Abstract

HAMILTON, BRIAN JOHN. Process and Product Data Management for Staple Yarn Manufacturing. (Under the direction of Dr. George Hodge and Dr. William Oxenham).

The contemporary cotton spinning mill is home to modern machinery capable of generating massive amounts of data. This data comes in the form of online data, which is real-time data created by the processing machinery, and offline data, which is created via laboratory testing of samples. Data mining is the use of advanced statistical tools to discover hidden relationships within a large set of data. When there is a situation in which a vast amount of data is available, such as in the US cotton yarn spinning industry, there is an opportunity for data mining.

This study applied data mining techniques to two data sets. One set was obtained from an open-end spinning plant. The other set is the results of a government research project. This analysis served to discover unknown trends within this data sample and to determine the potential value of data mining for the cotton spinning industry.

This research presents a perspective into the current state of data management in the cotton spinning industry obtained through interviews and observations of active spinning mills. It also details the data mining performed on the acquired data sets and suggests a data management model which facilitates effective data mining and enhanced decision making.

Process and Product Data Management for Staple Yarn Manufacturing

by
Brian John Hamilton

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Textiles

Raleigh, North Carolina

2010

APPROVED BY:

Dr. William Oxenham
Committee Co-Chair

Dr. George Hodge
Committee Co-Chair

Dr. Gilbert O'Neal

Dr. Kristin Thoney

Biography

Brian John Hamilton was born in the picturesque town of Tewksbury, Massachusetts in 1986. His education began in the Tewksbury public school system, where he learned about life and scholastics from K-12. After that, he attended the University of Massachusetts - Dartmouth as a Commonwealth Scholar, and his four years there culminated with a Bachelor's degree in Materials Technology with a minor in Business Administration. From there, Hamilton headed straight to Raleigh, North Carolina, where he participated in the masters program at the College of Textiles at North Carolina State University as a Fellow for the Institute of Textile Technology.

Acknowledgements

Acknowledgements go to ITT, my thesis committee, the companies that participated in this study, the USDA, NC State, my teachers, UMass Dartmouth, the Tewksbury Public School System, my entire family, and everyone that I'm friendly with.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

Cotton spinning has long been a major industry of the United States. The process of creating yarn has grown and advanced with the country. The complexity of modern spinning operations allows for the creation of monumental amounts of data.

This data is a product of machine settings, production yields, online quality data, and routine offline testing. More data is available today than ever before, and it is certain that the amount available will only increase.

While this in and of itself is not a major drawback, the problem lies in that this data could potential be put to much better use. The data capable of being accumulated at a modern spinning plant is a reflection of the investments that plant has made into modern machines and testing equipment. As a business, it is of paramount importance to take advantage of the investment and get the most from it. In the case of cotton spinning, much of the potential utility of modern data capabilities is not being tapped into.

It is important to convert this data into usable information. This is done by finding hidden relationships in the data. The process of extracting such relationships is known as data mining, and this is done with some specific computational techniques. Once more is known about all the data available and how different data items relate to one another, the business will be able to make better decisions about the process.

With this massive amount of data being created around the clock at a spinning mill, it could be counterproductive to employ a data management technique that attempts to record and analyze it all, especially when correlations between two or more measurements may make it unnecessary to manage both. If two data points are directly related, for example, the first could be used to predict the second, and it would only be necessary to manage one of the data points.

In this study, cotton spinning data is gathered and data mining techniques are applied in order to create recommendations for how a cotton spinning mill should be recording and monitoring data.

First, available data mining background and techniques are researched. This allows for the necessary data mining tools to be identified and utilized given the questions being asked of the data. Next, the current state of the US cotton industry is learned through first hand interviews and observations. A detailed knowledge of the industry ensures that the most relevant questions are being asked of the data.

Another step in this research is the procurement of the data itself. The data sets used in this research contain information and relationships that can lead to valuable knowledge. This study uses data mining to explore this information and reach conclusions that lead to recommendations for data management in cotton spinning mills. This is accomplished in a scientific fashion and is detailed in the upcoming chapters of this research.

Chapter 2 is the literature review. In this chapter, all the necessary literature needed to acquire suitable knowledge for this research is reviewed. This information provides the base on which the rest of the study was constructed.

Chapter 3 discloses the methodology of this research. It explains the goals that were set and the methods through which they were attained. The limitations for this particular study are also noted.

Chapter 4 discusses the first phase of the methodology, making industry visits in order to better understand the current state of data management in the cotton industry. This step also served to procure the data which is introduced and analyzed in the following chapters.

Chapter 5 deals with data understanding. The data that has been obtained is introduced and explained. It is also prepared for modeling and analysis.

Chapter 6 features analysis and results, which includes recommendations based upon the information gleaned from the preceding chapters. It explains the data mining process and interprets the results. The findings of the data analysis are formulated into data management strategies that are attainable given the state of the typical spinning mill.

Finally, Chapter 7 discusses conclusions and makes recommendations for future studies.

Chapter 2 - Literature Review

2.0 Introduction

This study revolves around multiple elementary fields, and it is these fields that were studied and summarized in this literature review. Firstly, this research is being done in hopes of advancing the cotton spinning industry. Therefore, one subject to be reviewed is the cotton spinning process. The history of spinning in the United States was researched, as well as the different processing steps. Also included is background on cotton fiber. This includes information on the intrinsic properties of the material, as well as the variables involved in growing and harvesting the crop.

The next aspect of this research lies within the data involved in modern day spinning processes. Research to this end included what data is measured directly on the machines of some prominent manufacturers as well as what laboratory equipment is available for offline data accumulation and the different measurements that each can make.

Another key component of this research is the analysis of such data. It has been determined that the cotton spinning industry is a candidate for data mining techniques (Braha, 2001). In order for these techniques to be applied, they would have to be reviewed and understood before implementation, so this literature also contains a section discussing the background of data mining and descriptions of many of the most effective data mining techniques that may be called for in this study.

This chapter also reviews some previous works that are similar to this study. Past works can be a strong tool as they can be learned from and built upon; and most include recommendations that can prove enhancive.

2.1 Cotton Properties

The reason for the popularity of cotton in the textile industry is in the inherent properties of the material. The characteristics of cotton fiber are outstandingly suited for the textile demands of modern civilization. Cotton clothing is generally regarded for its comfort, which comes from the moisture absorbing ability of the fiber (Geissler, 1993). Beyond feeling pleasant against one's skin, cotton fabric is ideal for apparel because it is durable and easy to wash, attributes apparent since the fiber first made its American debut centuries ago (Ouellette, 2007). While consumers are drawn to cotton because it feels nice to wear, the textile production industry is taken by its color-fastness, absorbency, heat resistance, and durability; which are all properties that enhance production rates and efficiency (McCreight, Feil, Booterbaugh, & Backe, 1997).

It is on the molecular level where cotton fibers share the most commonalities. Cotton fiber grows from the surface of the seed and each fiber is a long cellulose tube encompassing a feed channel, called the lumen, running through the center (Lord, 2003). Under a microscope, a mature cotton fiber appears as a flattened tube that is twisted and irregular. The flatness comes from the fact that once the cotton is harvested, it will lose moisture, causing the lumen to collapse (Morton & Hearle, 1993). Almost all United States cottons are of the Upland variety classification. Upland cotton fibers vary from 22 to 32 millimeters in length and uniformly have a diameter of 16 microns. Cotton fiber is classified as a short staple fiber because its length is less than 64 mm, making it suitable for a cotton spinning system (McCreight et al., 1997).

While all cotton shares the same basic characteristics, anyone involved with cotton production will note the vast differences between crop varieties, be it from year to year or from region to region. This is one glaring difference between natural fibers and synthetics, which can be precisely engineered and incredibly consistent. The irregularity of cotton is something that has always needed to be accounted for (Taggart, 1923). Even within a field of a single variety of cotton, there will be variations between plants as well as within a single plant. For instance, the fiber fineness within just a single plant can have a coefficient of variation of 15% (Lord, 2003).

The agricultural side of cotton alone offers endless agents of variation. Farmers have long known the impact of aspects such as seed character, soil character, land preparation, cultivation method, atmospheric conditions, weather, and irrigation, among others (Monie, 1904). The fact that the cotton plant has large, lush foliage and a long fruiting period makes it a prime target for pests. In all, there are over 500 species of insect with the cotton plant on their radar (Lord, 2003).

It speaks to the natural superiority of cotton that it remains prevalent even as science has continually produced innovations in synthetic textile fibers. There was a time when confident predictions were being maid about cotton's impending demise. Experts believed it would become relegated to luxury fiber status, on par with silk (Slater & Hoffmeyer, 1979).

In any kind of production environment, it is clear that having control over the final product is of utmost importance. In cotton spinning, the many facets involved and their ultimate impact on the resultant yarn have been studied. Machine settings, machine type, and fiber type can all dramatically influence the properties of the final yarn. One aspect that remains somewhat out of the realm of mastery is within the cotton crops themselves. The large variations present within this natural resource make predictions and consistency an inexact science (Lord, 2003).

One step of obvious importance in order to maintain consistency within the final product is the blending of cotton fibers. If enough fibers are spread adequately along the cross section of the final yarn, any variations among the different cotton varieties will tend to somewhat cancel out, making a more uniform final yarn. There are actually two factors that make proper fiber blending a high priority. Firstly, a yarn must have a strong appearance of uniformity, within itself and with the surrounding like-yarns in a fabric. Secondly, yarn needs to be consistent with similar yarns being produced on different sets of machines, and material produced one day must be fit to be processed with material that may have been produced the following day (Grosberg & Iype, 1999). Blending takes place throughout the cotton yarn production process, including during bale laydown, mixing, carding, and drawing.

Blending, while a crucial part of cotton yarn spinning, tends to hinder attempts at predicting final yarn quality based on fiber characteristics. In the past, successful predictive formulas have been derived to predict yarn quality based on the qualities of the input fibers for single-cotton yarn (Jayadeva, Guha, & Chattopadhyay, 2003). For instance, one study was able to determine that high fiber length distribution was correlated with improved yarn quality and spinning performance (Hequet & Ethridge, 2005). While such a study can help lead to the genetic development of better cotton, it does not have a great impact in the world of yarn production. This is because a blended cotton yarn does not simply reflect an average of the constituent cotton fibers, for interactions amongst the fibers must also be taken into consideration (Basu, 2009). This is significant because all major yarn manufacturers use blended cotton fibers for the sake of uniformity.

In addition to the variations naturally occurring in the cotton itself, the properties of cotton fibers are further altered during the yarn production process. These can be minimized if the proper technical controls can be applied as early as possible. One way to ensure that such variations are minimized is through proper training. A "Yarn quality prediction and control" seminar was held by the South India Textile Research Association in 2009 and covered such topics as predicting yarn elongation and hairiness based on cotton properties (The Textile Magazine, 2006).

2.2 Cotton Spinning History

The cotton plant has long has long been utilized as a textile resource. This natural cellulose fiber remains popular worldwide due to a dossier of desirable properties. Remaining relevant is no small feat for a fiber that has been known and used for thousands of years, anecdotally since beyond 2700 B.C., in the ancient civilizations of India and Peru (Morton, 1962).

In North America, the use of cotton dates back as far as the 1600s. Not long after catching on in England, several cotton varieties began to find their way to the colonies of New England, mainly via English plantations in the Caribbean. Colonists took to cotton fabrics as a desirable new alternative to wool (Ouellette, 2007). This time period represents the birth of the US cotton industry that has grown and prospered ever since.

Converting cotton fiber into cotton yarn is accomplished through the process of spinning. This process had been performed for as long as cotton yarns have existed, but the modern spinning encountered during this research represents years of honing and innovation.

Originally, all spinning was done by hand. The inceptive method simply involved the twisting of a bundle of fibers between a person's palm and thigh. The resultant thread was wound onto a hand spindle, which was a spinning stick with a momentum-producing weight attached to the bottom (Shaikh, 2005). This later gave way to the use of a manual wheel spinning system which, along with hand carding, represented a big step forward in terms of efficiency and production. Spinning wheels provided plenty of yarn to loom weavers over the years, but come 1733, the invention of the flying shuttle caused a vast increase in the rate of weaving production. This necessitated the creation of the spinning jenny, which allowed a single highly skilled operator to produce 8 spindles at a time (Morton, 1962). This technology eventually gave way to the spinning mule, and later the self-acting mule, a completely automatic machine capable of producing very fine yarns (Linton, 1963).

Mule spinning became the norm until being overcome by ring spinning, a modern method of which remains responsible for much of the cotton yarn in the market today. The advantages of ring spinning over the spinning mule are numerous. Firstly, ring spinning represents a continuous process as opposed to the jenny and the mule, which were utilized in intermittent spinning systems. The speed of the ring frame itself was a great improvement, as was its operating simplicity and smaller demand for floor space (G. C. Anderson, 1976). Today, however, the floor space demanded for ring spinning is actually proportionately large when compared to newer technology such as air-jet spinning (Shaikh, 2005).

2.3 From Crop to Yarn

2.3.1 Farming

Cotton fiber is derived from the cotton crop, which is cultivated on a farm. Challenges in growing cotton include controlling pests and choosing a species of cotton that best fits the growing conditions. Cotton is spaced in a field such that it allows for machinery such as pickers and sprayers to travel throughout the crops. The farmer must deduce the best time to harvest based on factors including weather. Once picked, the cotton is stored properly or sent to a gin for further processing (Lord, 2003).

2.3.2 Ginning

A cotton gin serves to clean the raw cotton and create the bales of cotton fiber that will act as the input material in the modern cotton spinning mill. The gin is meant to remove the courser impurities from the cotton fibers (Lord, 1974). First, heavy foreign matter such as rocks and unopened bolls are expunged through trap doors. Next, the fiber is heated and dried so that fine dirt and leaves can be removed by spiked rotating cylinders. This is followed by the gin stand, which breaks apart the seeds and separates them from the fiber. The last removal step is the lint cleaner, which takes out remaining seed fragments and other surviving trash. The ginned fiber is then compressed into bales that weigh about 500 pounds, strapped tight, and shipped out (Lord, 2003).

2.3.3 Blowroom

Once the cotton arrives at the cotton spinning facility, it enters the blowroom, or opening room. This is a series of machines connected by tubes through which the cotton fibers travel as they go through each step of processing. The first step in the blowroom is the plucking or feeding of cotton tufts from the bales. This is generally done with a top feeder, which skims across the top of a series of bales, plucking the fiber as it travels via the use of rotating spikes or teeth (Shaikh, 2005). It should be noted that the bales of fiber do not go directly into a laydown. Once a shipment of bales arrives, their packaging must be cut open, and time must be given for the bales to bloom. This process involves the fibers gaining size once they are no longer being held in compression by their straps or constraints. At this time, the moisture level of the fibers is able to reach equilibrium with the environment, allowing for more consistent production at the mill (Lord, 2003).

The order in which the bales are set up is actually of critical importance. The plant will have the data describing the average physical properties of each bale in their inventory. These bales will be of varying varieties that need to be blended together to produce a final blended yarn of minimized variability. A computer program actually produces a grouping of bales that best blends the different measured qualities in such a way that the average characteristics are similar in each unique laydown for a particular yarn type. These fibers will continue to be mixed in many of the following processes. Mix homogeneity will depend on the methods and machines used, but the goal is always to average out the characteristics of the constituent fibers (Singh & Kothari, 2009).

The number of bales used in a particular laydown is usually determined by figuring out how many bales are necessary to average out the proper fiber properties. With cotton, this usually means somewhere between 16 and 64 bales are needed for each laydown, but that number can be higher. The amount of bales used can obviously also be influenced by the production capacity and floor space of the particular mill. It is important that the bales are of similar height and density as well because this allows for even layers to be taken from each one as the bale plucker passes (McCreight et al., 1997).

Once the bales are plucked, their fibers begin traveling through a succession of opening, cleaning, and mixing machines via the aforementioned tubes of the blowroom. Opening machines serve to further reduce the size of the tufts of fabric that are taken off of the bales. Cleaning machines beat impurities and trash out of the cotton, and they also provide some additional blending of fibers. This additional cleaning must be done because ginning alone is not capable of removing all of the present trash. This is an example of a step that does not need to be used when dealing with synthetic fibers (Lord, 2003). Mixing machines are designed to create even more intimate blends by separating and then recombining the fibers in a measured fashion (Shaikh, 2005). An example of this can be seen in Figure 1.

There is actually a chance for additional cleaning in the blow room. Raw cotton contaminates can be detected and removed by a device such as the Barco Cotton Sorter. This system can be placed in an existing line without the need for additional fan capacity. It uses a high speed camera and a transparent detection zone. It is able to spot deviances in color or size and identify any material other than cotton. When such a contaminant is detected, it is simply removed by a high speed air gun (Singh & Kothari, 2009).

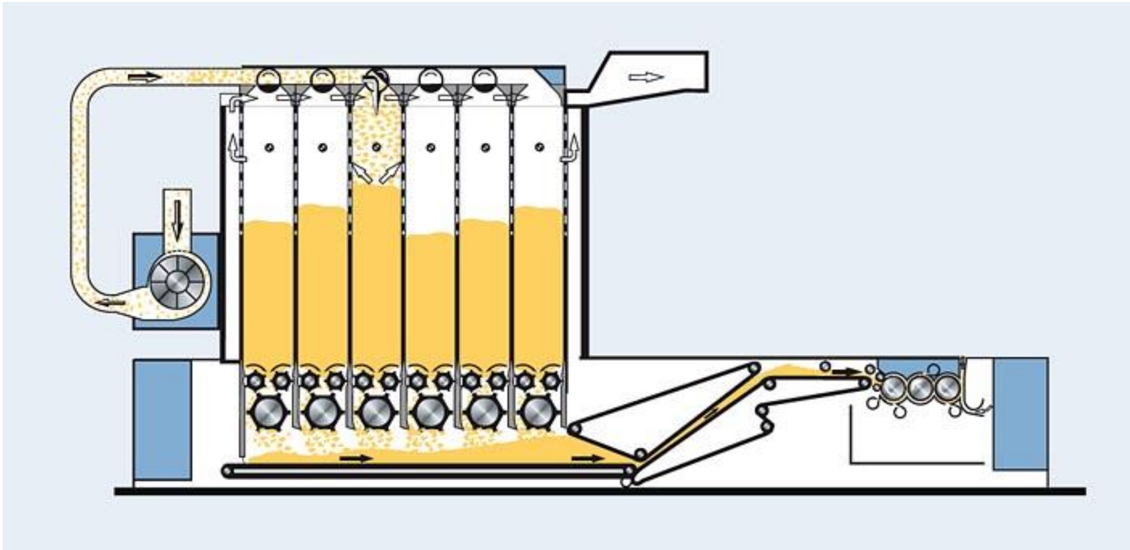


Figure 1 - Trützschler Integrated Mixer (Truetzschler.eu, 2009)

2.3.4 Carding

After the opening room, fibers make their way into the card room. Carding is a very important process and can have a powerful impact on the final yarn product; especially in eliminating neps, which are small tangled lumps of fiber that diminish the appearance of the eventual fabric (Linton, 1963). The goals of successful carding are to separate the tufts into individual fibers, further eliminate impurities, remove the shortest fibers, remove neps, stretch the fiber, and create a sliver (Shaikh, 2005).

This is one aspect of spinning that has certainly gained efficiency over the years. In fact, for typical carded yarns since 1955, card performance has grown by a factor of 20 as labor cost has concurrently dropped to only 10% of what it once was (Artzt, 2004). Modern cards provide an impressive degree of cleaning, achieving 80 - 95%, and when this is combined with the cleaning carried out in the blow room, the cotton should now be 95 - 99% cleaned, which is important because this is the last step in the entire process that provides a chance for cleaning (Shaikh, 2005).

Before entering the card itself, the fibers are made into a mat by the feed chute. The mass of the mat must be evenly dispersed with a consistent weight and openness throughout (McCreight et al., 1997). The card is the first step in the production process that serves to align the fibers, orienting them in a common direction. This is an important quality, as it is the aligned fibers from which yarn gets its strength. After entering the card as a mat, the fibers go through a series of rollers and teeth, essentially separating them into a sheet of individual fibers while also providing orientation and eliminating neps. This sheet finally passes through a trumpet, which shapes the mat into a sliver, which has a rope-like appearance. The auto-doffing machine coils this sliver into a can, in which it is transported to the ensuing processing steps (Lord, 2003).

2.3.5 Drawing

In the drawing process, multiple slivers are combined to create a single uniform sliver. This step provides additional blending and orientation of the fibers and is considered the last chance for quality improvement in the process (Shaikh, 2005). Drawing is able to correct for any weight variations and fiber misalignments that may exist in the card sliver. 6 to 8 card slivers, which are referred to as doublings in this process, pass through a series of rollers of increasing speed. This is what stretches out the fibers and allows multiple doublings to become a single sliver. The evenness and consistency of the sliver is accounted for by an autoleveling process. This is an automated process which senses the size variations of the sliver as it leaves the draw frame. It will automatically adjust the speed of the rollers to compensate for any detected variations. This sliver should be properly oriented, consistent with the other slivers being produced, show evenness throughout, and be the proper weight for the next processing step (McCreight et al., 1997).

2.3.6 Ring Spinning

This next step, however, is dependent on the spinning system being employed at the particular mill. As seen in Table 1and Table 2, there are some differences in the production cycles of ring-spun yarn and rotor-spun yarn. It should be noted that in ring spinning, there is sometimes an additional step after drawing known as combing. This process is typically done on high end cottons being made into quality yarns. Combing simply removes more short fiber and further improves fiber orientation (Lord, 2003).

Ring spinning produces a dense, flexible, and strong cotton yarn. Positive aspects of this type of spinning are that it is simple and can handle a variety of counts and fiber types. Drawbacks include the necessity of a couple of additional processes, both before and after the spinning frame, which reduce production while increasing cost and energy use (McCreight et al., 1997).

Table 1 - Cycle of Open-End Cotton Yarn (adapted from Shaikh, 2005)

Stage	Machine	Entry Material	Delivery Material	Package Form
Opening and Cleaning	Bale Plucker, Opener, Blender	Raw Cotton	Lap	n/a
Carding	Card	Lap	Card Sliver	Can
Drawing	Draw Frame	Card Sliver	Drawn Sliver	Draw Frame Can
Spinning	Open - End Spinning Frame	Drawn Sliver	(O-E) Yarn	Package Form

For ring spinning, the drawing sliver must become roving before it makes its way to the spinning frame. A roving machine will further draw out sliver and add a slight twist while winding the material around a bobbin. This further straightens the fiber and prepares it for the spinning frame (Shaikh, 2005). Roving may not seem like the most critical of the yarn spinning processes, however nearly half of all ring spinning end breaks are the result poor roving preparation (Lord, 2003).

Table 2 - Cycle of Ring-Spun Cotton Yarn (adapted from Shaikh, 2005)

Stage	Machine	Entry Material	Delivery Material	Package Form
Opening and Cleaning	Bale Plucker, Opener, Blender	Raw Cotton	Lap	n/a
Carding	Card	Lap	Card Sliver	Can
1st Drawing	Draw Frame	Card Sliver	Drawn Sliver	Sliver Can
2nd Drawing	Draw Frame	Drawn Sliver	Drawn Sliver	Roving Can
Roving	Roving Frame	Drawn Sliver	Roving	Roving Bobbin
Spinning	Ring Spinning Frame	Roving	Ring-Spun Yarn	Bobbin/ Spool/ Cheese
Post- Spinning Processes	Winding, Doubling, Singeing, Reeling, Twisting, Winding-off Machines	Yarn	Yarn	Various (Skein, Bobbin, Package)

Each bobbin of roving is connected to a single spindle/traveler system on a ring frame. Frames contain a large number of spindles, around 1000 on modern equipment. The roving bobbins are larger than the individual spindles, so the spindles must be replaced, or doffed, more often than the bobbins. The reason for the small size of the spindles is the limitations of the traveler in the spinning systems. Too large a ring would require faster traveler speeds, which would cause too much friction between the traveler and the ring. The system works by drafting the roving through the use of aprons, and then connecting the yarn to a traveler which circles around the ring. As this occurs, the ring is moving up and down the spindle. These actions lead to the yarn being twisted and wrapped around the spindle evenly. When the spindle is full, it is doffed automatically and moved to a separate winding machine (Lord, 2003).

A breakage in the yarn during this process causes an end-down, or end-break. It is of utmost importance to reduce ends-down in any spinning process. Not only does a prevalence of ends-down signify poor yarn strength, they can greatly reduce production, as a spindle with a broken end is no longer producing any product until it is pieced back together (Subramanian & Garde, 1974).

Winding is needed in ring spinning due to the small size of the spindles. It would not be productive to send such a small package downstream for further processing, as it would lead to weaving or knitting inefficiencies. This is why multiple spindles must be combined onto a single, larger package. The winder puts pressure on the yarn as it works in order to straighten out kinks and break weak spots in the yarn, thus removing them. The winder also contains a clearing zone, which detects defects and irregularities such as slubs, thick places, and thin places; and the clearer cuts out any such deformity (McCreight et al., 1997).

Spindle ends and cut yarns are connected at the joining zone of the winder via splicing, a joining process superior to the knots of that were employed in the past. In fact, a spliced joint is only 20% larger than the regular yarn diameter and is able to maintain 80% of the strength that the rest of the yarn possesses. The result of winding is a larger package with fewer deformities that is ready to become fabric (Karthikeyan, Senthilkumar, & Patil, 2009).

2.3.7 Open-End Spinning

Another popular spinning option is rotor spinning, or open-end spinning. In this spinning process, the yarn is formed within a rotating disk. The method is known for its speed and low cost. Rotor spinning systems can have over 10 times higher productivity than their ring spinning counterparts (McCreight et al., 1997). While rotor-spun yarn shows good consistency, it is of lower strength, has a harsher hand, and is more susceptible to pilling than is ring-spun yarn (Shaikh, 2005).

The open-end spinning process is accomplished by first separating the drawing sliver into individual fibers, which are then transported via air assistance through a fiber transfer tube. These fibers collect on a groove on the wall of the rapidly spinning rotor. The fibers then twist around the end of the yarn, which is being pulled from the rotor and wrapped around a package. This package can be much larger than the spindles used in ring spinning, so no additional winding step is needed (McCreight et al., 1997). An image of a rotor appears in Figure 2.

A recent international market study revealed a large drop in short-staple spinning investments worldwide. Of the three categories in the study, rotor spinning had by far the strongest decline. Investments in open-end spinning equipment were down 66% from the previous year worldwide. However, while sales were indeed very low in Asia and Europe, investments in rotor spinning actually showed an increase in North America (Melliand International, 2009).

Figure 2 - Rotor (Rieter.com, 2009)

2.3.8 Air-Jet Spinning

Beyond ring spinning and rotor spinning, there exist other spinning systems, the most noteworthy of which is air-jet spinning. This is the newest and most efficient means of spinning available. Air-jet spinning is achieved, like the name implies, by utilizing nozzles to dispense jets of air that entangle the fibers. These machines save on space, labor, and energy when compared to ring spinning systems. Air-jet yarns are weaker than ring-spun yarns and stronger than rotor yarns. They show great evenness, but display a harsh hand (Shaikh, 2005). The structure of these yarns includes a core of parallel fibers being surrounded by wrapper fibers, a configuration which lends itself to very low hairiness, a trait that is responsible for the low pilling tendency of air-jet yarns (Beceren & Uygun, 2008).

2.3.9 End-Uses of Yarn

Weaving is one means of producing a fabric out of cotton yarn. The Latin word for weaving is 'texere,' the term from which the word 'textile' is derived (Spencer, 2001). The weaving process involves interlacing perpendicular yarns. The yarns are crossed at right angles over and under each other. This is normally done on a machine called a loom, the most modern of which can contain up to 5000 individual parts. The properties of a woven fabric are derived from not only the weave structure, but the properties of the yarn from which it is created (Adanur, 2001).

Another process that uses cotton yarn as an input is knitting. A knitted fabric structure is based on continuous intersecting loops. This technique can produce an enormous array of different structures (Adanur, 2001). Knitting is known for its ability to create form fitting or shaped articles. Modern knitting machines are faster than ever and capable of producing fabrics too intricate to ever be made by hand. When producing a cotton yarn for the purpose of knitting, the desirable product would be strong, yet fine and smooth (Spencer, 2001).

The operating procedure of a mill can be influenced by whether the cotton yarn is being sold directly to an outside buyer or if the company will be keeping the yarn in order to create fabrics to sell further down the line. A mill that will be using its own yarn to weave or knit fabric will want to minimize cost while still providing the requisite amount of yarn needed for fabric forming operations. A mill creating yarn to sell directly to an outside buyer will have a goal of maximizing profit; possibly sacrificing quality in the name of higher production (Subramanian & Garde, 1974).

2.4 Data in the Spinning Mill

If any sort of advanced data analysis is going to be applied to yarn spinning data, it is crucial to have an understanding about all of the different data that is collected in a spinning mill. Obviously, the quality and quantity of data will vary by mill, with modern mills having access to much more data.

Modern textile equipment is often equipped with advanced online monitoring systems. These systems collect data as the process is taking place and show this data on a display located on the machine. Online data can also be sent to a central monitoring location. For example, Trützschler machines are often equipped with the Kit System. This system measures quality sliver data from drawframes or cards. A typical display on a modern Trützschler card is shown in Figure 3 (Truetzschler.eu, 2009).

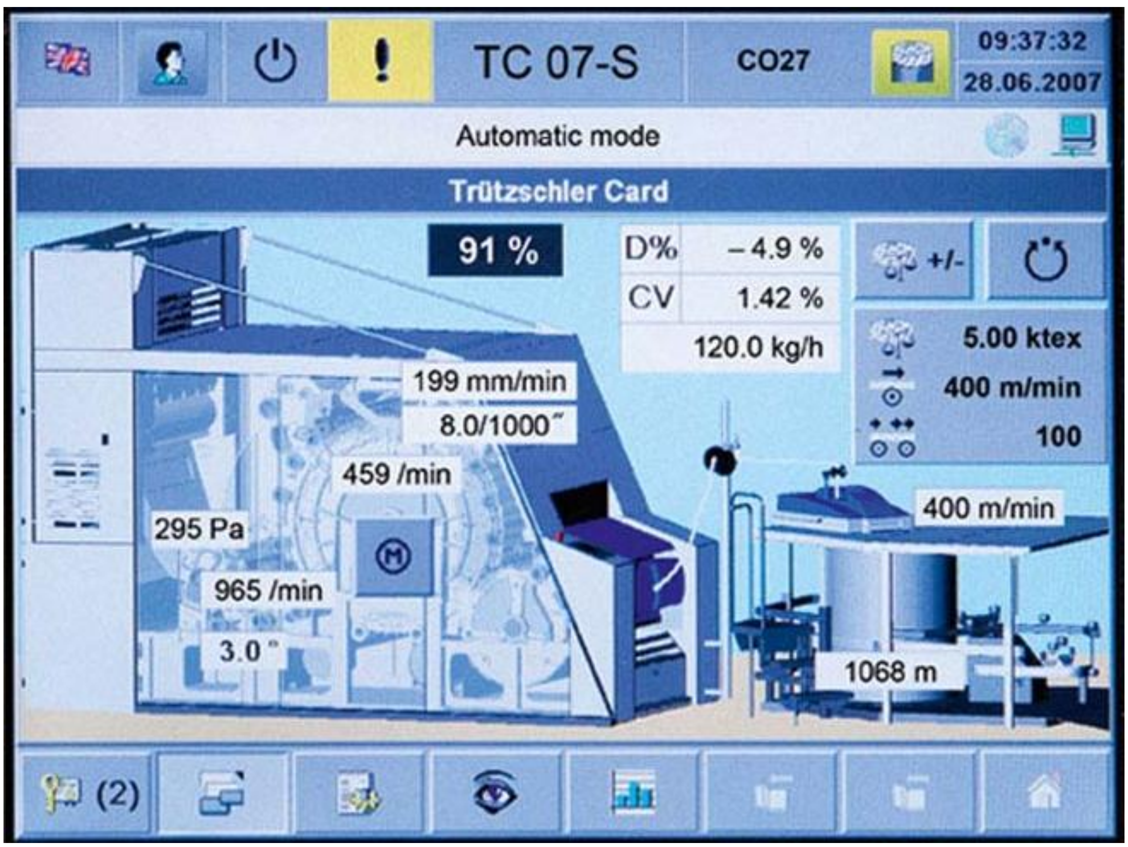


Figure 3 - Modern Card Display Panel (Truetzschler.eu, 2009)

Rieter is another example of a manufacturer of textile processing equipment. They produce machines for the entire spinning process. Their information and data collection system is known as SPIDERweb. The company reports that every machine from the blowroom to spinning frame can be linked into the Spiderweb system, where online data will be accumulated and stored. A company using this system will have better monitoring and require less offline testing. The system has a user-friendly computer interface that can be accessed locally at the plant, or distally via the internet (Rieter.com, 2009).

Offline data is the result of testing done to samples from the bale or plant floor. USTER is the manufacturer of most of the offline testing equipment used in the cotton spinning industry. The USTER HVI (High Volume Instrumentation) tester is the standard used to measure the properties of raw cotton fiber. The variables it measures in cotton fibers include length, uniformity, strength, micronaire, and trash content. It uses a variety of tools to measure these characteristics. Photoelectric sensors are used for length characteristics, air-flow instruments measure fineness properties, and clamps test the tensile measurements. USTER tries to ensure instrument stability and reduce variation between laboratories by including advanced automatic self-diagnostic features (Uster.com, 2009).

The United States Department of Agriculture (USDA) uses HVI to test bales of cotton. Their procedure calls for 4 ounce samples of fibers to be extracted from each side of the bale. This is done for up to 2 million bales per week (Cotton Incorporated, 2001).

The Engineered Fiber Selection (EFS) System produced by Cotton Incorporated is used to process millions of cotton bales every year. USDA uses the system to class all US cotton. The EFS system is managed through Cotton Incorporated's MILLNet software package. It uses USTER HVI data to enable the user to select and warehouse cotton bales more effectively. It saves historical bale data and produces reports used to select proper mixes of cotton. Bar codes are used on all bales at the mill to keep track of the current inventory and generate the best mix possible for the next laydown as well as plan for future laydowns and shipments (Cotton Incorporated, 2009).

The USDA uses EFS in order to set standards in the cotton industry. It follows that having a universal measurement system would be important in setting standardizations. These standards help to set market values for American Upland cotton as well as settle international trading disputes (United States Department of Agriculture, 2009).

USTER also makes the AFIS, which stands for Advanced Fiber Information System. This machine measures fibers from the bale, mat, sliver, and roving for things like neps, thick/thin places, and trash (Uster.com, 2009). In an AFIS machine, a combing roll individualizes cotton fibers while separating them from microdust and trash. The fibers are then linearly orientated and pass a near infrared optical sensor. They scatter light in proportion to their length and diameter, creating characteristic waveforms. These waveforms are interpreted by the machine and converted to actual measurements (Aarnink, 1996).

Additionally, USTER produces yarn testing equipment. Their latest is the USTER TESTER 5, which will measure yarn for qualities like mass coefficient of variation (Cv) and hairiness. For tensile and strength testing, many spinning mills have a USTER

TENSORAPID or USTER TENSOJET in their laboratory area to measure things like strength, force to break, and elongation of yarns. The USTER CLASSIMAT is used to classify thick and thin places in yarn and helps determine optimal clearing limits (Uster.com, 2009).

2.5 Data Mining

2.5.0 Introduction to Data Mining

Data mining refers to a relatively young process that has been around for barely 20 years. The computer science community has not yet come to an agreement on what exactly is encompassed by the term. Explanations can vary, going from simple to specific. The definitions range from, "Data mining refers to the analysis of the large quantities of data that are stored in computers (Olson & Shi, 2007)," to "Data mining is the process of extracting previously unknown, valid, and actionable information from large databases and then using the information to make crucial business decisions (Cabena, 1997)," and many in between. Some definitions even explicitly mention terms such as "machine learning," "pattern recognition," and "visualization" (Larose, 2005). It seems that all of these definitions cannot seem to agree on either the process or implementation of data mining, but they do all share the general idea of using specialized techniques to discover information that was previously unapparent in a vast collection of data.

2.5.1 History of Data Mining

Data mining arose out of necessity in the 1980s. Advances in computer technologies lead to a continued increase in the amount of data capable of being stored by companies. Because of this, companies began amassing more and more transactional data from their day-to-day operations. The copious amounts of stored numbers became too vast for the traditional statistical techniques at the time. Fortunately, the increase in computer power that lead to this problem also made it possible to utilize more complicated avenues through which knowledge might be extracted from this data. By 1995, data mining was really beginning to boom with Montreal's First International Conference on Data Mining and Knowledge Discovery followed by the 1997 launch of *Data Mining and Knowledge Discovery*, the first journal specialized to the subject (Britannica Online Encyclopedia, 2009).

Figure 4 shows data mining as a confluence of many technologies which were coming of age at the time. The synergy of these different disciplines is what makes data mining so effective at ascertaining information hidden within a compilation of data.

The process has continued being honed over time thenceforth. However, whatever progress is made in the world of data mining is usually matched with an equal boost to the amount of data being stored, which is ever-increasing. As time goes on, the complicity and quantity of data being stored keeps growing, but so too does the power of the data mining techniques that can be utilized to make sense of it all (U. M. Fayyad, 2002).

It has been said that data storage capacity is actually increasing faster than the rate at which computer power is increasing, which is said to double every 18 months according to Moore's Law. This Storage Law indicates that storage capacity is increasing at twice that rate, giving data accumulation the edge in this race, meaning that the field of data mining must continue to respond with further advancements (U. M. Fayyad, 2002).

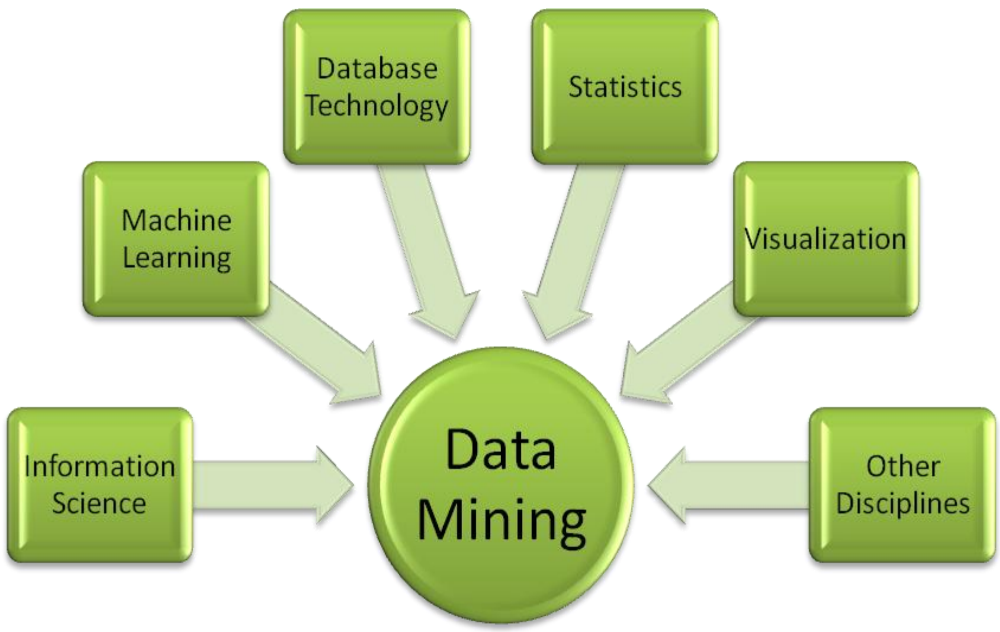


Figure 4 - Data Mining Graphical Model (adapted from Manik, 2009)

2.5.2 The Data Mining Process

Data Mining is not a specific action, but a process. Numerous models have been made to show the process, but most are similar to the CRISP-DM Data Mining Process displayed in Figure 5. CRISP-DM is an acronym which stands for CRoss-Industry Standard

Process for Data Mining (CRISP-DM.org, 2009). The process begins with *business understanding*. This means that the data analyst should be familiar with the business goals in order to effectively develop data mining goals and a project plan. Next comes *data understanding*, in which data is collected and assessed. The data quality is noted and the data miner gets an idea of what information could potentially be hidden within. The data can be found in what is called a data warehouse, which is a collection of databases. *Data preparation* follows. This is everything involved with selecting, cleaning, and formatting data so that it is ready to be mined (Olson & Shi, 2007).

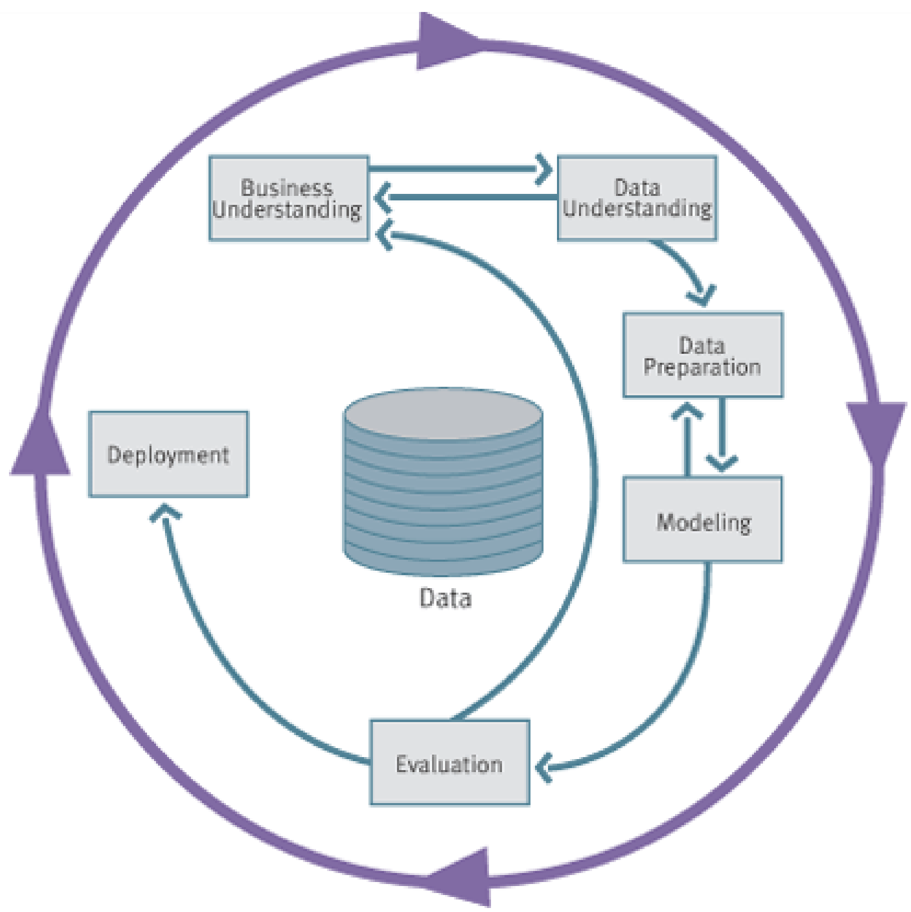


Figure 5 - CRISP-DM - Process Model (CRISP-DM.org, 2009)

There is perhaps no data mining problem as troublesome as having dirty data, which means that data preparation is arguably the most important aspect of the entire data mining process. If data is not properly cleaned, it could have a very significant impact on the results of the subsequent data mining. Cleaning data is not easy either, as it is the most time consuming step in the mining procedure. This act of improving data quality so that it can be mined can be expected to take at least 60% of all the time and resources associated with the project. It involves making alteration by means of both adding variables and modifying the values of variables that are already in the database so that the data mining tool will be able to distinguish the patterns within the data with minimal difficulty (Pyle, 2003).

One way in which data can be dirty or corrupted is duplication, which is when one record appears multiple times in the database. Missing data fields are another form of data corruption, and they can be caused by human error. Outliers are another cause of poor data quality and involve a number that is very different from the other numbers in its field. This could be due to a decimal point being in the wrong place or an extra zero accidentally being added to an input. Sometimes trouble can also be found in fields such as date or time, which can be expressed in many different formats. The data miner must be certain that the entrees for such a field agree with each other in regard to format, especially when combining multiple data sources (Sethi, 2001).

In the CRISP-DM data mining process found in Figure 2, *data preparation* is followed by *modeling*, which is when the cleaned data is put through one or more data mining tools so that information might be gleaned from it. The CRISP-DM figure shows an arrow pointing back at the *data preparation* phase because sometimes the form of the data may require alteration depending on the model being used (CRISP-DM.org, 2009).

There are many data mining tools available for modeling and the selection of which to use in a particular data mining problem is based on what data is available to work with and what the goals of the data mining are. There are a few typical goals normally associated with data mining. *Data processing* deals with the “selection, integration, filtration, sampling, cleaning and/or transformation of data” (Braha, 2001). *Verification* tries to fits models to data based on previously held hypotheses. *Regression* deals with finding any relationship that may exist between values in a database. Finding such relationships can lead to models that can then be applied to predict the values of future records. If *classification* is the goal, classes are created and then the records are assigned to the different classes based on their analysis. *Clustering* partitions data into groups based on similarities of characteristics. *Association* goes about finding rules to associate one attribute set to other sets of attributes. *Sequential pattern analysis* looks at the events over a time line than cause an eventual final event. *Model visualization* is concerned with displaying knowledge in a manner that is understandable and interpretable. *Deviation analysis* simply looks at deviation over time, which can entail deviation from the mean or deviation from a certain set standard (Braha, 2001; U. M. Fayyad, 1996).

After *modeling* in the CRISP-DM data mining process is *evaluation*. This is where the results of the modeling efforts are examined in the context of business goals and expectations. From this, a decision can be made as to whether more analysis might be needed, in which case more modeling can be performed. If the results meet the business and data mining objectives, decisions can now be made based on the results (Olson & Shi, 2007).

The final step of the process is *deployment*, in which the models obtained in the previous steps can be applied to the operations of the business. They can either verify that the business is functioning properly or can be used to help implement changes. They can also be used for making predictions and identifying key situations. It should not be assumed, however, that these models will hold true forever (Olson & Shi, 2007). Businesses change constantly, some more than others, and this means more data mining and newer models may be needed again down the road.

2.5.3 Data Mining Techniques

There is a large tool box of data mining techniques available to the modern data miner, ranging from fairly simply to very complex. The choice of tool will come down to the data mining goals and the data available.

Once the data has been chosen and cleaned, one of the simpler tools available is a structured query language (SQL) statement. This can be used to find information that may already be on the surface. It has been said that 80% of interesting information in a set of data can be found just using SQL. These statements simply allow the user to request certain values from the data set, thus making smaller, easier to understand tables (Gonzalez &

Kamrani, 2001). For example, if there was a database that contains statistics for the Boston Celtics 2009-2010 season, and the user wanted to be given the first and last name (assuming these are columns in the database) of all players who are averaging 2 or more steals per game (SPG), the SQL statement seen in Figure 6 would produce such a table. This would eliminate all but the columns asked for and display only the entities meeting the qualifying criteria.

SQL Statement:

```
SELECT [first name], [last name] from Celtics WHERE SPG >= 2
```

Resulting Table:

First Name	Last Name
Rajon	Rondo

Figure 6 - Sample SQL Statement and Resulting Table

The next tool at the data miner's disposal is statistical techniques, where statistical process controls (SPCs) such as histograms, scatter diagrams, and Pareto diagrams are used to plot the data and possibly fit the patterns to a curve or learn further information from the data. Even if no important relationships are discovered during the statistical analysis of the data, it should at least show which relations definitely do not exist, and this information can be useful should the data miner move on to more advanced techniques (Gonzalez & Kamrani, 2001).

Regression analysis enables the researcher to compare 2 variables, even when many more variables are present. The tool manages to hold all other variables artificially constant, so that the relationship between the two variables being examined can be focused on (Levitt & Dubner, 2006).

Visualization is another tool that can be applied to the data. Such tools continue to improve with technology, allowing for better and better graphical representations of the data by utilizing 3 dimensions and various colors. An example of this can be seen in Figure 7. This is a visualization provided by SAS Enterprise Miner. It allows for someone with the proper experience to get a feel for what relationships may exist in the data. It combines computer power with human pattern recognition skills (Cox, Eick, Wills, & Brachman, 1997). Visualization is often a first step, helping the data miner decide how to proceed with further mining.

The way in which the data is stored can also enhance the data mining process. This can be seen in online analytical processing (OLAP). This stores the data in such a way as to reflect the real dimensionality of the data. It actually reduces the data size and calculation requirements, thus allowing the user to access fast insightful information. It is like storing the data on a cube, thus adding a third dimension behind the data, so that the user can interact with the data and get quick answers to queries (Wu & Li, 2003).

A tool well suited for small problems is case-based learning. This is also known as the k-nearest neighbor algorithm. This is a way of classifying an unknown pattern. The k is a variable controlled by the user. If k=3, for example, the unknown pattern will be put into whichever class is most popular amongst its 3 nearest matches (Sethi, 2001).

Figure 7 - Visualization with SAS Data Miner (SAS.com, 2009)

Decision trees split the data into nodes. The data is branched out through a series of decision nodes, with each possible outcome becoming a branch which extends downward from the root node, and ends at a leaf node. Figure 8 shows a simple example of this concept. This tree allows for an accurate prediction to be made for whether a game will be played based on a few variables about the conditions. The branches terminate once games are either always played or always not played. This shows that that combination of decisions should always lead to the predicted outcome (Larose, 2005).

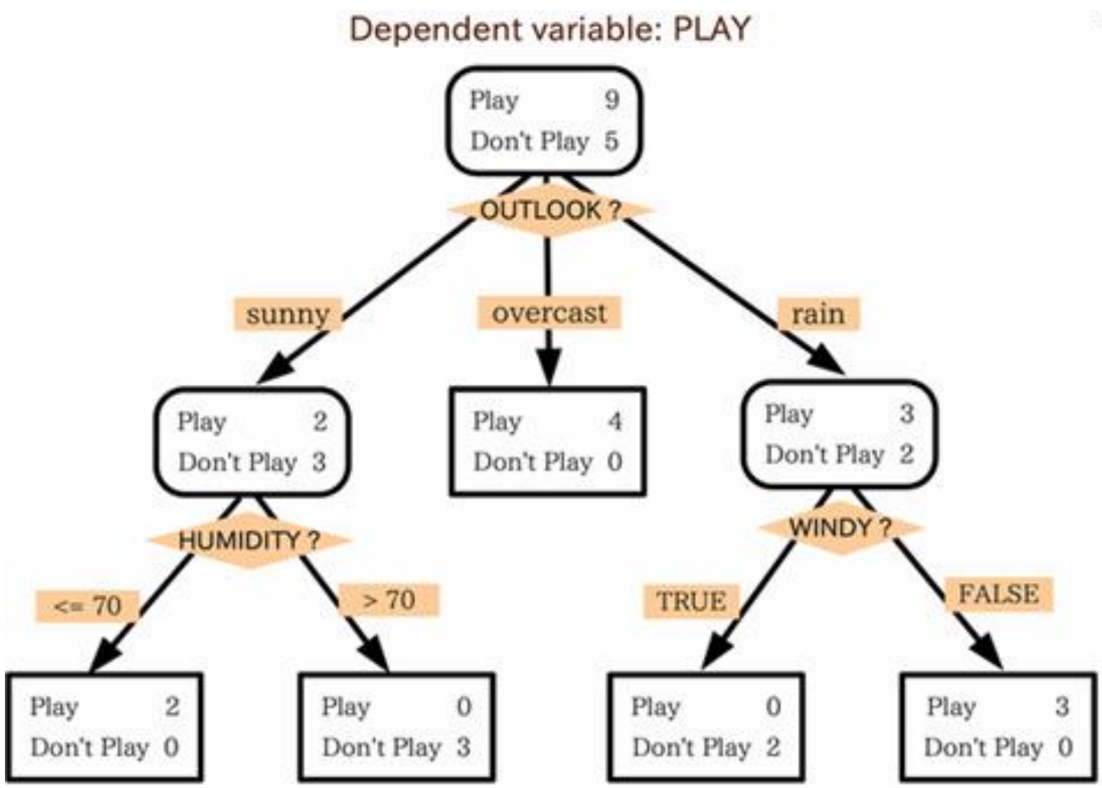


Figure 8 - Simple Example of a Decision Tree (LIS.Illinois.edu, 2009; Sethi, 2001)

Neural networks are another tool that data miners are familiar with. This particular tool was actually designed to mimic the action of the human brain, which function through closely connected sets of neurons. These are an effective tool for noisy data, which has many random fluctuations. Each neuron gathers information from other neurons then dispatches its output response. Neural networks consist of an input layer, one or more hidden layers, and then an output layer. As shown in Figure 9, every node in each layer is connected to each node in the following layer (Larose, 2005). This is actually a two-phase process. Before the network can make classifications and predictions, which is called decoding, it must be trained, or encoded. This involved putting examples with known outcomes through the network so that it can learn (Gonzalez & Kamrani, 2001).

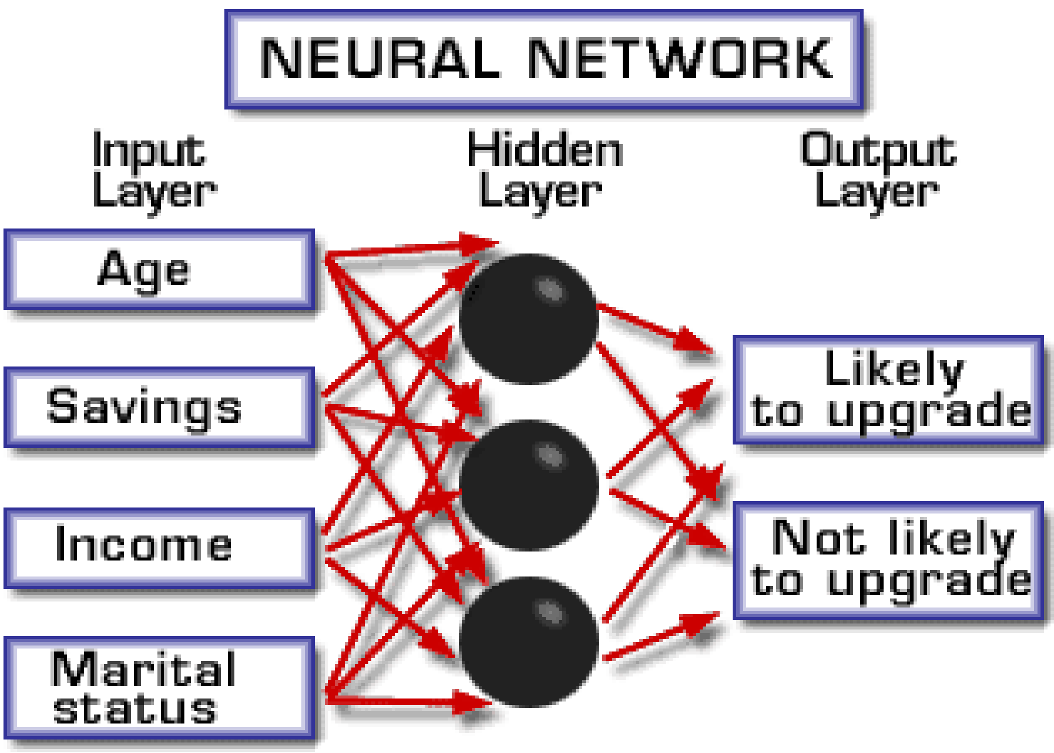


Figure 9 - Neural Network Representation (Siggraph.org, 2009)

Genetic algorithms, like neural networks, were also designed to imitate nature. In the case of genetic algorithms, the process of evolution is mimicked. They create a population of individuals that are tested for best fit. Those with peak fitness are used to produce the next generation. Crossovers and mutations are also used. Through many generations, this algorithm will produce individuals with a better and better fit (Liu, Yu, & Motoda, 2003).

The aforementioned data mining tools can all be utilized through data mining computer programs. These programs are designed to run the tools through a user-friendly interface, allowing complicated data mining to be done with minimal confusion. The programs are constantly being updated and newer versions are regularly released in order to stay on top of the latest data mining developments and to keep pace with ever-growing data warehouses. Popular software programs include PASW Modeler from SPSS, Salford Systems' assortment of programs (CART, MARS, TreeNet, RandomForests), RapidMiner, SAS Enterprise Miner, and Knowledge SEEKER from Angoss (The-Data-Mine.com, 2009).

2.5.4 Traditional Uses

Data mining's most obvious and perhaps most popular use is for market research. Companies are able to mine their data to categorize those customers who will potentially do more business with them and those who probably will not. This means that the company can focus its marketing toward those most likely to respond. One area of business that has taken advantage of data mining in recent years is the casino business. Most casinos will give out a card to customers with allows them to accumulate points for playing games, shopping, and even eating. These cards also allow the casino to amass data on the behavior of all of their customers. The patterns they discover in this data through data mining tells them which customers to market to in order to maximize revenue and even which customers may have lost in the past and might require a special offer to entice them to come back (Olson & Shi, 2007).

The other current most popular applications of data mining can be found in Table 3. This chart shows that the majority of the applications of data mining are used to make identifications. For the most part, these applications enable the company to learn more about their customers. These types of businesses have been taking advantage of data mining since the beginning. Noticeably missing from this chart, however, are manufacturing applications of data mining. This is one area where data mining appears to being vastly underutilized. However, with plenty of data being accumulated during manufacturing, some companies are now starting to improve their process and products through data mining in manufacturing.

Table 3 - Data Mining Application Areas (adapted from Olson & Shi, 2007)

Application Area	Applications	Specifics
Retailing	**Affinity positioning**	**Position products effectively**
	Cross-selling	**Find more products for customers**
Banking	**Customer relationship management**	**Identify customer value**
		Develop programs to maximize revenue
Credit Card Management	**Lift**	**Identify effective market segments**
	Churn	**Identify likely customer turnover**
Insurance	**Fraud Detection**	**Identify claims meriting investigation**
Telecommunications	**Churn**	**Identify likely customer turnover**
Telemarketing	**Online Information**	**Aid telemarketers with easy data access**
Human Resource Management	**Churn**	**Identify potential employee turnover**

2.5.5 Potential in Manufacturing

The application of data mining in manufacturing is infrequently used on a large scale. Decision makers are hesitant to commit to the complex broad-based integration that may be required to perform data mining. The opportunities for data mining are quite apparent. Firstly, relationships can be determined between internal factors that occur during the manufacturing process and external elements such as sales or product quality (Braha, 2001). Any relationship here may show how important a certain area of the process is. The company can, in turn, make sure to give this internal factor the proper attention. Other areas exist within manufacturing where data mining could be employed as well and can be seen in Table 4.

One study used data mining to optimize the testing of semiconductor wafers. Data mining was used to determine the best locations to test on each wafer. Once the correct locations are discerned, testing becomes faster and more accurate, as fewer measurements need to be taken to represent 100% coverage of the wafer. This is important in manufacturing because testing samples like this is a non-value added step and having a quick turnaround time is getting more important due to competition (Lee & Park, 2001).

The advent of computer integrated enterprise (CIE) goes hand-in-hand with the necessity for data mining. As more companies are linking all of their systems together, massive data warehouses are created (Cummins & Knovel, 2002). Contained within this data is information that will lead to knowledge of how to better run the business.

Table 4 - Data Mining Opportunities in Manufacturing (adapted from Braha, 2001)

Acquiring expert manufacturing knowledge for decision making
Adaptive machine interfaces
Diagnosing faults
Forecasting for supply and delivery
Generating operational control policies
Learning for robotics
Predicting defects
Process control
Quality control
Scheduling preventative machine maintenance
Summarizing complicated data

Manufacturing could appreciably benefit by taking advantage of the data mining process. Manufacturing facilities are at no loss for data, and all the numbers they accumulate may hold some insight into how they could improve their product or process (Braha, 2001).

2.5.6 Cotton Yarn Spinning Applications

Modern cotton spinning is an example of a manufacturing setting that generates the amount of data necessary to actualize data mining. There have been a couple of interesting studies done that show the capabilities of a neural network in terms of comparing fiber characteristics to final yarn characteristics. One study found that neural networks produced better results compared to previous mechanical and statistical models. The other used a neural network to rank the impact of fiber characteristics on different final yarn characteristics (Guha, Chattopadhyay, & Jayadeva, 2001; Jayadeva et al., 2003).

In the relatively short time that data mining has been around, it has already been honed into a very effective process. Many businesses have benefited from the knowledge they have gained through data mining. Such knowledge has led to many successful decisions that might not have otherwise been made. As computer technology continues to improve, the amount of data being created and stored will only continue to grow, which leads to ample chances to mine data (Olson & Shi, 2007).

2.6 Previous Works

There have been several works in the preceding decades that have explored data in the cotton spinning industry. This topic has long been an area of interest, particularly at the Institute of Textile Technology (ITT) and North Carolina State University (NCSU).

In a 1996 thesis entitled *Data Communications in the Computer Integrated Manufacturing for the United States Textile Spinning Industry*, Beck explores the current state of data management by analyzing the results of a survey conducted by the American Textile Manufacturers Institute. This research found that, at the time, the textile industry was lacking in data communication. The main problems were the lack of industry-wide communication standard, outdated machinery, and "islands of automation" caused by systems coming from different vendors not being able to communicate with each other and exchange information properly. It was also found that most plants were mainly interested in data communication as a means of improved control (Beck, 1996).

The following year, a similar thesis was completed by Rasmovich. *Information Technology and Information Engineering in the United Stated Textile Spinning Industry,* advocated the need for spinning companies to streamline their data management in order to get a clearer data architecture. Because massive volumes of data can hide and obscure the most necessary information, the decision making process can be delayed to the point of no longer being timely or effective, which is a condition known as information overload. The research conducts surveys in order to determine what data is most important to plant managers and develops a data architecture in which prudent information can be accessed quickly and easily; thereby facilitating successful decision making (Rasmovich, 1997).

In 2006, Karpe delved into textile-related information management with *Weave-Room Performance Decision-Making Process in Textiles: Mapping an Information Engineering Methodology*. This research used case studies of weaving plants to create "AS-IS" decision making models. From these, the greatest aspects of the current practices were used to create "TO-BE" decision making models, which represent the best possible approaches to decision making within the different management levels of a weaving mill. This research served to look at the day-to-day management of a weaving facility from an engineering point of view (Karpe, 2006).

In 2008, Echeverria completed a thesis named *Yarn Specifications and Performance Metrics for Short Staple Yarn Manufacturers.* This research looked into the different characteristics of yarn capable of being measured and the hierarchy of their importance in a final yarn product. The research found that different characteristics have varying importance depending on the respective end-use. Surveys determined companies in the supply chain that purchase yarn are generally foremost concerned with yarn type, yarn count, and fiber content; and that more detailed characteristics such as tensile properties are mostly of significance to the mill only for process control purposes(Echeverria Magariänos, 2008).

The initial attempt at using data mining in a yarn spinning context was Schertel's 2002 dissertation, *Data Mining and its Potential Uses in Textiles: A Spinning Mill*. This research used a case study to determine the viability of conducting data mining on a particular plant using the current data already flowing through the mill. The study showed insight as to the different points throughout the process where data was being amassed. A similar study completed in 2008 by Anderson was entitled *Applicability of Data Mining in Yarn Manufacturing*. Anderson expanded upon a model, first created by Schertel, that shows some of the online and offline data available as fiber moves through an open-end spinning process. This is presented in Figure 10. Unfortunately, not enough data was gathered for significant data mining in either research, and both were complicated by incompatible data formats coming from the different collection points. However, both obtained valuable insight into the drawbacks of current data management practices and outlined strategies for potentially gathering and cleaning sufficient data. Such strategies would enable for data mining-level analysis to be enacted in a typical plant; one which normally maintains a decent cache of data but tends to utilize it only for benchmarking and controlling quality (M. E. Anderson, 2006; Schertel, 2002).

Data mining was achieved with a larger set of data by Daley in the 2008 thesis, *Application of Data Mining Tools for Exploring Data: Yarn Quality Case Study*. This research attempted to use a variety of data mining tools to determine the main causes for ends-down in an open-ended spinning system. The study mainly focused on HVI data and cotton crop variety as variables responsible for ends-down. This research concentrated mostly on the data mining techniques themselves and their viability as a means of exploring textile data. The research did not delve very far into analyzing the data management practice of the US spinning industry (Daley, 2008).

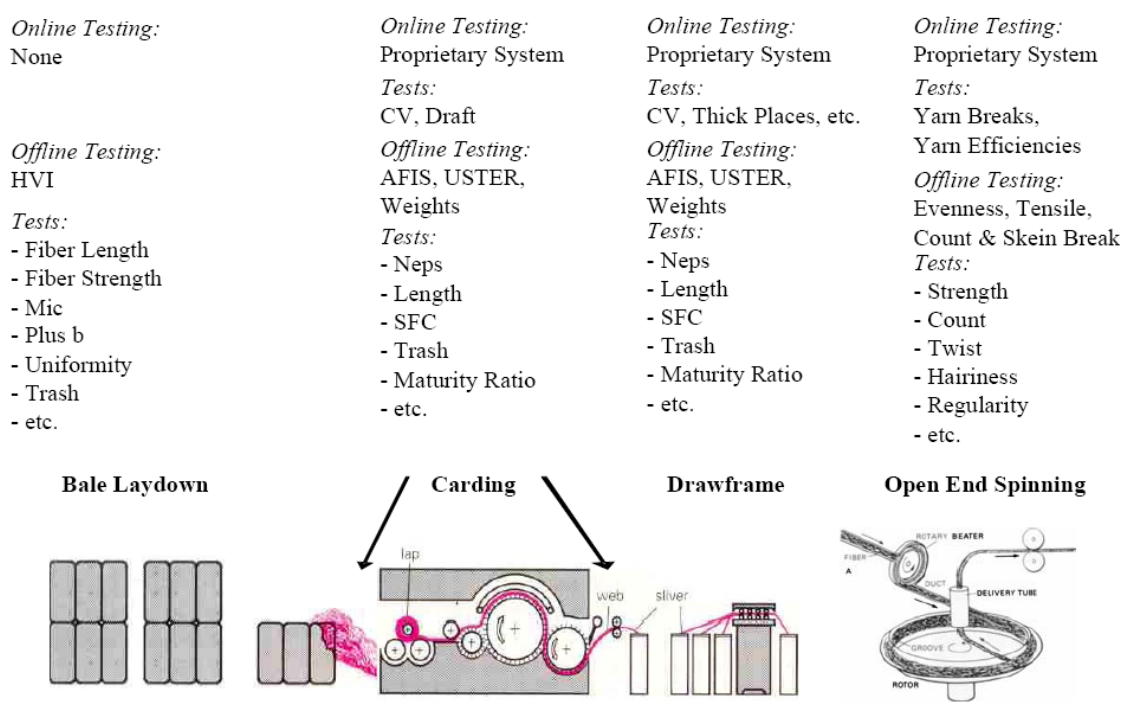


Figure 10 - Collection Points and Data Elements (M. E. Anderson, 2006)

Chapter 3 - Methodology

3.0 Introduction

This chapter will present the methodology used to accomplish this research. First, the problem and research goals will be clearly expressed. Next, the methods and procedures being employed to accomplish these goals will be explained.

The problem being explored in this study is one of potential. The cotton spinning industry as a whole is just not living up to its potential in terms of data management (Rasmovich, 1997). This is especially apparent when comparing yarn manufacturing to more modern, customer-oriented businesses (Olson & Shi, 2007). There is a massive amount of data available at a cotton mill at all times. This data comes in the form of such things as online monitoring and offline testing. The newer the machines being used, the more data the plant has at its disposal.

Unfortunately, most modern mills continue to operate as they did a generation ago, when the abundance of data and computing power were much more limited. For some plants, this is necessary because they are still using older equipment. Some mills, however, are making grand investments in newer technology and are not harnessing all of the value which they have purchased.

3.1 Research Goals

The primary goals on which this research was focused are presented in Table 5. These goals build upon each other and the achievement of the first 2 is vital to the completion of the third.

Table 5 - Research Goals

1. Determine the current state of data management in the American cotton spinning industry.
2. Collect and analyze data using data mining techniques to determine relationships that may exist in the data available at a mill.
3. Make a new model for cotton spinning data management based on the capabilities of a typical mill and the relationships in the data.

3.2 Methods

The first goal was accomplished by traveling to a number of US cotton spinning facilities and other industry related visits. The data capabilities of each plant were observed and noted. It was determined that seeing a sample of 8 - 12 plants would establish a clear picture of the state of the industry. Observations of the plant floor as well as interactions with personnel elicited the information from these travels that led to knowledge about the state and capabilities of the industry at large.

The means of collecting data was determined during the research travels. Factors influencing this process included deciding which plant would provide the willingness and means for the collection of a proper amount of data. If data was to be manually collected from a plant, the methods researched in the previous works section of the literature review could have been utilized in order to obtain an appropriate data set. Another possible outcome was that a plant that stores long-term data would be able and willing to provide a data set of recorded data from the recent past.

In either case, the data would have to be properly formatted and cleaned. After this, the data was to be analyzed using data mining techniques such as those mentioned in Chapter 2 of this study.

It was determined that the insight gained from accomplishing the first 2 goals should lead to the actualization of the third goal. The results of the modeling and analysis were used in conjunction with the understanding of the capabilities of the typical spinning plant to create a data management collection and utilization plan, which will potentially increase the productivity and efficiency of a mill because of enhanced decision making.

3.3 Limitations

This research is not without limitations. For one thing, some of the data being collected for analysis was real-world data. These were not the results of a science laboratory experiment in which the constants were held and everything was meticulously observed and recorded. There exist circumstances not necessarily reflected by the data. For instance, if a fiber is creating atypical data, adjustments may simply be made to the machines to even out the fluctuation. Ideally, such a spike in data would reveal relationships with data points downstream in the process, but this may not hold true at a work place where adjustments are made immediately. Also, many tests are taken on just samples of cotton, and not on every fiber. For example, on a 500 pound bale, only 8 ounces of fiber are tested to determine the properties of the entire bale.

Other limitations are due to cotton itself. Being a natural fiber, cotton crops vary depending on a copious slate of factors. Any relationships found between cotton fiber qualities in any given year may simply not hold true the following year with an entirely new crop. With these limitations in mind, the first goal of the methodology was undertaken.

Chapter 4 - Business Understanding

4.0 Introduction

The following chapter explains the different industry visits and research travels that were made in performing this research. These visits where make in order to accomplish Research Goal 1. In the CRISP-DM process model for data mining presented in Chapter 2, the first step is *business understanding*, and it is that step which these travels aim to address. The locations and events attended are presented and applicable findings are discussed.

In forging an expertise in the subjects crucial for this research, it was determined that additional learning beyond just a literature review would be essential. Not only can concepts be absorbed more clearly through a first-person experience, but it is the only way to get a real time as-is perspective into the contemporary state of the industry. Because of this, the industry visits performed over the course of this study served multiple purposes, which are shown in Table 6.

Table 6 - Purpose of Industry Visits

1. They enhanced the learning of the concepts studied in the literature review process.
2. They gave insight into the real-world data management practices currently being employed.
3. They served to put into context the data that was accumulated for analyzing.

A variety of travels were undertaken throughout the course of this research, and these experiences were vital to the understanding of the subject matter, the interpretation of results, and the formation of conclusions. The purpose of these travels, which covered many miles and six different states, varied depending on the phase the research was in at the time of the trip.

4.1 Non-Mill Travels

The first conference attended was the 7th Annual Six Sigma Forum, which took place at Research Triangle Park in Durham, North Carolina and was sponsored by IBM and the NC State College of Textiles. Six Sigma is a method of process management. It focuses on statistical techniques to make decisions and enhance quality. While this research does not aim to create or recommend a Six Sigma system for the cotton spinning industry, some of the techniques utilized in this style of management cross over to data mining. For this reason, many of the presentations at this conference were helpful to this research.

Additionally, since this research is not dealing with a terribly massive quantity of data, the software packages prevalently used in the presentations at the Six Sigma conference should be able to perform the necessary techniques in the data analysis section of this research. The two software programs obtained and learned for use in this study are SAS Jmp, a stand-alone analysis program, and SigmaXL, which is an add-on program that works through Excel. These will be discussed further when they are employed later in this research.

The second set of conferences attended with an aim to further this study was the 2010 Beltwide Cotton Conferences, which were held in New Orleans, Louisiana. This is the US cotton industry's largest gathering. It was a collection of conferences which dealt with all facets of the cotton industry. In addition to the presentations dealing with important cotton production factors such as irrigation and pest control, there were a couple of conferences tailor made for the subject matter of this research.

The two conferences at the Beltwide of most interest to this research were the 22nd Annual Cotton Quality Measurement Conference and the 18th Annual Cotton Utilization Conference, which featured a session on textile technology. Applicable subject matter covered at these conferences included comparing fiber testing methods, the influence and definitions of measurements such as short fiber content, and attempts at predicting selected properties in rotor yarns.

The telephone conversation held with Cotton, Incorporated was conducted during the earlier stages of research, and was used to supplement findings taken from literature and from the company's own website. The EFS system and its MillNet software were discussed, specifically the best way to involve EFS information in a data management system. It was explained that a newer version of the software, called MillNet 32, is capable of running on a Windows platform, making it easy to export data from the software into Excel. Since the different sources of data found throughout a spinning facility are not designed to communicate with each other, the next best way of getting multiple kinds of data into one location is if they are all capable of being exported to the same format. The standard for this is Excel. If all of a plants different electronic data can be easily moved into an Excel spreadsheet file, it can be combined and analyzed with data mining techniques. If two systems are not capable of being swiftly exported into the same format, manual input may be the only feasible solution.

The USDA's Southern Regional Research center provided helpful information and a first-hand look at a government research facility. However, the most important outcome of this particular visit turned out to be the human contacts established there, through which data would be obtained in the future. This data will be discussed further in the data understanding section of this research.

The textile museum helped to put the historical concepts of cotton spinning into proper perspective and deepened the understanding of how far the industry has progressed over time. The American Textile Museum in Lowell, Massachusetts gave a hands-on look at some of the information reviewed in the cotton spinning history section of the literature review in Chapter 2.

One location that was unique to the other spinning-related travels was a visit to a spinning machinery manufacturer. This trip included a look at some of the newer machines in action, as well as a chance to ask some question. One interesting note gathered from this visit was that 2009 had been a tough year for selling equipment. It was mentioned in

Chapter 2 of this study that the sales of spinning equipment are in a worldwide decline, and this insight was consistent with those findings.

This is relevant to this research because it shows that the cotton spinning industry, at least in the Southeast United States, is not actively acquiring the latest equipment with the most modern capabilities at this time. Most plants are content with the equipment they have had for years, and this must be kept in mind later in this study when new data management frameworks are suggested. Even the newest equipment that this machinery manufacturer demonstrated was not capable of compiling long term data right out of the box. The machine would have to be set up to load information into a host system for long term storage, elsewise it would be overwritten after a couple of shifts.

4.2 Spinning Mill Observations

4.2.0 Introduction

It was vital to observe a solid number of plants due to the nature of this research. In all, 10 spinning facilities were visited. The purpose of this study is not to simply streamline the data management of a particular plant, nor is it to solve the unique problems of any one mill. For this research to satisfy its vision, the data management of plants in general must be understood. An understanding must be achieved of not just the most technologically advanced or the least technologically advanced of plants, but of all spinning plants. The conclusions drawn in this study are meant to be valuable to the entire textile industry, and this can only be accomplished if a wide swatch of the industry is examined and scientifically scrutinized.

Obviously there are limitation to the time and space available for this research, so there is a necessity for a deal of focusing. For example, all of the facilities visited work mainly with cotton fibers, are in the United States, and are located in the Southeast region of the country. However, even within these confines, the visited facilities still spanned an array of attributes. Upon visiting so many plants, it was apparent how many variables there are within a yarn spinning plant. Table 7 shows some of the variables that tended to differ between each facility.

Table 7 - Variables Among Spinning Plants

Cotton Variety	**Management Philosophy**	**Software**
Customer	**Parent Company**	**Spinning Method**
Inventory Level	**Plant Size**	**Testing Equipment**
Location	**Quality Control**	**Yarn End-Use**
Machinery Manufacturer	**Run Size**	**Yarn Styles**

The following is a discussion of the results of this pilot study of data management in cotton yarn spinning. This information was obtained during the various industry visits mentioned above. The findings of these travels were obtained through discussions with a variety of personnel at each location and through the first-hand observations of the researcher. While no formal interviews were performed, the checklist of questions found in Table 8 was referenced to ensure a thorough visit that would enrich this study properly. These questions were compiled based on those used in research done by M. E. Anderson (2006) as well as additional curisities of this research.

Most visits began with a sit-down discussion with a supervisor or plant manager. This was useful to prepare for what was to be seen on the tour. It also provided a chance to discuss topics such as what quality characteristics are deemed most important at this particular mill, and to learn information about the size and production norms of the facility.

From here, either the supervisor or a second employee would lead a tour around the mill while answering questions and pointing out interesting features. After this, most plants introduced the researcher to a lab technician, who showed some of the routine testing being done at the lab, the testing equipment being used, and what was being done with the data accumulated in the lab. After this, a few more questions were usually asked to the supervisor and the visit was concluded.

Table 8 - Data Management Research Questions (adapted from M. E. Anderson, 2006)

What production machines are used and what are their capabilities?
What data is collected from the machines?
What is collected online and what is collected offline?
Where and how often is the data collected?
Where are the collection points?
What are the lag times between processes?
How is the data analyzed and utilized?
What is the data format?
For how long is the data stored?
What software is used?
For how long has data been managed this way?
What are the most important characteristics of product quality?
What testing is performed and what testing methods are used?

The goal of this section is to present the findings of these travels such that they are most conclusive, understandable, and helpful for the purpose of achieving the research objectives. The information is presented mainly as a summary because the most important aspects are those which seem to apply to most or all of the subjects. The goal was to obtain a snapshot of a typical operation. Outliers and extremes that were observed are noted when appropriate and discussed against what is deemed to be the norm.

The most suitable way to present the information compiled from these visits, for the sake of clarity and understanding, would be through a descriptive tour of a representative plant. The different constituents of a typical plant will be discussed individually, and for each, a divulgence of relevant information and noteworthy observations will be made. The following is based on the cumulative observations of 10 distinct spinning facilities.

4.2.1 The Manager's Office

The manager's office is normally located close enough to the plant that the manager would be able to routinely walk around the floor and quickly respond to alerts. The office usually has a computer, and at some locations this computer can link into some of the online monitoring systems of the plant. This allows the manager to monitor the current production levels of the plant. Otherwise, the manager can at least scan data reports through e-mail. These reports will be further discussed when the plant's laboratory is visited later.

4.2.2 Blowroom

Yarn production begins with the raw material, cotton fiber. Incoming 500 pound bales are placed as inventory in a warehouse, the size of which varied from plant to plant. There was a general trend amongst several of the mills toward reduced inventory size. For example, one mill is now holding only a quarter of the inventory it had in the past. This style of inventory is associated with lean manufacturing, which was the philosophy being specifically preached at one location. Such just-in-time delivery methods save on floor space and other expenses related to holding inventory.

Bales are moved from the warehouse to the laydown area. As described earlier, the bales are chosen and positioned to ensure a similar average of properties. For example, one mill's aim is to ensure that color, micronaire, length, and strength HVI measurements are close to constant at every laydown. This is accomplished at most mills using EFS and MILLNet software and having the computer configure each laydown. For efficiency and organization, the bales needed for the subsequent laydowns are placed around the perimeter of the area in order to reduce the time needed to set up the ensuing laydown.

The software used here is designed to save historical data. It can give a history of bale laydowns and the HVI data of all the bales that have been used for as far back in time as the mill chooses to store the data. The mill can check how properties have changed over time and also look at other historical aspects such as the cost per bale. For compatibility with other data in the plant, the MILLNet 32 version of the software is desirable as it runs on a Windows platform and can easily make Excel spreadsheets of bale data.

After the laydown comes the mixers and cleaners. This is not generally a location from which data is taken. These machines just typically do their job and are only paid attention to when something goes wrong as opposed to being actively monitored. The number of mixers and cleaners vary among plants and some also have additional equipment such as dedusters which provide further intensive cleaning of the cotton.

4.2.3 Carding

The carding areas are all generally similar, with the machines operating as they were described in Chapter 2. All but the oldest cards have an interactive display that will at least give basic production data. The newer cards will also send online data to a central monitoring system, located either right on the plant floor, on the manager's office computer, or both. Whether this data is saved depends on the practices of the mill and the capabilities of the particular machine and software.

As mentioned previously, most of the major machine makers such as Rieter or Trützschler have proprietary data systems that come with all of their newer equipment. It is normally possible to extract the data from these programs into another format, such as an Excel spreadsheet. Otherwise, such a task needs to be done manually. Most of these proprietary systems are only designed to save historical data for a short period of time, sometimes only a couple of shifts, so the onus is on the plant management to develop a system to extract and catalogue such data.

In a spinning mill, gathering data into one location is useful, be it for reports or analysis, but it fails to address the issue of the machines communicating with each other directly. The card or draw frame may be able to interpret its own online data to make adjustments, such as autoleveling or shutting itself down; but it will not be able to send such information downstream if the next level of machinery is not running a compatible online system. This leaves the plant with a situation of having islands of automation, which are groups of automated machinery that exist separately and cannot communicate with other groups. This seems to be a common problem in spinning mills, as most of the mills visited for this research housed varying makes of machinery.

Another data-related note from the card room deals with the wire on the card cylinder. When the wire wears down during the carding process, this issue must be addressed with either grinding or replacement so that the card can continue to perform its vitally important function at an acceptable performance level. In the past, such maintenance was done on a regular schedule. This was similar to making routine car repairs based on mileage. With all the data available in the modern age, however, many mills now base such maintenance decisions on nep levels or other performance metrics.

4.2.4 Drawing

Draw frames came equipped with online data systems much the same as those found on cards. The display panel will give online data and allow the operator to set control limits and monitor productivity. Draw frames also have the important autoleveling component, which has to make the necessary roller speed adjustments in order to maintain sliver evenness. At one plant that does not have an autoleveling system built into its draw frame, a scale is placed on the plant floor on which the sliver is manually checked routinely to ensure that it fits within the mill's weight per length specifications. If the weight is either heavy or light, the draw frame's settings will be adjusted manually. These measurements are hand-written on paper, and would need to be manually inputted into a computer to achieve any level of data analysis.

The amount of drawing performed on the cotton varies among plants. It depends on the type of spinning and the emphasis the plant places on the evenness and additional blending achieved through this process. Three-process drawing typically consists of breaker drawing, intermediate drawing, and finisher drawing. On the other end of the spectrum, one plant employs a newer system in which the draw frame is attached directly to the card. This increases efficiency but reduces the opportunity to make leveling adjustments, meaning that more importance must be placed on the quality control leading up to the card room.

4.2.5 Spinning

The mills visited for this research covered all of the most popular methods of spinning cotton yarn. Ring spinning, open-end spinning, and air-jet spinning were all represented during the industry visits. While the methods employed by these machines are quite different, the online monitoring and data systems tend to be similar. Like the other machines in the process, the quality of the online monitoring systems is far superior in the more modern machines. In ring spinning, evenness of the roving is very important, as it has a large impact on the final yarn. One plant manager pointed out that a 1% change in roving can result in a 4% change in the yarn, as the process will amplify variances.

One floor manager's office displays the online information of all of the spinning frames in the facility. The manager is able to see which frames are performing best and which were showing the most ends-down. The display will reveal the number of stops as well as the nature of each stop, be it quality, weight, or doffing. Parameters can also be defined on this system, such as setting what percentage of evenness variation will cause the winder to intervene, making a cut and splice. For instance, the manager can set these parameters to have a higher tolerance when smaller yarn is being spun because these evenness defects are less apparent in thinner yarns.

Again, the varying monitoring systems found on the spinning frames depend on the make and age of the machine. The storage of the data also depends on the particular system, but none are designed specifically for long-term data storage and management. The data would need to be exported into another system to be saved and analyzed on a long-term basis.

The online monitoring systems found during this research share the same basic principles. Most focused on quality control and not on accumulating data for long term analysis. Most systems only save data for a short period of time. This did not seem to pose a problem, as the majority of the spinning plants visited only use online data as a watchdog to signal production issues. The data can be used to make necessary adjustments, or to troubleshoot and diagnose problems.

Even though this data is not created with the intention of data mining, it can still be analyzed and potentially expose unknown trends and relationships. The main issue is getting all of the data coming from the different systems in the yarn creation process to end up in the same location, such as a data warehouse. Even if the data can be accumulated on a single spreadsheet, difficulties could still arise if the collection frequencies vary between operations, the lag time between machines is unknown, or if the time or date is input in differing formats.

4.2.6 Laboratory

The laboratory is generally the domain of the laboratory technician. Some mills have one person designated for the job, while others have multiple lab employees. The laboratory is generally the place where all of the mill's offline data is created. Every mill generally has at least a USTER TESTER to measure their yarn characteristics. Most also have an AFIS tester for cotton fiber characteristics, and the plants that do not possess one will generally send fiber samples to the company headquarters or a larger mill location for off-site testing. Other testing equipment found in labs included USTER TENSORAPIDs, USTER TENSOJETs, and USTER CLASSIMATs. Perhaps the most important device in the typical spinning mill laboratory is the scale. This classic instrument is used to measure the weights of lengths of yarn or sliver to determine the mass evenness level and whether they fit into the plants specified parameters.

The frequency with which the various tests are performed varied slightly but tends to be mostly similar throughout the industry. The weights for cotton coming off of the cards, draw frames, and spin frames are generally measured daily. The tensile testing of yarn tends to be a weekly test, meaning that by the end of the week, each section of each frame will have been tested once. Samples are tested each day, but they are coming from different locations on the frame. The most popular frequency for AFIS testing is monthly, although some mills conduct weekly AFIS testing.

The output of offline data created by this testing depends on the mill and on the equipment. Some plants put all of their offline data into one spreadsheet for future reference. Others are satisfied just collecting and filing the printouts that are created by some of the testing machines.

In addition to routine monitoring for any problems in the operation of the plant, the lab generally uses its data to accomplish other important tasks as well. One responsibility of the laboratory technician is normally to create weekly or monthly reports. There are generally three kinds of reports created, as seen in Table 9.

Table 9 - Reports Generated by a Spinning Mill Laboratory

1. In-house reports - to summarize the mill's performance for supervisors
2. Corporate reports - for management to keep track of the mill
3. Customer reports - to ensure customers that their product has desired quality

These reports are normally sent out as e-mail attachments, although some are sent in the mail the old fashioned way. In-house reports can be circulated to the different supervisors of a particular location, enlightening the supervisors about the performance of their realm. Corporate reports are sent to the company's headquarters so that the main decision-makers can be informed about how their facility is operating. Customer reports are sent to the customer to ensure them that their yarn has met any required specifications. This can often include a Certificate of Quality or Certificate of Analysis to verify the successful testing.

One insight discovered in this research is that yarn customers typically do not specify desired measurable characteristics for their yarn. They generally just want the yarn to run well in their knitting or weaving process. Some of the spinning plants will knit samples on the premises as a means of testing to ensure that the yarn is capable of knitting well before it is sent to a textile knitting customer.

4.3 Data Procurement

Aside from the task of determining the as-is state of data management in yarn spinning operations, the industry visits undertaken served the second purpose of determining how to acquire a suitable data set on which to perform analysis.

As the different plants were visited, the potential of each for providing data was observed and noted. There were two means of data acquisition to consider, so the questions found in Table 10 were addressed before concluding each visit.

Table 10 - Questions Pertinent to Data Procurement

1. Will the plant allow access to archived data?
2. If not, is there potential to manually extract a sufficient data set?

All things being equal, the first means of procuring data is superior because all of the data could be obtained at once. The benefit of manual extraction of data is that it enables the data set to be customized to this research. However, this second option would require a lengthy process of data collection, providing less time for analysis and potentially result in a smaller sample of data. Also of detriment to a manual collection plan is the potential for unforeseen hazards, which could cause an abrupt end or skip in the data collection timeline.

At the conclusion of the industry visits, a couple of plants were kept in contact in case a manual data gathering process was deemed the appropriate approach. However, one open-end spinning mill that practices some advanced data management processes was generous enough to provide a workable data set for this research. This allowed for sufficient data to be emailed all at once, ready for analysis. So a manual data accumulation plan would not have to be enacted.

As mentioned earlier, the USDA was also able to provide data from its own scientific research. These two data sets are to be further described and then analyzed in order to meet Research Goal 2 in the following chapters. This analysis will be married to the findings of the current state and capabilities of the typical spinning mill, as discovered in this chapter, to forge recommendations for data management in cotton yarn spinning.

Chapter 5 - Data Understanding

5.0 Introduction

The benefit of having volumes of data can be amplified if the proper information within is discovered. This is an important proposition considering the investments that these piles of data represent. Most of the data collected at a plant is typically either utilized to inform the supervisor about whether a particular machine is running properly, or it will indicate to a customer whether a yarn is likely to knit well.

Data mining is a tool that can help extract further information out of data. This is the information that can become knowledge that will lead to better decision making and strategies in the future. Table 11 lists some data mining requisites found in reviewing literature, which were considered throughout this process. Among the most vital are fully understanding the problem and providing a simple solution. In this research, definitive uncomplicated findings are most likely to produce strong practical recommendations.

Before analysis is carried out, it is vital to fully understand the two data sets being investigated. They will be referred to as Data Set 1 and Data Set 2, and the contents of these 2 data sets will drive the statistical analysis and data mining that follows.

Table 11 - The 10 Commandments of Data Mining (adapted from Pyle, 2003)

1. Select clearly defined problems that will yield tangible benefits.
2. Specify the required solution.
3. Define how the solution delivered is going to be used.
4. Understand as much as possible about the problem and the data set.
5. Let the problem drive the modeling.
6. Stipulate assumptions.
7. Refine the model iteratively.
8. Make the model as simple as possible - but no simpler.
9. Define instability in the model.
10. Define uncertainty in the model.

5.1 Data Set 1 - USDA Research Results

Data Set 1 was obtained from the Agricultural Research Service of the USDA. This data set was delivered to the researcher in Excel format which will work with the SigmaXL software, but can also easily be transferred in Jmp format for use with the SAS Jmp program. The data originates from a research project performed as the Southern Regional Research Center in New Orleans, which was visited earlier in this research. The project was called the Fiber Quality Evaluation Laboratory (FQEL) and took place between 2004 and 2008, during which time the results were used for several publications and conference proceedings (Campbell, Delhom, & Thibodeaux, 2005; Delhom, Cui, Campbell, & Thibodeaux, 2005; Delhom, Thibodeaux, & Rodgers, 2007).

The idea behind this research was to examine changes in fiber quality throughout processing as well as to relate fiber quality to textile end quality. This was done by producing both ring-spun and open-end single-cotton yarns under standardized conditions in a laboratory and taking fiber samples from a variety of processing points to deduce the changes undergone to the fiber throughout the process. HVI measurements were taken at the bale and AFIS measurements were taken on samples obtained from the processing points found in Table 12.

Once the fibers became either ring-spun or open-end yarn, they were subjected to tensile and evenness testing using the USTER TESTER and the USTER TENSORAPID. The yarn testing data, AFIS data, and HVI data are presented in 3 different data tables, which constitute Data Set 1. It is possible to see the changes in a fiber as it goes through each of these tests because the data is linked by a Fiber ID number designation. This means a fiber can be followed through the process.

Table 12 - Processing Points for AFIS testing in FQEL

1. From the bale	**5. From the card sliver**
2. After opening/cleaning	**6. From the breaker drawing sliver**
3. At the condenser	**7. From the finishing drawing sliver**
4. At the card chute	**8. From the roving**

The value of this research is that it was done in a scientific fashion, with settings such as target sliver weights, twist, and production speeds held constant. This is in contrast to the real-world mill environment, where such settings may be changed often during processing.

While the original purpose of the FQEL research project was to develop fiber quality measures for breeding purposes, the findings should also shed some light on the data available in a cotton mill as it pertains to this research. It should be noted that these tests were all done on single-cotton yarns, and not on blends. This is obviously in contrast to the yarns found in production plants. The content of this data set is summarized in Table 13.

Table 13 - Data Set 1 Summary

Table Name	# of Inputs	Data Make-Up
Fiber Data	139	Fiber ID, Length, Uniformity, Strength, Micronaire, Trash
Processing Data	518	Fiber ID, Process Stage, Neps, Length, Maturity, Trash
Yarn Data	145	Fiber ID, Yarn Type, Uniformity, Trash, Thick/Thin Places, Neps, Evenness, Elongation, Count

5.2 Data Set 2 - Open-End Plant Data

Data Set 2 is the data set obtained from the open-end spinning plant. The data was delivered to the researcher in Jmp files. These files can be used in the SAS Jmp program, but can be easily transferred to Excel files for use with the SigmaXL software. This data covers about a two month span of operation from March 2008 to May 2008, and is broken up into 9 data tables. The content of this data set is summarized in Table 14.

Table 14 - Data Set 2 Summary

Table Name	# of Inputs	Data Make-Up
HVI from bales	13270 (148 laydowns)	Laydown ID, Length, Uniformity, Strength, Micronaire, Trash
Card sliver weight	272	Project #, Weight (grains per yard)
Card sliver evenness	299	Project #, Evenness variation
Finisher sliver weight	477	Project #, Weight (grains per yard)
Finisher sliver evenness	183	Project #, Evenness variation
Yarn size	194	Project #, Cotton count, Size variation
Yarn tensile properties	207	Project #, Force, Elongation, Work
Yarn evenness	1655	Project #, Evenness variation, Thin/Thick places, Neps
Spinning ends down	52 (1 per day)	Spin Date, Ends-down

The common measure which enables these tables to be linked together is a project number. The one exception to this is the spinning ends-down, which are given by day. The mill that provided the data has used several methods to determine the lag time between the bale laydown and the spinning frame. These methods have independently arrived at an estimation of about a 2 day lag time for this particular plant. Understanding lag times is very important in this type of analysis, which tries to follow data through the process.

Both Data Set 1 and Data Set 2 contain moderate amounts of data; not necessarily the amount coveted for the deepest data mining, but certainly enough to reveal some trends and produce interesting results. Not all of the measurements recorded in these data sets will necessarily be utilized in this study. That is why only summaries are given at this time. Explanations of the measurements and their units will be further clarified during the analysis phase in Chapter 6.

Now that the data understanding and business understanding are accomplished, the data can be prepared and analyzed. Each analysis will use the data in order to answer a question. The totality of these answers will be the basis upon the recommendations and conclusions of this research are formed. The questions posed were determined using the modeling techniques and the data sets available to this study.

5.3 Data Preparation

As was discussed earlier, data preparation is of utmost importance. It is often the most time consuming of the data mining processes. While computers are capable of performing tremendously complicated functions to expose relationships in data, they are not able to interpret data that has not been properly prepared, so data must be cleaned manually in order to suit each data mining function being performed.

In this study, the data preparation phase is decidedly less intensive then it was in some of the previous works described in Chapter 2. Since both data sets were acquired already complete and in a single computer format, they did not need to be tediously hand-entered or reformatted into a common spreadsheet or database. All data was also fairly clean in terms of not having missing data fields. The few instances of missing data that did exist were cleaned using SigmaXL, which can automatically delete rows containing blank fields. This is useful as long as missing data is a rarity, so that only a very low percentage of rows are lost.

The two data sets came in differing file formats. Data Set 1 contains 3 data tables on a single Excel page, and Data Set 2 contains 9 data tables on 9 separate Jmp files. Conveniently, these two formats are very compatible and can be easily cut-and-pasted into one another's program. The only additional preparation for such a transfer consists of separating the column headings from the data, and telling Excel which time/date format to recognize. As the data goes through the modeling phase of the data mining process, the custom preparation of the data for each modeling method is described. For instance, data from different tables may have to be merged into a single table, a process that may take several steps because the tables all have different numbers of inputs.

Chapter 6 - Modeling & Results

6.1 AFIS Changes through Processing

One of tables from Data Set 1 is the Processing Data table. This table contains all of the basic AFIS measurements taken from samples that span the stages of processing. The USDA Southern Regional Research Center tested samples from 8 different process points over the course of a number of production runs. This data will be analyzed as part of Research Goal 2.

In order to investigate what impact these process steps have on AFIS properties, SigmaXL was used to compute basic descriptive statistical information, organized by process stage, for each of the following measurements: Nep count, trash count, maturity ratio, fineness, average fiber length, and short fiber content (SFC). These measurements are detailed in Table 15 and their full descriptive statistics charts created can be found in Appendix A.

From the descriptive statistics, it was decided that the statistical mean value of each measurement could be used to track the changes in AFIS properties through the yarn spinning process. This shows the average changes occurring to each fiber as it goes through the different processing steps. These statistical means were made into charts in order to visualize and understand the changes to the cotton on the fiber level as it progresses from bale to yarn.

Table 15 - AFIS Measurements and Descriptions (adapted from Uster.com, 2009)

Measurement	**Description**
Nep Count	**# of neps per gram**
Trash Count	**# of trash particles per gram**
Maturity Ratio	**Relative cellulose content of fiber cross-section**
Fineness	**Fiber fineness in millitex**
Mean Fiber Length	**Mean fiber length in inches**
Short Fiber Content (Number)	**Fibers shorter than 1/2 inch by number in %**
Short Fiber Content (Weight)	**Fibers shorter than 1/2 inch by weight in %**

The first property being investigated is the nep count per gram of fiber. A nep is a tangled clump of fiber that can effect spinning efficiency and the appearance of the end cotton product. Figure 11 displays the amount of neps measured in the fiber samples taken from each process phase. The figure shows an increase in neps through each blow room process. Once the fiber has been carded, the nep count is greatly reduced and remains low through roving. This data shows the importance of carding in terms of nep reduction. It is the only process that eliminates neps, and in this data, it reduced nep count by approximately

75%. Faulty carding will lead to weaker yarns with a diminished appearance due to a prevalence of neps.

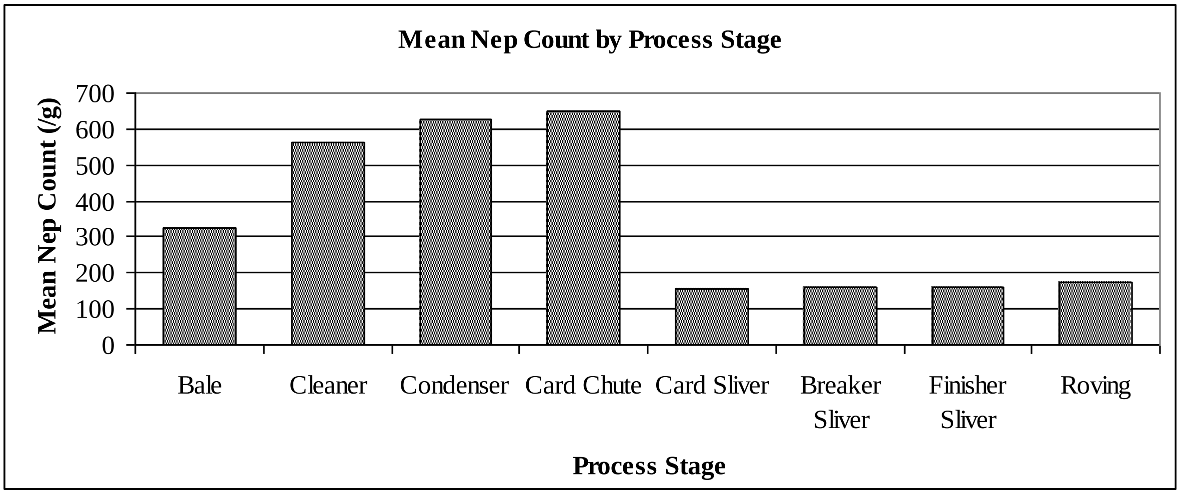


Figure 11 - Mean Nep Count by Process Stage

Trash count is similar to nep count in that minimization is the goal. Trash in fibers will obviously get in the way of spinning processes and diminish yarn properties. Figure 12 shows the AFIS measurement of trash count per gram for each of the 8 process points from which samples were taken. Unlike neps, trash content measurements are highest at the bale, before processing begins. This is because while neps can be created during processing, no new trash is introduced to the fiber once it enters the mill. The figure shows that two process steps reduce the trash count. Firstly, the cleaning step removes approximately 25% of all trash content. The card eliminates approximately 80% of the remaining trash content. This is another measurement that highlights the importance of carding amongst the processing stages of cotton yarn production.

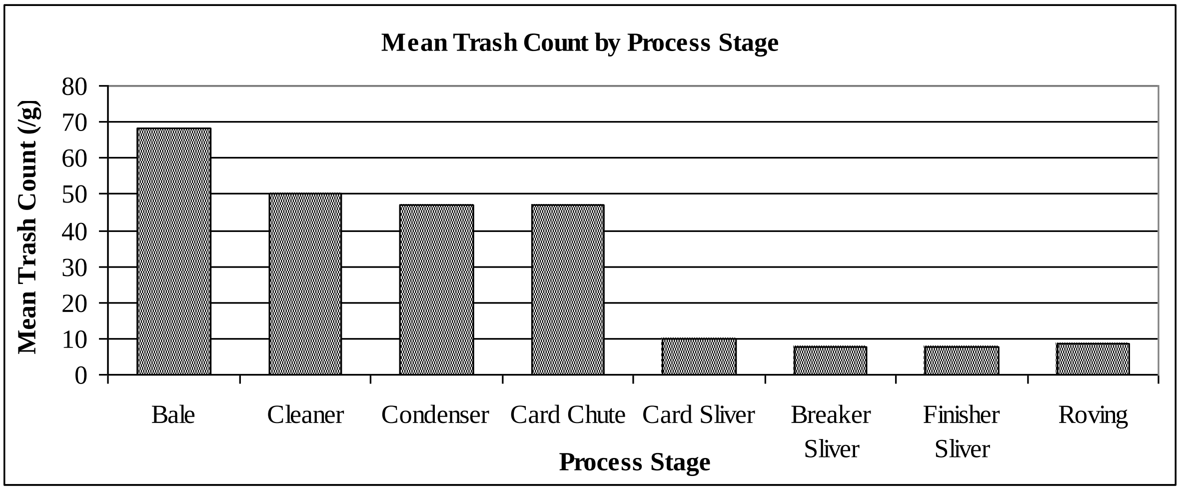


Figure 12 - Mean Trash Count by Process Stage

In contrast to the measurements of trash count and nep count, the maturity ratio of the cotton, as measured by AFIS, begins to increase at the card. As seen in Figure 13, the average maturity levels of the fiber samples slightly decrease through the blowroom stages. Beginning with carding, however, maturity ratio increases with each additional process stage that the fiber travels through. Maturity ratio is a measure of the amount of cellulose in the cross section of a cotton fiber, so it basically measures the thickness of the fiber wall. So, the increase here is likely due to the alignment of fibers that takes place in the last several stages. As the fiber is stretched out to be oriented and aligned, the lumen, which is the empty tube running through the middle of the cotton fiber, will likely decrease in cross-sectional size due to the collapsing of the fiber walls around it. This means more fiber in the cross section, thus a higher maturity ratio. The trend in maturity mirrors the trend in neps. Neps often form from immature fibers, so the removal of neps leads to a higher overall maturity ratio amongst the fibers.

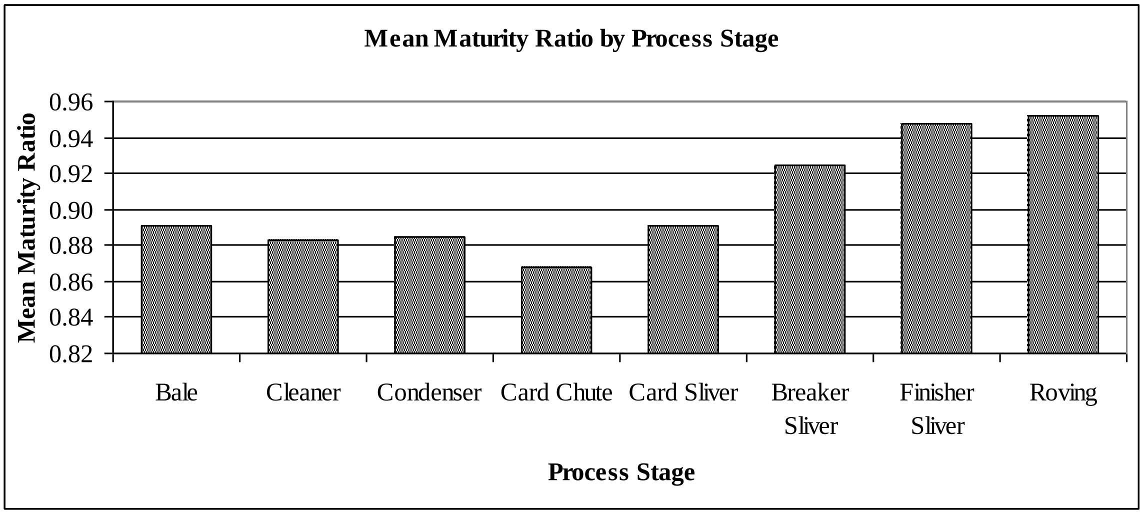


Figure 13 - Mean Maturity Ratio by Process Stage

Relating to maturity ratio is fiber fineness. Fineness and maturity are the two factors that go into determining the micronaire of cotton fibers. The reason the two measures are combined to give one micronaire value is that they tend to correlate, as high maturity means more fiber in the cross section and a high fineness means higher weight density, which tends to be greater when more cellulose is present.

As seen in Figure 14, this correlation holds true in the USDA's FQEL research project testing. The fineness of the cotton fiber increases at the last 4 process points, as the fibers are squeezed and stretched. The fineness does show more of a significant decline in the early process stages than does maturity. This can likely be attributed to the densest fibers being removed in the processes which are designed to remove impurities.

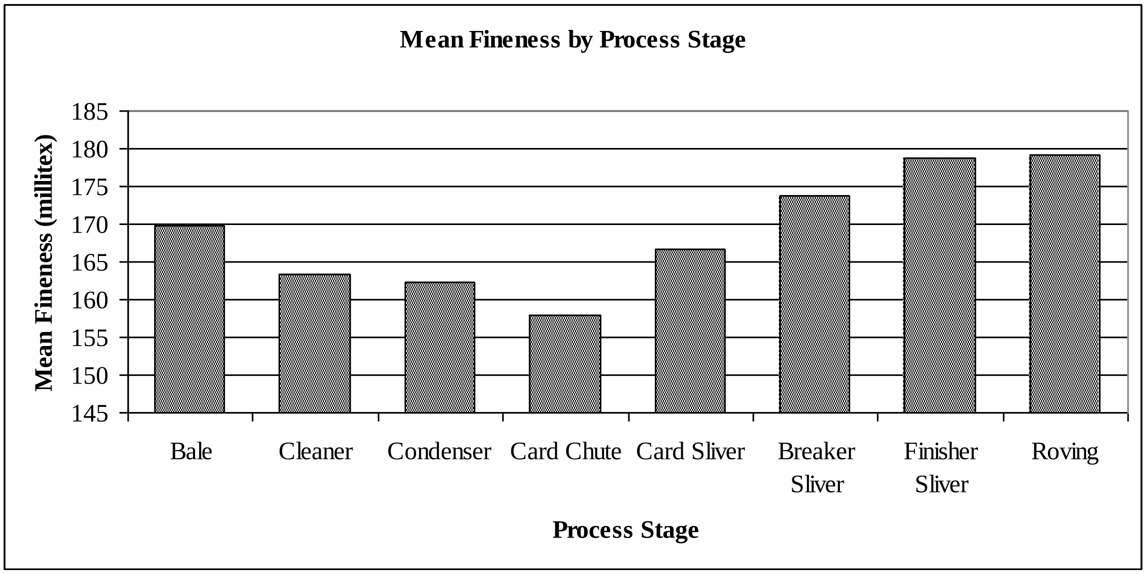


Figure 14 - Mean Fineness by Process Stage

Another measure in the AFIS repertoire is mean fiber length. This is simply the average length in inches of the fibers being tested. As shown in Figure 15, this measure shows a similar pattern to the previous two measurements, beginning with a slight decline and then increasing through the end of the production process.

The difference here is that the increase in fiber length begins one stage after the increases for fineness and maturity. The card, which began the increase for the previous 2 measurements, actually produces an additional drop for mean fiber length. This may be due to the utility of the card to separate tufts, which frees more of the shorter fibers for testing, or to more short fibers being created by the breaking of longer fibers in the carding process. The slight increases over the last 3 stages are due to the straightening of the fibers. When the end of a fiber is hooked, the optical sensor in an AFIS tester sees it as a shorter fiber, as it is only registering the linear length as the fiber passes by. When these hooks are straightened out during the drawing stages, the AFIS tester sees it as a longer fiber, when it is actually the same length.

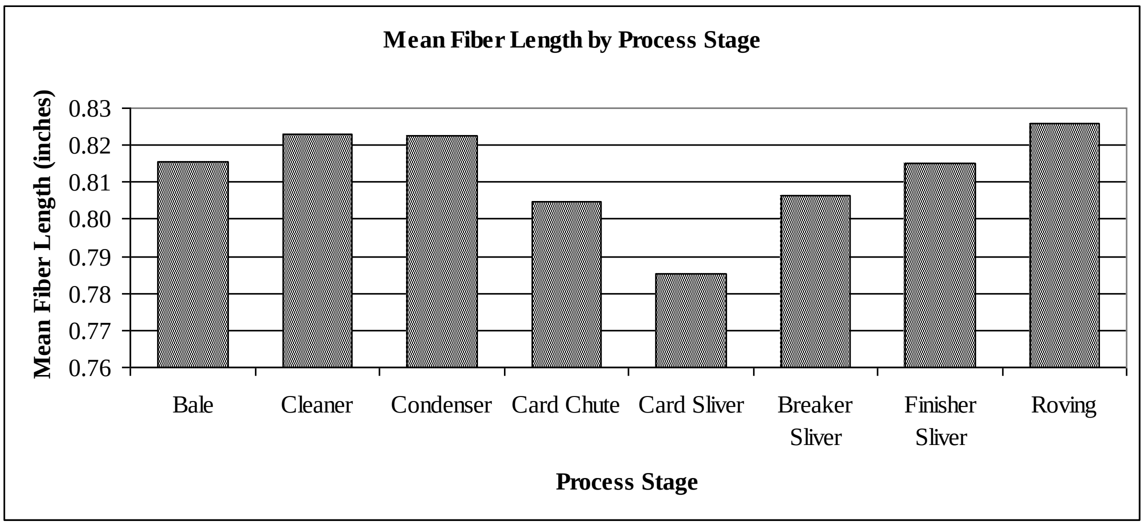


Figure 15 - Mean Fiber Length by Process Stage

Perhaps more important than mean fiber length is short fiber content (SFC). This measurement is the percentage of the cotton fibers that are classified as short fibers. As was discussed at the 2010 Beltwide Cotton Conferences, this is a controversial measurement. For one thing, not everybody agrees about the definition of a short fiber, especially when dealing internationally. Another issue is that there are multiple ways in which this property is measured.

There is both a weight-biased measurement and a number-biased measurement for short fiber content available on an USTER AFIS testing machine. The weight-biased measurement gives SFC as a percentage of total weight taken up by short fibers, whereas the number-biased measurement gives SFC as a percentage of fibers classified as short. For comparison sake, the results of both SFC measurements were placed beside each other in Figure 16.

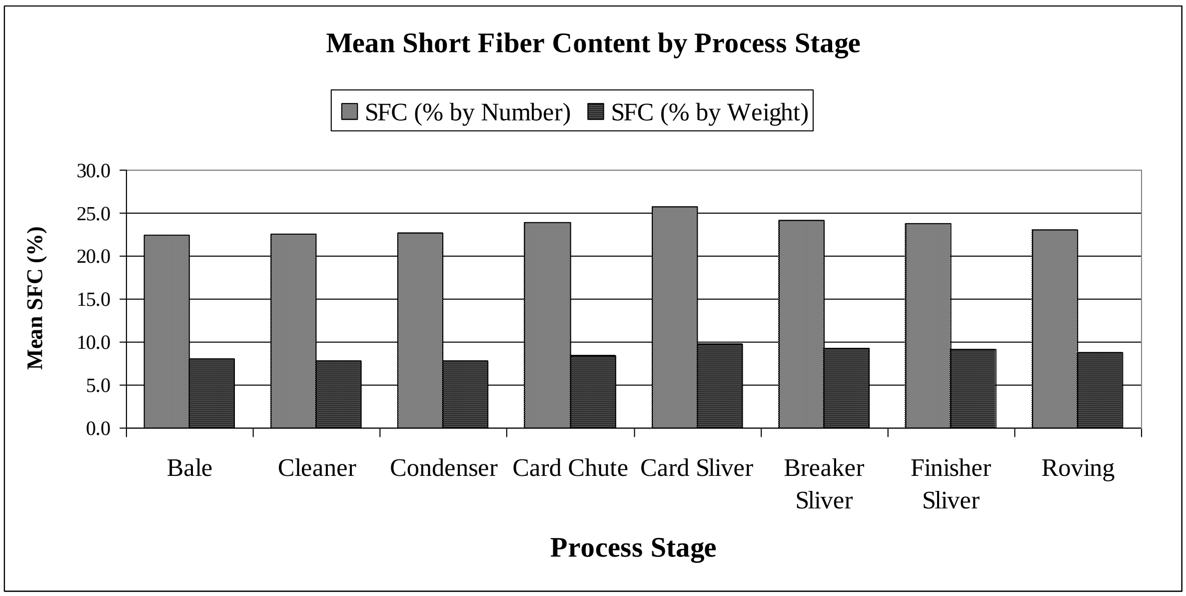


Figure 16 - Mean Short Fiber Content by Process Stage

This figure shows two trends. Firstly, the percentages computed based on number differ greatly from those based on weight. Secondly, both forms of SFC measures show the same slight trends through the spinning process. A closer look does reveal some differences, though. Looking at the increases in SFC between the condenser and the card sliver, the increase is 3.1 in SFC by number and 1.93 in SFC by weight. However, the proportional increase is significantly greater in the SFC by weight (24.7% proportional increase) because the average SFC is smaller than in SFC by number (13.8% proportional increase). This is something to be considered when comparing the two methods of measuring.

Although the changes are slight, SFC is shown to rise until drawing begins, at which point it decreases. The early increase can likely be attributed to the short fibers being created by the aggressive actions of the cleaning and carding processes. The eventual decrease could be caused by short fibers being lost if they do not span the gap in the drawing process and hooked fibers being straightened enough that they no longer are seen as short fibers by the AFIS tester. Not surprisingly, the trend of the short fiber content is something of an inverse of the average fiber length chart. The more short fibers are present, the lower the average fiber length is going to be.

In tracking AFIS measurements through the yarn spinning process, it was interesting to see the trends in the measurements as the fibers were followed from bale to roving. The data shows the impact that the different processes can have on AFIS properties. If these trends were to prove to be consistent over time within a spinning plant, the plant could basically just make their measurements at one or two of the points and be able to predict the results for all other points. This is an example of information being used to increase efficiency.

With the data available in Data Set 1, predictive equations were made for each measurement through the processing stages. This was done to decide if measurements could be taken at fewer of the stages, and these measurements could then be used to predict what the other measurements would have been. It was determined that for most of the fiber properties, an initial measurement at the cleaner could accurately predict the measurements for the condenser and the card chute. A second measurement of the card sliver could then be used to predict breaker sliver, finishing sliver and roving.

This was done using the Jmp software. For each measurement, a predictive equation was made to link each of the stages. The equations were chosen based on the lowest adjusted R squared value, which measures the power of the equation to make accurate predictions, making adjustments based on the number of terms used in the equation. The adjusted R squared values of these predictive equations can be found in Appendix B.

6.2 Mean and Median for Laydown HVI Properties

Upon becoming more familiar with Data Set 2, which was provided by the open-end spinning factory, it was observed that the HVI data from the laydowns was given for each individual bale. This provides an opportunity for closer study.

As was noted when looking at past studies and literature on this subject, most of the research that has been done that aimed to link HVI properties to yarn properties did so using just a single cotton yarn (Guha, Chattopadhyay, & Jayadeva, 2001; Jayadeva et al., 2003). Indeed, Data Set 1 contains HVI data, but the USDA only used single cotton yarn in the FQEL research project. The data found in Data Set 2 can perhaps give more insight into the impact of mixing different fiber varieties, which each have unique characteristics, to create a yarn of blended cotton. The HVI measurements from Data Set 2 utilized in this research are described in Table 16.

Table 16 - HVI Measurements and Descriptions (adapted from M. E. Anderson, 2006; Daley, 2008)

Measurement	**Description**
Micronaire	**Measure of fineness and maturity of the fiber**
Fiber Length	**The upper half mean length of fiber**
Fiber Strength	**Force in grams required to break a 1 tex bundle of the sample**
Plus B	**How much yellow color is in the sample**
Rd	**How light or dark the sample is in percent of brightness**
Trash Area	**Percent of surface area covered by trash**
Uniformity	**Percent ratio of upper half mean length to overall mean length**

Bale laydowns are typically set up using EFS systems to maintain uniformity from laydown to laydown. The computer will try to mix the bales in inventory so that some key HVI measurements are consistent in each laydown. However, since the interactions between these different cottons remains somewhat mysterious, it is not necessarily certain that the mix will behave like the average of its constituents. It is possible that the blend will only be as strong as its weakest bale.

For this research, each laydown was observed in two ways. One looked at the mean of each the HVI bale measurements, the second used median values in place of the mean. This provides a look at how different methods of balancing laydowns could perhaps yield varying degrees of yarn consistency.

Data Set 2 contains data for 13,270 bales that were used in 148 bale laydowns over a period of about 2 months. This means about 90 bales went into each laydown. Using the HVI measurements from these bales, the mean and median values were found for each laydown, numbered 1 through 148. To visualize the potential differences within this data, plots were created for 7 different HVI measured characteristics. In 4 of these plots, the median and mean produced very similar results. In these cases, either method of balancing yarn blends would likely yield similar product consistency. These 4 plots can be found in Appendix C.

Of the plots that displayed deviances between mean and median results, perhaps the greatest difference was found in micronaire values. This is a measure of the maturity and fineness of the cotton fibers. Higher fineness and lower maturity levels will give a lower micronaire. Low micronaire cotton is desirable for spinning stronger yarns, due to having more fibers in each yarn cross-section. However, fine fibers break more easily, so slower processing speeds are necessary (Cotton Incorporated, 2001).

As displayed in Figure 17, the measured median micronaire value was consistently greater than the measured mean micronaire value. In this case, it is clear that different bales laydown mixes may be used if it was sought to balance the median micronaire values. It is interesting that a majority of the laydowns had a median of exactly 4.4 micronaire, yet the mean was more consistent from laydown to laydown. The overall spread here is quite small, however, going from about 4.3 to 4.5 micronaire.

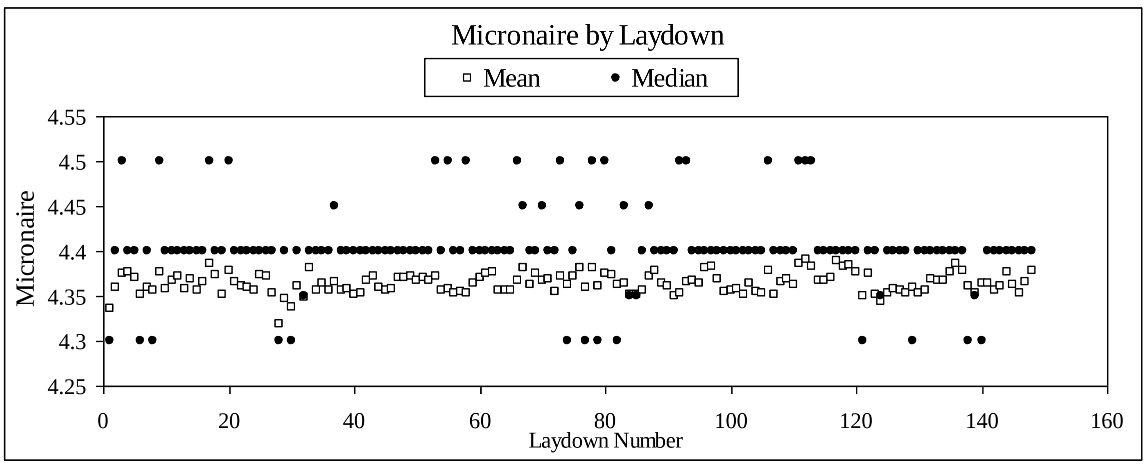


Figure 17 - Micronaire by Laydown

The plot of fiber length by laydown, as seen in Figure 18, also shows a propensity for median values to be greater than their mean counterparts. This sort of trend is the result of an imbalance in the values above and below the median. The shorter fibers are below the median to a greater degree than the longer fibers exceed the median. This causes the mean to be dragged below the median value. By far the most popular mean is 1.09 inches, which is greater than almost every laydown mean.

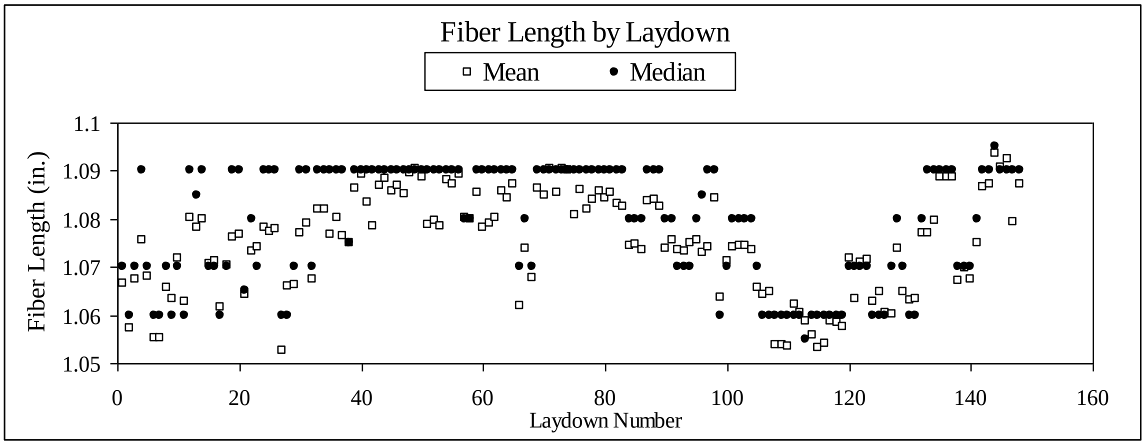


Figure 18 - Fiber Length by Laydown

Another interesting HVI measurement is the Plus B measurement. This is a color measure of the cotton fiber which tends to relate to the other properties of the fiber. This is perhaps because the difference in color is caused by other factors which directly impact the behavioral characteristics of the cotton. The plot of Plus B measurements is shown in Figure 19. The median value of this measurement often fluctuates wildly between laydowns, while the mean maintains consistency.

These comparisons shed light on the fact that because fiber interactions in cotton blends are complex and difficult to approximate, simply creating laydowns based on either the statistical means or statistical medians of HVI properties may not be the optimal method. While no experiments were done to test the results of different mixing protocols, this data does show that the laydowns would likely have been different had differing philosophies been used. It would be interesting to see what method of balancing HVI measurements among laydowns would lead to the greatest yarn uniformity. This method could seek consistent means, medians, or even other values such as variations or maximums. This would depend on which has the greatest effect on blended yarn properties, however that is unknown at this time.

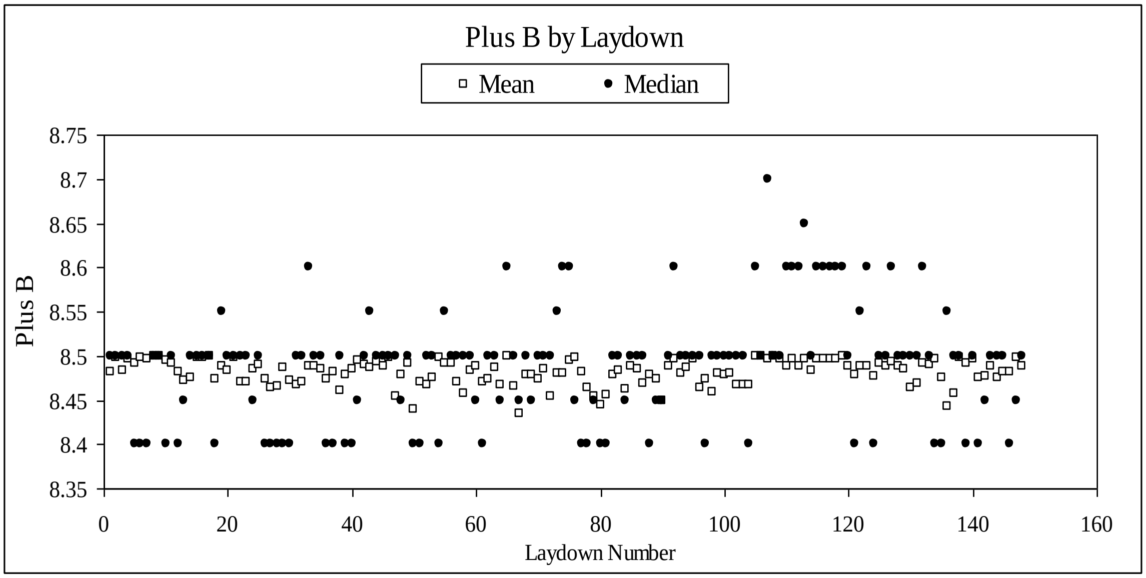


Figure 19 - Plus B by Laydown

6.3 HVI Measurement Relationships

To gain more insight into HVI measured properties and their relationships to one another, further data mining was done using the Jmp software. The statistical mean and median values for each of the measurements for each laydown were analyzed with pairwise correlations in order to discover relationship between the measurables. Knowledge of such relationships could be used to eliminate redundant data and perhaps lower testing demand.

If two properties are directly correlated, only one would need to be kept track of because it could be used to predict the other.

A pairwise correlations chart examines a set of continuous data variables and determines which are strongly correlated, positively or negatively. The correlation scale goes from -1 to +1, with -1 being a perfect negative correlation and +1 representing a perfect positive correlation. This study looked at possible correlations between both statistical mean HVI values and statistical median HVI values taken from the cotton bales from each laydown. These are the same measurements described in Table 16. The significance probability value on these charts refer to the likelihood that the result is incorrect. So the lower this value is, the more confidence can be put in the result.

If any two factors show a high correlation, it demonstrates a strong relationship. This is important because an efficient data management plan would not need to include both such measures. Any results from this technique, such as the discovery of a strong correlation, can be further explored with additional techniques.

The result of the pairwise correlation testing using the mean values is shown in Figure 20. Of the correlations with a statistically significant probability of being correct, trash area and length are highly positively correlated. Trash area is also involved in the highest negative correlation with Plus B. Uniformity is highly correlated with both length and strength, which in turn, are also positively correlated with each other.

Pairwise Correlations

Variable	by Variable	Correlation	Signif Prob
Mean(TrashArea)	Mean(Length)	0.8406	<.0001*
Mean(Uniformity)	Mean(Length)	0.7587	<.0001*
Mean(Uniformity)	Mean(Strength)	0.7318	<.0001*
Mean(Strength)	Mean(Length)	0.6112	<.0001*
Mean(Uniformity)	Mean(TrashArea)	0.4996	0.0002*
Mean(TrashArea)	Mean(Strength)	0.2867	0.0435*
Mean(Rd)	Mean(Mic)	0.2724	0.0557
Mean(PlusB)	Mean(Mic)	0.1225	0.3966
Mean(Rd)	Mean(PlusB)	0.1219	0.3989
Mean(TrashArea)	Mean(Mic)	0.0865	0.5504
Mean(Length)	Mean(Mic)	-0.0587	0.6854
Mean(PlusB)	Mean(Strength)	-0.0998	0.4905
Mean(Uniformity)	Mean(Mic)	-0.2091	0.1451
Mean(TrashArea)	Mean(Rd)	-0.2576	0.0709
Mean(Strength)	Mean(Mic)	-0.2855	0.0444*
Mean(Rd)	Mean(Strength)	-0.3211	0.0230*
Mean(Rd)	Mean(Length)	-0.3359	0.0171*
Mean(Uniformity)	Mean(Rd)	-0.3953	0.0045*
Mean(Uniformity)	Mean(PlusB)	-0.4187	0.0025*
Mean(PlusB)	Mean(Length)	-0.5457	<.0001*
Mean(TrashArea)	Mean(PlusB)	-0.5868	<.0001*

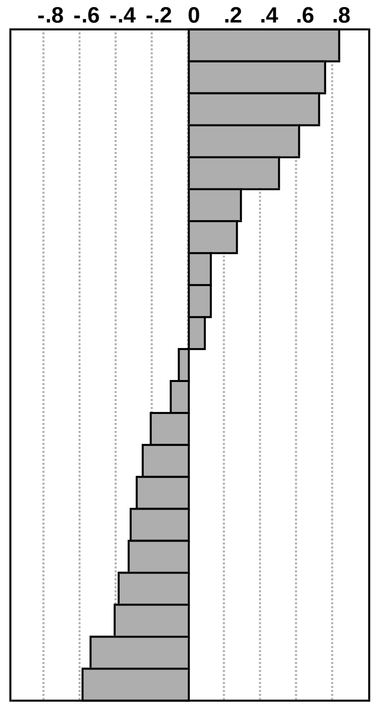


Figure 20 - Pairwise Correlations for Bale Laydown Mean HVI Values

The result of the pairwise correlation testing using the median values is shown in Figure 21. The highest statistically significant positive correlation is between trash area and length, which echoes the statistical mean value correlations. However, the correlation is not quite as strong in this case. The same is true for uniformity by strength and uniformity by length. An interesting difference in this figure is seen in the negative correlations. The median values reveal the strongest negative statistically significant correlation between uniformity and both of the measures of color, Rd and PlusB. These color measures also both influence length negatively.

Pairwise Correlations			
Variable	**by Variable**	**Correlation**	**Signif Prob**
Median(TrashArea)	Median(Length)	0.7538	<.0001*
Median(Uniformity)	Median(Strength)	0.6431	<.0001*
Median(Uniformity)	Median(Length)	0.5749	<.0001*
Median(Rd)	Median(PlusB)	0.5390	<.0001*
Median(Strength)	Median(Length)	0.3674	0.0087*
Median(Uniformity)	Median(TrashArea)	0.2844	0.0453*
Median(Rd)	Median(Mic)	0.2386	0.0952
Median(PlusB)	Median(Mic)	0.1880	0.1910
Median(TrashArea)	Median(Strength)	0.1284	0.3742
Median(Strength)	Median(Mic)	-0.0533	0.7132
Median(PlusB)	Median(Strength)	-0.0811	0.5755
Median(TrashArea)	Median(Mic)	-0.0875	0.5456
Median(Uniformity)	Median(Mic)	-0.1132	0.4339
Median(Length)	Median(Mic)	-0.1681	0.2433
Median(Rd)	Median(Strength)	-0.2299	0.1082
Median(TrashArea)	Median(PlusB)	-0.2649	0.0630
Median(TrashArea)	Median(Rd)	-0.2710	0.0570
Median(Rd)	Median(Length)	-0.3161	0.0253*
Median(PlusB)	Median(Length)	-0.4004	0.0040*
Median(Uniformity)	Median(PlusB)	-0.4369	0.0015*
Median(Uniformity)	Median(Rd)	-0.4735	0.0005*

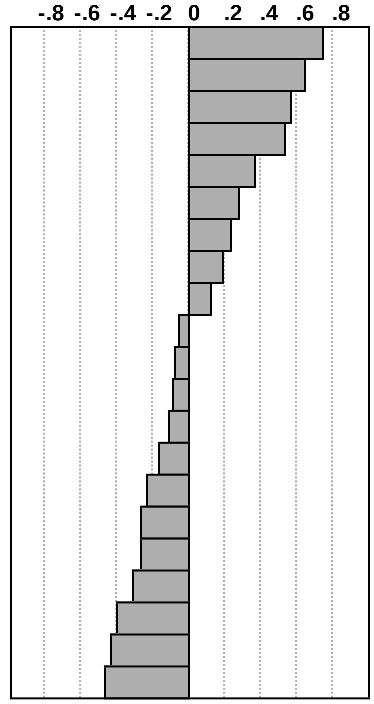


Figure 21 - Pairwise Correlations for Bale Laydown Median HVI Values

This technique revealed some relationships within HVI data measurements and also showed that the same data can be interpreted differently depending on how it is prepared. In this case, the bale laydown properties were calculated by both the statistical mean of each bale and the statistical median of each bale. These are only two of many ways the blend properties can be calculated, and perhaps two of the most similar, and yet they reveal differences in the overall data.

6.4 Spinning End-Breaks

Because the open-end spinning plant provided ends-down statistics along with the HVI data, it is possible to use data mining techniques to explore the prediction capabilities the different HVI measurements on spinning end-breaks. Limiting end-breaks is of paramount importance to a cotton spinning company as productivity is lost in the time it takes to repair the yarns. Ends-down are also a sign of weak yarn that may not perform well in knitting or weaving.

Having looked into this relationship before, the plant was able to provide an educated estimate of a two day lag time between the time a laydown took place and the time those fibers were being spun into yarn. This lag time was verified using multiple methods of estimation. Having a known lag time enabled the laydown HVI and ends-down tables to be joined into a single table based on date. This prepares the data for data mining.

Neural networks, an advanced data mining technique, can be constructed through the Jmp software to predict whether a certain laydown will result in an above-average or below-average amount of end-breaks. The end-break totals for 50 days of open-end yarn spinning operation are included in Data Set 2.

The end-breaks for each day is given in ED/MRh units. This is the number of ends-down for each 1000 hours of rotor operation on that particular day. Of those 50 days, the highest 25 days of end-breaks, or top 50th percentile, were labeled as above-average end-break days. The lower 25 were deemed below-average. These labels simply compare the data within Data Set 2 and do not reflect the long term average of ends-down at this particular facility. In this data, the below-average category is more desirable for the company, as it means less end-breaks.

Neural networks were created using Jmp for both the mean laydown HVI measurements and the median laydown HVI measurements. A diagram of a neural network is shown in Figure 22. Because of the small sample size of 50, the basic Jmp neural network settings were used, using a K-fold cross validation with a K of 5, and with 3 hidden nodes (H1, H2, and H3 in the figure). This means overlapping portions of the data were used to act like a larger data set where some data can be withheld in order to judge the fit of the model. The results for both models showed that, given this HVI data, a neural network can construct a model to predict whether it will be an above-average or below-average day for end breaks. The model using statistical means performed slightly better, obtaining an R squared value of 0.993, while the median model had an R squared of 0.988.

Categorizing the end breaks into just 2 categories simplifies things and creates more accurate models, as there are only two possible predictions. When similar neural networks were made to predict the actual end-break numbers, which was given in units of ends-down per 1000 rotor hours, rather than just above-average or below-average, the power of the models was lower. Figure 23 shows the actual by predicted plots. The plot on the left hand side (R squared = 0.901) shows the predictions made by a neural network based on mean bale properties and the plot on the right hand side (R squared = 0.849) is for median bale properties. It can be noted from these plots that the models have about the same power for predicting actual amounts for above-average amounts and below-average amounts, as seen by how close the dots are to the line as the amount increases.

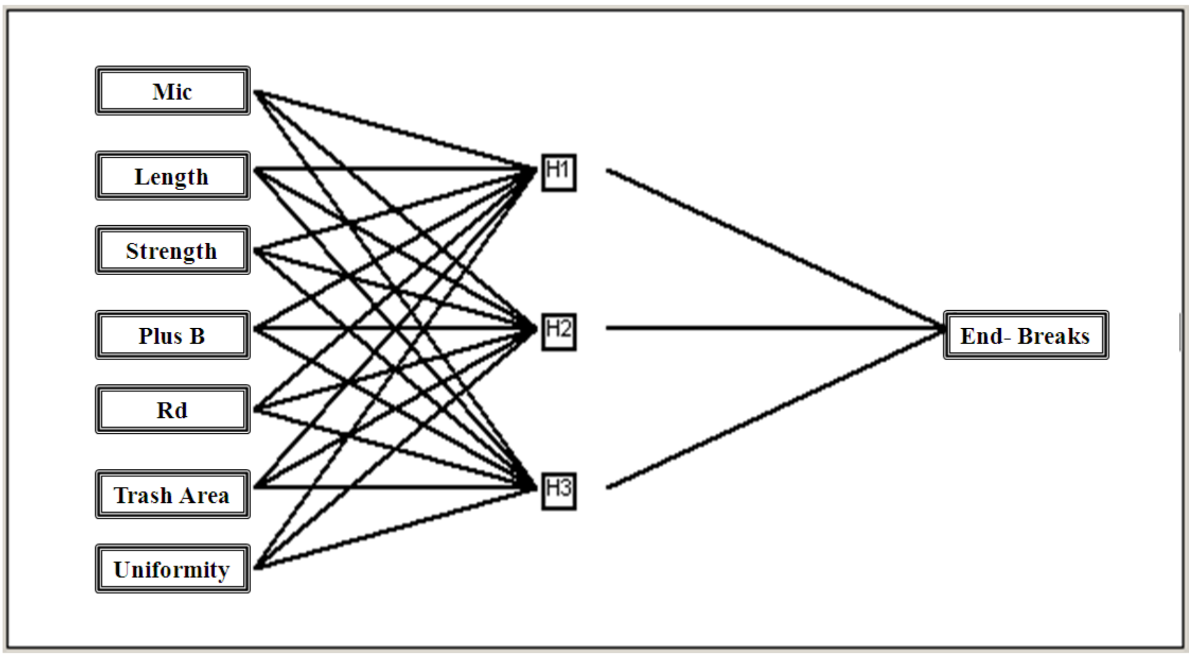


Figure 22 - Neural Network Diagram, Laydown Medians

By exploring neural networks, an advanced data mining tool, it is seen that when given the right amount of significant factors and properties, accurate predictions can be made. The existence of a tool such as this should be a strong incentive for a cotton spinning company to want to possess a competitive data management system. Data techniques can be performed that lead to better decision making, provided that the right data is kept.

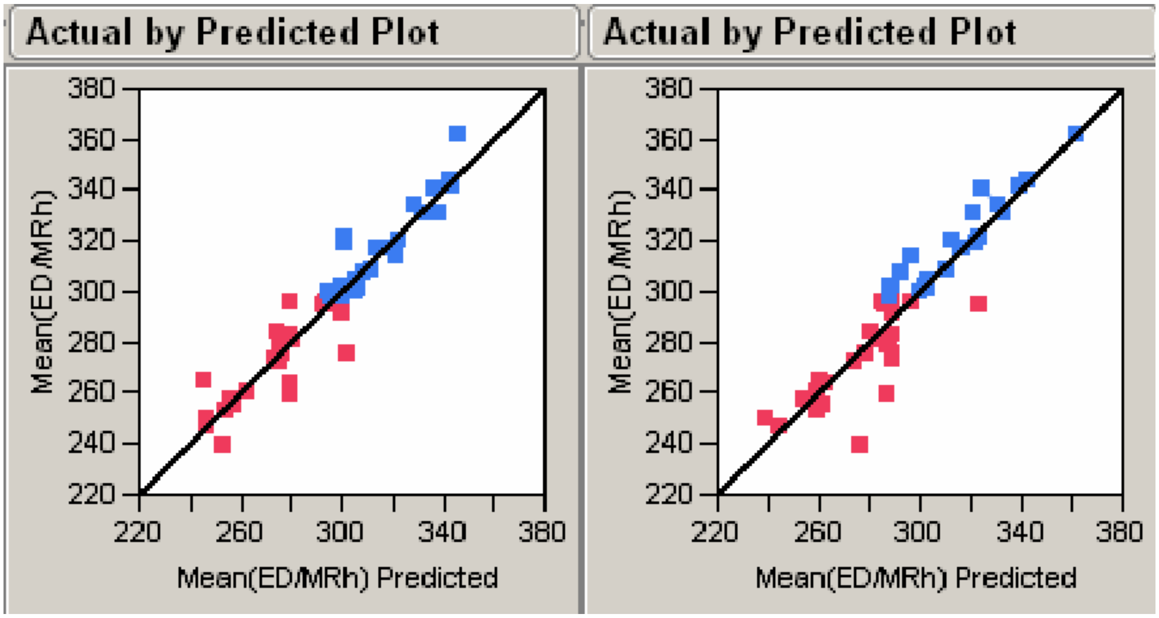


Figure 23 - Actual Ends-Down by Predicted Ends-Down Plots

To get more specific results for the open-end spinning mill data found in Data Set 2, further neural network models were created using mean laydown properties to predict the number of end-breaks. After the initial model was created using 7 properties, 6 additional models were made, with each subsequent model using one less property than the previous model. For each model, the combination of properties which gave the highest R squared value was used. Table 17 shows the power of each model and the properties utilized by each.

Table 17 - Neural Network Results

# of Inputs	Properties Used	R Squared
7	Micronaire, Length, Strength, Plus B, Rd, Trash Area, Uniformity	0.901
6	Micronaire, Length,Plus B, Rd, Trash Area, Uniformity	0.878
5	Micronaire, Length,Plus B, Rd, Trash Area	0.876
4	Micronaire, Length, Rd, Trash Area	0.804
3	Length, Rd, Trash Area	0.784
2	Length, Rd	0.752
1	Rd	0.665

Based on the results of these networks, a decision can be made on which properties are more valuable for predicting end-breaks. The model loses power each time an input is taken away. However, little power is lost between using 6 inputs and using 5 inputs. The model with 5 inputs is a good compromise for having enough inputs for a powerful predictive neural network model without requiring every single property to be tracked. This follows Data Mining Commandments 7 and 8 from Table 11, as the model was refined iteratively and made as simple as possible without being overly simple.

It is interesting to note that of the models created using just 1 input, the model which used Rd produced a much greater R squared than any other single-input model. In fact, it was more than double the next closest single-input model. This may be because while Rd is a measure of brightness, this measurement may be influenced by properties such as wax content. This could impact cohesion of fibers, leading to a change in end-breaks, as end-breaks are often caused by poor fiber cohesion and not necessarily fiber breakage.

6.5 Yarn Properties

Obtaining two data sets, Data Set 1 and Data Set 2, provided the opportunity for some direct comparisons. The main difference between the yarns being produced in these data sets is the cotton blend. The yarn being spun in Data Set 1 is made from single cotton, which helps to control variables in the research environment. The yarn being spun in Data Set 2 is made from a blend of many cottons, which provides consistency in the mill's products over time.

Both data sets contain some measurements of yarn evenness properties. Data tables were made from each data set, featuring the common yarn evenness measurements. Of the yarns the USDA spun in Data Set 1, some are open-end yarns, while others are ring-spun. To provide the best comparison with Data Set 2, only the open-end yarn tests were used.

This produces a table of 53 rotor yarn entities, which is much smaller than the table produced from Data Set 2, which included 1655 unique entities. For better comparison, the similar columns in these tables were given matching titles. These tables were then analyzed using pairwise correlations to see if blending influences the correlations of these properties. The yarn evenness measurement descriptions and the modeling results are found in Appendix D.

The analysis showed no great difference between the two data sets. Because all of these measurements relate to evenness, they show a large deal of statistically significant positive correlations. Data Set 1, which has the single cotton yarn, shows a nearly perfect correlation (0.98) between thick and thin place measurements. Whereas the thick to thin relationship in Data Set 2, which has the blended cotton yarn, has a correlation of only 0.87. It should be notes that the sample size was vastly greater for Data Set 2 (1655 samples) than Data Set 1 (53 samples).

Appendix E shows the descriptive yarn evenness statistics, created with SigmaXL, for both Data Set 1 and for Data Set 2. In comparing the statistical breakdown of each, Data Set 1 has lower coefficients of variation, thick place, thin places, and neps for nearly all measurements. This could be due to the smaller production runs used in the experiment or even the quality of the yarns involved. It could also be due to incomplete mixing of the blended yarn.

The two data sets available for this research had a lesser amount of common yarn tensile measurements than they did yarn evenness measurements. However, of the measurements available, the results exhibited some notable differences. Both data sets contain the variables of work, force, and elongation, which are defined in Table 18, so the correlations between these measurements were created with Jmp for each data set. The results appear in Figure 24.

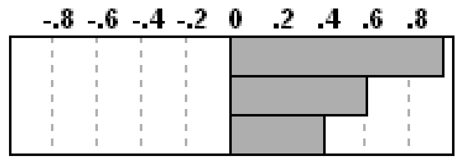
Table 18 - Yarn Tensile Measurements and Descriptions (adapted from Uster.com, 2009)

Measurement	**Description**
Elongation	**% the sample has increased in length before breaking**
Force	**Maximum force value measured to break yarn**
Work	**The area below the force/elongation curve at the time of yarn break**

Both sets feature a very strong statistically significant positive correlation, however, the members of the correlation differ between data sets. Data Set 1 has the strongest relationship between work and force, while Data Set 2 has the strongest relationship between work and elongation. While this disparity could be the result of different testing protocols, it may also be because of the cotton make-up of the yarn being produced. The single-cotton yarn acts differently than the blended yarn. This perhaps gives further reason why predictive models based on single-cotton yarn experiments may not necessarily hold much value for yarn spinning facilities that use blended yarns. In either case, it is likely not necessary for an efficient data management plan to track all 3 of these measures.

Pairwise Correlations *Data Set 1*

Variable	by Variable	Correlation	Signif Prob
Work	Force	0.9659	<.0001*
Elongation	Work	0.6195	<.0001*
Elongation	Force	0.4262	0.0015*



Pairwise Correlations *Data Set 2*

Variable	by Variable	Correlation	Signif Prob
Work	Elongation	0.9306	<.0001*
Work	Force	0.7414	<.0001*
Elongation	Force	0.4877	<.0001*

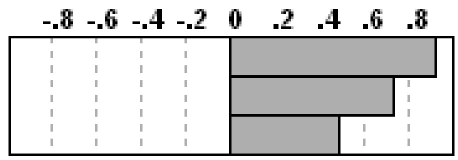


Figure 24 - Pairwise Correlations for Yarn Tensile Measurements

When analysis was performed to determine correlations between yarn evenness and tensile measurements, no such correlation was found in Data Set 2, which came from an open-end spinning plant. Only 1 relationship was found in Data Set 1, which came from the USDA study and had a very small sample size. This was a linear relationship between yarn uniformity and force to break (R squared = .826).

Also using the admittedly small sample size from Data Set 1, a table was created that combined HVI properties from the bale and final open-end yarn properties. This was accomplished by linking the data using the Fiber ID field found in this data set. This table was used to find equations that could predict yarn qualities using fiber qualities measured at the bale. One HVI property that seemed to have an impact on the final yarn is micronaire. Micronaire could make fairly accurate predictions for yarn uniformity (adjusted R squared = .759) and work to break (adjusted R squared = .756). The other HVI property that led to a couple of fairly accurate predictive equations is short fiber index (SFI). This is a measure of short fibers in the bale made by the USTER HVI. SFI was used in an equation for predicting yarn uniformity (adjusted R squared = .759), and a more impressive equation for work to break (adjusted R squared = .881). The predictive power of these two measurements was kept in mind in creating a data management model based on the data of this research.

6.6 Data Management Recommendations

Based on the results of the data analysis and what is known about the typical cotton spinning plant, a data management model was created in accordance with Research Goal 3. This model suggests a system of data management that stores the data in such a way that it is easily accessible.

Based on the data available at each plant, there is room for variation on this model. For some plants, most data is electronic and can simply be sent to the same location. For others, with written data or incompatible online systems, setting up a suitable database may require regular manual data entry. This model can be customized to match the needs and capabilities of any facility.

The basis of this model is the linking of data. For data mining purposes, it is best if all the data in a plant can be related by common identifiers. The industry visits of this study revealed that there could be difficulty in accomplishing this. It would require a mill to estimate all of the lag times in the production process. These times are needed to link data being measured at different sides of the plant. It must be decided when the same fiber being measured in one spot is going to be measured elsewhere. Once the plant makes such estimations, this study suggests they work backwards from the final yarn.

An example of an effective database design was created and is diagrammed in Figure 25. This design included some of the key data points as discovered in this research. Each data source comes with an identification designation. These sources are all linked to the final product. This provides a way to data mine for interactions between all of the data in a plant, which would be a very powerful ability, as it can help the mill to eliminate redundant data. This also enables a customer to see every piece of data that was produced for their product, if they so desire. Once the mill has linked their data and figured out how to insert it all into a single database, the possibilities are massive.

Along with the database design, this research also created a sample mock-up for a program that would allow plant personnel to easily access this database. The idea is to create a user-friendly interface that allows all personnel to be able to access customized information from the available data. This program is linked directly to the mill’s database, and uses structured query language (SQL) statements, one of the simpler data mining tools, to provide custom tailored tables. Such a program requires minimal computer knowledge, because it creates its SQL statements behind the scenes based on the user's inputs.

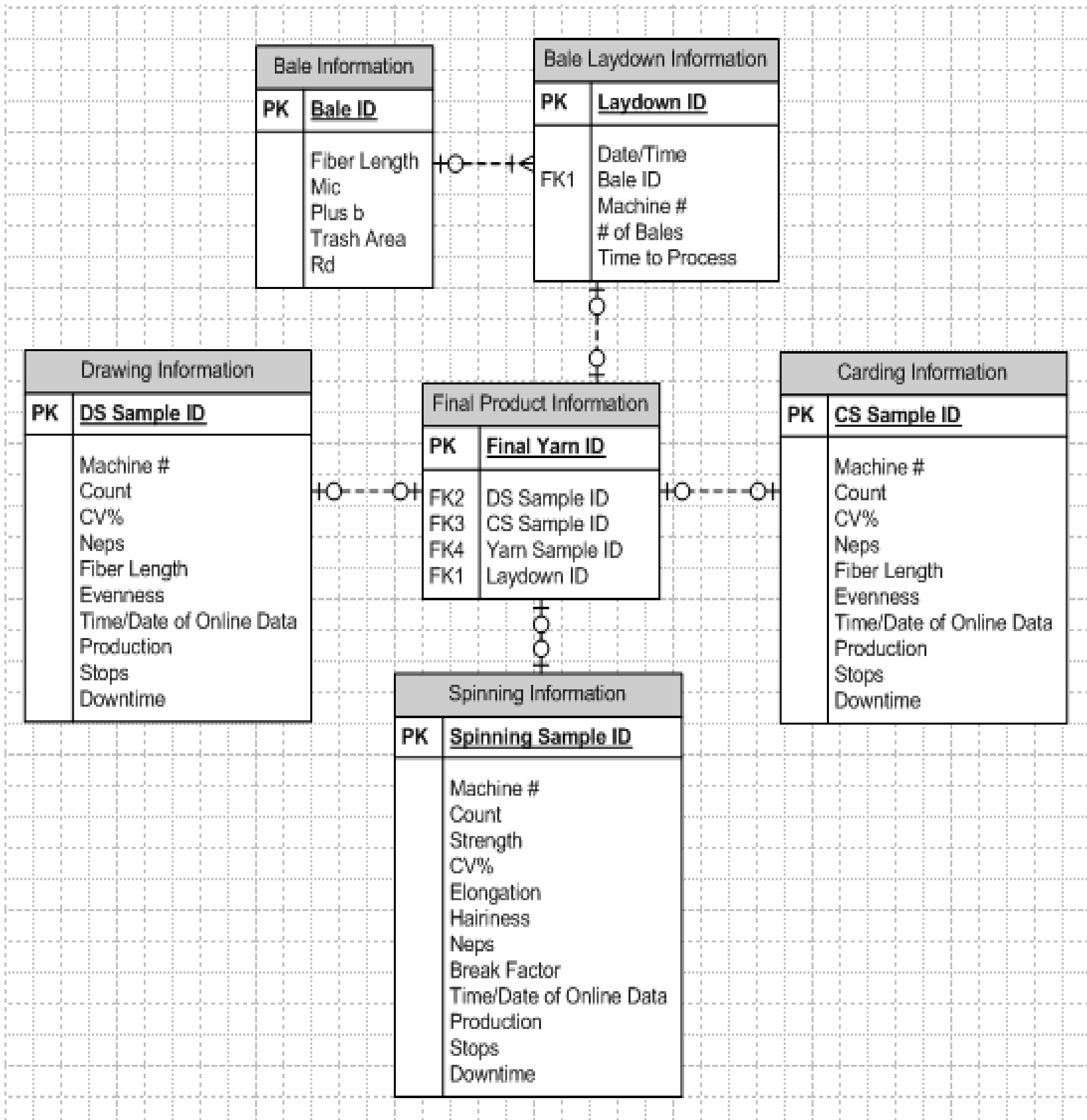


Figure 25 - Sample Database Design

One function of the program is shown in Figure 26. This screen enables an employee to choose what data to examine. For example, the program can be told to display the performance of a certain machine for any date in the database history.

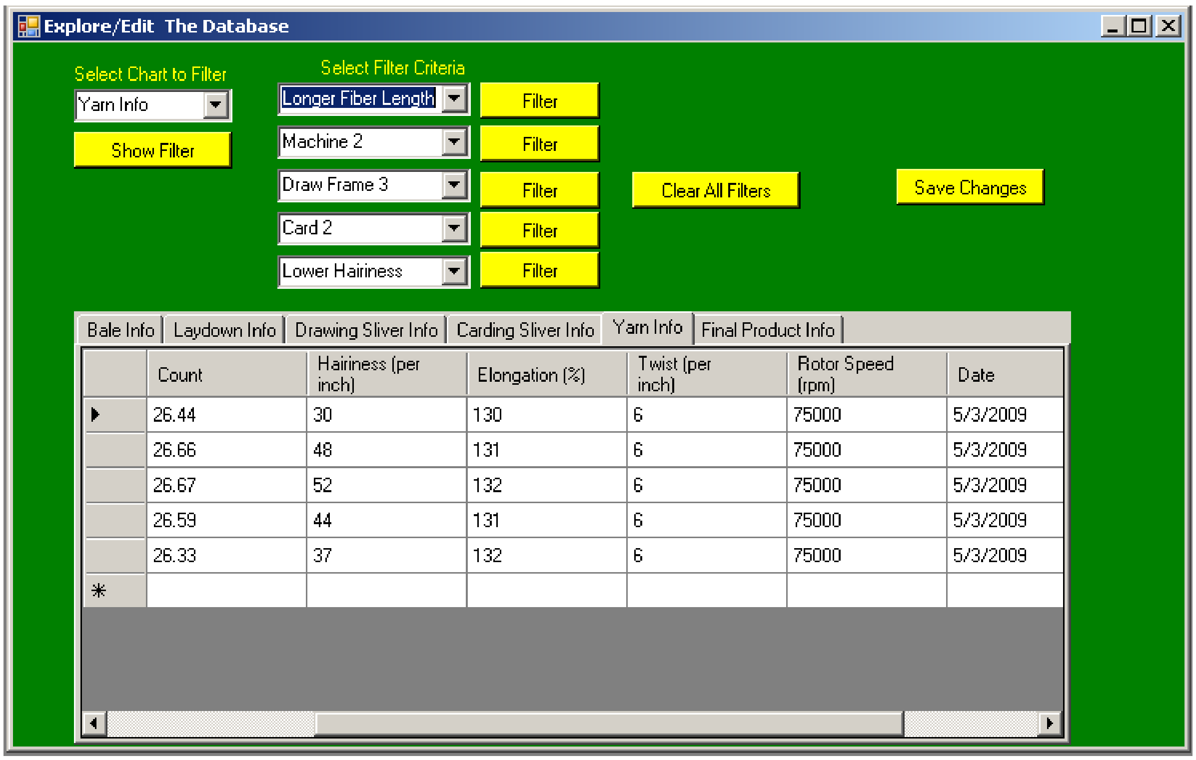


Figure 26 - Data Management Program Screen 1

The data can also be explored using a variety of filters. For instance, the program can be asked to show only data for cotton yarns with higher hairiness for the current week. The employee can then see all of the data for such yarns going back to the bale if the correct lag times have been decided. This may point to a certain draw frame or laydown date, for example. This screen can also be used to manually input or edit every data entity. A password could be used to keep such a function exclusive to those with the necessary authority.

The other function of this program is a screen that allows the user to input a final product code and to see all of the data created from it at once. This could be shown to the customer if they have a question about their purchase or used internally to troubleshoot defects. This screen would look something like the sample shown in Figure 27.

Figure 27 - Data Management Program Screen 2

A program such as this represents just one method of extracting value from an organized accumulation of data. It allows custom information to be given to plant personnel that will help augment everyday decision making and troubleshooting. It also serves to increase the overall data knowledge of the work force.

The rest of the value of this data can be found in the hidden trends and relationships within. These represent the sort of information that data mining can bring to the surface and can lead to larger-scale decisions that impact the overall value of the plant, and even the industry at large.

Chapter 7 - Conclusions & Recommendations for Future Studies

7.1 Conclusions

For effective decision making in any business, it is vital to have access to the appropriate information as quickly as possible. Obviously, it is impossible to be able to get all of the necessary information for decision making if a company's data is spread out or not stored over time. However, just stockpiling every piece of data created in one massive warehouse would also hinder effective management. When too much data is available, it is difficult to locate that which is germane to the decision at hand. This is a condition known as information overload.

To reach Research Goal 1, this study observed that a well managed database of important cotton spinning data was absent from most modern-day plants. While many plants archive their data in some fashion, they rarely take advantage of modern technology. Some facilities save only paper copies of testing reports, while others use computers but copy over old data after a period of time. With the computer memory space and advanced data mining techniques available, it can only enhance the business of a cotton spinning mill to develop a plan to harness the potential of all the data being generated every day.

As this research discovered, the key to finding valuable knowledge in data is to be able ask the right questions. However, having data limitations inhibits the amount of questions that can be asked. To address Research Goal 2, this study worked with two data sets, and neither was very large nor created for the purpose of data mining. However, this data was able to address some questions and provide interesting results.

Data mining could possibly inform yarn spinning plants of the best lprocessing points from which to take samples for certain tests. It may allow the plant to find a more effective way of mixing bales for laydowns. It may help a plant look back to discover the cause of faulty yarn. It could even save time and money by eliminating unnecessary processes or tests. Any one of these outcomes would represent a competitive advantage over other members of the cotton yarn industry.

Even if a plant cannot initially reach any of these achievements, it is certain that computing power will continue to grow. Higher quantities of data will be generated. Computer storage capacities will continue to rise. Data mining software will improve, and innovations will be made in the field of data mining. At some point, a cotton spinning mill will need to update its data management. This is recommended because the earlier this happens, the sooner an effective data warehouse can be built. Such a management model was made in accordance with Research Goal 3.

Such a change may be easier for some plants than others. As was seen in the business travels of this research, plants have a varying degree of modern technology. Some are only a few steps away, while others may require an investment in computers or newer machinery. Most of the plants visited seem content with their data management style, and few indicated a strong urge to change. From what this study learned about the capabilities of data mining in other industries, it is recommended that updating data management be made a higher priority.

In analyzing the data that was available in this study for Research Goal 2, some specific conclusions in regards to creating an effective data management system were reached. Firstly, data mining was used to determine that AFIS measurements taken at just the cleaner and the card sliver could be used to predict the fiber properties for all other points in the yarn manufacturing process. Also, a hierarchy of AFIS properties with regards to predicting spinning end-breaks was decided using neural networks. It was determined that the following 5 properties could lead to accurate predictions of end-breaks: micronaire, fiber length, plus b, Rd, and trash area. Lastly, micronaire and short fiber index were determined to be the best HVI properties to use for predicting final yarn characteristics such a uniformity and work to break.

Having known predictive equations can be doubly beneficial. Firstly, the equations can be used to determine ahead of time if the final product is going to be of substandard quality. If this is known, the proper adjustments and decisions can be made. Secondly, the equations can be used backwards. This would allow the cotton spinning mill to determine what fiber properties are needed to achieve a given end product result. This would allow for better controls and material selection early in the process.

7.2 Recommendations for Future Studies

This study revealed several opportunities for which further research should prove valuable. Firstly, this study narrowed its focus to cotton spinning mills located in the Southeast United States. Similar studies could be performed for synthetic yarn spin mills, and for facilities in different geographic locations. Studies could also be conducted for other facets of the textile industry, such as weaving or knitting plants.

While this study discusses a model for data management designed to make things simpler and possibly expose hidden trends, it does not look into the economic value of such a data management model. Future research could be done that looks into the costs involved with implementing such a plan, and its potential effects on the plant's bottom line and the economic competitiveness of the industry.

Also, as has been the case in similar studies in the past, there is always room for a larger and more effective data set on which to perform data mining. This could be done by identifying additional plants possessing sufficient data warehouse, or by setting up an intensive data collection plan.

It would also be interesting to perform a study on blended yarns, in which the bales are mixed on differing criteria in each laydown. This could give a look into how the fibers interact and how the properties of the blend reflect those of its components. It would also be of interest to perform a controlled test on yarn spinning similar to the FQEL research project done by the USDA, except using blended yarn rather than single-cotton.

References

Aarnink, H. W. M. (1996). *AFIS based, quality monitoring and control in textile processing*.

Adanur, S. (2001). *Handbook of weaving*. Lancaster, Pa.: Technomic Pub. Co. Inc.

Anderson, G. C. (1976). *The evolution of spinning : The Edward R. Schwarz lecture presented at the annual conference of the textile division of the American society of mechanical engineers, Charlotte, North Carolina, September 14-15, 1976*. New York: American Society of Mechanical Engineers.

Anderson, M. E. (2006). *Applicability of data mining in yarn manufacturing* (Masters thesis, North Carolina State University). Retrieved from http://www.lib.ncsu.edu/theses/available/etd-05162006-235654/unrestricted/etd.pdf

Artzt, P. (2004). Short staple spinning. *International Textile Bulletin English Edition, 50*(3), 17.

Basu, A. (2009). Interaction of the properties of individual cotton fibres in a blend. *Journal of the Textile Institute, 100*(8), 722.

Beceren, Y., & Uygun, B. (2008). Comparison of the effects of cotton yarns produced by new, modified and conventional spinning systems on yarn and knitted fabric performance. *Textile Research Journal, 78*(4), 297.

Beck, N. T. (1996). *Data communications in computer integrated manufacturing for the United States textile spinning industry* (Masters thesis, North Carolina State University).

Braha, D. (2001). *Data mining for design and manufacturing : Methods and applications*. Dordrecht ; Boston: Kluwer Academic.

Britannica Online Encyclopedia. (2009). *Data mining*. Retrieved 11/08, 2009, from http://search.eb.com.www.lib.ncsu.edu:2048/eb/article-9437561

Cabena, P. (1997). *Discovering data mining: From concept to implementation.* Upper Saddle River, NJ: Prentice Hall.

Campbell, J., Delhom, C., & Thibodeaux, D. (2005). FQEL: Laboratory Data Validation. *Proceedings of the Beltwide Cotton Conferences*.

Cotton Incorporated. (2001). *The classification of cotton*. Retrieved 6/08, 2009, from http://www.cottoninc.com/ClassificationofCotton/

Cotton Incorporated. (2009). *EFS system*. Retrieved 6/08, 2009, from http://www.cottoninc.com/EFS/

Cox, K., Eick, S., Wills, G., & Brachman, R. (1997). Brief application description; visual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery, Volume 1*(2), 225-231. Retrieved from http://www.springerlink.com.www.lib.ncsu.edu:2048/content/gk180510n2370602

CRISP-DM.org. (2009). *CRISP-DM - process model*. Retrieved 10/29, 2009, from http://www.crisp-dm.org/Process/index.htm

Cummins, F., & Knovel. (2002). *Enterprise integratio*n. New York: John Wiley & Sons. Retrieved from http://www.lib.ncsu.edu/cgi-bin/proxy.pl?server=knovel.com/knovel2/Toc.jsp?BookID=1418

Daley, C. M. (2008). *Application of data mining tools for exploring data* (Masters thesis, North Carolina State University). Retrieved from http://www.lib.ncsu.edu/theses/available/etd-10292008-165755/unrestricted/etd.pdf

Delhom, C., Cui, X., Campbell, J., & Thibodeaux, D. (2005). Determination of Fiber and Product Quality Through Small-Scale Processing Trials: Fiber to Yarn. *Proceedings of the Beltwide Cotton Conferences*.

Delhom, C., Thibodeaux, D., & Rodgers, J. (2007). Relating Fiber Quality to Yarn Quality Utilyzing Data From Multiple Years. *Proceedings of the Beltwide Cotton Conferences*

Echeverria Magariänos, C. (2008). *Yarn specifications and performance metrics for short staple yarn manufacturer*s (Masters thesis, North Carolina State University). Retrieved from http://www.lib.ncsu.edu/theses/available/etd-07102008-061810/unrestricted/etd.pdf

Fayyad, U. M. (2002). Evolving data mining into solutions for insights-introduction. *Communications of the ACM, 45*(8), 28.

Fayyad, U. M. (1996). *Advances in knowledge discovery and data mining.* Menlo Park, Calif.: AAAI Press : MIT Press.

Geissler, J. E. (1993). *Development and evaluation of cotton, wool, and mohair blend fabrics and assessment of consumer response*.

Gonzalez, R., & Kamrani, A. (2001). A survey of methodologies and techniques for data mining and intelligent data discovery. In D. Braha (Ed.), *Data mining for design and manufacturing* (pp. 41-59). The Netherlands: Kluwer Academic Publishers.

Grosberg, P., & Iype, C. (1999). *Yarn production: Theoretical aspect*s. Manchester England: Textile Institute.

Guha, A., Chattopadhyay, R., & Jayadeva. (2001). Predicting yarn tenacity: A comparison of mechanistic, statistical, and neural network models. *Journal of the Textile Institute, 92*(1), 139.

Hequet, E., & Ethridge, M. (2005). Impacts of fiber length distribution on market value and yarn quality: Implications for U.S. cotton. *Textile Topics, 2005*(3), 2.

Jayadeva, Guha, A., & Chattopadhyay, R. (2003). A study on the capability of a neural network ranking fibre parameters having an influence on yarn properties. *Journal of the Textile Institute, 94*(1), 186.

Karpe, Y. S. (2006). *Weave-room performance decision-making process in textiles: Mapping an information engineering methodology* (Doctoral dissertation, North Carolina State University). Retrieved from http://www.lib.ncsu.edu/theses/available/etd-08312006-095818/unrestricted/etd.pdf

Karthikeyan, P., Senthilkumar, M., & Patil, C. (2009). Ring yarn quality traits on cone winding machines. *The Indian Textile Journal, 119*(7), 34.

Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data minin*g. Hoboken, N. J.: Wiley-Interscience.

Lee, J. H., & Park, S. C. (2001). Data mining for high quality and quick response manufacturing. In D. Braha (Ed.), *Data mining for design and manufacturing* (pp. 179-205). The Netherlands: Kluwer Academic Publishers.

Levitt, S. D., & Dubner, S. J. (2006). *Freakonomics: A rogue economist explores the hidden side of everything.* New York: William Morrow.

Linton, G. E. (1963). *The modern textile dictionary* (3rd ed.). New York: Duell, Sloan and Pearce.

LIS.Illinois.edu. (2009). *Decision tree*. Retrieved 11/08, 2009, from http://gautam.lis.illinois.edu/monkmiddleware/public/analytics/decision_tree.png

Liu, H., Yu, L., & Motoda, H. (2003). Feature extraction, selection, and construction. In N. Ye (Ed.), *The handbook of data mining* (pp. 409-423). New Jersey: Lawrence Erlbaum Associates, Inc.

Lord, P. R. (1974). *Spinning: Conversion of fiber to yarn.* Raleigh: School of Textiles, North Carolina State University.

Lord, P. R. (2003). *Handbook of yarn production: Technology, science and economic*s. Boca Raton; Cambridge : Woodhead Publishing: CRC Press; Textile Institute.

Manik, S. M. (2009). *Data mining: History of data mining*. Retrieved 11/08, 2009, from http://dataminingwarehousing.blogspot.com/2008/10/data-mining-and-warehousing-history-of.html

McCreight, D., Feil, R., Booterbaugh, J., & Backe, E. (1997). *Short staple yarn manufacturing.* Durham, N.C.: Carolina Academic Press.

Melliand international. (2009). Strong decline in short-staple spinning investments. *Melliand International, 15*(4), 132.

Monie, H. (1904). *The cotton fibre and the mixing of cotton.* Manchester: Heywood.

Morton, W. E. (1962). *An introduction to the study of spinning* (3rd ed.). London: Longmans, Green.

Morton, W. E., & Hearle, J. W. S. (1993). *Physical properties of textile fibre*s (3rd ed.). Manchester, UK: The Textile Institute.

Olson, D. L., & Shi, Y. (2007). *Introduction to business data minin*g. Boston: McGraw-Hill/Irwin.

Ouellette, S. (2007). *US textile production in historical perspective: A case study from Massachusett*s. New York: Routledge.

Pyle, D. (2003). Data collection, preparation, quality, and visualization. In N. Ye (Ed.), *The handbook of data mining* (pp. 365-391). New Jersey: Lawrence Erlbaum Associates, Inc.

Rasmovich, J. A. (1997). *Information technology and information engineering in the United States textile spinning industry* (Masters thesis, North Carolina State University).

Rieter.com. (2009). *SPIDERweb - innovative mill monitoring from blowroom to spinning machines*. Retrieved 6/08, 2009, from http://www.rieter.com/en/textile/short-staple-yarn/mill-information-systems/spiderweb/

SAS.com. (2009). *Data mining software, SAS® enterprise miner | SAS*. Retrieved 11/4, 2009, from http://www.sas.com/technologies/analytics/datamining/miner/index.html#section=4

Schertel, S. L. (2002). *Data mining and its potential use in textile*s (Doctoral dissertation, North Carolina State University). Retrieved from http://www.lib.ncsu.edu/theses/available/etd-05072002-181204/unrestricted/etd.pdf

Sethi, I. K. (2001). Data mining: An introduction. In D. Braha (Ed.), *Data mining for design and manufacturing* (pp. 1-40). The Netherlands: Kluwer Academic Publishers.

Shaikh, I. A. (2005). *Pocket spinning expert: A practical handbook on textile yarn manufacturing, includes basic & advanced information.* Lahore, Pakistan: Textile Info Society.

Siggraph.org. (2009). *Neural network*. Retrieved 10/29, 2009, from http://www.siggraph.org/education/materials/HyperVis/applicat/data_mining/images/neural.gif

Singh, R., & Kothari, V. (2009). Developments in blowroom, card draw frame. *The Indian Textile Journal, 11*9(7), 17.

Slater, K., & Hoffmeyer, F. (1979). Economic and aethetic factors in the future of cotton. In P. W. Harrison (Ed.), *Cotton in a competitive world : Papers of the 63rd annual conference of the textile institute organized in collaboration with the textile association of India, January 18th-23rd, 1979, Indian institute of technology, Delhi* (pp. 8). Manchester, Eng.: The Institute.

Spencer, D. J. (2001). *Knitting technology: A comprehensive handbook and practical guide* (3rd ed.). Cambridge, England; Lancaster, Pa.: Woodhead Publishing Limited; Published in North America by Technomic Pub. Co.

Subramanian, T. A., & Garde, A. R. (1974). *End breaks in ring spinning.* Ahmedabad, India: Ahmedabad Textile Industry's Research Association.

Taggart, W. S. (1923). *Cotton mill management: A practical guide for managers, carders and overlookers.* London: Macmillan.

The Textile Magazine. (2006). SITRA's programme on yarn quality. *The Textile Magazine, 47*(11), 78.

The-Data-Mine.com. (2009). *Most popular data mining software*. Retrieved 11/4, 2009, from http://www.the-data--mine.com/bin/view/Software/MostPopularDataMiningSoftware

Truetzschler.eu. (2009). *Trutzschler - product range*. Retrieved 6/08, 2009, from http://www.truetzschler.eu/product-range/product-range/

United States Department of Agriculture. (2009). *Understanding standardization cottons*. Retrieved 12/13, 2009, from http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELDEV3099536

Uster.com. (2009). *USTER products*. Retrieved 6/08, 2009, from http://www.uster.com/UI/textile-Products-2-2.aspx

Wu, T., & Li, X. (2003). Data storage and management. In N. Ye (Ed.), *The handbook of data mining* (pp. 393-407). New Jersey: Lawrence Erlbaum Associates, Inc.

Appendixes

Appendix A - AFIS Descriptive Statistics Charts

Nep Count /g				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	324.58	564.08	626.16	649.17
Stdev	232.13	363.73	397.04	377.89
Range	805	1214	1220	1111
Minimum	42	62	205	203
25th Percentile (Q1)	156.50	240	256.50	310.50
50th Percentile (Median)	235	445	486	550
75th Percentile (Q3)	436	1026	1065.5	1068
Maximum	847	1276	1425	1314
95.0% CI Mean	268.82 to 380.34	461.78 to 666.38	506.87 to 745.44	545.01 to 753.33
95.0% CI Sigma	198.82 to 278.94	304.34 to 452.15	328.69 to 501.54	317.18 to 467.55
Anderson-Darling Normality Test	2.924	3.001	2.834	2.716
p-value (A-D Test)	**0.0000**	**0.0000**	**0.0000**	**0.0000**
Skewness	1.002059	0.639739	0.751008	0.629926
p-value (Skewness)	**0.0015**	0.0571	**0.0375**	0.0565
Kurtosis	0.009787	-1.138	-0.954606	-1.0843
p-value (Kurtosis)	0.8231	**0.0018**	**0.0398**	**0.0036**
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	156.98	158.98	159.38	173.35
Stdev	135.78	143.96	150.02	152.41
Range	544	483	519	543
Minimum	10	23	21	21
25th Percentile (Q1)	49	39.500	42	33
50th Percentile (Median)	103	104	92	136
75th Percentile (Q3)	261	286.50	241.75	324
Maximum	554	506	540	564
95.0% CI Mean	120.96 to 193.01	117.63 to 200.33	115.81 to 202.94	126.45 to 220.25
95.0% CI Sigma	114.63 to 166.57	120.05 to 179.85	124.89 to 187.91	125.66 to 193.71
Anderson-Darling Normality Test	2.666	3.467	3.441	2.131
p-value (A-D Test)	**0.0000**	**0.0000**	**0.0000**	**0.0000**
Skewness	1.035189	0.93197	1.09401	0.788565
p-value (Skewness)	**0.0027**	**0.0096**	**0.0035**	**0.0332**
Kurtosis	0.24101	-0.57185	-0.112977	-0.550575
value (Kurtosis)	0.5513	0.3454	0.9743	0.4155

Trash Count /g				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	68	51	45	53
Mean	68.191	49.961	47.156	47.132
Stdev	99.684	23.954	18.821	26.854
Range	547	97	89	184
Minimum	2	4	13	2
25th Percentile (Q1)	27.500	37	36	32.500
50th Percentile (Median)	50	48	46	43
75th Percentile (Q3)	72	62	57.500	57.500
Maximum	549	101	102	186
95.0% CI Mean	44.062 to 92.320	43.224 to 56.698	41.501 to 52.810	39.730 to 54.534
95.0% CI Sigma	85.292 to 119.97	20.043 to 29.777	15.581 to 23.775	22.540 to 33.225
Anderson-Darling Normality Test	11.344	0.445129	0.298429	1.766
p-value (A-D Test)	**0.0000**	0.2729	0.5721	**0.0001**
Skewness	4.063	0.30226	0.547484	2.647
p-value (Skewness)	**0.0000**	0.3480	0.1178	**0.0000**
Kurtosis	16.927	-0.134667	0.77159	12.961
p-value (Kurtosis)	**0.0000**	0.9888	0.2380	**0.0000**
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	10.05263158	7.633	7.667	8.628
Stdev	5.572	4.595	4.874	6.392
Range	21	19	23	24
Minimum	1	1	0	1
25th Percentile (Q1)	6	4	3.250	4
50th Percentile (Median)	10	7	7	7
75th Percentile (Q3)	14	11	10	10
Maximum	22	20	23	25
95.0% CI Mean	8.574 to 11.531	6.313 to 8.952	6.251 to 9.082	6.661 to 10.595
95.0% CI Sigma	4.704 to 6.836	3.832 to 5.740	4.058 to 6.105	5.271 to 8.124
Anderson-Darling Normality Test	0.566675	1.049822562	1.158	2.508
p-value (A-D Test)	0.1359	**0.0084**	**0.0045**	**0.0000**
Skewness	0.358572	0.804077	1.077501	1.405
p-value (Skewness)	0.2448	**0.0222**	**0.0039**	**0.0007**
Kurtosis	-0.744741	0.10507	1.816	1.29
p-value (Kurtosis)	0.1139	0.7017	**0.0400**	0.1053

Maturity Ratio				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	0.890870	0.882941	0.884444	0.867736
Stdev	0.048408	0.047678	0.050161	0.053803
Range	0.240000	0.200000	0.230000	0.250000
Minimum	0.760000	0.760000	0.760000	0.720000
25th Percentile (Q1)	0.865000	0.850000	0.860000	0.835000
50th Percentile (Median)	0.880000	0.890000	0.890000	0.870000
75th Percentile (Q3)	0.920000	0.920000	0.920000	0.900000
Maximum	1	0.960000	0.990000	0.970000
95.0% CI Mean	0.879241 to 0.902498	0.869532 to 0.896351	0.869374 to 0.899515	0.852906 to 0.882566
95.0% CI Sigma	0.041463 to 0.05817	0.039892 to 0.059267	0.041526 to 0.063364	0.04516 to 0.066569
Anderson-Darling Normality Test	1.030562864	0.525670	0.624968	0.229538
p-value (A-D Test)	**0.0097**	0.1721	0.0973	0.7983
Skewness	0.039821	-0.438907	-0.495628	-0.33297
p-value (Skewness)	0.8857	0.1793	0.1539	0.2944
Kurtosis	0.02075	-0.433113	-0.06478	0.021013
p-value (Kurtosis)	0.8074	0.5320	0.9049	0.8016
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	0.890702	0.924898	0.947500	0.951860
Stdev	0.05669	0.055981	0.057111	0.063142
Range	0.250000	0.210000	0.230000	0.230000
Minimum	0.750000	0.800000	0.800000	0.820000
25th Percentile (Q1)	0.850000	0.885000	0.920000	0.890000
50th Percentile (Median)	0.880000	0.920000	0.940000	0.960000
75th Percentile (Q3)	0.940000	0.980000	1	1.01
Maximum	1	1.01	1.03	1.05
95.0% CI Mean	0.875660 to 0.905744	0.908818 to 0.940977	0.930917 to 0.964083	0.932428 to 0.971293
95.0% CI Sigma	0.047861 to 0.069545	0.046684 to 0.069937	0.047543 to 0.071537	0.052063 to 0.080254
Anderson-Darling Normality Test	0.436180	0.696242	0.607351	0.869696
p-value (A-D Test)	0.2881	0.0648	0.1082	**0.0235**
Skewness	0.051509	-0.15551	-0.27907	-0.263939
p-value (Skewness)	0.8646	0.6318	0.3980	0.4458
Kurtosis	-0.568011	-0.827938	-0.499969	-0.96145
p-value (Kurtosis)	0.3024	0.0880	0.4513	**0.0436**

Fineness (millitex)				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	169.79	163.36	162.24	157.94
Stdev	19.788	21.671	21.663	20.095
Range	69.600	65.800	67.400	67
Minimum	133.60	133.60	133.80	133.80
25th Percentile (Q1)	153.80	143.80	143.50	142.60
50th Percentile (Median)	168.80	154.20	152.80	149.80
75th Percentile (Q3)	189.80	185.80	185.30	174.80
Maximum	203.20	199.40	201.20	200.80
95.0% CI Mean	165.04 to 174.55	157.27 to 169.46	155.73 to 168.75	152.40 to 163.48
95.0% CI Sigma	16.949 to 23.778	18.133 to 26.939	17.934 to 27.365	16.867 to 24.863
Anderson-Darling Normality Test	0.981497	2.145	1.924	3.155
p-value (A-D Test)	**0.0128**	**0.0000**	**0.0001**	**0.0000**
Skewness	0.094035	0.327298	0.461789	0.904895
p-value (Skewness)	0.7347	0.3106	0.1821	**0.0091**
Kurtosis	-0.926717	-1.487	-1.324	-0.590757
p-value (Kurtosis)	**0.0095**	**0.0000**	**0.0001**	0.2958
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	166.72	173.85	178.67	179.09
Stdev	22.707	23.868	24.815	27.254
Range	73.600	73.200	76.600	76
Minimum	134.20	137.80	139	140.60
25th Percentile (Q1)	148.90	153.60	156.80	152.60
50th Percentile (Median)	155.20	165.80	173.90	169.80
75th Percentile (Q3)	190.30	201.50	206.95	208.40
Maximum	207.80	211	215.60	216.60
95.0% CI Mean	160.69 to 172.74	166.99 to 180.70	171.47 to 185.88	170.70 to 187.48
95.0% CI Sigma	19.171 to 27.856	19.904 to 29.818	20.658 to 31.083	22.472 to 34.640
Anderson-Darling Normality Test	2.813	2.065	1.775	2.523
p-value (A-D Test)	**0.0000**	**0.0000**	**0.0001**	**0.0000**
Skewness	0.592719	0.310706	0.237412	0.174642
p-value (Skewness)	0.0625	0.3436	0.4707	0.6120
Kurtosis	-1.154	-1.482	-1.47	-1.728
p-value (Kurtosis)	**0.0005**	**0.0000**	**0.0000**	**0.0000**

Average Fiber Length (inches)				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	0.815507	0.822941	0.822444	0.804717
Stdev	0.051264	0.034017	0.033177	0.04956
Range	0.380000	0.160000	0.130000	0.280000
Minimum	0.510000	0.730000	0.760000	0.600000
25th Percentile (Q1)	0.795000	0.800000	0.790000	0.790000
50th Percentile (Median)	0.810000	0.820000	0.830000	0.810000
75th Percentile (Q3)	0.850000	0.850000	0.850000	0.840000
Maximum	0.890000	0.890000	0.890000	0.880000
95.0% CI Mean	0.803192 to 0.827822	0.813374 to 0.832509	0.812477 to 0.832412	0.791057 to 0.818377
95.0% CI Sigma	0.043909 to 0.061602	0.028463 to 0.042286	0.027466 to 0.041909	0.041598 to 0.061319
Anderson-Darling Normality Test	2.966	0.629125	0.839567	1.583
p-value (A-D Test)	**0.0000**	0.0957	**0.0281**	**0.0004**
Skewness	-3.168	-0.32301	-0.142518	-1.793
p-value (Skewness)	**0.0000**	0.3168	0.6723	**0.0000**
Kurtosis	18.018	-0.112113	-0.974015	5.246
p-value (Kurtosis)	**0.0000**	0.9789	**0.0332**	**0.0003**
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	0.785088	0.806327	0.815000	0.825814
Stdev	0.049212	0.047201	0.054422	0.04797
Range	0.210000	0.180000	0.250000	0.160000
Minimum	0.660000	0.720000	0.670000	0.740000
25th Percentile (Q1)	0.750000	0.765000	0.780000	0.780000
50th Percentile (Median)	0.780000	0.810000	0.810000	0.830000
75th Percentile (Q3)	0.825000	0.850000	0.860000	0.880000
Maximum	0.870000	0.900000	0.920000	0.900000
95.0% CI Mean	0.772030 to 0.798146	0.792769 to 0.819884	0.799198 to 0.830802	0.811051 to 0.840577
95.0% CI Sigma	0.041548 to 0.060371	0.039362 to 0.058968	0.045304 to 0.068167	0.039553 to 0.06097
Anderson-Darling Normality Test	0.736547	0.585506	0.382443	1.114
p-value (A-D Test)	0.0519	0.1204	0.3850	**0.0058**
Skewness	-0.157846	0.147254	-0.133061	0.052676
p-value (Skewness)	0.6027	0.6499	0.6842	0.8779
Kurtosis	-0.651709	-1.00999	-0.239555	-1.374
p-value (Kurtosis)	0.2013	**0.0159**	0.8431	**0.0000**

SFC (% by number)				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	22.497	22.551	22.667	23.885
Stdev	6.687	5.656	5.496	6.515
Range	44.100	20.100	18.200	35.600
Minimum	12.300	12.800	12.800	11.200
25th Percentile (Q1)	17	16.400	16.650	18.050
50th Percentile (Median)	23.300	23.800	23.700	25.600
75th Percentile (Q3)	26.450	27.700	27.100	28
Maximum	56.400	32.900	31	46.800
95.0% CI Mean	20.891 to 24.104	20.960 to 24.142	21.016 to 24.318	22.089 to 25.681
95.0% CI Sigma	5.728 to 8.036	4.732 to 7.030	4.550 to 6.942	5.469 to 8.061
Anderson-Darling Normality Test	1.305	1.214	1.333	1.280
p-value (A-D Test)	**0.0020**	**0.0033**	**0.0016**	**0.0023**
Skewness	1.726	-0.24225	-0.295436	0.338996
p-value (Skewness)	**0.0000**	0.4497	0.3851	0.2860
Kurtosis	8.3	-1.222	-1.324	1.594
p-value (Kurtosis)	**0.0000**	**0.0003**	**0.0001**	0.0504
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	25.788	24.180	23.775	23.023
Stdev	7.024	6.969	7.188	7.249
Range	22.800	21.600	24	23.100
Minimum	13.500	12.800	11.200	12.200
25th Percentile (Q1)	19.450	16.550	16.475	14.500
50th Percentile (Median)	27.600	25.400	24.350	23
75th Percentile (Q3)	31.400	30.300	29.400	30
Maximum	36.300	34.400	35.200	35.300
95.0% CI Mean	23.924 to 27.652	22.178 to 26.181	21.688 to 25.862	20.792 to 25.254
95.0% CI Sigma	5.930 to 8.617	5.812 to 8.706	5.983 to 9.003	5.977 to 9.213
Anderson-Darling Normality Test	1.878	1.218	0.844032	1.626
p-value (A-D Test)	**0.0001**	**0.0032**	**0.0276**	**0.0003**
Skewness	-0.449831	-0.290494	-0.220029	-0.150018
p-value (Skewness)	0.1489	0.3750	0.5033	0.6627
Kurtosis	-1.188	-1.224	-1.15	-1.512
p-value (Kurtosis)	**0.0002**	**0.0004**	**0.0023**	**0.0000**

SFC (% by weight)				
Descriptive Statistics	**Bale**	**Cleaner**	**Condenser**	**Card Chute**
Sample Size	69	51	45	53
Mean	8	7.825	7.822	8.413
Stdev	3.157	2.309	2.134	2.889
Range	23.500	9.800	7.400	17.700
Minimum	3.900	4	4.100	3.600
25th Percentile (Q1)	5.700	5.500	5.550	6
50th Percentile (Median)	8.100	8	8.200	8.500
75th Percentile (Q3)	9.250	9.800	9.550	10
Maximum	27.400	13.800	11.500	21.300
95.0% CI Mean	7.242 to 8.758	7.176 to 8.475	7.181 to 8.463	7.617 to 9.210
95.0% CI Sigma	2.704 to 3.794	1.932 to 2.870	1.767 to 2.696	2.425 to 3.575
Anderson-Darling Normality Test	2.716	0.787743	0.857221	1.058194353
p-value (A-D Test)	**0.0000**	**0.0383**	**0.0254**	**0.0081**
Skewness	3.347	0.125063	-0.141123	1.481
p-value (Skewness)	**0.0000**	0.6944	0.6753	**0.0001**
Kurtosis	20.216	-0.624661	-1.218	6.361
p-value (Kurtosis)	**0.0000**	0.2658	**0.0011**	**0.0001**
Descriptive Statistics	**Card Sliver**	**Breaker Sliver**	**Finisher Sliver**	**Roving**
Sample Size	57	49	48	43
Mean	9.753	9.214	9.152	8.770
Stdev	3.113	3.013	3.183	3.072
Range	11.800	9.900	12.300	10
Minimum	4.500	4.500	3.900	4.300
25th Percentile (Q1)	7.050	6	5.925	5.200
50th Percentile (Median)	10.1	9.400	9.200	8.900
75th Percentile (Q3)	12.150	11.650	11.500	11.600
Maximum	16.300	14.400	16.200	14.300
95.0% CI Mean	8.927 to 10.579	8.349 to 10.07963	8.228 to 10.07619	7.824 to 9.715
95.0% CI Sigma	2.628 to 3.819	2.512 to 3.764	2.649 to 3.986	2.533 to 3.905
Anderson-Darling Normality Test	1.033121891	0.898351	0.518062	1.272
p-value (A-D Test)	**0.0094**	**0.0202**	0.1793	**0.0023**
Skewness	-0.148539	-0.142195	0.048137	-0.04705
p-value (Skewness)	0.6241	0.6611	0.8828	0.8909
Kurtosis	-1.01764	-1.145	-0.839764	-1.437
p-value (Kurtosis)	**0.0067**	**0.0021**	0.0848	**0.0000**

Appendix B - Adjusted R Squared Values for AFIS Predictive Equations

Nep Count Predictive Equations

Cleaner

- Condenser (.988 R^2 Adj)
- Card Chute (.995 R^2 Adj)

Card Sliver

- Breaker Sliver (.990 R^2 Adj)
- Finisher Sliver (.990 R^2 Adj)
- Roving (.935 R^2 Adj)

Trash Count Predictive Equations

Cleaner

- Condenser (.948 R^2 Adj)
- Card Chute (.929 R^2 Adj)

Card Sliver

- Breaker Sliver (.757 R^2 Adj)
- Finisher Sliver (.466 R^2 Adj)
- Roving (.688 R^2 Adj)

Maturity Ratio Predictive Equations

Cleaner

- Condenser (.925 R^2 Adj)
- Card Chute (.927 R^2 Adj)

Card Sliver

- Breaker Sliver (.965 R^2 Adj)
- Finisher Sliver (.949 R^2 Adj)
- Roving (.939 R^2 Adj)

Mean Fiber Length Predictive Equations

Cleaner

- Condenser (.954 R^2 Adj)
- Card Chute (.965 R^2 Adj)
- Breaker Sliver (.920 R^2 Adj)
- Finisher Sliver (.796 R^2 Adj)
- Roving (.775 R^2 Adj)

Card Sliver

Short Fiber Content (by Weight) Predictive Equations

Cleaner

- Condenser (.969 R^2 Adj)
- Card Chute (.957 R^2 Adj)
- Breaker Sliver (.864 R^2 Adj)

Card Sliver

- Finisher Sliver (.730 R^2 Adj)
- Roving (.782 R^2 Adj)

Short Fiber Content (by Number) Predictive Equations

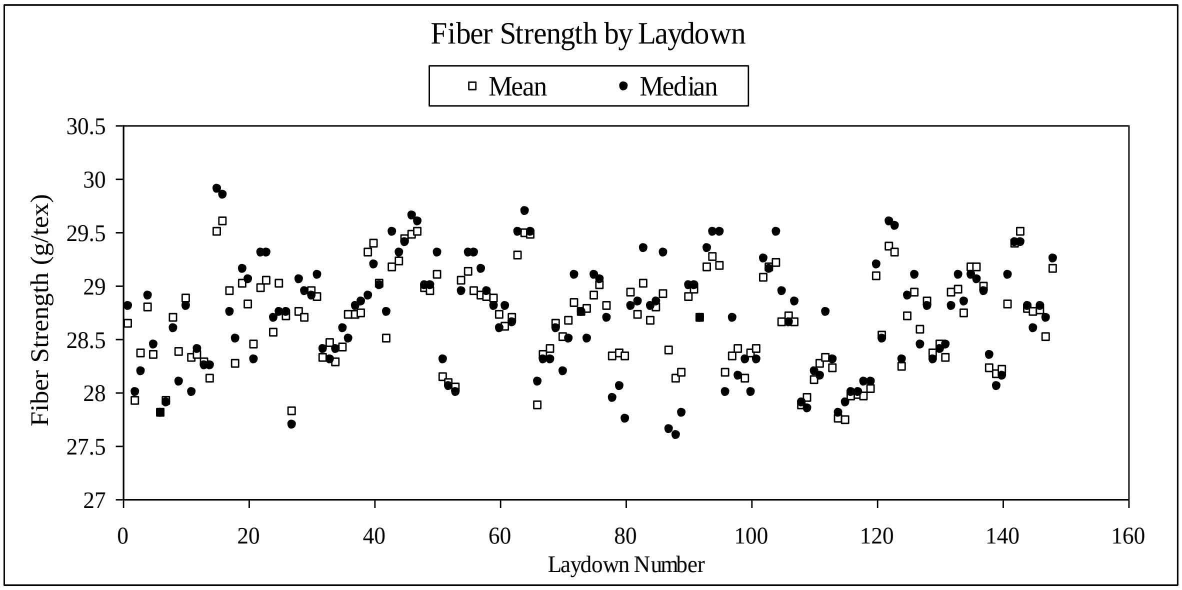
Cleaner

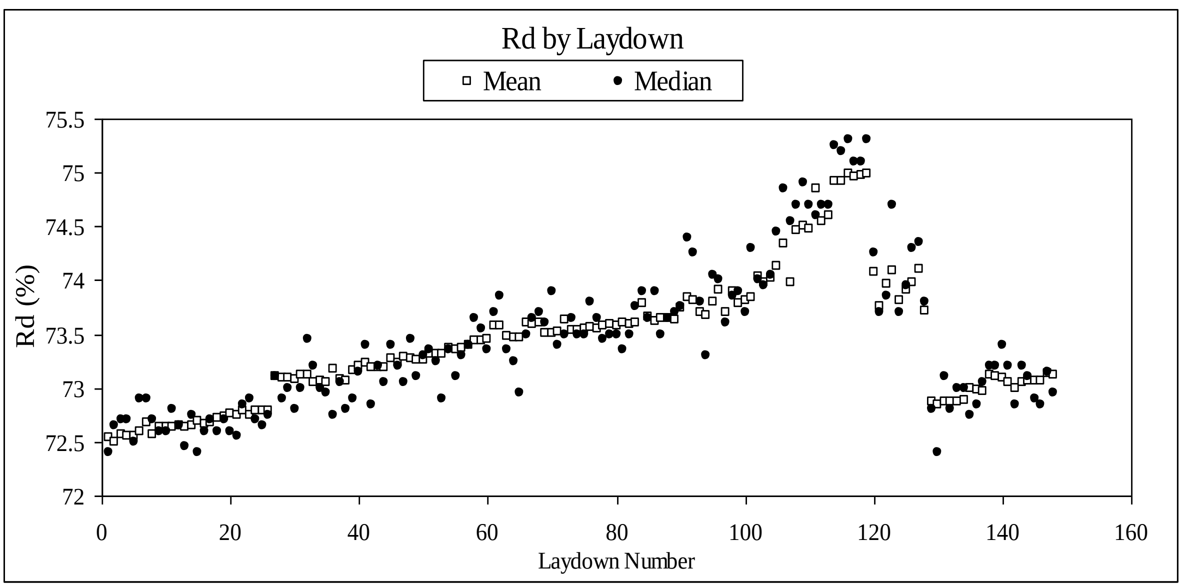
- Condenser (.956 R^2 Adj)
- Card Chute (.969 R^2 Adj)
- Breaker Sliver (.871 R^2 Adj)
- Roving (.833 R^2 Adj)

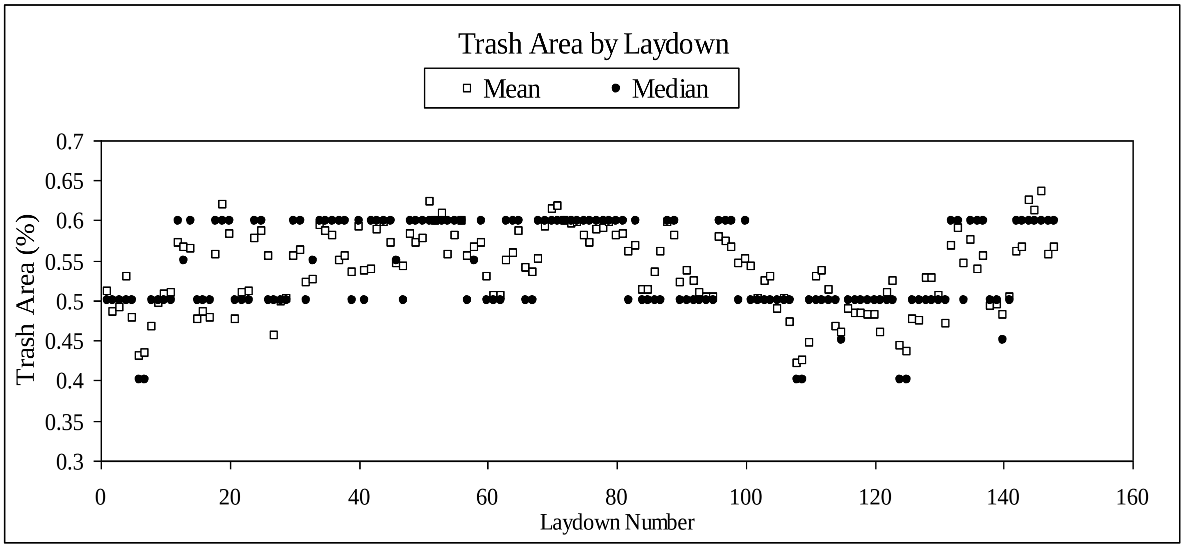
Card Sliver

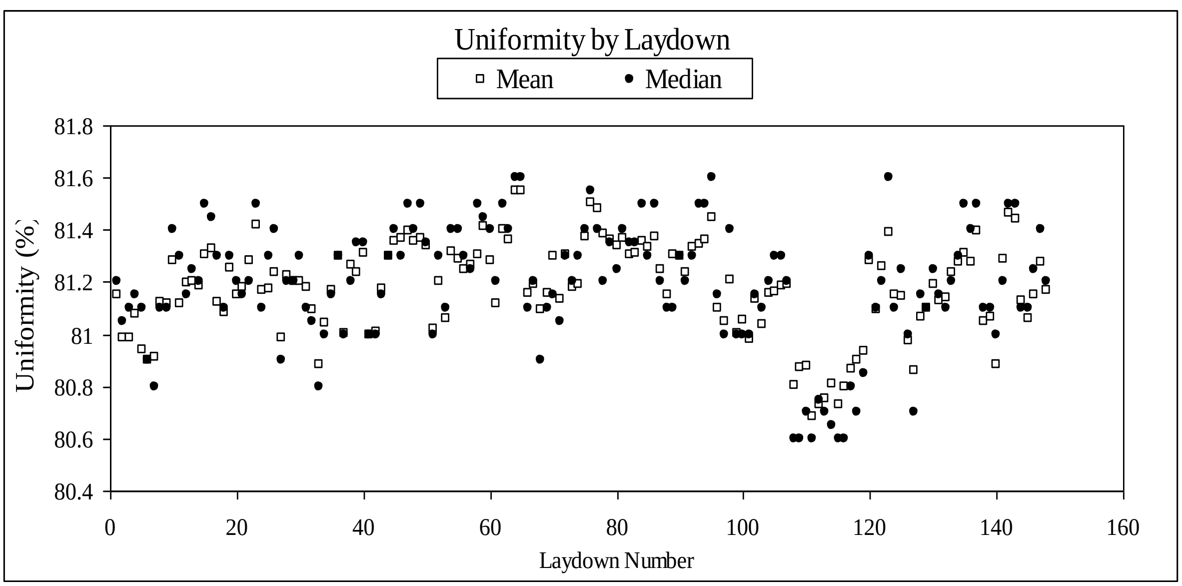
- Finisher Sliver (.687 R^2 Adj)

Appendix C - Mean and Median Laydown Charts






Trash Area by Laydown
Mean
Median
Trash Area (%)
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0
20
40
60
80
100
120
140
160
Laydown Number


Uniformity by Laydown
Mean
Median
Uniformity (%)
81.8
81.6
81.4
81.2
81
80.8
80.6
80.4
0
20
40
60
80
100
120
140
160
Laydown Number

Appendix D - Yarn Evenness

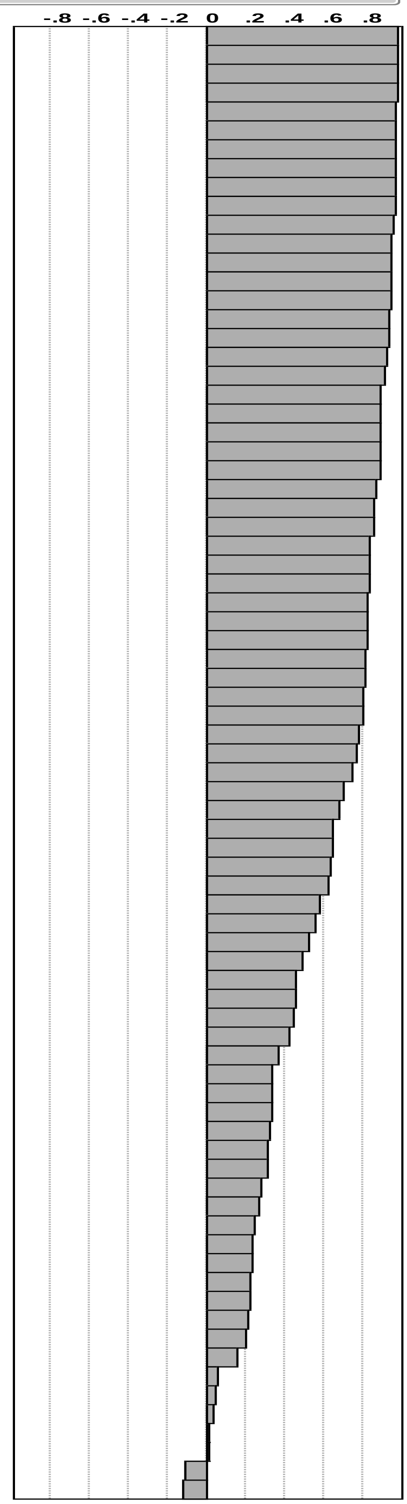
Yarn Evenness Measurements and Descriptions (adapted from Uster.com, 2009)

Measuremnent	**Description**
CV Value	**Mass coefficient of variation for 100 meter sample length**
CVm 1%	**Mass coefficient of variation for 1 meter sample length**
CVm 3%	**Mass coefficient of variation for 3 meter sample length**
CVm 10%	**Mass coefficient of variation for 10 meter sample length**
Thin -30% /kyd	**Count of places 30% thinner than the average diameter**
Thin -40% /kyd	**Count of places 40% thinner than the average diameter**
Thin -50% /kyd	**Count of places 50% thinner than the average diameter**
Thick +35% /kyd	**Count of places 35% thicker than the average diameter**
Thick +50% /kyd	**Count of places 50% thicker than the average diameter**
Thick +70% /kyd	**Count of places 70% thicker than the average diameter**
Neps +140% /kyd	**Count of neps 140% thicker than the average diameter**
Neps +200% /kyd	**Count of neps 200% thicker than the average diameter**
Neps +280% /kyd	**Count of neps 280% thicker than the average diameter**

Data Set 1

Pairwise Correlations

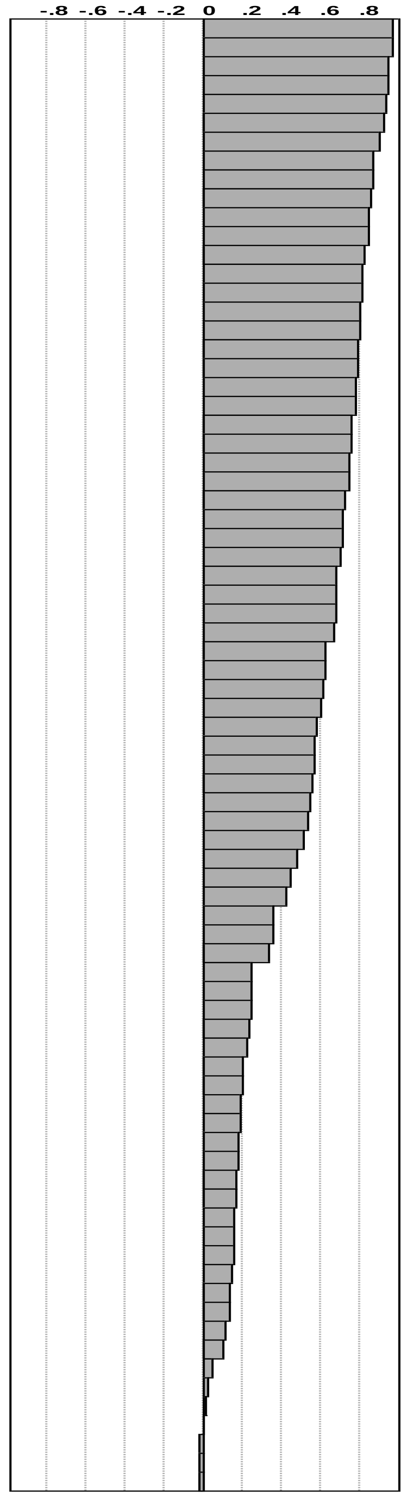
Variable	by Variable	Correlation	Signif Prob
CVm 3 %	CVm 1 %	0.9884	<.0001*
Thick +50% /kyd	Thin -40% /kyd	0.9808	<.0001*
CV Value	Thin -30% /kyd	0.9799	<.0001*
Neps +140% /kyd	Neps +200% /kyd	0.9792	<.0001*
Thick +35% /kyd	Thick +50% /kyd	0.9774	<.0001*
Thick +35% /kyd	Thin -30% /kyd	0.9748	<.0001*
CV Value	Thick +35% /kyd	0.9732	<.0001*
Thin -30% /kyd	Thin -40% /kyd	0.9728	<.0001*
Thick +70% /kyd	Thin -50% /kyd	0.9690	<.0001*
Thin -40% /kyd	Thin -50% /kyd	0.9689	<.0001*
Thick +35% /kyd	Thin -40% /kyd	0.9608	<.0001*
Thick +50% /kyd	Thick +70% /kyd	0.9541	<.0001*
Thick +50% /kyd	Thin -30% /kyd	0.9536	<.0001*
Thick +50% /kyd	Thin -50% /kyd	0.9508	<.0001*
Thick +70% /kyd	Thin -40% /kyd	0.9455	<.0001*
CVm 10 %	CVm 3 %	0.9406	<.0001*
Neps +200% /kyd	Neps +280% /kyd	0.9340	<.0001*
CV Value	Thin -40% /kyd	0.9284	<.0001*
CV Value	Thick +50% /kyd	0.9215	<.0001*
Thick +35% /kyd	Thick +70% /kyd	0.8958	<.0001*
CVm 10 %	CVm 1 %	0.8958	<.0001*
Neps +140% /kyd	Neps +280% /kyd	0.8942	<.0001*
Thick +35% /kyd	Thin -50% /kyd	0.8902	<.0001*
Thin -30% /kyd	Thin -50% /kyd	0.8900	<.0001*
Thick +70% /kyd	Thin -30% /kyd	0.8755	<.0001*
Neps +140% /kyd	Thin -30% /kyd	0.8592	<.0001*
Neps +140% /kyd	Thin -40% /kyd	0.8589	<.0001*
Neps +200% /kyd	Thin -50% /kyd	0.8403	<.0001*
Neps +200% /kyd	Thin -40% /kyd	0.8374	<.0001*
Neps +200% /kyd	Thick +70% /kyd	0.8352	<.0001*
CV Value	Thin -50% /kyd	0.8261	<.0001*
Neps +140% /kyd	Thick +50% /kyd	0.8229	<.0001*
Neps +140% /kyd	Thin -50% /kyd	0.8225	<.0001*
CV Value	Thick +70% /kyd	0.8199	<.0001*
Neps +200% /kyd	Thick +50% /kyd	0.8133	<.0001*
Neps +140% /kyd	Thick +70% /kyd	0.8081	<.0001*
Neps +200% /kyd	Thin -30% /kyd	0.8056	<.0001*
Neps +140% /kyd	Thick +35% /kyd	0.7838	<.0001*
CV Value	Neps +140% /kyd	0.7729	<.0001*
Neps +200% /kyd	Thick +35% /kyd	0.7463	<.0001*
CV Value	Neps +200% /kyd	0.7062	<.0001*
Neps +280% /kyd	Thick +70% /kyd	0.6784	<.0001*
Neps +280% /kyd	Thin -50% /kyd	0.6464	<.0001*
Neps +280% /kyd	Thin -40% /kyd	0.6464	<.0001*
Neps +280% /kyd	Thin -30% /kyd	0.6375	<.0001*
Neps +280% /kyd	Thick +50% /kyd	0.6270	<.0001*
CVm 10 %	CV Value	0.5863	<.0001*
Neps +280% /kyd	Thick +35% /kyd	0.5613	<.0001*
CV Value	Neps +280% /kyd	0.5252	<.0001*
CVm 10 %	Thin -30% /kyd	0.4957	0.0002*
CVm 1 %	CV Value	0.4545	0.0006*
CVm 3 %	CV Value	0.4537	0.0006*
CVm 10 %	Thick +35% /kyd	0.4518	0.0007*
CVm 10 %	Thin -40% /kyd	0.4245	0.0015*
CVm 10 %	Thick +50% /kyd	0.3690	0.0065*
CVm 10 %	Thin -50% /kyd	0.3391	0.0130*
CVm 3 %	Thin -30% /kyd	0.3330	0.0148*
CVm 1 %	Thin -30% /kyd	0.3300	0.0158*
CVm 1 %	Thick +35% /kyd	0.3272	0.0168*
CVm 10 %	Thick +70% /kyd	0.3179	0.0204*
CVm 3 %	Thick +35% /kyd	0.3146	0.0218*
CVm 1 %	Thin -40% /kyd	0.2821	0.0407*
CVm 3 %	Thin -40% /kyd	0.2715	0.0492*
CVm 1 %	Thick +50% /kyd	0.2509	0.0699
CVm 1 %	Thin -50% /kyd	0.2380	0.0862
CVm 3 %	Thick +50% /kyd	0.2303	0.0971
CVm 1 %	Thick +70% /kyd	0.2223	0.1097
CVm 10 %	Neps +140% /kyd	0.2195	0.1143
CVm 3 %	Thin -50% /kyd	0.2133	0.1252
CVm 3 %	Thick +70% /kyd	0.1993	0.1525
CVm 10 %	Neps +200% /kyd	0.1559	0.2650
CVm 3 %	Neps +140% /kyd	0.0530	0.7064
CVm 1 %	Neps +140% /kyd	0.0495	0.7248
CVm 10 %	Neps +280% /kyd	0.0331	0.8137
CVm 1 %	Neps +200% /kyd	0.0132	0.9253
CVm 3 %	Neps +200% /kyd	0.0075	0.9572
CVm 3 %	Neps +280% /kyd	-0.1092	0.4365
CVm 1 %	Neps +280% /kyd	-0.1200	0.3920



Data Set 2

Pairwise Correlations

Variable	by Variable	Correlation	Signif Prob
CVm 3 %	CVm 1 %	0.9753	0.0000*
Neps +200% /kyd	Neps +140% /kyd	0.9679	0.0000*
Thin -40% /kyd	Thin -30% /kyd	0.9535	0.0000*
Thin -30% /kyd	CV Value	0.9490	0.0000*
Thick +35% /kyd	CV Value	0.9396	0.0000*
Thick +50% /kyd	Thick +35% /kyd	0.9291	0.0000*
Thin -40% /kyd	CV Value	0.9044	0.0000*
Thick +35% /kyd	Thin -30% /kyd	0.8728	0.0000*
Thick +50% /kyd	CV Value	0.8696	0.0000*
Neps +280% /kyd	Neps +200% /kyd	0.8637	0.0000*
Neps +140% /kyd	Thin -30% /kyd	0.8545	0.0000*
Thin -50% /kyd	Thin -40% /kyd	0.8493	0.0000*
Neps +140% /kyd	CV Value	0.8242	0.0000*
Thick +35% /kyd	Thin -40% /kyd	0.8173	0.0000*
Neps +200% /kyd	Thin -30% /kyd	0.8108	0.0000*
Neps +200% /kyd	CV Value	0.8096	0.0000*
Neps +200% /kyd	Thick +35% /kyd	0.8003	0.0000*
Neps +140% /kyd	Thick +35% /kyd	0.7915	0.0000*
Neps +280% /kyd	Neps +140% /kyd	0.7889	0.0000*
Neps +200% /kyd	Thick +50% /kyd	0.7824	0.0000*
Thick +50% /kyd	Thin -30% /kyd	0.7769	0.0000*
Neps +140% /kyd	Thin -40% /kyd	0.7612	0.0000*
Thin -50% /kyd	Thin -30% /kyd	0.7606	0.0000*
CVm 10 %	CVm 3 %	0.7466	<.0001*
Neps +140% /kyd	Thick +50% /kyd	0.7438	<.0001*
Thick +50% /kyd	Thin -40% /kyd	0.7239	<.0001*
Neps +200% /kyd	Thin -40% /kyd	0.7206	<.0001*
Neps +280% /kyd	Thick +50% /kyd	0.7152	<.0001*
Thin -50% /kyd	CV Value	0.7051	<.0001*
Thick +70% /kyd	Thick +50% /kyd	0.6854	<.0001*
Neps +280% /kyd	Thick +35% /kyd	0.6785	<.0001*
CVm 10 %	CVm 1 %	0.6775	<.0001*
Neps +280% /kyd	CV Value	0.6708	<.0001*
Neps +280% /kyd	Thin -30% /kyd	0.6290	<.0001*
Thick +35% /kyd	Thin -50% /kyd	0.6217	<.0001*
Thick +70% /kyd	Thick +35% /kyd	0.6201	<.0001*
Neps +280% /kyd	Thick +70% /kyd	0.6084	<.0001*
Thick +70% /kyd	CV Value	0.5836	<.0001*
Neps +200% /kyd	Thick +70% /kyd	0.5685	<.0001*
Neps +140% /kyd	Thin -50% /kyd	0.5653	<.0001*
Neps +280% /kyd	Thin -40% /kyd	0.5557	<.0001*
Thick +50% /kyd	Thin -50% /kyd	0.5518	<.0001*
Neps +200% /kyd	Thin -50% /kyd	0.5408	<.0001*
Neps +140% /kyd	Thick +70% /kyd	0.5122	<.0001*
Thick +70% /kyd	Thin -30% /kyd	0.4837	<.0001*
Thick +70% /kyd	Thin -40% /kyd	0.4461	<.0001*
Neps +280% /kyd	Thin -50% /kyd	0.4295	<.0001*
Thick +70% /kyd	Thin -50% /kyd	0.3552	<.0001*
CVm 1 %	CV Value	0.3527	<.0001*
CVm 3 %	CV Value	0.3388	<.0001*
Neps +280% /kyd	CVm 1 %	0.2512	<.0001*
Thick +70% /kyd	CVm 1 %	0.2482	<.0001*
Neps +280% /kyd	CVm 3 %	0.2414	<.0001*
Thick +70% /kyd	CVm 3 %	0.2364	<.0001*
Thick +50% /kyd	CVm 1 %	0.2197	<.0001*
Thick +50% /kyd	CVm 3 %	0.2060	<.0001*
Neps +200% /kyd	CVm 1 %	0.2026	<.0001*
Thick +35% /kyd	CVm 1 %	0.1935	<.0001*
Neps +200% /kyd	CVm 3 %	0.1895	<.0001*
Thick +35% /kyd	CVm 3 %	0.1814	<.0001*
Neps +140% /kyd	CVm 1 %	0.1741	<.0001*
Neps +140% /kyd	CVm 3 %	0.1635	<.0001*
Thin -30% /kyd	CVm 1 %	0.1626	<.0001*
Thin -40% /kyd	CVm 1 %	0.1611	<.0001*
Thin -40% /kyd	CVm 3 %	0.1538	<.0001*
Thin -30% /kyd	CVm 3 %	0.1533	<.0001*
Neps +280% /kyd	CVm 10 %	0.1439	<.0001*
Thin -50% /kyd	CVm 1 %	0.1343	<.0001*
Thin -50% /kyd	CVm 3 %	0.1306	<.0001*
CVm 10 %	CV Value	0.1123	<.0001*
Thick +70% /kyd	CVm 10 %	0.0993	<.0001*
Neps +200% /kyd	CVm 10 %	0.0472	0.0549
Thick +50% /kyd	CVm 10 %	0.0168	0.4951
Neps +140% /kyd	CVm 10 %	0.0072	0.7691
Thin -50% /kyd	CVm 10 %	0.0017	0.9455
Thin -40% /kyd	CVm 10 %	-0.0173	0.4826
Thin -30% /kyd	CVm 10 %	-0.0201	0.4139
Thick +35% /kyd	CVm 10 %	-0.0201	0.4128



Appendix E - Descriptive Statistics for Yarn Evenness

Data Set 1

Descriptive Statistics	CV Value	CVm 1%	CVm 3%	CVm 10%
Sample Size	53	53	53	53
Mean	14.570	4.304	3.590	2.542
Stdev	1.414	0.624017	0.607454	0.473309
Range	5.940	2.360	2.470	2.190
Minimum	12.030	3.430	2.730	1.820
25th Percentile (Q1)	13.605	3.805	3.100	2.150
50th Percentile (Median)	14.480	4.120	3.510	2.560
75th Percentile (Q3)	15.645	4.925	4.115	2.840
Maximum	17.970	5.790	5.200	4.010
95.0% CI Mean	14.180 to 14.959	4.132 to 4.476	3.423 to 3.757	2.412 to 2.673
Descriptive Statistics	**Thin -30% /kyd**	**Thin -40% /kyd**	**Thin -50% /kyd**	**Thick +35% /kyd**
Sample Size	53	53	53	53
Mean	2795.6	445.19	39.321	744.96
Stdev	1443.6	440.33	65.991	348.53
Range	5870	1789	270	1557
Minimum	706	22	0	244
25th Percentile (Q1)	1603	119.50	3	506
50th Percentile (Median)	2429	291	14	673
75th Percentile (Q3)	3724.5	652.50	51	942.50
Maximum	6576	1811	270	1801
95.0% CI Mean	2397.7 to 3193.5	323.82 to 566.56	21.131 to 57.510	648.90 to 841.03
Descriptive Statistics	**Thick +50% /kyd**	**Thick +70% /kyd**	**Neps +140% /kyd**	**Neps +200% /kyd**
Sample Size	53	53	53	53
Mean	90.755	4.774	1417.2	149.11
Stdev	76.221	6.387	1221.1	196.04
Range	363	30	4724	831
Minimum	16	0	289	6
25th Percentile (Q1)	40	1	546.50	31.500
50th Percentile (Median)	68	2	844	55
75th Percentile (Q3)	122.50	6	1708.5	149
Maximum	379	30	5013	837
95.0% CI Mean	69.746 to 111.76	3.013 to 6.534	1080.7 to 1753.8	95.079 to 203.15

Data Set 2

Descriptive Statistics	CV Value	CVm 1%	CVm 3%	CVm 10%
Sample Size	1655	1655	1655	1655
Mean	15.480	4.585	3.806	2.326
Stdev	0.391740	0.254005	0.255802	0.264538
Range	3.540	1.900	1.870	2.070
Minimum	14.320	3.860	3.100	1.600
25th Percentile (Q1)	15.220	4.410	3.630	2.140
50th Percentile (Median)	15.500	4.570	3.790	2.300
75th Percentile (Q3)	15.740	4.740	3.970	2.500
Maximum	17.860	5.760	4.970	3.670
95.0% CI Mean	15.461 to 15.499	4.573 to 4.597	3.793 to 3.818	2.314 to 2.339
Descriptive Statistics	**Thin -30% /kyd**	**Thin -40% /kyd**	**Thin -50% /kyd**	**Thick +35% /kyd**
Sample Size	1655	1655	1655	1655
Mean	3687.8	507.89	25.801	1041.3
Stdev	392.12	107.97	12.404	142.42
Range	3987	1627	303	1137
Minimum	2688	271	4	621
25th Percentile (Q1)	3438	440	19	955
50th Percentile (Median)	3700	505	24	1049
75th Percentile (Q3)	3931	566	31	1134
Maximum	6675	1898	307	1758
95.0% CI Mean	3668.9 to 3706.7	502.69 to 513.10	25.203 to 26.399	1034.5 to 1048.2
Descriptive Statistics	**Thick +50% /kyd**	**Thick +70% /kyd**	**Neps +140% /kyd**	**Neps +200% /kyd**
Sample Size	1655	1655	1655	1655
Mean	141.89	7.056	2714.8	326.94
Stdev	34.874	3.697	565.21	111.07
Range	255	25	4001	820
Minimum	46	0	1435	92
25th Percentile (Q1)	118	4	2312	245
50th Percentile (Median)	142	7	2695	319
75th Percentile (Q3)	165	9	3050	385
Maximum	301	25	5436	912
95.0% CI Mean	140.21 to 143.57	6.877 to 7.234	2687.5 to 2742.0	321.59 to 332.30