, tend to produce a more even roving (lower CV%). There might not be a direct relationship among these three fiber characteristics. However, the results suggest that there might be some underlying factor, such as degree of maturity or friction properties of fiber that is not measured. The factor not measured, may be represented by the fiber properties of Rd and +b. Thus, Rd and +b might act as a proxy for the missing variable, which has a significant influence on the evenness of roving. Further investigation is needed to confirm this conclusion.

Other than fiber properties, there might be some other factors in the spinning process of a plant that can be measured that will have an influence on spinning performance or yarn quality. Temperature and relative humidity were two factors considered in this study. The results suggest that these two factors do not have significant effects on the roving or yarn quality measures. It may be that these two variables are controlled to such a degree that they will not cause any variation or influence on the output measures. This might not be true for a plant with less controlled environmental conditions. It must be pointed out that all the conclusions are only good for this specific plant and that the variables are plant dependent. A unique or general model is not intend to be proposed in this study. However, a model can be developed using the same technique and data from a plant selected for study.

## References

1. Dillon, W. R. and Goldstein, M. "Multivariate Analysis." John Wiley & sons Inc. New York, NY. (1984).

2. Dunteman, G. H. "Principal Components Analysis." Sage Publication Inc. Newburry Park, PA. (1989).

3. El Mogahzy, Y. E. and Broughton, R. M. Jr. "Diagnostic Procedures for Multi-collinearity Between HVI Cotton Fiber Properties." Textile Research Journal, **59** (8), 440-447. (1989).

4. Johnson, R. A. and Wichern, D. W. "Applied Multivariate Statistical Analysis." 3rd ed., Prentice-Hall International Inc. Englewood Cliff, NJ. (1992).

5. Rogers, C. D. and Yu, W "The Use of Principal Component Analysis to Optimize Spinning Performance." 1995 Beltwide Cotton Conference San Antonio, TX., Proceedings, 1404-1405. (1995).

6. "Uncscambler User's Guide." Ver. 5.5., CAMO A/S. Trondheim, Norway. (1994).