Abstract

ANDERSON, MARY ELIZABETH. *APPLICABILITY OF DATA MINING IN YARN MANUFACTURING. (Under the direction of Dr. George Hodge and Dr. William Oxenham.)*

Information technology has become a priority in many businesses today. Understanding the data collected in information systems has allowed companies to stay competitive in different industries. Data mining is a set of statistical techniques for discovering previously unknown trends and patterns in large datasets. Data mining is becoming a significant tool in today's competitive business world and helping business leaders to make more informed decisions.

The purpose of this research has been to understand the yarn manufacturing process, the data collected, and the different data collection systems in order to determine the potential application of data mining in cotton yarn spinning. Plant interviews and a case study with a cotton open-end spinning plant were conducted to understand how data mining could be used within yarn manufacturing. Data was collected from the case study plant and analyses were performed to determine relationships and trends.

This research provides an overview of the data collection requirements for textile spinning and the different data elements collected throughout the spinning process. A model of collection points and data elements is presented. The different data collection systems used for monitoring these elements are discussed as well as the quality of data being collected. Analyses are performed to determine the applicability of using data mining techniques on the many data sets to improve both process and product quality.

APPLICABILITY OF DATA MINING IN YARN MANUFACTURING

by
MARY ELIZABETH ANDERSON

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

2006

APPROVED BY:

Nancy Powell
Committee Member

Dr. George Hodge
Co-Chair of Advisory Committee

Dr. William Oxenham
Co-Chair of Advisory Committee

To my parents for their endless love and support

Biography

Mary Elizabeth Anderson was born on May 10, 1982 in Greenville, North Carolina. She moved to Wilson, NC at the age of 3 where she grew up and graduated from high school in the spring of 2000. She attended North Carolina State University where she earned her Bachelor of Science in Textile Technology with a concentration in design and graduated in 2004. Beth also studied abroad one semester of her undergraduate degree at the University of Manchester in Science and Technology located in Manchester, England. Beth was accepted into the Institute of Textile Technology program at NC State and began work on her master's degree in the fall of 2004. She worked as a design intern at Glen Raven Custom Fabrics during the summer of 2005.

Acknowledgements

I would like to thank my co-advisors, Dr. George Hodge and Dr. William Oxenham, for all their guidance and contribution throughout my graduate program. I would also like to thank Professor Nancy Powell for all her support through this research project as well as during my undergraduate degree. I would also like to thank Mr. Chris Moses who has been very helpful, understanding, and patient during this graduate program.

The author would like to thank the companies that were willing to participate in the research project. All the employees were very helpful and their contributions were valuable to the success of this project.

I would like to thank the Institute of Textile Technology and all the Textile Advisory Committee members for this amazing opportunity. The knowledge and skills learned in this program will be such an asset in all my future endeavors.

The author would especially like to thank all her family and friends who supported her throughout the graduate program. Their love and support were the foundation of her motivation to finish and do well. I feel so lucky to have such special people in my life and hope they know how much they are loved.

Table of Contents

List of Figures

List of Tables

1 Introduction

Information technology has become important in the success of today's fast paced business world. Understanding data collected in information systems could be a key to gaining a competitive advantage over companies in an industry. With the ever-changing textile industry and many manufacturers moving overseas, the data mining technology could lead to the competitive edge the industry needs by enabling business leaders to make more informed decisions.

Data mining is a set of statistical techniques for discovering previously undetected or unknown trends and patterns in large datasets. Data mining is becoming a significant tool in today's competitive business world and helping business leaders to make more informed decisions. It is currently applied in banking, marketing, insurance, fraud detection, telecommunications, and retail, but has not yet found wide applications in textiles. Exploring the potential use of data mining to analyze the data available in cotton spinning plants will provide a better insight into the key parameters affecting yarn quality and efficiency.

In modern cotton spinning plants a vast amount of data is available from routine online and offline testing and monitoring equipment. This research provides an overview of the different data elements collected throughout the spinning process, the different data systems used for monitoring these elements, the quality of data being collected, and determines the applicability of using data mining techniques on the many data sets to improve both process and product quality.

The specific objectives of this research are to:

1. Define data collection requirements for textile spinning
2. Define data quality issues for textile spinning
3. Identify which data elements are needed for monitoring and controlling product quality
4. Explore exchange of data between the different data analysis systems
5. Investigate relationship between process performance and final product quality

Applying data mining techniques in yarn manufacturing will enable spinners to make more effective use of existing resources and at the same time provide them the opportunity at improving both process and product quality. Should this technique prove successful in cotton spinning it may then be possible to extend the work to cover other areas such as knitting, weaving, and finishing where again it may be possible to identify those parameters that are responsible for possible fabric defects such as barré or unwanted dye stripes.

In Chapter 2 the literature review discusses the yarn manufacturing process. An overview of the different steps taken to produce yarn from cotton fibers is explained. An introduction to data mining is also provided, along with a few examples of where it is used how it has been applied in other industries. The different data mining techniques are presented as well as discussion of data quality and preparation.

Data mining is not commonly used in manufacturing and has been used primarily in industries with a strong customer focus such as retail and marketing. There has been little research in the application of data mining techniques in a manufacturing environment.

Chapter 3 states the research objectives and explains the tasks that are taken in order to meet the objectives. The research was conducted in three phases. A data mining process model is presented and the steps of this model are explained. Limitations to the research are discussed.

Chapter 4 discusses the results of completing Phase I and Phase II of the research. The tasks involved with the plant interviews are explained as well as planning steps for the next phase. A data model of collection points is presented.

Chapter 5 discusses the results of Phase III's case study and data analysis. Steps to collect and clean/prepare the data for analysis are explained. The data quality issues of the different data sets are presented as well as steps to resolve them. Data analysis was performed on specific data sets and results are explained.

Chapter 6 states conclusions from the research and recommendations for future studies in this field.

2 Literature Review

The textile spinning process is a series of steps that manufacturers yarn from individual fibers. In the yarn manufacturing process, there is data produced from offline testing, which monitors yarn quality information. Other data describing machine performance and quality are generated from most of the processing machinery and displayed on a machine panel or stored in a data collection system. Although there are several sources of data collection in the spinning process, it is evident that this data is often of poor quality and needs "cleaning" before any analyses are possible. Several studies have suggested that well over 80% of the effort in a data mining project is in the effort to clean/prepare and organize the data (Nisbet, 2004).

The main problems associated with manufacturing data include: time scale of data collection (establishing a timeline between data gathered at different parts of the processing operation); missing data; standardization of data entries (e.g. does 0 mean 0 value, or simply no data collected); and incompatibility of data from different measuring systems (Braha, 2001).

The technologies for generating and collecting data have been advancing rapidly. There is an increasing inability to extract useful information from these massive amounts of data. Poor data quality is costing U.S. businesses more than $600 billion annually (Dubois). Poor data quality results in deficient, ineffective data analyses. Technologies and tools enable process data to be automatically transformed into useful information, thus enhancing the quality of data.

Data mining, therefore, has become a research area with increasing importance. Many companies have recognized data mining as an important technique that will have an immediate impact on their performance. For manufacturers, making sense of the millions of bytes of data generated daily by many factory operations has in the past been a daunting task. Today this task is becoming far less challenging, due to advances in data mining software. Data mining searches for any valuable information that exists within the large volumes of available data. The process uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions (Two Crows, 1999).

In a textile spinning plant large amounts of data are generated during processing. Integrating these systems has been part of the efforts in computer-integrated manufacturing. Standards have been developed for communicating between the plant floor equipment and higher-level business analysis systems. Data profiles have recently been developed to standardize the terms and definitions used in textile manufacturing. However, vendors do not allow access to the raw data collected which they consider part of their proprietary systems. The integrated systems are useful for reporting and monitoring overall plant performance. However, these standards were not developed with applications like data mining in mind.

While there have been several studies to attempt to relate yarn properties back to specific fiber properties, there has not been many studies incorporating all available data elements in an attempt to identify the most significant parameters affecting different aspects of processing and yarn quality. There is no doubt that yarn strength is important and receives the most attention of researchers; however, other aspects such as spinning efficiency (including end break rates), carding speeds, and fiber loss also play dominant roles in modern competitive spinning plants.

2.1 Introduction to Cotton Yarn Manufacturing

The actual process sequence required in cotton yarn manufacturing is influenced by several factors. These factors include the initial raw material characteristics, the spinning system on which the yarns are to be spun, the specifications of the fabrics to be manufactured from the yarns, the fabrication methodology, and the end use planned for the resultant fabrics (McCreight, 1997). Although the process steps may vary by plant, there are common basic process functions that must be accomplished and are done so in the opening room, card room, and spinning room.

2.1.1 Opening Room

In cotton yarn manufacturing, bales of cotton are positioned together at the initial processing stage in a bale laydown. The number of bales in a laydown can range anywhere from several bales up to 100 bales depending on the plant. The smaller laydowns are typically used when running a polyester/cotton mix. The bales are conditioned for at least 24 hours in controlled temperature and relative humidity before placed in a laydown. Once the laydown is positioned, an opening machine will run back and forth across the bales picking fibers from the top. Cotton fiber tufts are opened which aids in the mixing and necessary cleaning of stock that must be accomplished from later processes. From the bale laydown, cotton fiber goes through a series of blending and cleaning machines (McCreight, 1997). Because of the significant variability in fiber properties within and between bales, these processes are important in cotton yarn manufacturing in order to produce a consistent product.

2.1.2 Card Room

Opened, mixed, and cleaned fiber tufts are transported by air ducts from the final stage of the Opening Room processes to a chute that prepares a fiber mat for feeding the card. The function of the chute is to form a continuous and even mat of small fiber tufts to be fed into the card (McCreight, 1997). The card separates, aligns, cleans, and mixes the fibers, producing sliver.

Once sliver has been formed, it goes to the next process of drawing. The purpose of drawing is to reduce the weight of fiber and therefore forcing them into a more parallel position. To compensate for the reduction in weight and to further mix the fibers, a drawing machine combines several slivers, usually six to eight from different machines, to form one sliver. The combining of different slivers is called doubling.

2.1.3 Spinning Room

The final process step in yarn manufacturing is the actual spinning of sliver into yarn. Rotor spinning involves the separation of fibers by vigorous drafting, and then recollection and twisting of the fibers in a rotor (McCreight, 1997). In open-end (or rotor) spinning, each spinning machine contains multiple rotor positions. Finisher sliver from the drawing frames is fed to opening rollers, where separation into individual fibers occurs. The fibers are then transported with air to the rapidly revolving rotors. Forces created from the high rotational speeds cause the fibers to collect along the wall of the rotor, forming a ring. Untwisted fibers are formed into a bundle and with each rotation of the rotor, are converted into yarn and pulled out of the rotor. The yarn is then automatically wound onto a package, eliminating the need for a separate process step. Rotor spinning production is much faster than ring spinning and also enables for larger packages due to the twist insertion by the rotors (McCreight, 1997).

2.2 Introduction to Data Mining

Business Intelligence (or BI) is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. Data mining is just one application of BI, and can be used when trying to improve process and product quality. Discovering unknown trends and patterns in large datasets increases the understanding of certain processes and therefore the ability to predict and affect the outcome. Data mining focuses on the detection and correction of data quality problems and the use of algorithms that can tolerate poor data quality. Some examples of techniques used in data mining are classification, estimation, and cluster analysis.

Data mining is an integral part of knowledge discovery in databases (KDD), which is the overall process of converting raw data into useful information (Tan, 2006). Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions (Two Crows, 1999). Data mining can help provide decision-making information about the future. This technology has emerged in response to a need from industry for effective and efficient analysis of large data sets. It has become increasingly important in today's business world, as it is being used to increase revenues and reduce costs, allowing many companies to stay competitive.

"Know your customers and give them what they want" is the fundamental principle of marketing (Shearer, 2004). Marketers cannot always know and anticipate customer thoughts or needs for the future. Many organizations are using data mining to help manage all phases of the customer life cycle, including acquiring new customers, increasing revenue from existing customers, and retaining good customers (Two Crows, 1999). Organizations generate and collect vast amounts of data in the ongoing process of doing business. The problem is knowing what to do with the data and knowing how to interpret the data. Data mining offers many advantages to a company. It provides information about business processes including the customer and market behavior; it takes advantage of data that may already be available in databases, data marts, and the data warehouse; and it provides patterns of behavior which can be seen in the data that can increase business knowledge and the ability to foresee and shape future events. The goal of data mining is to improve the quality of the interaction between the organization and their customers.

Although data mining is primarily used in areas with a strong customer focus, the interest has been growing among manufacturing companies across many industries regarding the potential of data mining for changing business performance. A collaborative effort of domain experts (designer, production manager), data exports (IT professionals), and data mining experts is essential to the success of the data mining integration within manufacturing environments. Successful implementation of the data mining process includes six important stages. The first stage involves understanding the problem and elements to which the data mining is applied. The second step includes the selection, integration, and checking of the target data that may be stored in various databases. The third step is data preprocessing which includes data transformation, handling of missing or unknown values, and data cleaning. In the fourth step, the actual data mining takes place involving model and hypothesis development, selection of appropriate data mining techniques, and extraction of desired data. The fifth step is the analysis and interpretation of any found patterns and relationships. In the final step, results are reported and a knowledge maintenance mechanism can be set up (Braha, 2001).

2.3 Data Warehouse

Much of the data mining literature starts with the assumption that a data warehouse exists or at least a collection of databases, such as may be found in an enterprise resource planning (ERP) or customer resource management (CRM) systems (Hodge, 2002). These techniques are useful where the data is available in a data warehouse of cleaned data. The barriers or limitations to collect and clean the data may be too high to produce this data warehouse if it is done as an addition to existing systems.

Usually the process of data mining begins with the extraction of data from the data warehouse into a data mining database or data mart. A data warehouse does not sort the data into useful trends, relationships or profiles; it is just a database filled with potential information (Schertel, 2002). The purpose of a data warehouse is to collect data from various sources throughout an organization and make it available to those who need it. The method of data warehousing is known as structured query language (SQL) tools (Agosta, 2004).

A data warehouse combines databases across an entire enterprise, whereas data marts are usually smaller and focus on data for one division or one workgroup within an enterprise ("The business intelligence"). For example, a geographic data mart might contain geographic-based data for demographic analysis (Henderson, 1998). A data warehouse gathers all available information and then specialized data marts are created with subsets of this information. These data marts are easier to use because they only have the particular information the specific user group needs ("The business intelligence"). Figure 2.1 shows how a data warehouse can be divided into different data marts for different subjects. Data that is stored in a data warehouse has usually already been cleaned and is easier to use when data mining. Setting up a large data warehouse resolves data integrity problems. However, data can be mined from the data sources.

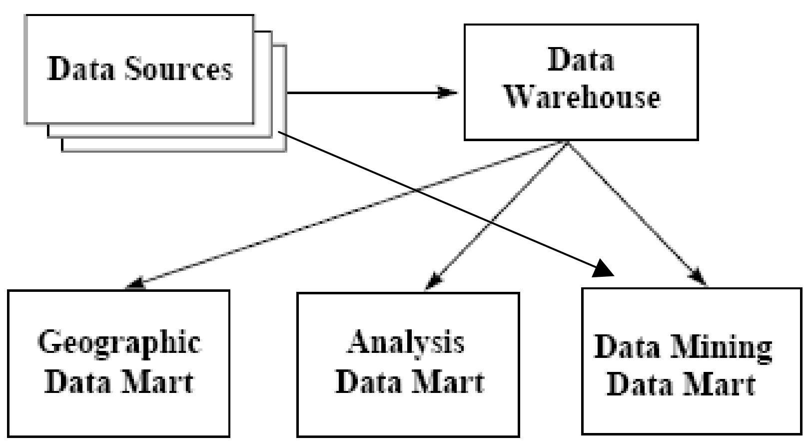


Figure 2.1: Data mining data mart extracted from a data warehouse (Two Crows, 1999)

Although data warehouses are popular for data mining applications, they are not imperative since most software today can extract, clean, and filter the collected data. The data warehouses may contain valuable external data, such as regulations, demographic, or geographic data, which when combined with internal organizational data offers a firm foundation for data mining (Moss, 2003).

2.4 Data Mining and OLAP

All databases provide a variety of query tools for users to access information stored in the database. In dimensional databases, the query tools are online analytical processing (OLAP) tools. The tools allow users to look at multiple views of data to answer specific questions. Finding answers from data in OLAP is similar to calculations in a spreadsheet. OLAP tools are basically visualization tools that can help users learn patterns in the data. The tools do not learn anything from data, whereas data mining tools obtain answers by learning the relationships between different attributes of database records. OLAP tools are useful for data mining because of their capabilities to visualize relationships between different data dimensions, however they are not a substitute for data mining (Braha, 2001).

2.5 Data Quality

Data quality is a challenge for any organization. Quality data must consist of information that a company needs to be able to make informed decisions. Improving data quality improves the quality of the resulting analysis. Data can never be perfect and problems can include human error, limitations of measuring devices, or flaws in the data collection process (Tan, 2006). There are many different factors that influence the quality of data. Issues that often need to be addressed include the presence of outliers; missing, inconsistent, or duplicate data; and data that is biased or unrepresentative of the sample (Kuonen, 2005). It is also important to know how the data was collected, if the collection times were consistent, the format of the data, and how it is stored. Preprocessing steps are taken to transform raw input data into an appropriate format for analysis.

2.6 Data Preparation

Data preparation consumes 60 to 90% of the time needed to mine data and contributes 75 to 90% to the mining project's success. Poor or nonexistent data preparation can be 100% responsible for a project's failure (Ye, 2003). Data preparation involves obtaining the most value out of the available data; while data mining is discovering any meaningful patterns in the data. Data preparation is not an optional activity as it is necessary as a mining tool to see some of the patterns. Data preparation starts with actually collecting the data and making sure to select the right data. Four fundamental principles involved in selecting the right data are select relevant elements, choose redundant elements, select records randomly from the source data, and ensure that the records represent the full range of within-variable and between-variable behaviors (Ye, 2003).

Condition refers to the state of readiness of the data for analysis. Data that require minimal time and cost to clean before reliable analysis can be performed are well conditioned. Data that involve a substantial amount of time and cost are ill conditioned. Smaller amounts of data are typically cleaner, and thus better conditioned. Large amounts of data are an outgrowth of today's digital environment, which generates data flowing continuously from all directions at unprecedented speed and volume, and which almost always require cleansing. Data cleaning is used to ensure that the data are of a high quality and contain no duplicate values. The data-cleaning process involves the detection and possible elimination of incorrect and missing values.

2.7 Data Mining vs. Statistics

Data mining differs from other methods of analysis mainly in the approach used to explore the data. Many analytical tools are "user driven" and support a verification-based approach. This is where the user hypothesizes about the data relationships and then uses tools to support or disprove those assumptions. Data mining is "data driven" and uses a discovery-based approach where algorithms are utilized to determine the key relationships in the data. Other differences include all of the following: statisticians develop their own equations to match their hypothesis, whereas data mining algorithms in the tool can automatically develop the equations; statistical analysis uses only numerical data, but data mining tools can use different types of data and are not limited to numerical data; statisticians can find and filter dirty data during their analysis, whereas data mining depends on clean, well-documented data; and statisticians interpret their own results while data mining results are not as easy to interpret and usually need the help of a statistician (Moss, 2003).

2.8 Technology Hierarchy

In the technology hierarchy, data warehousing is generally considered an architecture for data management. When implemented, a data warehouse is a database that provides information from across a company. Data mining is a process for knowledge discovery. Predictive analysis is an application that builds on these two predecessor technologies, using the data to predict future trends. Statistical processing has been useful in data preparation, model construction, and validation, however predictive analytics used to obtain knowledge from the model, are then implemented in business applications. In general, tools in predictive analytics employ methods to identify and relate independent and dependent variable to provide a pattern and a model for the behavior of the future variables (Agosta, 2004).

2.9 Data Mining Techniques

Data mining techniques have been useful in a wide variety of industries. Telecommunications and credit card companies apply data mining to detect fraudulent use of their services, insurance companies use techniques to reduce fraud, and the effectiveness of surgical procedures and medical tests can be predicted. Financial companies can determine market and industry characteristics as well as predict individual company and stock performance, and retailers can decide which products to stock in particular stores and the determine the effectiveness of promotions and coupons (Two Crows, 1999). Data mining techniques have been applied successfully in many areas, however it is not as commonly used in manufacturing.

Knowledge discovery is a type of data mining. Knowledge discovery can either be directed, where the task is to explain the value of a selected target field, or undirected, where there is no target field and the computer just identifies patterns in the data that may be significant (Berry, 1997). Directed knowledge discovery, the most common type, is a way of using the past to build a model of the future; trying to learn from past mistakes. These techniques incorporate past experience to manipulate data in order to yield more useful results. Classification, estimation, and prediction are types of directed data mining. Undirected knowledge discovery is used as a way of generating unknown relationships that can be verified using more directed methods. Affinity grouping and association, clustering, and description and visualization are types of undirected data mining (Berry, 2000).

Companies today collect and refine substantial quantities of data. Data mining techniques can be implemented to analyze these massive databases and deliver answers to questions that will aid in the prediction of trends (Thearling, 2004). The basis of data mining is to build a model for one situation in which the answer is known and apply it in a situation where the answer is unknown. Data mining techniques build such models and allow for prediction of future events.

2.9.1 Classification

Classification is the most common data mining task and consists of examining the features of a newly presented object and assigning it to one of a predefined discrete (or categorical) set of classes (Berry, 1997). The purpose is to build some sort of model to be used on unclassified data in order to classify it. Decision trees and memory-based reasoning are two examples of classification.

Decision trees are tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Decision trees are one of the most popular methods of predictive modeling for data mining purposes because they provide rules and logic statements that enable more intelligent decision-making. A decision tree partitions data into smaller segments called terminal nodes or leaves. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Memory-based reasoning uses known instances as a model to make predictions about unknown instances. This can be done by looking for the nearest neighbors in the known instances and combining their values to assign classification or prediction values. The two key elements in memory-based reasoning are the distance function used to find the nearest neighbors and the combination function that combines values to make a prediction (Berry, 1997).

2.9.2 Estimation

Estimation is often used to perform a classification task. Given some input data, estimation is used to come up with a value for some unknown continuous variable (Berry, 1997). Neural networks are an example of estimation tasks.

Artificial neural networks, one of the most common data mining techniques, are non-linear predictive models that learn through training and resemble biological neural networks in structure. They are simple models of neural interconnections in brains, adapted for use on digital computers. One of the chief advantages of neural networks is their wide applicability. However, there are two major drawbacks. The first is the difficulty in understanding the models they produce. The second is their particular sensitivity to the format of incoming data (Berry, 1997).

2.9.3 Prediction

Prediction is the same as classification or estimation except that the records are classified according to some predicted future behavior or estimated future value (Berry, 1997). The only way to validate the prediction is to wait and see the results. This is a way of using the past to learn about the future. Popular techniques falling in this area are market basket analysis, memory-based reasoning, decision trees, and artificial neural networks. Genetic algorithms can be used to support memory-based reasoning and neural networks.

The task of affinity grouping is to determine which things go together (Berry, 1997). An example of this is thinking what items would be grouped together in a shopping cart at the grocery store. This is how the term “market basket analysis” can be best understood. Association rules can be used to specify the affinity grouping. A typical market basket analysis would look at all combinations of products, and characterize them based on several statistics (Coppock, 2003). This technique is popular in the retail industry to follow customer patterns on their purchases.

Genetic algorithms are optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution. Genetic algorithms are similar to statistics, in that the form of the model needs to be known in advance. Genetic algorithms can be used to solve the same types of problems as other data mining techniques, as well as to enhance memory-based reasoning and neural networks (Berry, 1997).

2.9.4 Clustering

Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or clusters (Berry, 1997). Clustering is basically data grouping or partitioning. It is similar to classification except that there are no predefined classes, just groupings based on the similarities found in the data mining tool. Within a cluster the members should be very similar, but the clusters themselves should be very dissimilar. Clustering is used for problems such as detecting manufacturing defects or finding affinity groups for credit cards (Moss, 2003).

2.9.5 Description and Visualization

Sometimes the purpose of data mining is simply to describe what is going on in a complicated database (Berry, 1997). A good description can lead to finding an explanation. Visualization is using visual methods to describe what is going on in the database or a particular process.

Data visualization is the visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships. Three components are essential for understanding a visual model of the data. Representation is basically the visual form in which the model appears. A high-quality representation displays the model in terms of visual components that are already familiar to the user. Interaction is the ability to view the model in such a way that the user can "play" with the model. The user can see the forecasting results from different scenarios. Integration is the ability to display relationships between the model and alternative views of the data, providing the user with a holistic view of the data mining process. Interaction is essential, as the user should be able to interact with the data to discover information hidden in it.

In a recent KDnuggets poll of "data mining/analytic techniques you use frequently", out of 784 votes decision trees/rules was at the top with 14%, followed by clustering with 13%, then regression with 11%, statistics with 10%, visualization and neural nets with 8%, and association rules with 7% ("Kdnuggets poll", 2005).

2.10 Data Mining Software

The popularity of data mining has continued and will continue to grow (Beal, 2005). The amount of data that require processing is becoming too much for human analysts to handle. Human analysts with no special tools can no longer make sense of enormous volumes of data that require processing in order to make informed business decisions ("Data mining"). Computers are designed to handle the enormous amounts of data and determine trends and relationships that otherwise would not have been found. Computers are commonly used to produce effortlessly precise calculations. Using computers is also cheaper than hiring trained professional statisticians and can produce results quicker. While data mining does not eliminate human participation in solving the task completely, it significantly simplifies the job. Advances in database technology, computer processors, and the push towards electronic commerce have meant that our ability to generate, collect, and store data has outstripped our ability to discern new and valuable information from existing data.

According to the IDC, a leading market research firm, the worldwide market for data mining solutions is estimated at $539 million in 2002 and is expected to continue increasing to $1.85 billion in 2006. Today, SAS Institute, IBM, SPSS, Microsoft, and Oracle are the leading vendors of the data mining products ("Data mining industry," 2005). When it comes to market leadership in data mining, Cary, N.C.'s SAS Institute is number one, followed by Chicago's SPSS Inc., then the major database companies (Beal, 2005). Two of the most significant challenges driving the changes in data mining are scalability and performance. Organizations want data mining to become more powerful so they can, for example, analyze and compare multiple data sets, instead of the traditional individual large data sets. They also want to break up data into smaller and smaller categories for analysis. Scalability is critical because databases continue to increase in size. Terabyte-class databases have become more common today, particularly as the cost of storage has decreased and the amount of e-commerce has increased. As companies collect more data, managing and mining the information becomes more complex (Leavitt, 2002).

2.11 Previous Research on Data Mining in Yarn Manufacturing

A PhD dissertation by Stacey Schertel entitled *Data Mining and its Potential use in Textiles: A Spinning Mill* under the direction of Dr. George Hodge and Dr. William Oxenham, of North Carolina State University, established a model for investigating data mining in textile manufacturing and identified issues in data collection and data quality. In this dissertation, a case study was done with one spinning plant, where data was collected and mined using the SAS Enterprise Miner.

Schertel's dissertation was helpful to identify limitations to this research as well as possible opportunities. One major limitation of Dr. Schertel's study was not having enough data to analyze. As is discussed later, there are many possible data collection points throughout the spinning process. The model of collection points in Schertel's research was useful to visually represent the data collected throughout the process and where. This model is shown and discussed later in section 4.2.1

Dr. Schertel's study is the only known trial of applying data mining software in textile spinning. The process and results were very helpful in this research as a guide and understanding of the data elements.

3 Research Methodology

This chapter states the research objectives, how these objectives are met, and discusses the limitations of the research. The research objectives are:

RO1: Define data collection requirements for textile spinning

RO2: Define data quality issues for textile spinning

RO3: Identify which data elements are needed for monitoring and controlling product quality

RO4: Explore exchange of data between the different data analysis systems

RO5: Investigate relationship between process performance and final product quality

3.1 Objectives and Overview of the Research

The purpose of this research is to determine the applicability of data mining techniques in a textile manufacturing environment. This research is divided into the five research objectives mentioned above. The applicability of data mining is not as simple as just using the software on data. The manufacturing process needs to be understood as well as the data, data systems, and data storage. There are so many factors to be considered before the data mining steps can be implemented. These entail the observation of data collection in spinning plants, including the data parameters and the storage systems; understanding the current level of analysis on existing data and limitations due to data quality; and understanding the accessibility of data on the machines, between the machines, and between the different data systems.

Some more specific questions that are answered in order to meet the research objectives are:

- What data is collected throughout the yarn manufacturing process?
- What data is available at each process?
- What is the accessibility of the data?
- What is the quality of data collected?
- What is the level of data analysis in a plant?
- Where does opportunity exist for data mining?

This research will discuss visits and interviews with four textile spinning companies, as well as an in depth case study with one cotton yarn spinning plant. The information obtained from the four interviews was helpful in the preparation for the case study. The goal of the plant interviews and case study were to provide insights and understanding of the yarn manufacturing process, data collection systems, data elements, data storage, and opportunities for data analysis. During the case study, data collected from online and offline sources was compiled and analyzed to determine any data quality issues and to uncover possible trends and relationships in the spinning process.

3.2 Experimental Procedure

This research is divided into three sections, Phase I, Phase II, and Phase III. Phase I consisted of gathering information from secondary sources, setting up plant interviews, developing interview questions, and interviewing Cotton Incorporated about their software and its uses in yarn manufacturing. The purpose of Phase I is to provide an initial understanding of the yarn manufacturing process and to prepare for the plant interviews.

Phase II consisted of gathering information from primary sources, such as spinning plant visits, interviews, and preparation of interview questions for the case study. The purpose of the interviews in Phase II, as well as any information learned from Phase I, is to define data collection requirements (RO1) and data quality issues for textile spinning (RO2) and explore the accessibility of data between the different systems (RO4).

Phase III consisted of a case study with one cotton yarn spinning plant, analysis of the process stages and different data systems, collection of data, and data analysis. The importance of conducting a case study with one plant is to define data quality issues for textile spinning (RO2), identify which data elements are needed for monitoring and controlling product quality (RO3), explore the accessibility of data between the different collection systems (RO4), and investigate the relationship between process performance and final product quality (RO5).

3.3 Case Study

Before visiting the company for a case study, interview questions were prepared as well as process and data flow diagrams, and an organized approach to collecting information was formed. An initial walk through of the plant was useful to observe equipment and any monitoring systems. A process flow diagram was constructed based on the particular plant layout. Interviews with managers and technicians revealed the amount of data collected, where it is collected, how it is collected, and what analysis is done with the data. Based on this information, a data flow diagram was completed to visually represent and understand the different data collection points and data storage systems. For collecting the actual data, it was helpful to be prepared with some means of storing large amounts of data such as compact discs or flash drives. Data was collected from several different process steps and from different data collection and/or storage systems.

3.3.1 Data Analysis

Once the interviews, observations, and data collection tasks had been completed for the case study, data analysis could be carried out. Before the analysis of data, it is important to understand the manufacturing process and data elements. This understanding is the basis to the failure or success of actually data mining in the manufacturing environment. Data mining is rarely used in manufacturing and in order to find relationships between the different process elements, the lag times must be determined. Lag times are the times between each process in yarn manufacturing. Understanding the process and data flows helps determine the lag times.

Data mining, as seen from the literature review, involves many steps other than actually mining. The model shown in Figure 3.1 and obtained from Braha shows a structured six step methodology outlining the different steps and stages of the data mining process. The arrows going down show the order of the stages. The dotted arrows going up show that at the particular stage, it is possible to go back a previous stage if needed. This model was used for data analysis to investigate relationships between process performance and product quality.

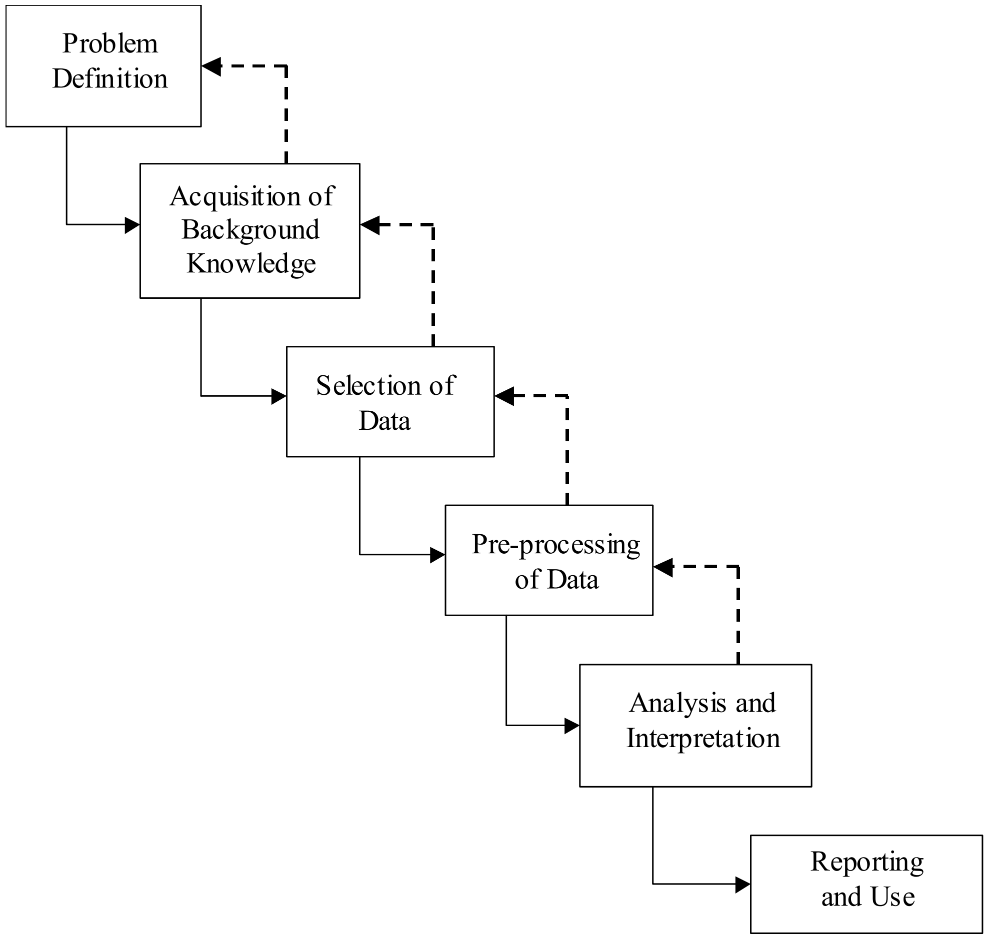


Figure 3.1: Data mining process model (Braha, 2001)

The problem definition phase is the first stage of the data mining process. This stage identifies the objectives of using data mining on a problem. For this study, the problem definition is to determine any useful trends and relationships between data collected in yarn manufacturing and process and product quality. Acquisition of background knowledge is the exploration of secondary literature. This stage is critical because prior information can provide limitations of the data and any already known relationships with the data. The next stage is selection of data that will be used and analyzed to give an answer to the problem under consideration. In this case, online and offline data from the yarn spinning process will be collected. Pre-processing of data is important since data usually needs to be cleaned before it can be mined. This stage was especially important since the data to be gathered would come from different systems, and possibly be in different formats. Once the data has been collected and cleaned, the actual mining can take place. This is the analysis and interpretation stage. The purpose of this stage is to find certain patterns, similarities, and other interesting relations between the available data. Once results of the data mining are determined, then they are reported and used.

3.4 Limitations of the Research

There are several possible limitations to this research. One limitation is that only four spinning plants are interviewed. This is a very small sample size, but due to time constraints, is the most feasible. This sample is used as the representative sample for all cotton spinning plants. There is also only one in depth case study that is performed. This means the data collected is limited to their specific data collection systems and methods.

Another limitation is the amount and quality of data collected. Data mining is typically performed on large data sets. It is known that data is collected during the spinning process, but in order to use this data to extract any meaningful trends, there must be a lot of data and the data must be of good quality. It is not known what data can be obtained, the format of the data, or from which data collection systems. Accessing data to be used in analyses could prove to be very difficult.

When performing analysis on datasets, the results are more accurate and dependable with a larger number of data points. There could prove to a sufficient amount of data for analysis, however if not, then the results and any relationships found could be inconclusive.

4 Results Phase I and Phase II

This chapter discusses the results from Phase I and Phase II. The results for the case study in Phase III are discussed in chapter 5, which includes the case study and data analysis.

4.1 Phase I

Phase I consisted of gathering information from secondary sources, such as companies using data mining, and Cotton Inc.; setting up plant interviews; and developing interview questions.

4.1.1 Exploration of Secondary Sources

In order to become more familiar with the data mining and the software, it was important to attend a SAS Data Mining Conference and training class for SAS Enterprise Miner. The purpose of this 3 day conference was to be able to network with people in the data mining field and get the most up-to-date information in the industry from data mining's top thought-leaders, visionaries and practitioners. What was learned from this conference was the wide applicability and success of data mining techniques in many industries. In a recent survey of CIO Magazine of subscribers involved in the purchase of business intelligence (BI) and/or data mining software, a full 88% indicated BI is a business priority for their organizations (Smith, 2005). However, as stated in the literature review, it is not commonly used in manufacturing environments. When data mining is used in manufacturing, it is focused on customer management and not the actual manufacturing process.

The World Wildlife Fund (WWF) uses analytics/modeling to increase the volume of online mailings, increase the response rate, and decrease costs. Data mining techniques will allow WWF to analyze and retain their most profitable donors, analyze direct response and online e-mail campaigns to increase response rates and drive additional revenue, and to stop pursuing donors that are not giving. Figure 4.1 shows a hierarchy structure created by the WWF to show the steps taken to achieve business intelligence and their ease/difficulty of implementation.

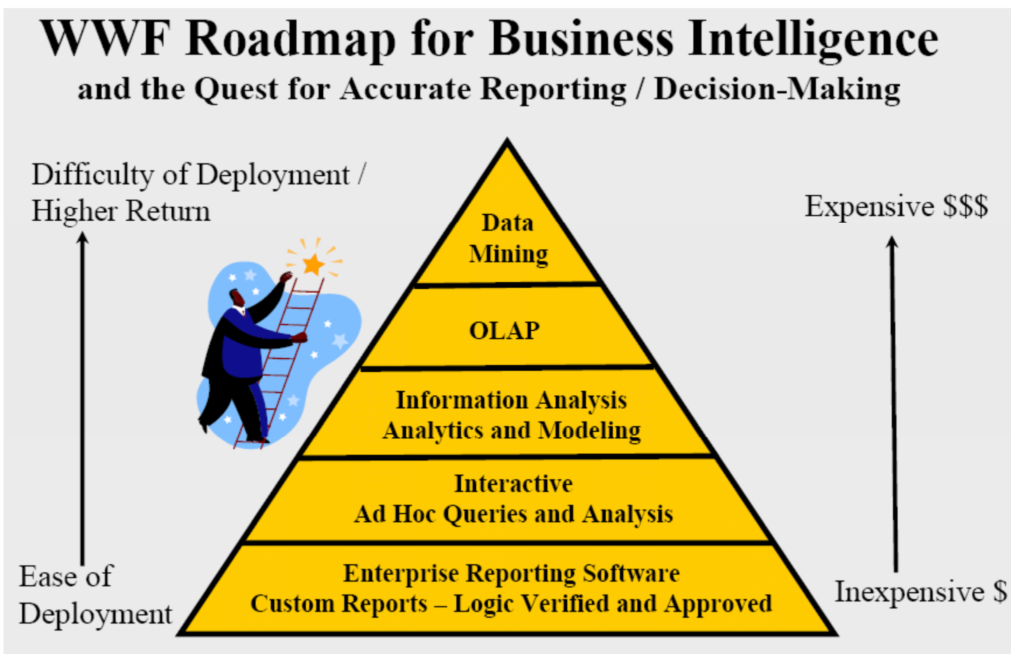


Figure 4.1: WWF roadmap for business intelligence (Smith, 2005)

Data mining is at the top of this pyramid, meaning it is the most difficult and expensive to implement, but has the highest return. WWF has worked through this roadmap and is in the process of implementing data mining by training staff and launching the software. This roadmap shows that data mining has major benefits in an organization, if the company can spend the time and money to get there.

Customer Intelligence utilizes information about a customer to determine relationships that generate greater profitability. It is also used to target new customers based on matching key attributes of the best customers. Figure 4.2 shows a hierarchy, obtained from Black, of business intelligence and customer intelligence needs.

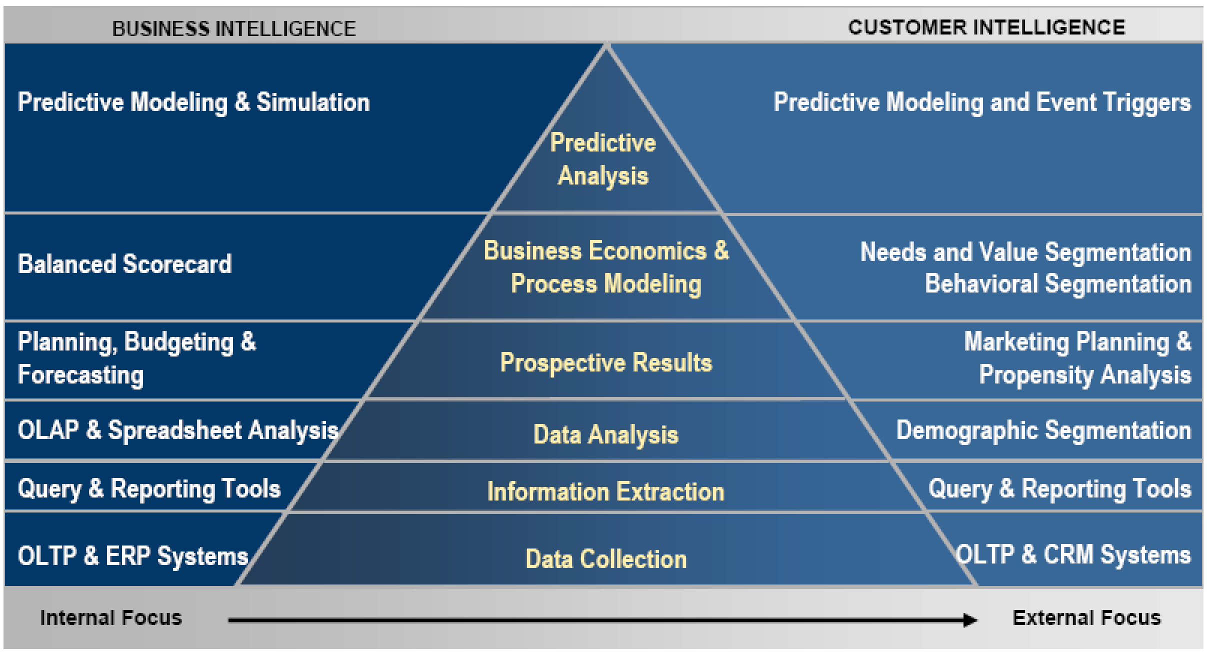


Figure 4.2: Hierarchy of business intelligence and customer intelligence needs (Black, 2005)

This figure is similar to Figure 4.1 because it shows all the different levels of business intelligence that precede data mining. Many companies stay in the lower levels of business intelligence and never actually achieve predictive modeling, or data mining. The companies that are operating on the predictive analysis level are mainly customer focused and not manufacturing.

In yarn manufacturing, most companies are at the data collection and information extraction levels. Some spinning plants are at the data analysis level and few are at the prospective results level. There have not been many known successful implementations of predictive analysis in yarn spinning.

One company that is focusing on finding relationship with fiber properties is Cotton Incorporated. Cotton Incorporated is a non-profit company that focuses on the research and promotion of cotton. The company has developed the Engineered Fiber Selection (EFS) System that is the world's leading cotton management system and covers 90% of the market. The EFS® System software enables cotton handlers to make accurate inventory, evaluation, and handling decisions from ginning to spinning (Cotton Inc). The system sorts the bales in a spinning plant's warehouse according to the fiber properties. In cotton spinning, mixing and blending are crucial to produce a consistent product. The system organizes the bales into categories set by the spinning plant, and then plans the laydown according to the different categories. The EFS® System software programs developed by Cotton Inc. for the cotton industry include MILLNet, MILLNet32™, EFS®-USCROP ™, and QRNet32. Figure 4.3 shows the information flow within the EFS® system.

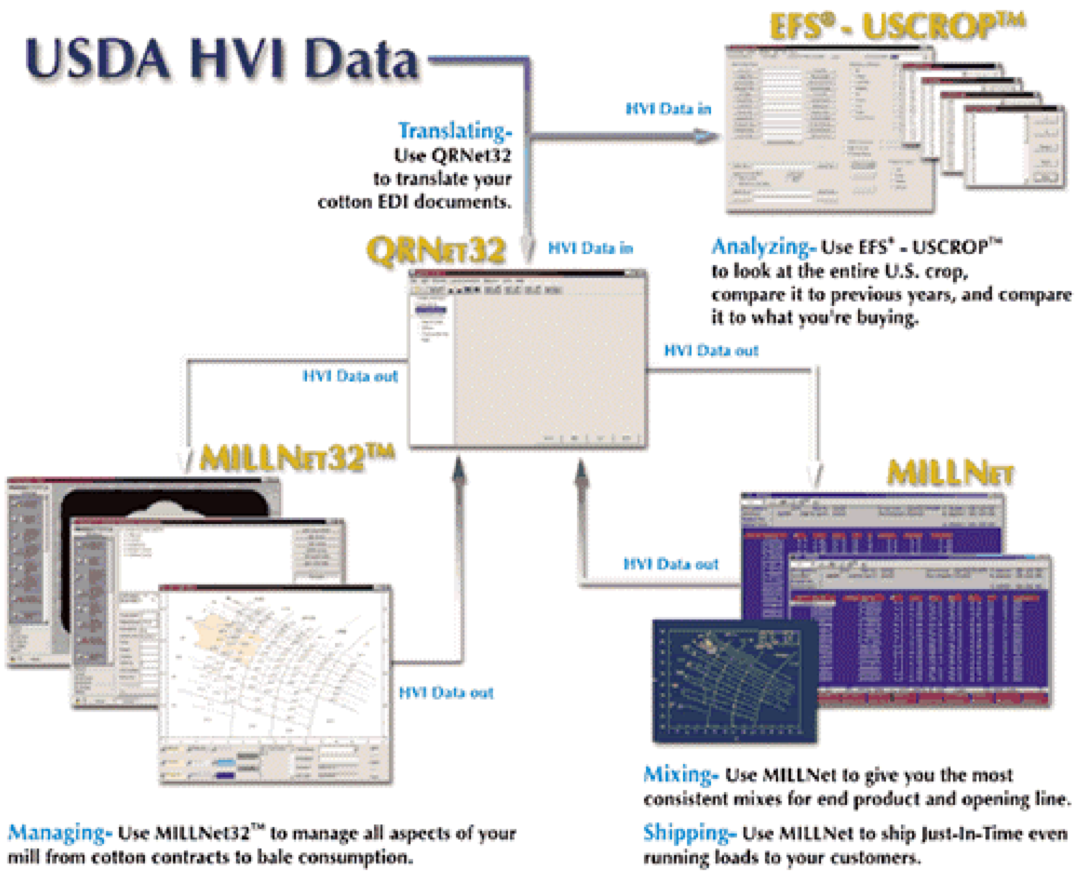


Figure 4.3: EFS System cotton flow (Cotton Inc.)

The United States Department of Agriculture (USDA) develops cotton grade standards and cotton classification systems. These uniform cotton standards were established to “eliminate price differences between markets, provide a means of settling disputes, make the farmer more cognizant of the value of his product, and, therefore put him in a better bargaining position, and in general be of great benefit to the cotton trade” (Cotton Inc). The purpose of cotton classification is to measure the physical attributes of raw cotton that affect the quality of the finished product and/or manufacturing efficiency (Cotton Inc).

Once cotton is harvested, it is sent to a gin where it is cleaned and put into bales. Two samples are taken from each bale of cotton, one from the top and one from the bottom. These samples are then sent to the United States Department of Agriculture (USDA) where they use High Volume Instrumentation (HVI) to test all properties of each bale of cotton. The data is returned and a Permanent Bale Identifier (PBI) is placed on each bale until consumed. This PBI contains a bar code with information that matches back to the USDA database. Bales are then stored in warehouses until their time of consumption. Each spinning plant sets up an account with USDA in order to be able to download this HVI data for their bales. Most spinning plants use the EFS® system to store and organize this data, as well as to give an optimal laydown. Table 4.1 contains information obtained from Cotton Inc., that can also be downloaded from the USDA website, that shows the universal classification data format.

Table 4.1: USDA universal classification data format (USDA)

FIELD NAME	COLUMN
Gin Code Number	01-05
Gin Bale Number	06-12
Date Classed	13-20
Module, Trailer, or Single Bale	21
Module/Trailer Number	22-26
Bales in Module/Trailer	27-28
Official Color Grade	29-30
Fiber Staple Length (32^{nds} of an inch)	31-32
Micronaire	33-34
Strength (grams/tex)	35-37
Leaf Grade	38
Extraneous Matter	39-40
Remarks	41-42
Instrument Color Grade	43-44
Color Quadrant	45
Color Rd	46-48
Color +b	49-51
Non-Lint Content (Trash Percent Surface)	52-53
Fiber Length (100^{ths} of an inch)	54-56
Length Uniformity Index (percent)	57-59
Upland or Pima	60
Record Type	61
Record Status	62
CCC Loan Premiums and Discounts	63-67

The information also contained a list of definitions to help understand each of the HVI properties. A compiled version of these definitions can be found in Appendix A.

Cotton Inc. provided information on the different fiber properties that are tested and their importance in the yarn manufacturing process, as well as provided a better understanding of the EFS® System. Cotton Inc. also provided a list of data elements collected for cotton fiber. This information was useful for RO2 and is discussed later in this chapter. The visit also generated useful background knowledge of cotton and the spinning process, which was important for the preparation of the subsequent plant visits.

4.1.2 Plant Contacts and Interview Questions

At the end of Phase I plant contacts were established for the initial investigative visits. A list of companies for possible interviews and case studies was compiled. This list was limited to yarn manufacturers in North Carolina and South Carolina that would be willing to cooperate and were available in the given time frame. Initial contacts were made through e-mail and the companies were chosen based on responses. Based on the responses received and due to time constraints, four spinning plants were visited. These plants were within reasonable proximity, were willing to discuss collection systems, and were available on the specified dates. In addition the companies varied in size and market share, which was thought to prove useful when comparing the amount of data collected and stored.

Interview questions were developed based on the previous research done by Drs. Hodge, Oxenham, and Schertel, as well as information gathered from secondary sources. Interview questions were focused on the yarn spinning process and included:

- What data is collected?
- What data is stored?
- Where are the data collection points?
- What data is online and what data is offline?
- What is the format of the data?
- How often is the data collected?

- What level of analysis is done with the data?

These questions were used to determine how much data these plants are collecting and how much data is being used for analysis. The purpose of the interview questions was to define data collection requirements and data quality issues for textile spinning.

4.2 Phase II

Phase II consisted of gathering information from primary sources. This involved spinning plant visits and interviews. There is much overlap between the different plants as far as data collected and data systems; and for this reason, the plants are discussed as a compilation and not as individual interviews. This information was then used for the preparation of interview questions for the case study.

4.2.1 Plant Interviews

Phase II began with visits and interviews with the four cotton open-end spinning companies. Most of these visits involved a walk through of the plant to look at process flow, equipment, and any data storage systems. The model in Figure 4.4, created by Stacey Schertel, was used as a model of collection points throughout the spinning process. The purpose of this model was to verify what data is collected, not collected, and at which stages in the process. This model includes some data elements that Schertel found to be commonly collected in spinning plants. Although this model does not include all possible data elements, it is useful to visualize the process flow and collection points. At the bottom of the model are estimated times at each process and an estimated overall time from beginning to end including the lag times.

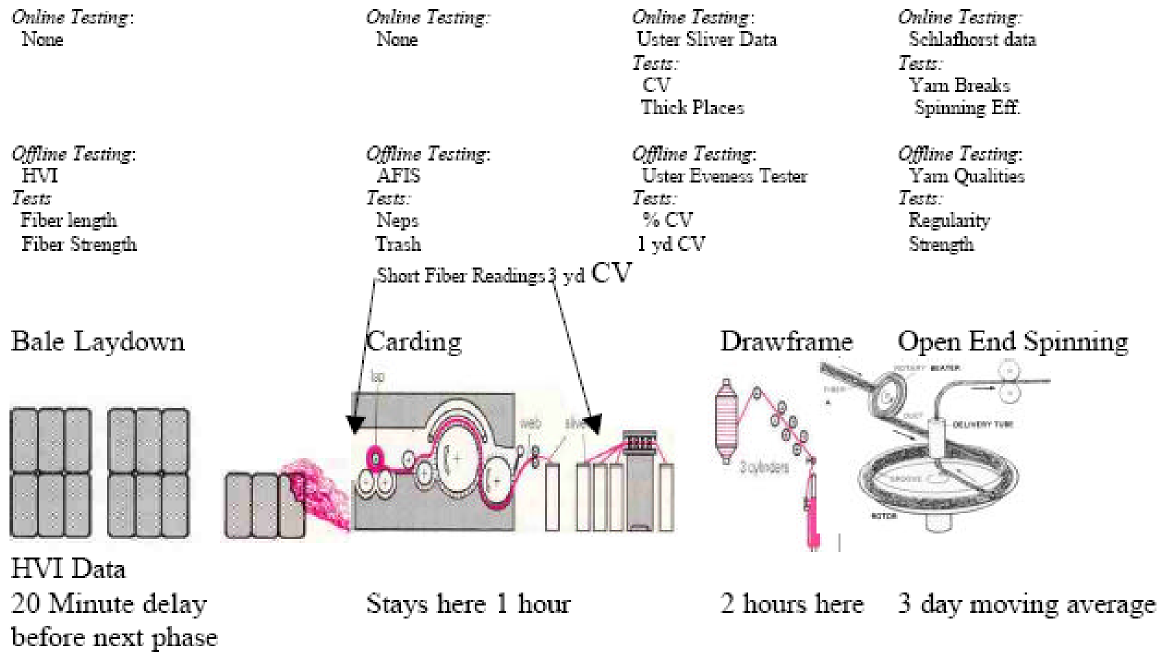


Figure 4.4: Flow of data through a spinning plant (Schertel, 2002)

Informal interviews with managers and technicians were conducted to answer the previous mentioned questions. These interviews were helpful in discussing data quality concerns, data storage, and limitations of data systems.

4.2.2 Machine Manufacturers and Data Collection Systems

There are several vendors for machinery and software used in yarn spinning. Most of the visited companies were processing yarn with equipment from leading manufacturers such as Trützschler, Rieter, and Schlafhorst. The plants collected online and offline data at different times in the process according to their specific data collection systems.

The systems used to collect the data varied from plant to plant depending on equipment and software packages as well as the different versions of software. Although there are several options for software and collection methods, the following sections discuss only the four interviewed companies. As stated, much of the information is redundant between different plants, so the findings are discussed as a composite of all four plants.

4.2.2.1 Fiber and Bale Laydown

All four of the plants used Cotton Inc.'s EFS® System to predict their optimal bale laydowns. As shown earlier in figure 4.3, there are several software packages that are part of the EFS® System. MiLLNet32 is a cotton management system that helps select, acquire, and warehouse cotton in inventory more effectively as well as analyze the HVI data more efficiently. Cotton merchants to make more accurate inventory for evaluation and for handling decisions most commonly use MiLLNet DOS. It is used for the purpose of producing statistically uniform cotton mixes based on the HVI properties. EFS – USCROP can be used to look at HVI properties of all cotton bales that have been tested and classified by the USDA from all over the country. QRNet32 is a file format translator program designed to support ANSI X12 standards (Cotton Inc).

The EFS® system will keep a storage history for any length of time in which it is set. All of the data accessible through the EFS® system can also be exported to Excel spreadsheets for reports, storage, etc. Some examples of this data include mincronaire, length, uniformity, strength, Rd, +b, leaf, grade, trash % area, and extraneous matter.

4.2.2.2 Mixing and Cleaning

There were no online data collection systems for mixing and cleaning at any of the four plants visited. There was also no offline data collection at these processes. Some data, such as machine settings and any malfunctions, was displayed on machine control panels.

4.2.2.3 Carding

An example of online monitoring at the cards is the Trützschler's Sliver Info System, also referred to as the Kit system, which collects production and quality data from the cards. The Trützschler equipment displays this data on machine panels for the current shift and previous shift, but also downloads the data so it can be seen from a central computer. This data is seen as instantaneous data on the computer, but the data can also be accessed from several previous shifts. Examples of data collected include sliver count, delivery speeds, production, draft, CV%, and downtime. The purpose of this system is to monitor the performance of the cards and notice any quality parameter issues with the sliver. This system is used mainly to resolve any problems that may be occurring due to machine malfunctions or settings. There are limitations with this Kit system, since the data cannot be put into a database for storage or be printed.

For offline quality data, lab equipment for measuring sliver properties can be connected to a system called TechWare Logbook on a central computer. Data that is generated can be sent directly to Logbook so there is no manual entry required. However, there is capability for manual entries. This data can be archived by machine and kept for several years. Examples of data collected are sliver count, tensile, CV%, and AFIS data such as neps, short fiber content, and trash. Any or all of this data can be exported to a database and is commonly used for trend reports and quality reports.

4.2.2.4 Drawing

Trützschler's Sliver Info System can also collect production and quality data from the drawframes. The Kit system is used the same way for drawframes as it is for the cards, to monitor drawframe performance as well as sliver quality. One limitation of this system, however, is the number of thick places in the sliver is not stored. This information can be seen on the drawframe panel, but is not recorded by the Kit system.

Another data collection system is Rieter's SpiderWeb system on the drawframes. The Rieter drawframes have machine panels, which display production and quality data for the current shift and previous shift. This data is sent directly to a central computer and can be seen as instantaneous data as well as can be archived back several years. This system is more flexible than Trützschler's kit system because the data can be exported to a database such as Excel for easier viewing, reports, and data can be printed. SpiderWeb is used similarly to the Kit system and as a way of monitoring machine performance and quality issues. This data is used in order to correct any maintenance problems with the machines or when the sliver properties are out of specifications.

For offline quality data, Techware Logbook can be connected to the testing equipment, such as AFIS and USTER. Logbook is used for drawing, the same way it is used for carding.

4.2.2.5 Spinning

Schlafhorst spinning machines have machine panels that display production and quality data for rotor spinning. The spinning frames contain a Barco clearing device on each individual rotor. Barco is used on rotor spinning machines as a yarn clearing and quality monitoring device (Barco). This quality data is displayed on the spinning frame machine panel along with the production data. Using Schlafhorst's CoroLab, the production and quality data displayed on the machine panels is sent to a central computer to be shown as instantaneous data or to be compiled for each shift and stored for a longer period of time. The individual rotor data is stored only a few previous shifts and then is averaged and stored by spinning frame. This data is used to pinpoint any problems with the machine, individual rotor, or when the yarn is not meeting the quality specifications. Most data from this system can be exported to a database such as Excel or Access. However, the more detailed quality cut information is not archived and cannot be exported easily.

For offline quality data, Techware Logbook can be hooked up to the testing equipment. Logbook is used for spinning, the same way it is used for carding and drawing. The only difference is that yarn data is being measured and stored instead of sliver data.

4.2.3 Analysis of Plant Interviews

It is important to understand that every spinning plant is not the same and does not always collect the same data. Each open-end spinning plant varied some with the number of machines at each process, the process steps, and the data systems. Each company also differed with the data collected, the amount of data collected, and the use of the data. Some of the companies relied more on the offline data, and some relied more on the online data. Some companies used the data to help resolve any problems with the quality, while others used the data to resolve problems as well as prevent/predict problems. Although a lot of data is collected, most plants are unsure of which data elements are the most important to collect and study in order to predict and control process and product quality. Some data is collected and stored with no future plans of use.

What was learned from these visits is that defining data collection requirements and data quality issues for textile spinning is not a simple task. In the spinning process, it is not required that data be collected at every process stage. Collection of data varies from plant to plant according to management standards and machine vendor standards. Management of a plant can decide which data is important to collect and how it can be used. What data is collected, how it is collected, and how the data is stored also depends strongly on the machine vendor equipment and data systems. The vendors have already set equipment to record certain data elements for production, quality, or both. Often these elements cannot even be accessed, let alone manipulated in any way.

Data quality issues also varied for different plants and the different data collection systems. With some data collection systems, data was transferred from the machine or equipment automatically to a central computer, whereas other data needed to be input manually. Inputting data manually leaves room for operator error, which could result in data quality issues. Also there are limitations of the different data collection systems. Data from most of the systems could be exported to Excel, but usually this results in formatting issues. For example, some columns may contain no values, cells may be formatted as text when they should be formatted as date, some cells might contain a "0" when it really means the machine was stopped, etc.

It was also evident that there is little to no accessibility between the different data collection systems. Software and equipment vendors like to hold any data collected as proprietary to their system and do not allow for access to other systems. Most spinning plants like to use several equipment vendors for their processing machinery. The use of different equipment, leads to the use of different software packages from various vendors. Table 4.2 shows examples of the different online and offline data collection and/or storage systems for each process stage.

Table 4.2: Online and offline data collection systems

Process	Online	Offline
Bale Laydown	EFS System	N/A
Mixing	N/A	N/A
Cleaning	N/A	N/A
Carding	Trutzschler Sliver Info System	TechWare Logbook
Drawing	Trutzschler Sliver Info System Rieter SpiderWeb	TechWare Logbook
Spinning	Schlafhorst CoroLab	TechWare Logbook

These systems are limited because they cannot communicate between each other. An example of this is a plant that has Trutzschler cards and drawframes, but Schlafhorst spinning frames. In this particular example, data from the cards and drawframes is automatically sent to the Trutzschler Kit system and data from the spinning frame is sent to the Schlafhorst CoroPilot system. Although the data is accessible from a central computer, the systems will not read between one another. For example, the machine or central computer cannot automatically change settings for the spinning frames when there are problems with a card or drawframe. Many of the data systems allow for the export of data a database, such as an Excel spreadsheet. From Excel, however, it is difficult to combine data from all the systems due to different formatting, collection times, machines, and data properties measured. Further conclusions of data collection requirements and data quality issues for textile spinning are discussed in the following case study.

These visits and interviews also proved useful in identifying data elements needed for monitoring and controlling product quality. Some data elements were listed in Schertel's model shown in Figure 4.4. As stated earlier, the data collection varies at each plant. Through the plant interviews, more data elements were uncovered. All of this information was used to compile an initial listing of data elements for each process stage. These data elements were helpful in the case study research as discussed in the next section.

The visits and interviews established that data is not collected, stored, and used consistently throughout all spinning companies. There are a variety of options for manufacturing equipment and data monitoring systems. For this reason, a case study at only one plant would provide results based on the specific data collection routines and systems for that particular plant. The results from this study can then be applied to other plants according to their specific process flow, machinery, and data collection and storage systems.

With the information obtained from the plant interviews, it was important to decide which company would be used for the case study. The focus of this research is on 100% cotton open-end spinning and the plant chosen should collect a reasonable amount of data, the data should be of good quality, and the company should be willing to cooperate and share their data. It was beneficial to select a company that collects both online and offline data in order to get a larger amount of data. The choice of company was made based on the given criteria.

4.2.4 Preparation for Case Study

As a result of the initial plant visits and discussion with Cotton Inc., a structured interview notebook was prepared to keep questions and data organized as well as make sure that all goals were accomplished during the case study. This notebook was divided into sections by spinning processes for open-end spinning (bale laydown, mixers, cleaners, carding, drawing, and spinning) and each contained color-coded pages for the different areas of focus. This book is only set up to be applied in open-end spinning so if it is to be used in ring spinning, for example, processes and number of processes will change. The color-coded pages were separated into the areas of machine parameters, online production data, online quality data, offline data, data storage, and definitions. These subcategories were generated as a result of information gathered from initial visits to the four spinning plants as well as from literature sources. A general flow diagram of the processes and the different data collection points was prepared:

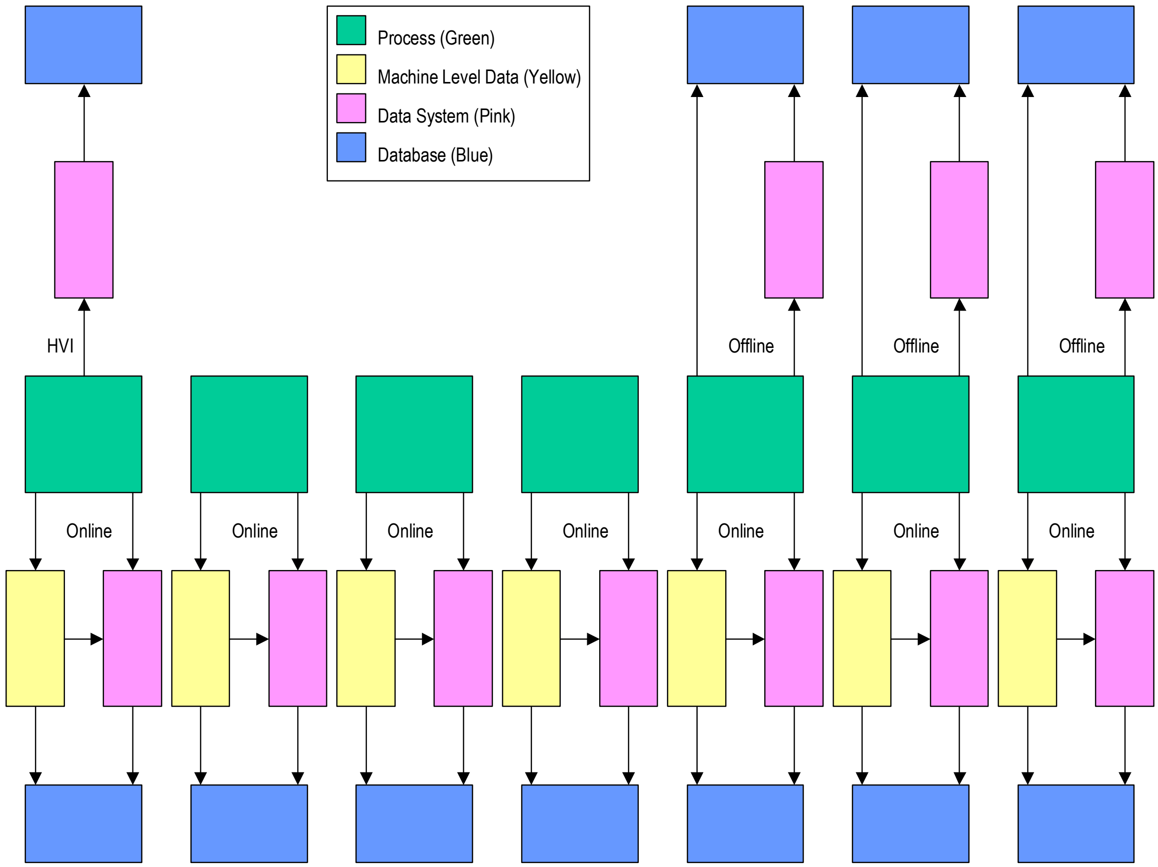


Figure 4.5: Data flow in a spinning plant

In this diagram, there are green blocks, which indicate a process step in yarn manufacturing. The yellow blocks symbolize machine level data. This means that there is a panel on the machine from which some sort of data can be read. The pink blocks represent a data system. A data system means software provided by a certain vendor to organize and store data from a machine or equipment. This data could be automatically dumped from the machine or equipment or it could be manually entered. The blue blocks represent a database, meaning some type of data storage, whether it is electronic or paper.

At each process step there could be online or offline data stored. Online data refers to data at the machine panel or a system that collects and stores data from the machines. Offline data refers to any measuring of fiber, sliver, or yarn done in a lab and not automatically generated by a machine in the production process. Online data can consist of just machine panel information, machine panel information that is taken manually and put into a database, machine panel information that is connected to some sort of data system automatically, or a data system that allows export to a database. Offline data can consist of measurements that are manually input into a database or a data system that is connected to measuring equipment, or a data system that can be exported to a database.

This diagram is useful when visiting the plant because it can be filled in according to their specific collection points and systems. The diagram is meant as a way of visually representing the flow of data in a spinning plant. Understanding the flow of data is important in order to uncover any opportunities to data mine. It is useful to know what data is collected and where it is collected to determine any relationships in the data throughout the process.

Each process section was divided into subcategories depending on the area of questions. The data elements included for each process were compiled from the four plant interviews and were used as initial lists. The case study was conducted in order to further investigate these elements and identify which are needed for monitoring and controlling product quality. Tables 4.3 through 4.8 show the subcategories and elements for each process stage and were included in the notebook.

The list of machine parameters contains elements that are general to all of the processes as well as elements specific to the process. The headings are used as a way of organizing the different questions. For example, machine parameters contains elements that have to do with the way the machine is set up to run. Machine parameters for bale laydown, as shown in table 4.3, contain questions about the number of bales and laydowns, etc. This is information that is important when determining the material flow and lag times in production. When trying to understand the process flow, it is important to know how many bales are in a laydown, the time to process a laydown, the weight of a bale, and the production speed.

The categories for online quality data and the machine panel contain no elements but were included so that notes could be taken upon observation. Online production data only contained two elements, production and production efficiency, leaving room to add any additional elements found at the machine.

The data storage category and elements were included for all process steps. It was unexpected that any online or offline data would be stored at bale laydown, so this was primarily focused on the EFS® System.

There are many HVI properties measured and recorded by the USDA, but previous research has shown these HVI properties listed to be of the most importance when predicting and improving quality in the process. These HVI properties were more or less a checklist of what is collected from the USDA information. It was to verify which properties the plant felt were more important.

Table 4.3: Bale Laydown

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage	Main HVI Properties
• Machine Name/Model • Year of Make • Machine Speed • Production Speed • Number of Bales • Weight of Bale • Number of Laydowns • Time to Process • Continuous or Discrete • Machine Settings • Manual or Automatic		• Production • Production Efficiency		• Online or Offline • Data System or Database • Production or Quality Data • Manual or Automatic • Real-time or Archived • Machine Level or Separate System • Frequency Collected • Storage Time • How Data Archived • Any Calculated Values • Data Format • Can Export Data • Show Graphs/Charts • Limitations • Use of Data	• Micronaire • Length • Uniformity • Strength • Rd • +b • Leaf • Grade • Trash % Area • Extraneous Matter • Remark

For machine settings in table 4.4, the list of elements included information about how the machine operates. This also included information specific to the process, which in this case was the number of chambers. This information was gathered to understand how much fiber could be held at this stage and the time it took for fiber to go through the mixing chambers.

As with bale laydown, the categories for online quality data and the machine panel contain no elements but were included so that notes could be taken upon observation. Online production data only contained two elements, production and production efficiency, leaving room to add any additional elements found at the machine.

Data storage and elements were included for mixers, although it was unexpected that there is any data recorded or kept at this process stage.

Table 4.4: Mixers

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage
• Machine Name/Model • Year of Make • Machine Speed • Production Speed • Number of Machines • Lag Time from Bale Laydown • Process Time • Number of Chambers • Machine Settings		• Production • Production Efficiency		• Online or Offline • Data System or Database • Production or Quality Data • Manual or Automatic • Real-time or Archived • Machine Level or Separate System • Frequency Collected • Storage Time • How Data Archived • Any Calculated Values • Data Format • Can Export Data • Show Graphs/Charts • Limitations • Use of Data

For machine parameters in table 4.5, the elements were general and did not include any elements specific to the process.

As with bale laydown and mixers, the categories for online quality data and the machine panel contain no elements but were included so that notes could be taken upon observation. Online production data only contained two elements, production and production efficiency, leaving room to add any additional elements found at the machine.

As with the mixers, data storage and elements were included for cleaners, although it was unexpected that there is any data recorded or kept at this process stage.

Table 4.5: Cleaners

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage
• Machine Name/Model • Year of Make • Machine Speed • Production Speed • Number of Machines • Lag Time from Mixers or Cleaners • Process Time • Machine Settings		• Production • Production Efficiency		• Online or Offline • Data System or Database • Production or Quality Data • Manual or Automatic • Real-time or Archived • Machine Level or Separate System • Frequency Collected • Storage Time • How Data Archived • Any Calculated Values • Data Format • Can Export Data • Show Graphs/Charts • Limitations • Use of Data

For machine parameters in table 4.6, general elements were included as well as elements specific to carding. These elements were critical in understanding the material flow and process times. Knowing the can size, number of cans, amount of sliver in each can, time to fill one can, and manual or automatic doff of cans, was important in figuring out the time the cotton is processing at the cards and lag times after the cards.

There are online systems that can be used at carding and that record production and quality data. The elements listed, were possible data elements that could be measured by an online system and were meant as a checklist. There was also room to include other elements.

The machine panel subcategory was included so that notes could be taken upon observation.

Offline quality data can be collected at the cards and the elements listed are possible elements that are collected from standard offline testing. They were included as a checklist of what the particular plant measures in the lab and any other elements not on the list, could be added.

Online and offline data can be collected at the cards and the data storage list was included as a way of understanding what data is collected, where it is collected, how it is collected, and how it is stored. This information is important when trying to understand capabilities of the data systems as well as any limitations.

Table 4.6: Carding

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage	Offline Quality Data
• Machine Name/Model • Year of Make • Machine Speed • Production Speed • Number of Machines • Lag Time from Cleaners • Chute Feed • Process Time • Can Size • Number of Cans • Amount in 1 Can • Time to Fill 1 Can • Machine Settings • Manual or Automatic Doff	• Sliver Count • CV % • Draft	• Production • Production Efficiency • Delivery Speed • Downtime • Total Doffs • Average Time per Doff • Stops		• Online or • Offline • Data System or Database • Production or Quality Data • Manual or Automatic • Real-time or Archived • Machine Level or Separate System • Frequency Collected • Storage Time • How Data Archived • Any Calculated Values • Data Format • Can Export Data • Show Graphs/Charts • Limitations • Use of Data	• Card Number • Sliver Count • CV% • Sliver Neps • % Elimination • Dust • Foreign Matter • Trash • Fiber Length • Fineness • Maturity • Short Fiber Content • Evenness • Immature Fiber Content

For machine parameters in table 4.7, general elements were included as well as elements specific to drawing. Like with carding, these elements were critical in understanding the material flow and process times. Knowing the can size, number of cans, amount in each can, time to fill one can, and manual or automatic doff of cans, was important in determining the time cotton is processing at the drawframes and lag times after drawing. There are different can sizes and the size of the cans going into the drawframe can differ from the size of the cans collecting the finisher sliver.

There are online systems which can be used at drawing and which record production and quality data. The elements listed, were possible data elements that could be measured by an online system and were meant as a checklist. It was expected there would be more elements measured, so these could be added to the list.

It was not known what information the machine panel displays, so this subcategory was included so that notes could be taken upon observation.

Offline quality data can be collected at the drawframes and the elements listed are possible elements that are collected from standard offline testing. They were included as a checklist of what the particular plant measures in the lab and any other elements not on the list, could be added.

Online and offline data can be collected at the drawframes and the data storage list was included as a way of understanding and organizing the data and data systems.

Table 4.7: Drawing

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage	Offline Quality Data
• Machine Name/Model	• Sliver Count	• Production		• Online or Offline	• Drawframe Number
• Year of Make	• CV %	• Production Efficiency		• Data System or Database	• Sliver Count
• Machine Speed	• Thicks/ Thins	• Delivery Speed		• Production or Quality Data	• CV%
• Production Speed		• Downtime		• Manual or Automatic	• Sliver Neps
• Number of Machines		• Total Doffs		• Real-time or Archived	• % Elimination
• Number of Passages		• Average Time per Doff		• Machine Level or Separate System	• Dust
• Lag Time from Carding		• Stops		• Frequency Collected	• Foreign Matter
• Process Time		• Duration of Stops		• Storage Time	• Trash
• Can Size				• How Data Archived	• Fiber Length
• Number of Cans				• Any Calculated Values	• Fineness
• Amount in 1 Can				• Data Format	• Maturity
• Time to Fill 1 Can				• Can Export Data	• Short Fiber Content
• Machine Settings				• Show Graphs/Charts	• Evenness
• Manual or Automatic Doff				• Limitations	• Immature Fiber Content
• Draft Settings				• Use of Data	• Evenness
• Autoleveller					

For machine parameters in table 4.8, general elements were included as well as elements specific to spinning. As with carding and drawing, these elements were critical in understanding the material flow, process times, and machine capabilities.

There are online systems that can be used at spinning and that record production and quality data. The elements listed, were possible data elements that could be measured by an online system and were meant as a checklist. It was expected there would be more elements measured, so these could be added to the list.

It was not known what information the machine panel displays, so this subcategory was included so that notes could be taken upon observation.

Offline quality data can be collected at the spinning frames and the elements listed are possible elements that are collected from standard offline testing. They were included as a checklist of what the particular plant measures in the lab and any other elements not on the list, could be added.

Online and offline data can be collected at the spinning frames and the data storage list was included as a way of understanding and organizing the data and data systems.

Table 4.8: Spinning

Machine Parameters	Online Quality Data	Online Production Data	Machine Panel	Data Storage	Offline Quality Data
• Machine Name/Model	• Yarn Count	• Production		• Online or Offline	• Spinning Frame
• Year of Make	• Length	• Production Efficiency		• Data System or Database	• Yarn Count
• Machine Speed	• CV%	• Rotor Speeds		• Production or Quality Data	• Strength
• Production Speed	• Neps	• Downtime		• Manual or Automatic	• CV%
• Number of Machines	• Thicks/ Thins	• Total Doffs		• Real-time or Archived	• Elongation
• Rotors per Machine	• Diameter	• Total Quality Stops		• Machine Level or Separate System	• Friction
• Lag Time from Drawing	• Foreign Fibers	• Yarn Breaks		• Frequency Collected	• Hairiness
• Process Time	• Moiré	• Clearer Cuts		• Storage Time	• Thicks/ Thins
• Machine Settings	• Twist	• Style		• How Data Archived	• Neps
• Manual or Automatic Doff	• Draft			• Any Calculated Values	• Break Factor
• Piecers per Machine				• Data Format	
				• Can Export Data	
				• Show Graphs/ Charts	
				• Limitations	
				• Use of Data	

At the end of the notebook there contained a section with definitions. These definitions were obtained from the ISO 10782-1 book of international standards and were included in case there was any question as to the meaning of any word (ISO, 1998). The list of definitions included:

- **Can Type & Size** – Shape and main dimensions of cans to be filled with sliver in spiral wraps, net mass.
- **Doffer** – Unit for the automatic replacement of full cans/bobbins for empty ones.
- **Automatic Piecers** – Traveling devices for the automatic detection and repair of thread breaks on spinning machines.
- **Runtime Efficiency** – Time running/Total time x 100%.
- **Stop/Go Percentage** – Ratio of cumulative running time to operating time of the production line x 100%.
- **Production Rate** – Amount produced per unit time.
- **Delivery Speed** – The speed of process material at the delivery of the appropriate process stage relevant for measuring production.
- **Downtime** – Time elapsed during interruptions of use of production equipment.
- **Draft, actual** – Ratio of the linear density of a textile material before a drawing process to that after.
- **Draft, nominal** – The draft as listed on a spinning schedule.
- **Sliver count** – Mass of sliver per unit length under standardized conditions of measurement.
- **Twist** – Number of turns in a yarn or sliver per unit length.

- **Coefficient of Mass Variation (CV%)** – Variation of mass of a fiber formation (slivers, yarns) depending on a base length, in %.

- **Thick Places, Long or Short** – Number of objectionable, shorter or longer thick places recorded by the monitor in relation to unit length.

- **Thin Places** – Number of relatively short, objectionable thin places recorded by the monitor, in relation to unit length.

- **Yarn Imperfections** – Term for number of tolerable yarn irregularities: “neps”, “thicks”, “thins” in a reference length.

- **Moiré** – Periodical yarn fault, producing a “chimney” in the spectrogram.

- **Neps** – Very short, thick section in the web or yarn, consisting of a knot of entangled fibers.

- **Nep Mat** – Number of neps per gram entering the card (within in the card mat).

- **% Elimination** – Percent reduction of neps, dust, and trash at the card (Mat-Sliver)/100.

- **Dust** – Number of dust particles per gram.

- **Yarn** – A product of substantial length and relatively small cross-section consisting of fibers and/or filaments, with or without twist.

- **Fiber** – Textile raw material generally; a unit of matter characterized by flexibility, fineness and high ratio of length to thickness.

- **Sliver** – A coherent assembly of fibers, normally without twist, and typically in the count range 2.5-25ktex.

- **Tenacity** – The tensile force per unit linear density corresponding with the maximum force on a force/extension curve.

- **Elongation** – The increase in length of a specimen during a tensile test, expressed in units of length.

- **Tensile Strength** – The maximum tensile force recorded in extending a test piece to breaking point.

5 Results Phase III Case Study

The initial visits to the four companies were provided a general overview of each plant, and the case study was conducted in order to get more details and be more specific. Phase III consisted of a case study with one plant, analysis of the process stages and different data systems, collection of data, and data analysis.

5.1 Case Study Process Analysis

When visiting the company, an initial walk through of the plant with a manager was useful to look at the equipment for each process, understand the process flow, identify any data systems, and ask any general questions about the process.

A process flow diagram, shown in Figure 5.1, was created once the walk through was completed. This diagram was important to understand all possible options for the flow of cotton. The case study plant had two lines of cleaners that each fed a line of cards. From the cards, canned sliver was manually taken to the drawframes, and then cans of finisher sliver are manually taken to the spinning frames.

One of the limitations of connecting the data from each process is not being able to determine precisely where cotton is in the process at a given time. A main objective of cotton yarn manufacturing is making sure the cotton is well blended to ensure a more uniform product. This makes it difficult when trying to correlate data measured at the beginning of the process to data measured in the middle or at the end. The only common variable in the many datasets is time, so it is important to understand the process flow and lag times between processes.

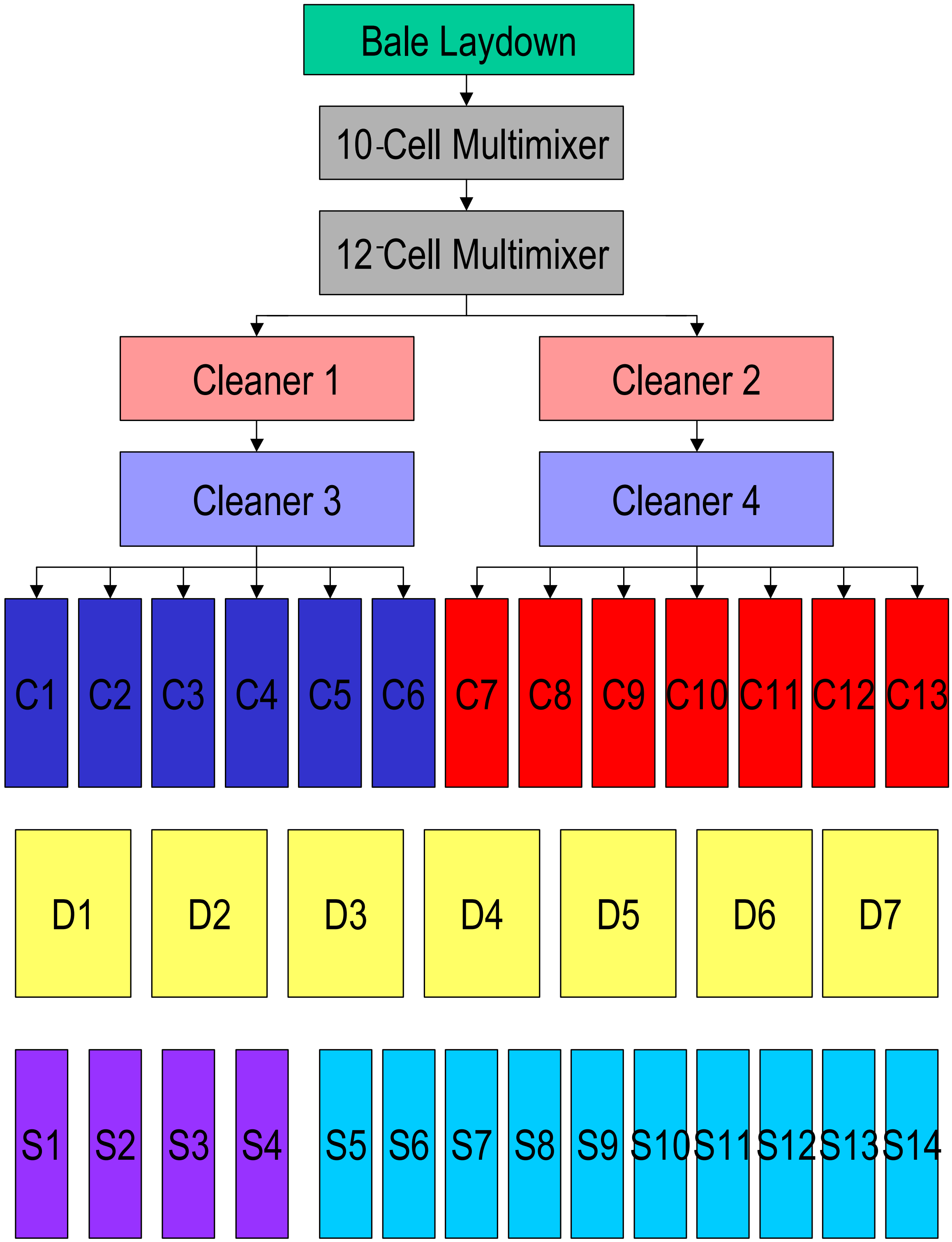


Figure 5.1: Process flow of case study plant

The data collection notebook described in chapter four was prepared as a means of collecting information in an organized way. The following sections will explain the information gathered from the case study through interviews and the use of tables 4.3 through 4.8.

5.1.1 Fiber

Cotton Inc.'s EFS System was used to plan and manage bale laydowns. All of the HVI data was stored in the system and could be exported easily to Excel. The archive time was set to keep the previous 150 laydowns, which resulted in almost six months of data. Some of this data included mincronaire, length, uniformity, strength, Rd, +b, leaf, grade, trash % area, and extraneous matter.

5.1.2 Bale Laydown

Yarn manufacturing begins with bale laydown. Different bales, each about 500 pounds, are placed side by side in a long line according to specifications given through Cotton Inc.'s EFS system. This system utilizes data obtained from fiber measurement to give an optimal bale laydown. The purpose of this is to acquire a blend of the various bales of cotton.

A machine, such as the Trutzschler Blendomat, has a mechanical arm that moves across the laydown picking fibers off the top of each bale. The depth at which the arm removes fibers is set and changed according to machine performance and run speeds. This plant chooses to run at a low depth so the machine will run more frequently. The bale opener ran only when fiber needed to be distributed to the other processes, usually resulting in about a 16 hour process time. This ensures there is always enough fiber blowing to the mixers, but at the same time not sending too much. These collected fibers are transported to the next process stage using air ducts. A machine panel shows machine settings and production data for the current shift. Some of this data includes advance (depth), height (arm is running), remaining runtime, efficiency, number of malfunctions, and downtime. This data is not collected from the machine panel or stored, and is only used to check for problems. As far as online quality or production data existing beyond the machine panel, there was none.

5.1.3 Mixers

From the bale laydown, fiber is blown through pipes to the next process, mixing. Mixing is important in cotton spinning plants in order to obtain a more uniform product. For mixing, this plant uses what is called a multimixer. A multi-mixer consists of different number of chambers side by side, usually 10 to 12 that are used for the mixing and blending of fibers. Fiber is blown into the multimixer and dropped down into the different chambers. When the fiber is released out the bottom, it is done so that a little comes out of each chamber and is mixed. This plant chose to have more than one multimixer so that more blending occurs. First the fiber went through the 10-cell multimixer and from there went through a 12-cell multimixer. There are machine panels that display some data, such as machine settings, but nothing is usually done with this data. There is no offline testing done at this process and no data is collected or stored. The machine panel shows data for the current shift and is used by operators to help resolve any problems with the machine or more specifically, the chambers. A typical 10-cell multimixer can hold about 600 lbs of fiber and a 12-cell multimixer can hold about 800 lbs of fiber.

5.1.4 Cleaners

From the mixers, the fiber is blown to the next process, cleaning. Cleaning is important in order to produce a good quality yarn. The machines work to remove trash, neps, seed coats, dust, foreign fibers and short fibers. Usually there is more than one stage of cleaning. For example, fibers are blown through a Trutzschler Cleanomat and into a Dustex cleaner. This is to ensure that as much trash is removed before the carding process. Cleaning machines use air to blow the fibers around, allowing the trash to fall out the bottom. The goal is to optimize cleaning and minimize fiber loss. There are machine panels on the cleaners that display data such as machine settings, but this data is typically not used for any analysis. There is no offline testing done at this process and no data is collected or stored. The machine panel shows data for the current shift and is used by operators to help resolve any problems with the machine.

5.1.5 Carding

From the cleaners, fiber is blown to the next process, carding. Carding opens, cleans, and blends the fiber and produces sliver. The cards in this plant are set to run at the same speed all the time. Cans are lined up at the end ready to be filled with sliver. With automatic doffing, once a can becomes full it is removed and an empty one replaces it. It usually takes around 45 minutes to fill one can with sliver.

There is a machine panel, which shows machine settings as well as production and quality data for that shift. Some data included is production, yards and pounds run in that shift, sliver breaks, stops, can changes, downtime, CV% yards, fineness (nominal and actual), web thickness (nominal and actual), and draft. This panel has no printer and can only be accessed at the machine. There was no online data collection system to access card data from a central computer. However, once a day some information was manually collected off the machine panel, as well as sliver weights twice a shift, and recorded in an Excel spreadsheet.

Offline data was collected at the cards. A sample/samples were collected at the back of the card, known as the mat and some samples were taken at the front of the card, known as sliver. This plant had two lines of cards, so one mat sample was taken for each line of cards and measured against the output samples of each individual card. This is done to calculate the percent eliminations of neps, short fiber content, dust, trash, and foreign matter. The samples were taken to the testing lab where tests such as AFIS and Uster evenness could be run. AFIS measures sliver count, CV% for length, evenness, and count, neps, seed coat neps, %elimination values, foreign matter, dust, trash, fiber length, fineness, maturity ratio, immature fiber content %, and short fiber content. AFIS properties were measured on mat from a line of card and sliver from each card every other week, sliver weights were taken daily and twice a shift, and Uster evenness was done weekly. The data was automatically sent from the testing equipment into a software system where the data could be more easily read, charts and graphs could be generated, and the data could be stored. This allows data to be stored back several years, or in this case, back to the date of installation. Averages for each machine were stored, especially since the AFIS test runs five repetitions. From this storage system, all the data can be exported to Excel. This data is used for the weekly quality reports. The weekly quality reports showed the trends that were going on with different properties and machines.

5.1.6 Drawing

After carding, the cans are moved manually to the drawframes. The estimated lag time before drawing was about an hour. The drawframes will support up to eight cans of sliver going in, but this plant only used seven cans of sliver at each drawframe. The purpose of drawing is to ensure more blending and to reduce the sliver. Carded sliver from several different cans from several different cards is combined to make one sliver. The drawframes in this plant were set to run at the same speeds. As with carding, cans are lined up at the end of the drawframe ready to be filled with sliver. The time to fill one can with sliver is about 45 minutes. Automatic doffing removes full cans and replaces them with empty ones. There is a machine panel, which shows machine settings, autoleveling settings, production data, and quality data. Information is shown for the current shift, as well as for the previous six shifts. The attainable data on the drawframe includes delivery speed, efficiency, pounds produced, target and actual sliver weight, stops, and A% stop limit and length. This plant did have an online data collection system so that all data collected from the machine is dumped directly into this system. Data is collected every six minutes and this detailed information is kept about a week. After a week, only shift averages are stored. Archives exist from initial launch of the system.

This plant collected offline data from finisher sliver. Finisher sliver samples, collected at the back of the drawframe were taken to the testing lab where Uster evenness test were run. At this particular plant, no AFIS testing was done on finisher sliver, Uster evenness was checked daily, and weights were recorded daily and twice a shift. Although no AFIS data was measured, the plant planned to start these measurements soon. The data that is collected is sliver count, CV% for count and evenness, weights, and evenness. As with the carding tests, the equipment is hooked up to a storage system so that all data is sent directly as it is measured. This data can be kept as long as needed. Each measured value is recorded and stored in the software. The data can be exported to Excel and used for weekly quality reports.

5.1.7 Spinning

Cans from the drawframes are taken manually to the spinning frames. The estimated lag time before spinning was about one hour. The cans are spread out randomly to different spinning frames to increase the amount of blending. The purpose of open-end spinning is to turn sliver into yarn and wind it on a package. There are many rotor positions on a spinning frame in order to create many packages at the same time. The function of a rotor is to collect the fibers and then draw them off as yarn. The number of rotor positions will vary according to the machinery manufacturer and type of machine. Some will hold up to 300 positions, meaning it can make up to 300 packages of yarn simultaneously. The machines run at different settings and speeds according to the different yarn styles. For example, a thick yarn might require that the spinning speed to be slower and a thin yarn could be spun much faster.

For this reason, it could take anywhere from less than one day to several days to spin one yarn on one spinning frame.

There was a machine panel, which has machine settings, production data, and quality data. Some data shown included yarn count, rotor speeds, length, yarn breaks, production in pounds, twist, and draft. There was a mini printer at the machine panel, but can only print very limited amounts of data. This plant did collect online data from the spinning frames. Data from the machine was directly sent into a software system that stored the data. This data included style identification running, shift efficiency, production, doffings, speed, CV%, neps, thins, thicks, short thicks, long thicks, thins, moiré, thin slivers, thick slivers, foreign fibers, quality stops, and production stops. At this plant, the data system kept individual rotor position data from the previous 19 shifts. Anything earlier is averaged and can be seen by individual spinning frame and shift. One limitation that was found with this particular data system is when the yarn count changes in the middle of a shift, the data is assigned to one shift or the other and the change is not recorded. Most of the data can be exported to Excel.

This plant did perform offline testing of yarn at the spinning process. This plant checks six samples from each spinning frame every time there is a count change and weekly. The samples are taken to the testing lab where tests such as evenness, tensile, and count & skein break are run. Data that is collected includes yarn twist, count, friction, strength, elongation, hairiness, thicks/thins, neps, break factor, and single end break. The yarn testing equipment is connected to a data storage system so that the data is sent automatically. This data can be kept for any length of time. Individual testing numbers and averages are stored. This data can be exported to Excel and used for weekly quality reports.

5.2 Analysis of Data Collection Systems from Case Study

Machine and software vendors determine the data elements needed for monitoring and controlling product quality. This plant had different online data systems for each process and the same offline data system for carding, drawing, and spinning. For the online systems, the elements measured were standard to each vendor. The elements measured for each process were mentioned in the previous section. This data is automatically collected, using the systems, and then it is up to the discretion of the plant to decide how to use the data. For offline data, the plant decides which lab tests need to be run on the sliver and yarn. The equipment used in the lab has standard data elements that can be measured. In the selection of equipment, the data elements measured are selected as well. The data system is set up to store all of the data measured from the different equipment. In this situation, the equipment vendors decide the data elements that need to be measured and the data system stores the information.

There is limited accessibility between the different data collection systems. The online systems were from different vendors and therefore each hold their information in a proprietary format. The systems are not set up in a way to communicate data between different systems and vendors. For example, the Rieter SpiderWeb system for drawing and the Barco system for spinning cannot exchange any data. There is no way of setting up the spinning system to run based on data changes in the drawing system. If this were possible, then problems caught at the drawframes might help prevent any problems at the spinning frames. All of the data systems could export data to Excel, however it is difficult to compare the different data sets because the data elements and frequencies the data was collected varied. Each system was designed specifically for each process and was not meant to communicate the data to or with another system.

There was one offline system that contained data from carding, drawing, and spinning products. This means that all the data from three different processes is accessible on one system. However, offline data is a measurement of product quality, and is not really used for any diagnostic problems. Usually when samples are measured and a problem is determined, it is too late to correct in the process because the final product will already have been produced. In this case, the data can be accessed easily for three different processes, but only in archived form.

Data is accessible from all of these before mentioned data systems, but trying to access data between the systems is not an option.

5.3 New Model of Collection Points

As a result of the case study, the model used from Schertel's dissertation was revised to reflect the recent findings. This model, shown in figure 5.2, shows the collection points of data elements needed for monitoring and controlling product quality. Not all data elements are represented and a list of these can be found in Appendix B. Instead of using the plant names for the online systems, it was changed to read "proprietary system". Data elements measured for the online and offline tests were added to the appropriate process stage, as well as the actual tests performed. Also listed are the estimated process times at each process stage. These times were used in the estimation of the overall lag time from bale laydown to spinning.

Online Testing:
None

Offline Testing:
HVI

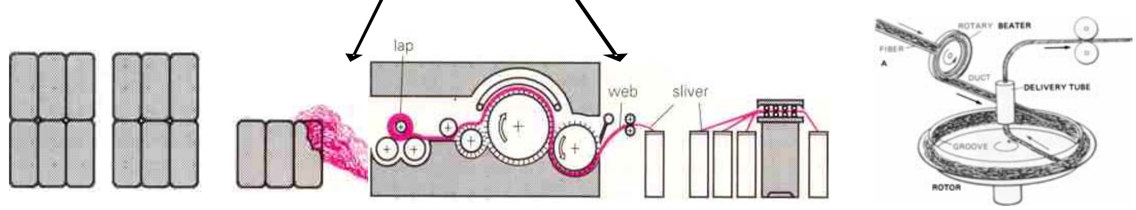
Tests:
- Fiber Length
- Fiber Strength
- Mic
- Plus b
- Uniformity
- Trash
- etc.

Online Testing:
Proprietary System
Tests:
CV, Draft
Offline Testing:
AFIS, USTER,
Weights
Tests:
- Neps
- Length
- SFC
- Trash
- Maturity Ratio
- etc.

Online Testing:
Proprietary System
Tests:
CV, Thick Places, etc.
Offline Testing:
AFIS, USTER,
Weights
Tests:
- Neps
- Length
- SFC
- Trash
- Maturity Ratio
- etc.

Online Testing:
Proprietary System
Tests:
Yarn Breaks,
Yarn Efficiencies
Offline Testing:
Evenness, Tensile,
Count & Skein Break
Tests:
- Strength
- Count
- Twist
- Hairiness
- Regularity
- etc.

Bale Laydown **Carding** **Drawframe** **Open End Spinning**



~ 16 hours ~ 1 hour ~ 1 hour ~ 3 Day Moving Average

Figure 5.2: Data collection in case study plant

5.4 Data Collection and Preparation

Data sets were collected in Excel spreadsheets from the EFS System, cards (offline), drawframes (online and offline), and spinning frames (online and offline). The purpose of collecting the data was to define data quality issues for textile spinning (RO2). Before any data analysis could be performed, the quality of data needed to be checked and any preparation/cleaning of the data needed to be completed.

5.4.1 HVI Data from Bales

HVI data was obtained in Excel for the previous 150 laydowns. In this case results included almost six months of data. The only data elements collected from the entire dataset were consume date, file ID, micronaire, length, uniformity, strength, Rd, +b, trash leaf, classer grade, trash % area, extraneous matter, and remark. The database in the EFS System contains many other elements concerning bale properties and laydown information, but only the few mentioned were selected for analysis.

It was discovered that the consume date is recorded when the bales are put into the central computer. This is not necessarily the date that the bales are positioned in the laydown and consumed, but is based on the demand for fiber. This presents a data quality issue since the date recorded is not always the actual consume date.

The consume data column was formatted as text, and the date was in quotation marks. The quotation marks were removed and the cells formatted to date. The extraneous matter column was removed since it contained only values of "0".

Looking through the data, there were no missing values for any of the properties, no extreme values, and no incorrect values such as text where numbers should be. Besides the formatting problem with the date column, no other data quality issues were discovered that needed to be resolved.

5.4.2 AFIS Data from Cards

Data from the AFIS equipment was compiled for the previous two years into an Excel spreadsheet. The data included in this spreadsheet was everything AFIS measures: project, machine, unit, process, product, material, test date, track number, reference, comment, nep size, neps, length by weight, length by weight CV%, upper quartile length by weight, short fiber content by weight, length, length CV%, short fiber content, length span 5%, total count per grain, trash size, dust, trash, visible foreign matter, seed coat neps, fineness, immature fiber content, and maturity ratio.

For each date, there was one row of data for the mat (measured at only one card but used to represent all cards in the line) and several rows of data for sliver measured from each card in the line. Since these measurements were done every other week, the line of card measured each week alternated. For example, one week line one is measured and the next week, line two is measured.

The unit column was deleted since the data only came from one particular plant. The reference and track number columns were deleted because they contained no data. The product and material columns were deleted as well because all the values were the same and were of no importance in evaluation. The test date column was formatted as text cells, so this was converted to date cells using the DATEVALUE function for Excel 2000.

The comment column stated whether the data was from the mat or the sliver. This column was manually entered and contained a few incorrect values. Examples of incorrect entries were: 3LIVER, MATVER, and MAT 208. These were corrected accordingly based on the other data. The seed coat neps, fineness, immature fiber content, and maturity ratio columns contained only 41 values each out of 737 possible entries. Further investigation revealed that the plant only began testing these properties recently.

5.4.3 Online Data from Drawframe

Online data from the drawframe was compiled for the past six months from the online system into Excel. The data for machine stops was separated into files according to the machine number. This data included: date, sliver break input, sliver break output, sliver monitor, autoleveling, lap up creel, draft roller lap, web funnel jam, sliver jam coiler, A%, CV%, no empty can, can truck missing, spectrogram, thick place, and shift total. All of these elements are reasons for the machine to stop, whether it be a machine problem or a quality problem. Another file showed the production efficiencies of each machine. This data was machine, date, efficiency, and shift total. A separate file showed the production per shift in pounds. This data included machine, date, production, and shift total production.

5.4.4 Online Data from Spinning

The online spinning data collection system contains two types of data, one that can be exported to Excel and one that cannot be exported to Excel. This data that cannot be exported shows information for individual spinning frames back the previous 19 shifts and then it is overwritten. This data was collected by copying screen shots and pasting them into Microsoft Word. Each file was a different shift containing screen shots from each spinning frame. Figure 5.3 shows an example screen shot of this data.

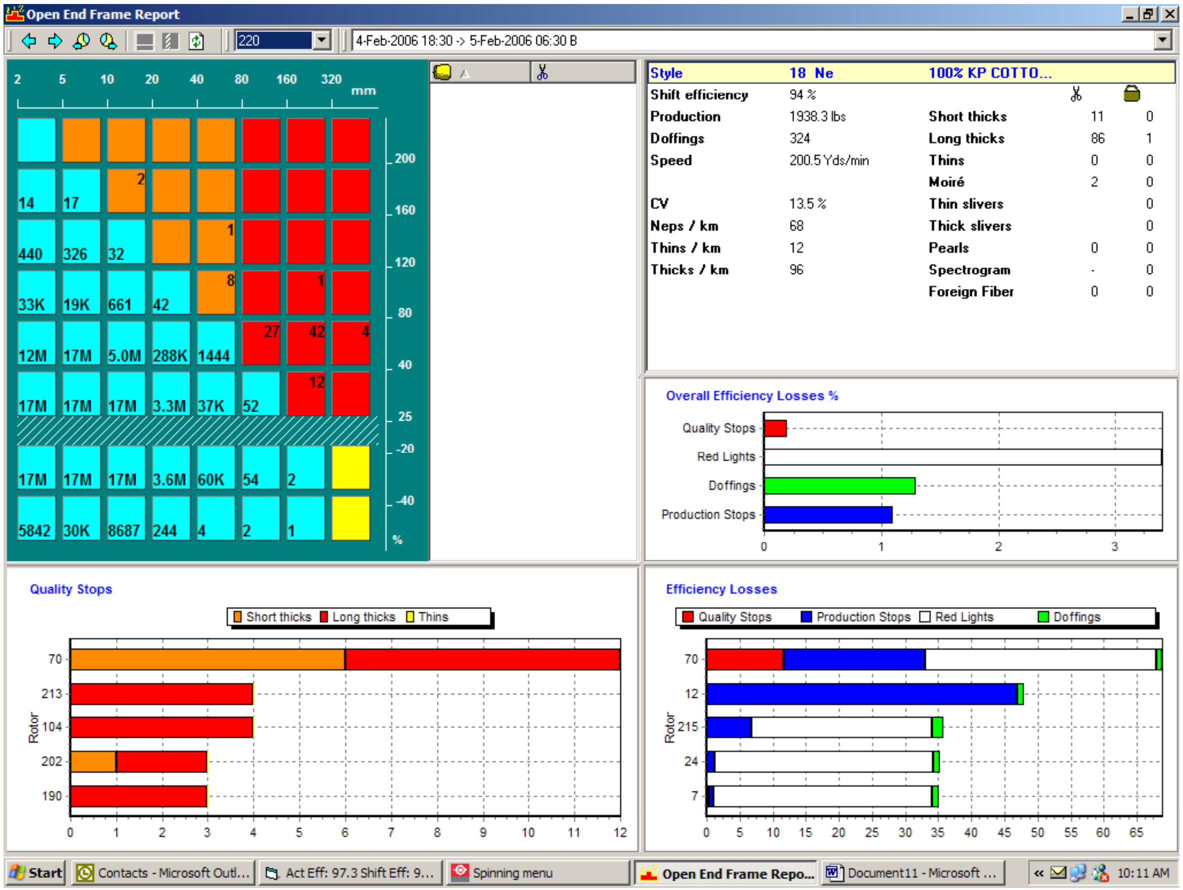


Figure 5.3: Screen shot of online spinning data

This data is recorded through the optical clearing device on the spinning rotors and gives average data for all the rotor positions as well as shows the worst five rotors. This data cannot be exported to a database and can only be put into table form manually, and therefore is not archived.

Data from the online system was collected for the last six months. This data was in Excel and was split into several different files by month. The data was compiled into one file and sorted by date. The data included: start date, machine, style, count, production, total stop time, quality stops, quality stop time, production stops, production stop time, doffings, doffing time, red lights, red light time, short thick stops, long thick stops, thin stops, moiré stops, thin sliver stops, thick sliver stops, foreign fiber stops, piecer attempts, IPI: neps/km, IPI: thins/km, IPI: thicks/km, IPI: CV, machine stop time, doffings/1000rhr, quality stops/1000rhr, production stops/1000rhr, red lights/1000rhr, out production time, average minutes per red light, total all stops, total stops/1000rhr, piecer ratio, adjusted efficiency, and absolute efficiency. Each date had two rows of measurements compiled for each shift, and recorded at 6:30am and 6:30pm.

The start date column format was set to custom and showed the date and time as such: m/d/yyyy h:mm. This column was reformatted as date to read mm/dd/yy. However, by doing this, the time of day the measurements were taken would not be able to be identified. There were also rows that contained totals for each month that needed to be removed.

Based on visual checks this data set contained the most quality issues. When sorting the data by count, it was noticed that many of the count data points were “0”. Count is the size of the yarn, so it is impossible to have a zero count yarn. One solution was to look at the production for the same days. It was found that all of the days where the yarn count was zero, the production was zero as well. However, looking at other elements, it was seen that when no yarn was being produced there were still values several parameters like short thick stops, long thick stops, and doffings. These values were determined to be incorrect. Further investigation showed that these problems all occurred within the same time range suggesting there were machine or data system problems at that time.

Looking at all the columns that contained times in hours, such as production stop time, doffing time, and red light time, proved that many of these values were incorrect. Many of the values were really extreme numbers and made no sense with the other data. It was concluded that the timers must not have been reset properly.

All of these data quality issues suggest that some of the data will not be useful in the analysis and could cause errors in the results. This kind of issue is common when dealing with large amounts of data. It is assumed that not all values are suitable for use in the analysis.

5.4.5 Offline Data from Spinning

Offline yarn data for several different yarn styles was collected for the last two years. A yarn style is the type of yarn that is being produced, including the size and amount of twist. It was difficult to acquire two years of data on the same yarn because this plant changes the yarns they run so regularly. Each style was its own separate file. The data included: project, machine, test date, unit, track number, reference, product, material, process, operator, count, count CV, single end, single end CV, elongation, elongation CV, CVm, thin (+50%0, hair, neps (+280%), neps (+140%), neps (+200%), CV (1yd), CV (3yd), CV (10yd), CV (50yd), work to break, skein, and break factor.

The date column was formatted as general text and was converted to date format using the datevalue function in Excel. The unit column was deleted because all yarns measured came from the same unit. The reference and break factor columns were deleted since they both contained no data, the process column was deleted since all yarns came from spinning, and the operator column was deleted because it would not be used in analysis for the study.

5.5 Data Analysis

As can be seen from the previous sections, there are many data quality issues in textile spinning, which include: missing values, irrelevant data, incorrect values, and wrong formats. These are all problems that exist within the data sets. Each data set contains the date in which the elements were measured. In order to compare the data between different data sets, the common variable must be time. However, each data set contains a different number of measurements for a given date, the collection times are different, and the frequency of measurements is different. For this reason, it is difficult to combine the data sets for analysis because it would create more data quality issues. The only way to combine the data sets is to use time as the common variable. Determining the lag times will allow for more accurate analysis between the data sets.

5.5.1 Level I Analysis

Because of the complexity of data mining, a preliminary study using Microsoft Excel was done to verify the data. This qualitative analysis involved a comparison of properties over time, properties over time with a moving average, and xy charts. The first level of analysis was creating scatter plots for individual elements in a data set over time. The second level was adding a moving average trend line to these charts to pick up general trends and determine lag times. For further investigation, xy charts were produced to represent the relationship between two elements at a time. Analysis was done on a few of the individual data sets before they were combined and studied across the whole spinning process.

5.5.1.1 AFIS Data

For the AFIS data, the properties chosen to analyze were neps, short fiber content, and dust. These properties were measured for the mat going into the card and the sliver coming out of the card. Mat and sliver properties were compared by card line to see the amount of elimination. Figures 5.4 and 5.5 show a comparison of neps from the mat and then from the sliver. There are fewer mat data points since only one measurement was taken for the entire line of cards, compared to sliver measurements taken from each individual card in the line. It was expected that there would be more neps before carding and fewer after carding and the data shows this result. Neps are considered imperfections and as each process stage is completed, the number of imperfections should decrease.

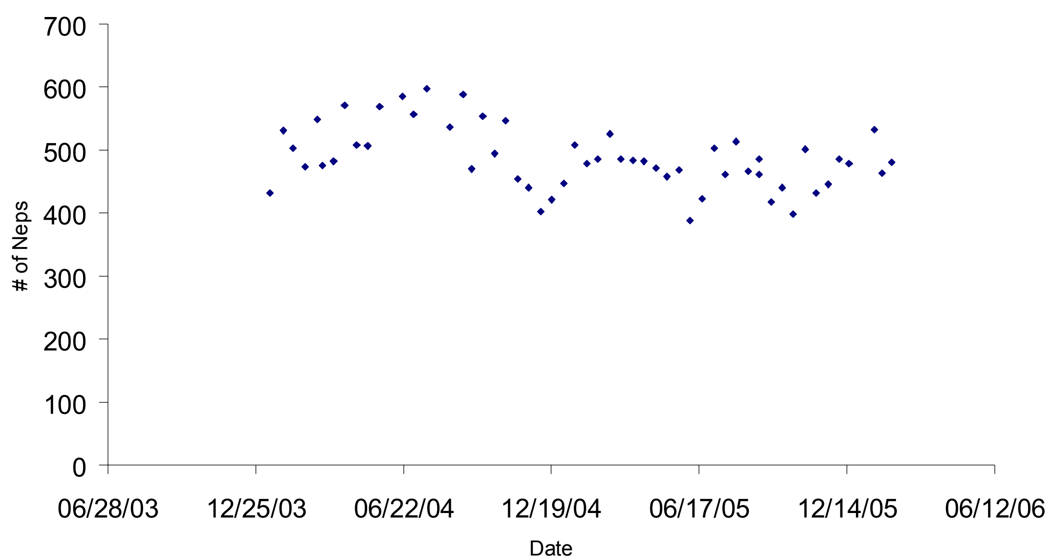


Figure 5.4: Mat neps for card line 1

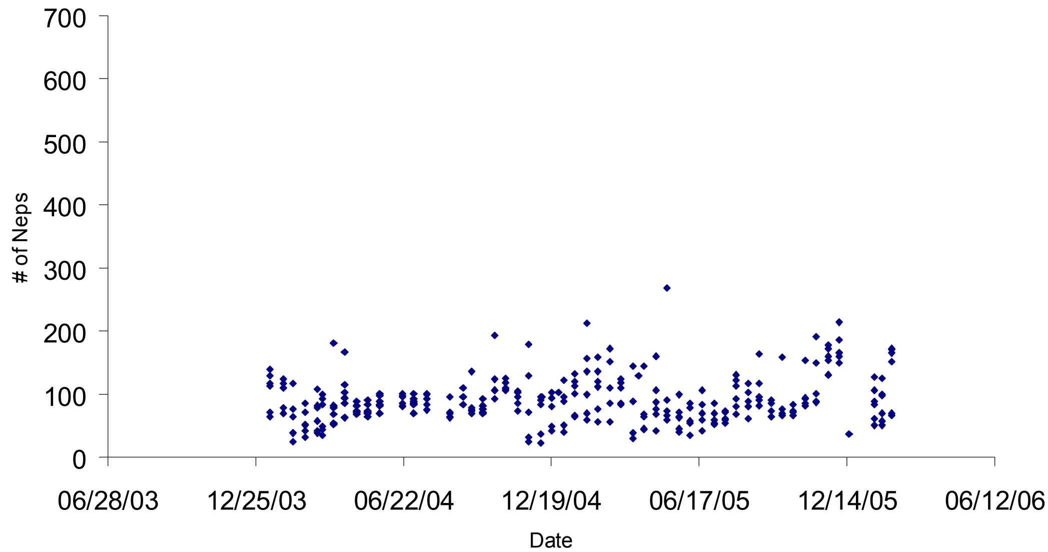


Figure 5.5: Sliver neps for card line 1

The same comparison was done for short fiber content and dust, as shown in Figures 5.6, 5.7, 5.8, and 5.9. The mat and sliver for each line of cards was compared for these elements. Short fiber content was expected to decrease after carding, but this was not the case. There was some unusual activity happening in the last few weeks of the data. Further investigation showed that this change in data corresponded with a change in AFIS equipment, meaning this could be one explanation to the large increase. However, no other properties seemed to be affected by this change in equipment. Since this was evident in data for both the mat and card sliver, there could have been a change in cotton resulting in the increase of short fiber content.

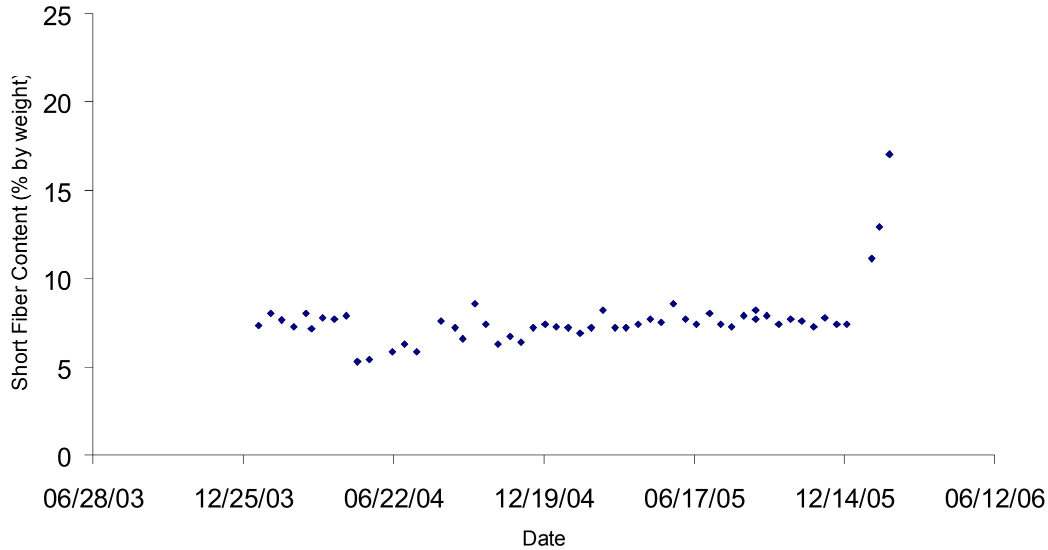


Figure 5.6: Mat short fiber content (% by weight) for card line 1

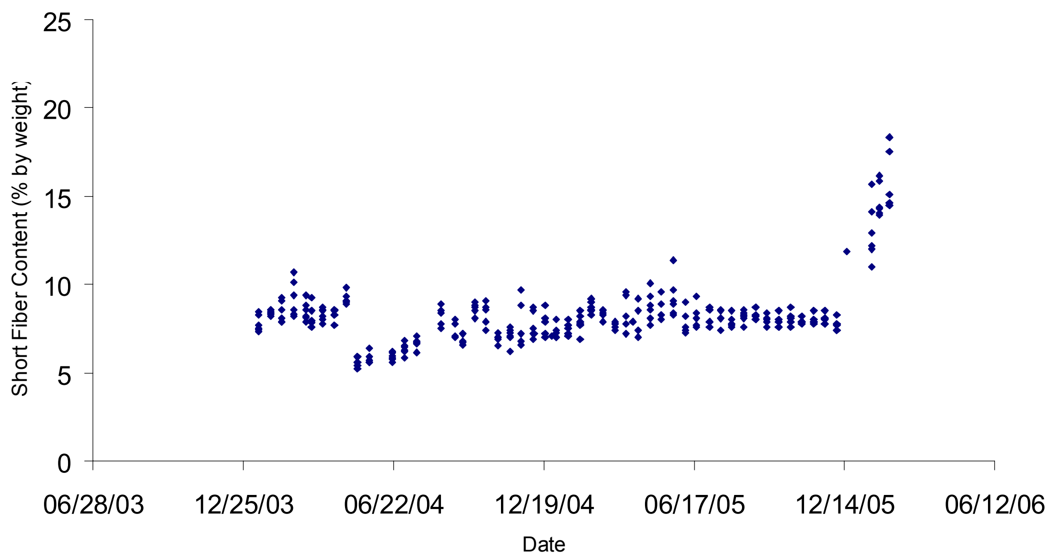


Figure 5.7: Sliver short fiber content (% by weight) for card line 1

Figures 5.8 and 5.9 show a decrease in dust content as was expected. There should a decrease in dust after carding. There were a few outliers, which could have been attributed by card problems, measuring equipment problems, or a bad section of fibers.

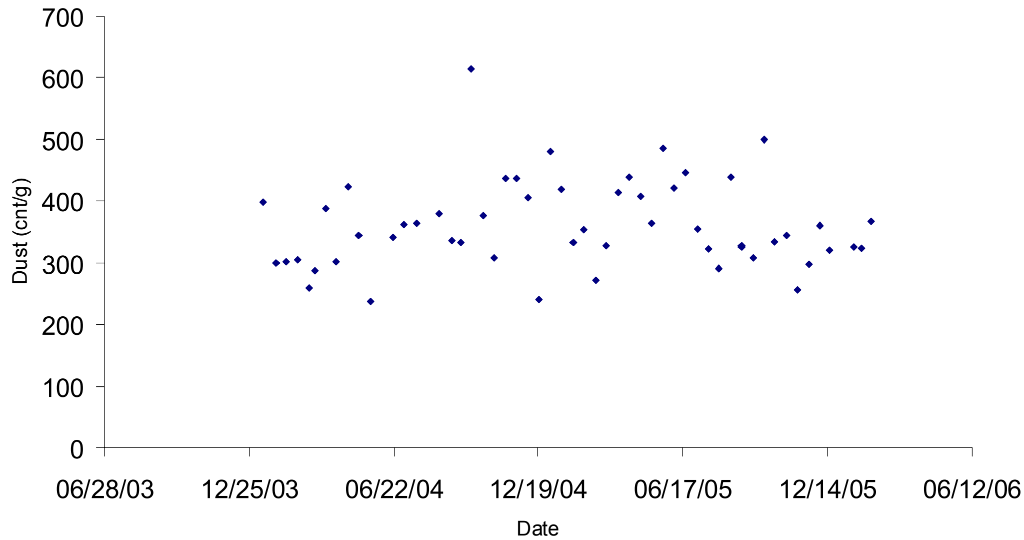


Figure 5.8: Mat dust (cnt/g) for card line 1

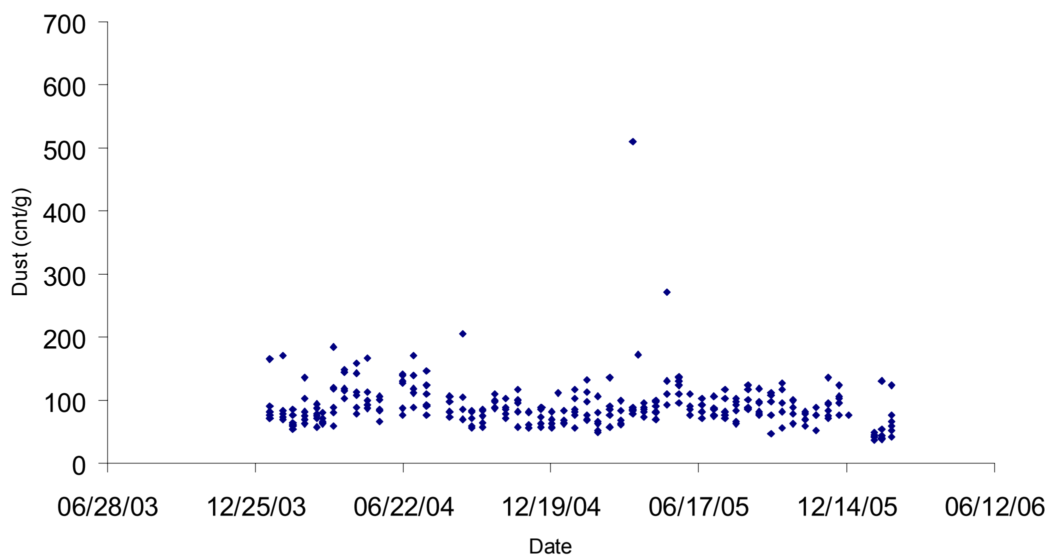


Figure 5.9: Sliver dust (cnt/g) for card line 1

Nep data for sliver was then investigated for several cards on each line. Figures 5.10 and 5.11 show data from two different cards, one in each line. A moving average trend line was added in order to better see the increases and decreases over time.

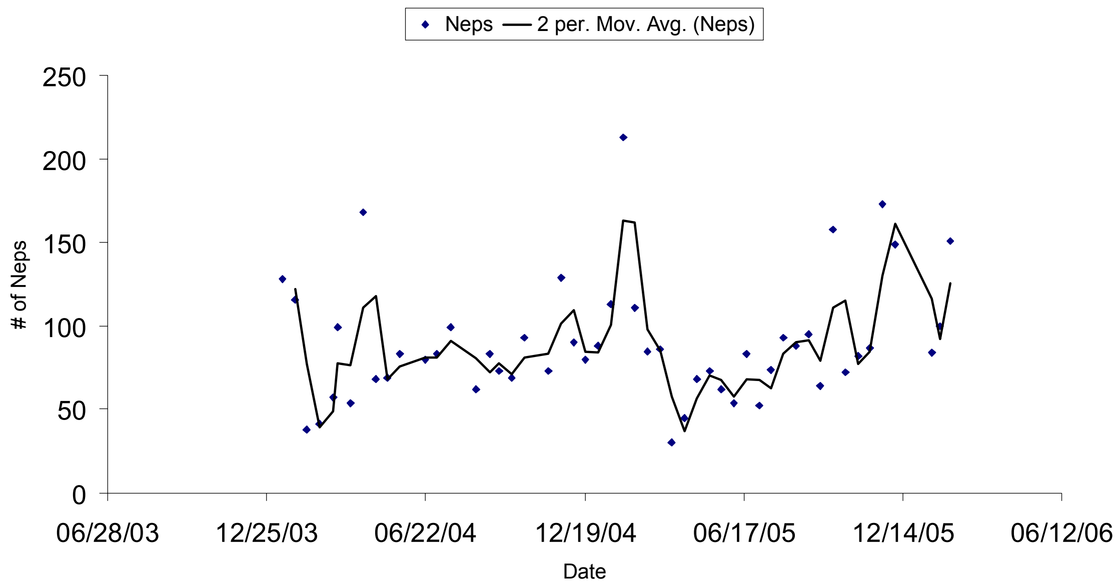


Figure 5.10: Neps from a card in line 1

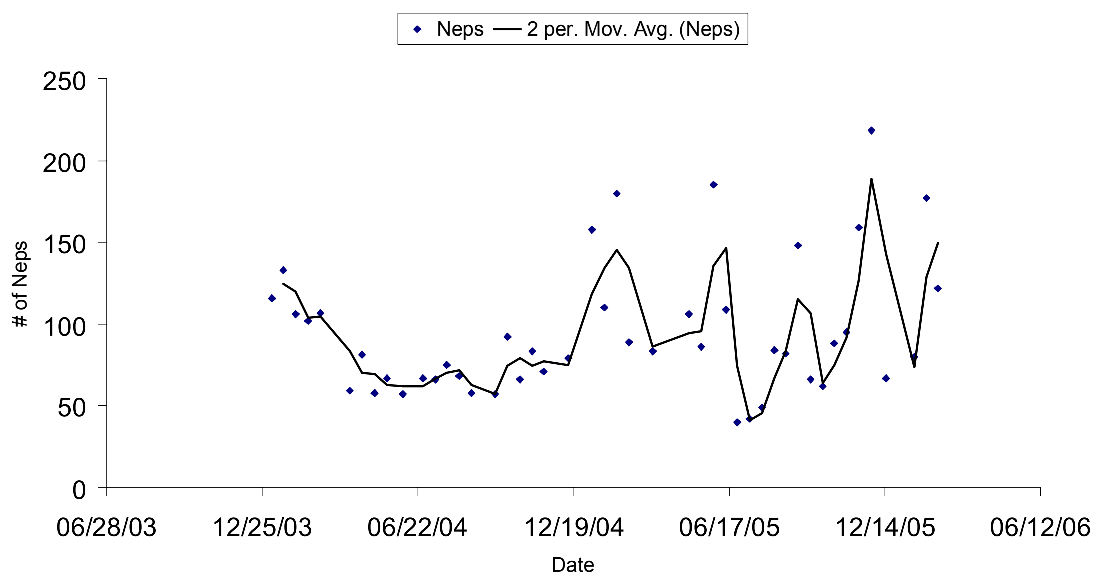


Figure 5.11: Neps from a card in line 2

The purpose of these graphs was to look at the trend in neps specifically for several machines. This would give more detail than comparing by an entire line of cards. These two graphs represent the general trend in neps for all the cards. The trend is that neps gradually increase over time until the card is re-clothed. Re-clothing a card means replacing the card wire at designated intervals based on the number of pounds run. The decreases in the neps correspond with dates that the particular card was re-clothed. The card in figure 5.10 was last re-clothed on 3/16/05 and the card in figure 5.11 was last re-clothed on 6/21/05. The only information available was from the dates of the last re-clothing. This proves that the number of neps removed is directly related to the machine performance.

5.5.1.2 Online Spinning Data

For the online spinning data, the parameters studied were short thick stops, long thick stops, moiré stops, and quality stops. This data shows information from several different spinning frames running different yarn styles and counts. Moiré stops and quality stops are discussed here, but examples of graphs for short thick stops and long thick stops can be found in Appendix D.

Figures 5.12 and 5.13 show moving averages for moiré stops on two different spinning frames over the last six months. Both graphs show trends of increases and decreases, although not necessarily at the same times. The data showed no unusual activity that needed to be investigated further. There is a period of time when there are no data points, which corresponded with the plant shutting down over the holidays.

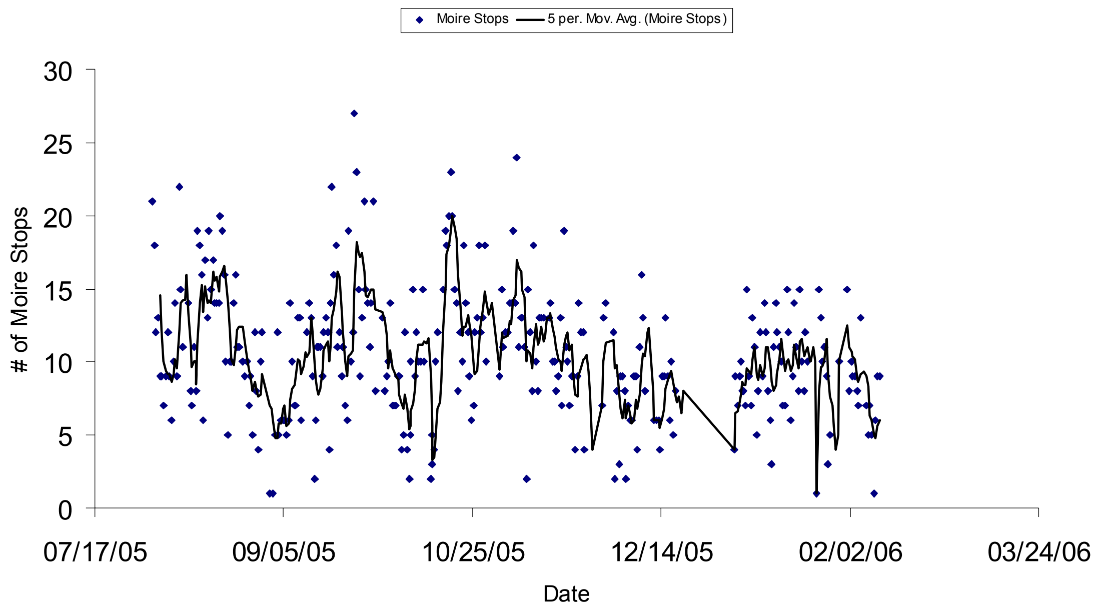


Figure 5.12: Moiré stops for spinning frame A

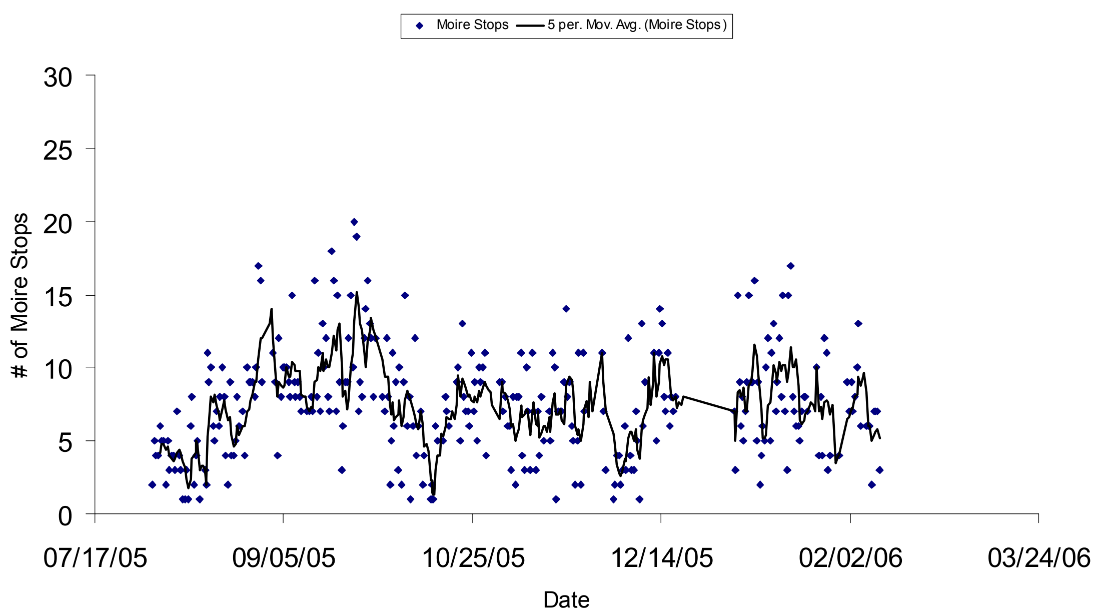


Figure 5.13: Moiré stops for spinning frame B

Figures 5.14 and 5.15 show quality stops for two different spinning frames. Both graphs show outliers at or around the same day. This is unusual and could be contributed to several factors. There could have been a problem with the monitoring system, could have been machine malfunctions, or most likely, the sliver from the cards had quality problems. Looking at the card data for the same week, there was no unusual activity noted to explain this large of an increase in quality stops for spinning. However, it is difficult to support any assumptions since the AFIS data is only collected once a week for one card line.

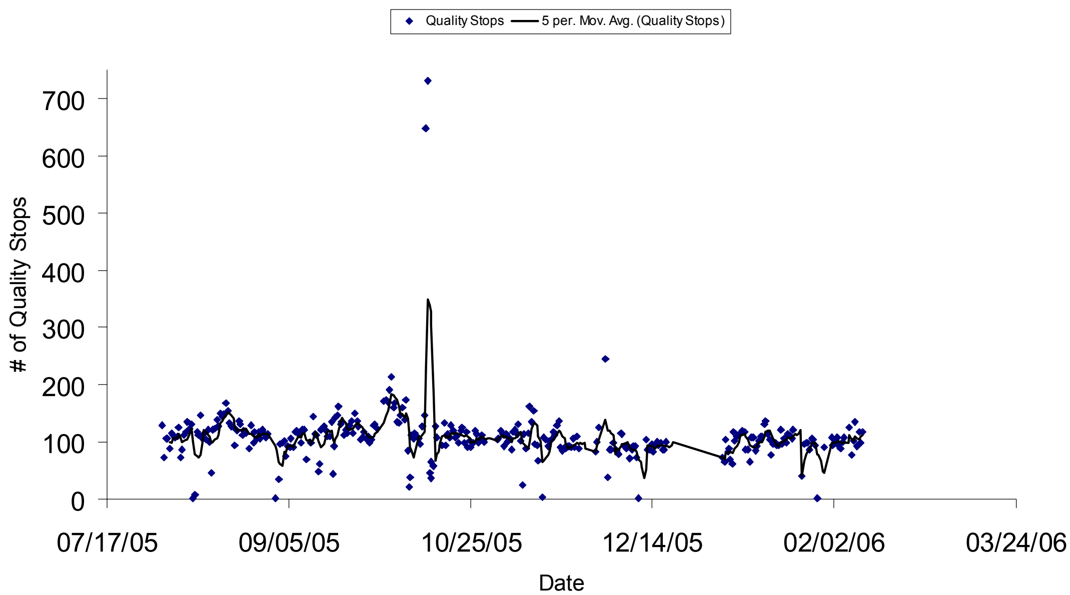


Figure 5.14: Quality stops for spinning frame A

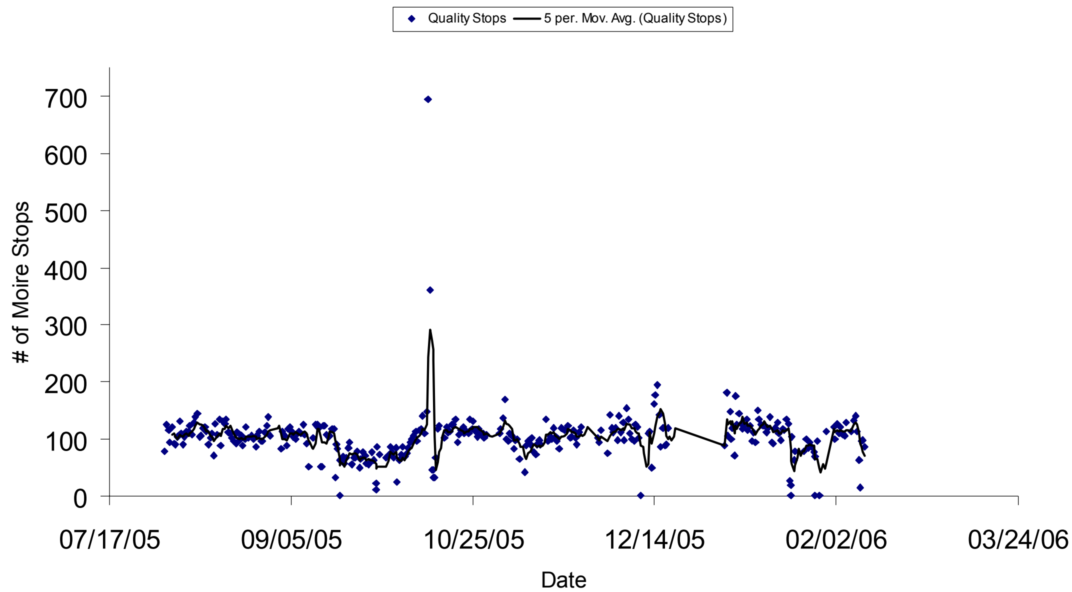


Figure 5.15: Quality stops for spinning frame B

5.5.1.3 Combining the Data

In order to investigate the relationship between process performance and final product quality (RO5) a method of comparing the different data sets was determined. A table was created in Excel that would combine elements from several data sets. To do this, it was decided to use the online spinning data as the standard data to be analyzed since it was collected every day. There are various measurements taken at different dates and times for the different processes. In order to plot data against time, all measurements taken each day in a data set needed to be averaged.

For the online spinning data, more cleaning was required before the data could be used further. Since there were two measurements for each day, these numbers would be averaged together. The problem observed is the yarn counts change frequently and the data for each yarn count varies. Also when averaging for a day, yarn counts cannot be averaged. The data was sorted and counted to determine which yarn style had the most data points. The data from this one yarn was the only data used from the online spinning data for this analysis. This data needed to be checked visually and quality standards set. The data was first sorted by total stop time. These times varied from 0 (none) to 12 (the entire shift) hours. It was decided that all rows with a stop time of more than 5 hours would be deleted. The data was then sorted by yarn count. Any values that contained a "0" for yarn count as well as "0" for production were removed. A quality stops/1000lb and moiré stops/1000lb column was created so as to take into account the variation in production each day. The averages of the properties for each date were calculated, and these numbers put into a separate table and sorted by date. The machine column was then removed because data was collected from all of the spinning frames and the machine numbers could not be averaged.

There was two years of data from the AFIS, so only the data from the last six months was used, in order to match the online spinning data time frame. The data from the AFIS contained several measurements in a given day, one for mat and one sliver for each card. The mat data was deleted, leaving only data from sliver. All the data points for each property were averaged for each date. Since AFIS was only measured once a week, there would only be data points for one day each week. Averaging the data from all the cards is useful in analysis since the sliver from different cards is mixed and blended at each drawframe. These averages were copied and pasted into the table containing the spinning data averages. The AFIS values had to be placed in the appropriate dates matching the spinning data.

The problem when comparing data from the offline testing against data from the online testing is the frequencies of measurement are different. The AFIS testing was only performed once a week, whereas the spinning data was recorded every day. This means there is more spinning data than offline card data.

The HVI data was averaged for each day. Since there were 96 bales for each laydown, 96 rows of values were averaged for each day there was a laydown. The averages were useful since the purpose of the bale laydown is to mix the different properties from each bale. These averages were input into the table next to the card AFIS data at the corresponding dates to match the online spinning dates. The same was done for the yarn offline data of that particular yarn style. This data was put into the master table beside the HVI data.

Each data set contains a different number of measurements for a given date, the collection times are different, and the frequency of measurements is different. The various measurements taken at different dates and times is shown in table 5.1.

Table 5.1: Frequency data sets collected and data points averaged for each date

Data Set	Frequency	Data Points Averaged
HVI	Each laydown (anywhere from one to several days)	96
AFIS from Cards	Once a week	6-7
Online Spinning	Twice daily	2
Offline Spinning	Once a week	6

These four datasets were combined over the time period from August 2005 to February 2006. The online spinning data for this time period was for several different yarn styles and counts. The data used was from the two yarn styles run most frequently, and therefore would have the most data points. Figure 5.16 shows a partial sample taken from this combined dataset.

Start date	Style	Ne	lbs	Total stop time	Quality stops	QStops/1000lb	Nep Size [um]	Card Neps	L(w) [in]	L(w) CV [%	UQL (w) [in	SFC(n) [%]	HVI Mic	HVI Length	HVI Unif	HVI Str
08/01/05	2216	12	2355.75	0:19	111.25	47.18071412	565	101.33333	0.9267	32.216667	1.125	19.583333				
08/02/05	2216	12	2364.225	0:16	112.25	47.46904801										
08/03/05	2216	12	2352.15	0:20	114.5	48.66091208										
08/04/05	2216	12	2318.25	0:30	105	45.27637333										
08/05/05	2216	12	2331.075	0:26	101.75	43.63642362										
08/06/05	2216	12	2306.85	0:33	112	48.5125322										
08/07/05	2216	12	2338.925	0:24	120.25	51.32535276										
08/08/05	2216	12	2351.075	0:20	124	52.70023358	549.4285714	88.166667	0.92	32.971429	1.1257143	20.842857	4.5567	1.0720619	812.062	28.577
08/09/05	2216	12	2346.675	0:21	125.5	53.49305716							4.567	1.0668557	811.546	28.497
08/10/05	2216	12	2340.125	0:24	118.75	50.71093365										
08/11/05	2216	12	2328.8	0:27	114	48.97439288							4.5485	1.0669072	813.711	28.966
08/12/05	2216	12	2023.025	0:52	90.75	44.86718142							4.5557	1.0684536	809.33	28.803
08/13/05	2216	12	2032.15	0:46	98	48.02363224										
08/14/05	2216	12	2040.575	0:43	75.75	37.24863075										
08/15/05	2216	12	2094.25	0:25	90.25	43.18295477	545.6666667	88.166667	0.9267	32.716667	1.1283333	20.35	4.5505	1.0752062	811.804	29.175
08/16/05	2216	12	1955.6	1:15	97.75	49.24483244							4.55	1.0807732	810.052	29.482
08/17/05	2216	12	2086.4	0:28	111.75	53.51203882							4.5552	1.0695361	812.062	28.761
08/18/05	2216	12	2069.325	0:34	103.25	50.39328611							4.5438	1.077268	813.608	28.908

Figure 5.16: Combined data sets

To reflect an estimated lag time of 3 days, another data set was created from the original shown in figure 5.16. All of the online and offline spinning data was shifted down 3 rows to represent a 3 day process time from bale laydown to yarn. The AFIS data was shifted down 2 rows to represent a lag from the bale laydown to the end of the card. The HVI data was kept the same since it is the beginning of the manufacturing process.

Analysis was done to compare yarn data elements against data elements from the bales and card. This was to determine if any relationships could be seen from the fiber and sliver that might affect the yarn performance. When viewing the graphs, it was important to keep in mind the effect of lag time. There is an estimated lag time of 3 days from beginning bale laydown to the finished yarn. Any changes in the HVI or card properties would have a delayed effect on the yarn properties. To notice any direct relationships affecting yarn properties, the elements were compared using xy charts. A horizontal line represents no relationships between elements, an upward line represents a positive relationship, and a downward line represents a negative relationship. The yarn properties chosen were quality stops, quality stops per 1000 pounds, CVm, and moiré stops. The relationships with fiber and sliver properties are shown in the following figures as well as in Appendix D.

Figure 5.17 shows the relationship of quality stops and neps over a six month period. The number of neps after carding should affect the quality stops in spinning since neps increase the number of yarn breakages, but this is not evident in the graph. The red circle shows what an expected trend should look like when there is a spike in neps, it is followed by an increase in quality stops. Figure 5.18 shows the direct relationship of quality stops and neps. From what is known, this should be a positive relationship, so as neps increase the quality stops should increase. The graph shows no relationship between the elements. Figure 5.19 shows the xy chart with the lag also demonstrating no relationship. This could mean that the data is incorrect, there is not enough data, or the lag was estimated incorrectly. These are just a few possible assumptions.

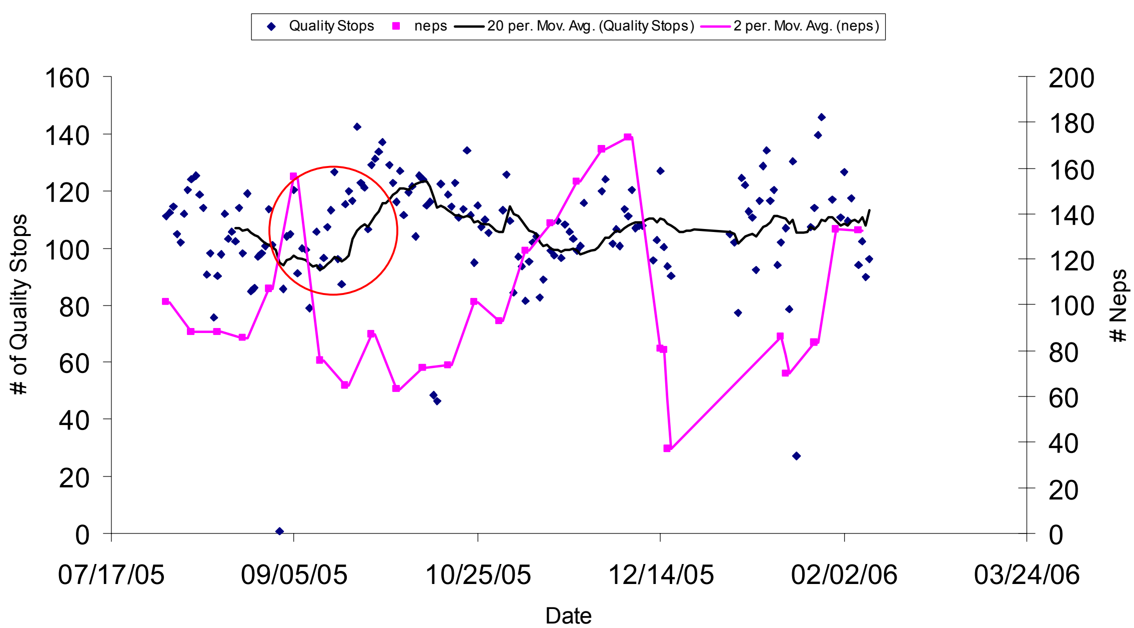


Figure 5.17: Quality stops & card neps vs. time

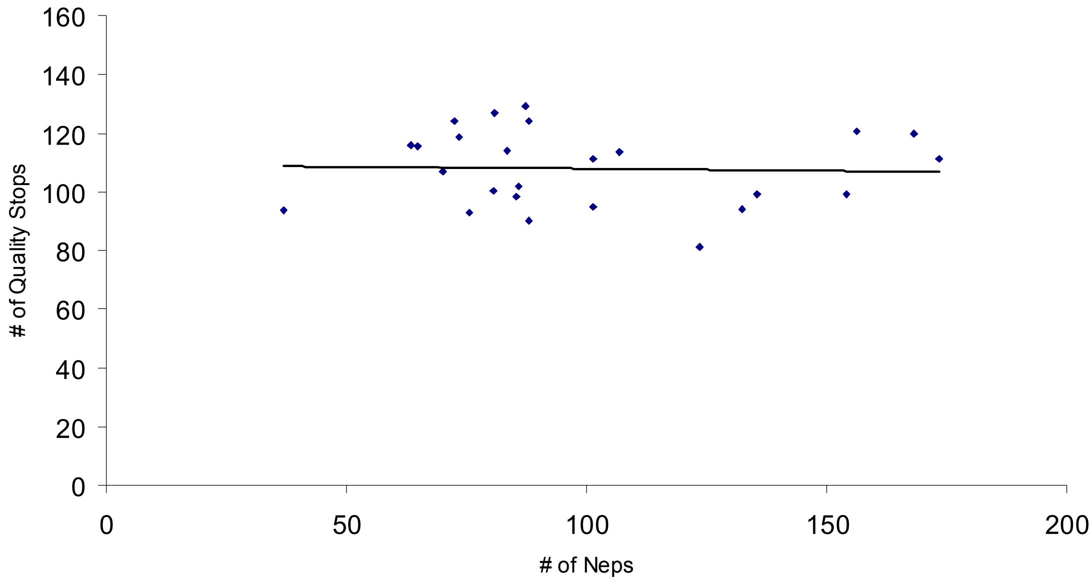


Figure 5.18: Quality stops vs. card neps

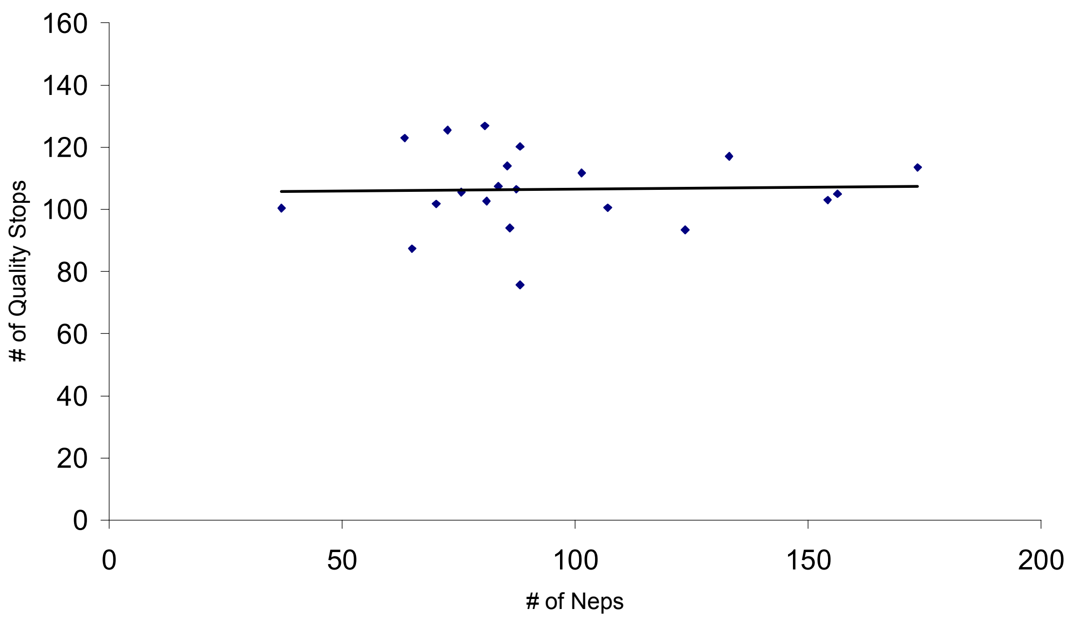


Figure 5.19: Quality stops vs. card neps with lag

Figure 5.20 shows the relationship of quality stops per 1000 pounds and HVI trash leaf over six months. The trend lines seem to almost follow the same increases and decreases. This indicates there may be a relationship between the two elements. This was further investigated by plotting a xy chart. Figure 5.21 shows no direct relationship between quality stops per 1000 pounds and HVI trash leaf due to the horizontal line. Figure 5.22 also shows no relationship with the lag included. The data could mean there is no relationship between these two elements, or there could be other factors such as not enough data.

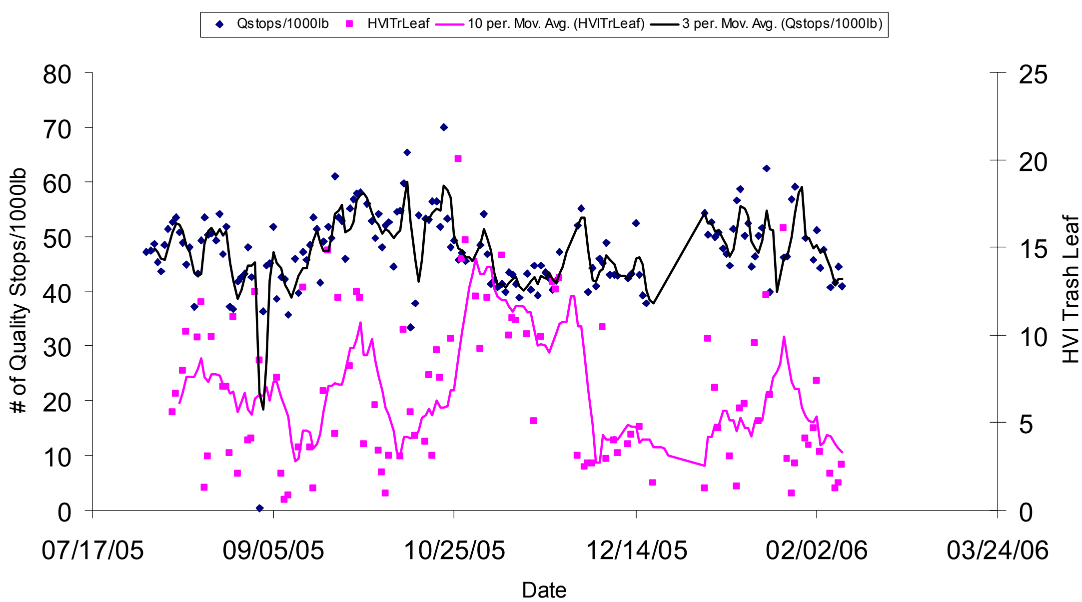


Figure 5.20: Quality stops/1000lbs & HVI trash leaf vs. time

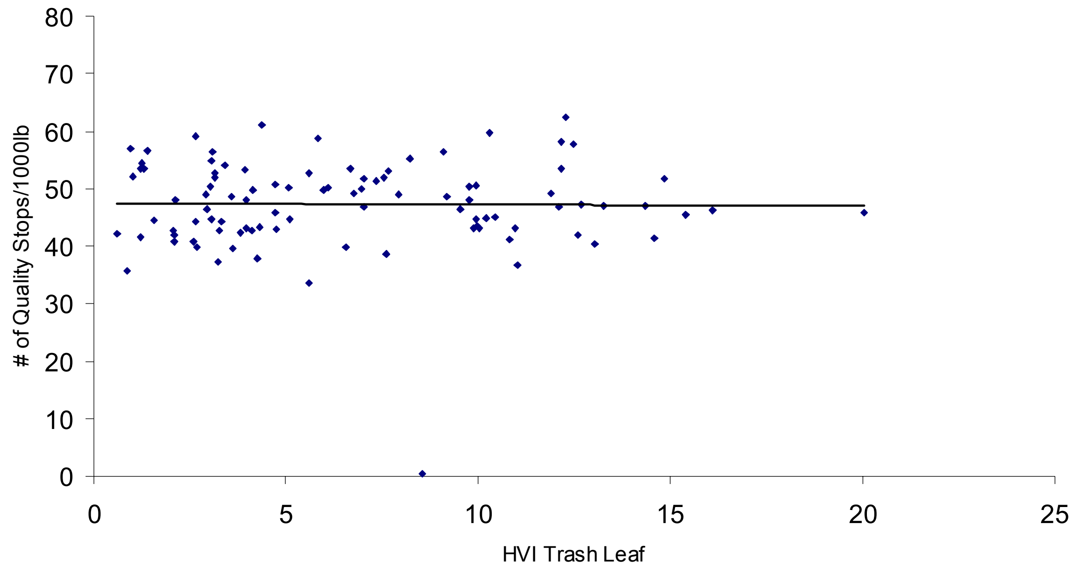


Figure 5.21: Quality stops/1000lbs vs. HVI trash leaf

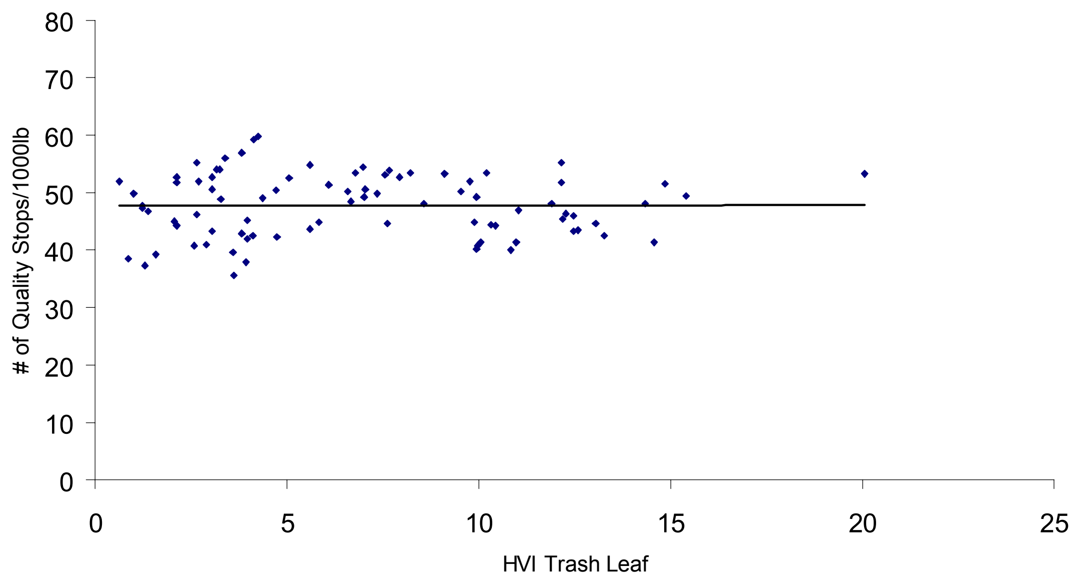


Figure 5.22: Quality stops/1000lbs vs. HVI trash leaf with the lag

Figure 5.23 represents a slight relationship between the yarn coefficient variation of mass and card trash over time. There should be an increase in variation of mass when there is an increase in trash. Figure 5.24 proves that there is a strong positive relationship between the two elements. As card trash increases, the variation in mass of the yarn increases. A xy chart representing the lag could not be constructed because many of the dates did not match up in the shifted dataset, leaving only 2 points to graph.

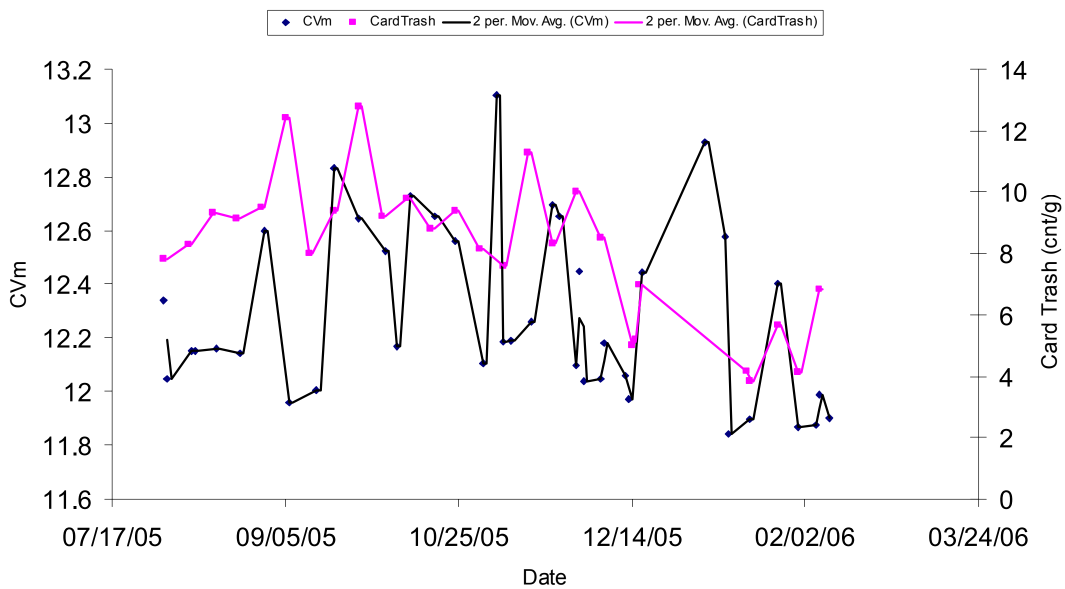


Figure 5.23: CVm & card trash vs. time

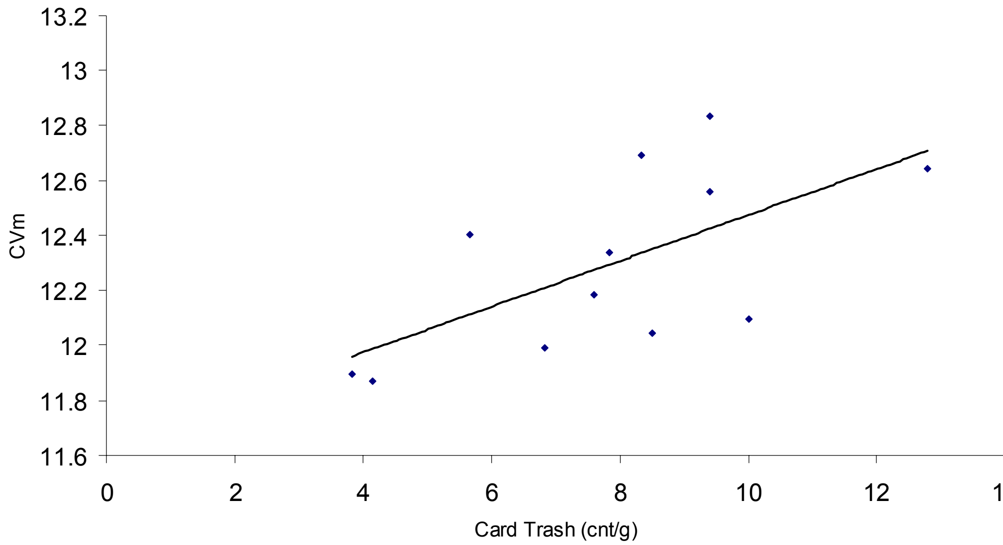


Figure 5.24: CVm vs. card trash

Figure 5.25 shows the relationship between moiré stops and card neps over six months. There is no obvious relationship found in the trend lines. Figure 5.26 shows the direct relationship of the two elements. The horizontal line demonstrates that there is no relationship between the two elements. Figure 5.27 shows a slight positive relationship with the lag. This means that as the number of neps increase, the number of moiré stops in spinning also increases slightly.

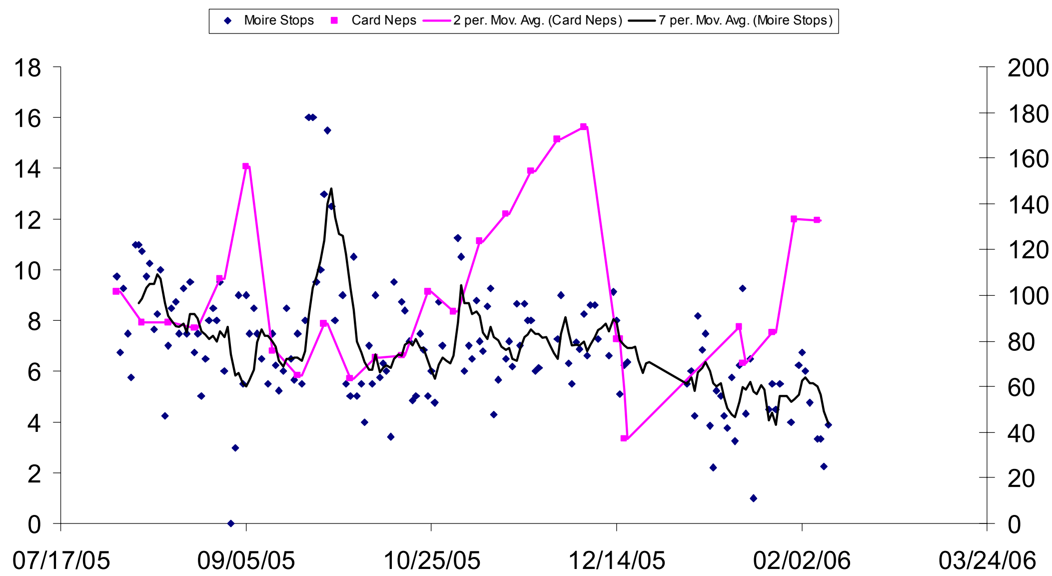


Figure 5.25: Moiré stops & card neps vs. time

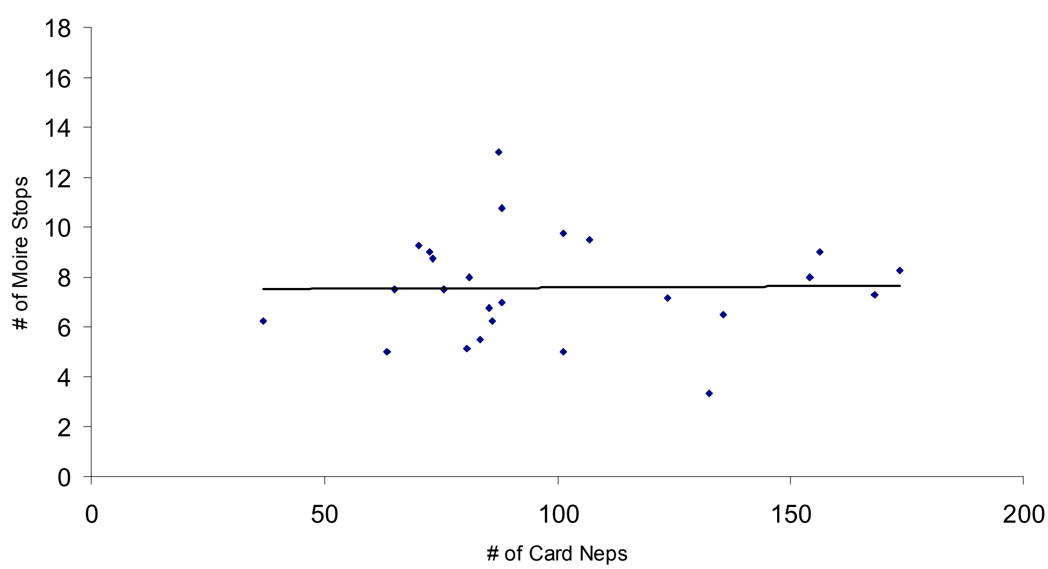


Figure 5.26: Moiré stops vs. card neps

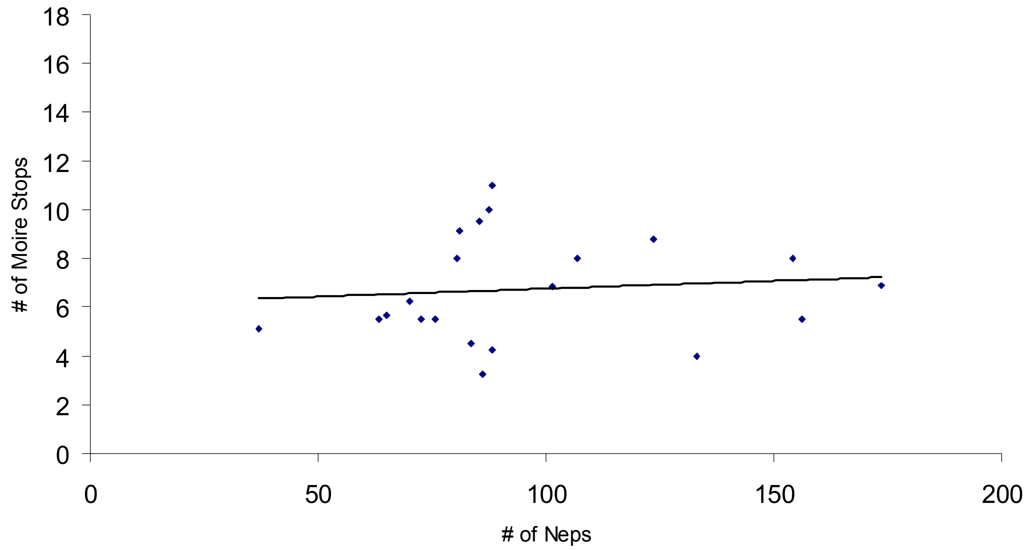


Figure 5.27: Moiré stops vs. card neps with the lag

Figure 5.28 shows moiré stops and card trash over a six month period. Both moving averages seem to follow the same general trend. Figure 5.29 shows that there is a direct relationship between the two elements. A positive relationship here means that as card trash increases, moiré stops increase. Figure 5.30, with the lag, also shows a positive direct relationship, although not as strong.

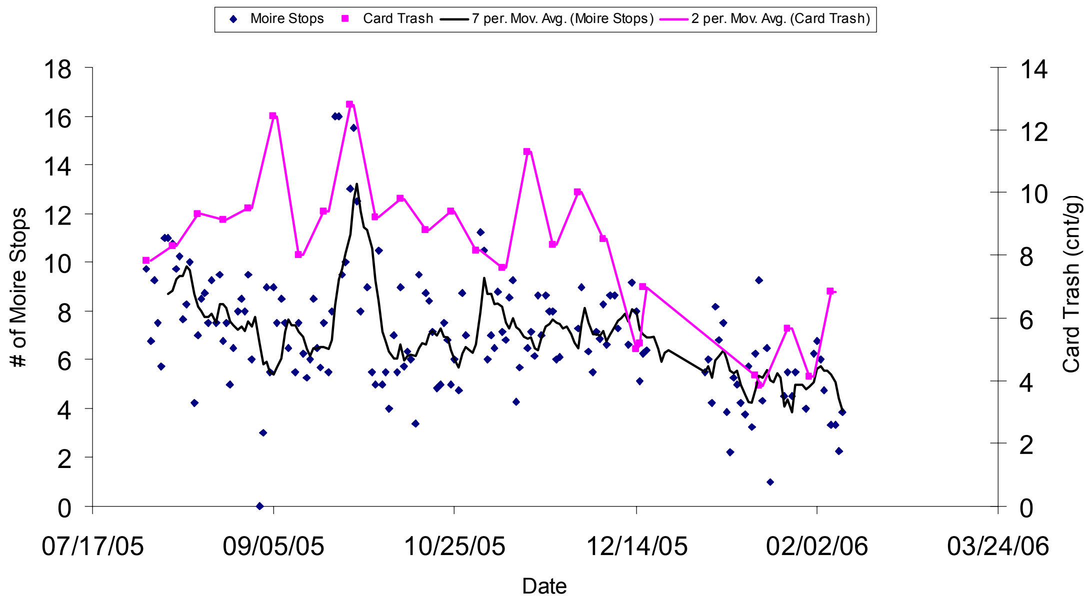


Figure 5.28: Moiré stops & card trash vs. time

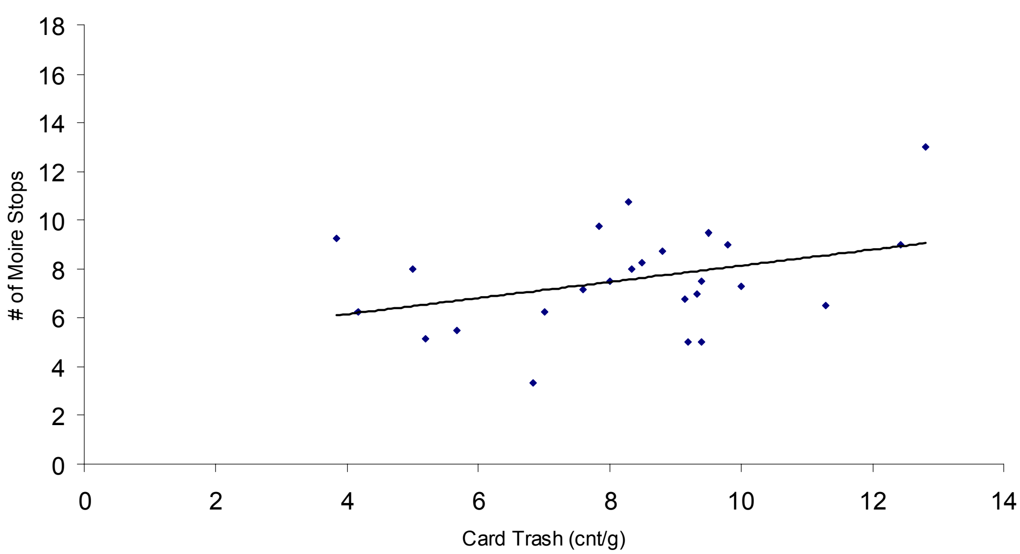


Figure 5.29: Moiré stops vs. card trash

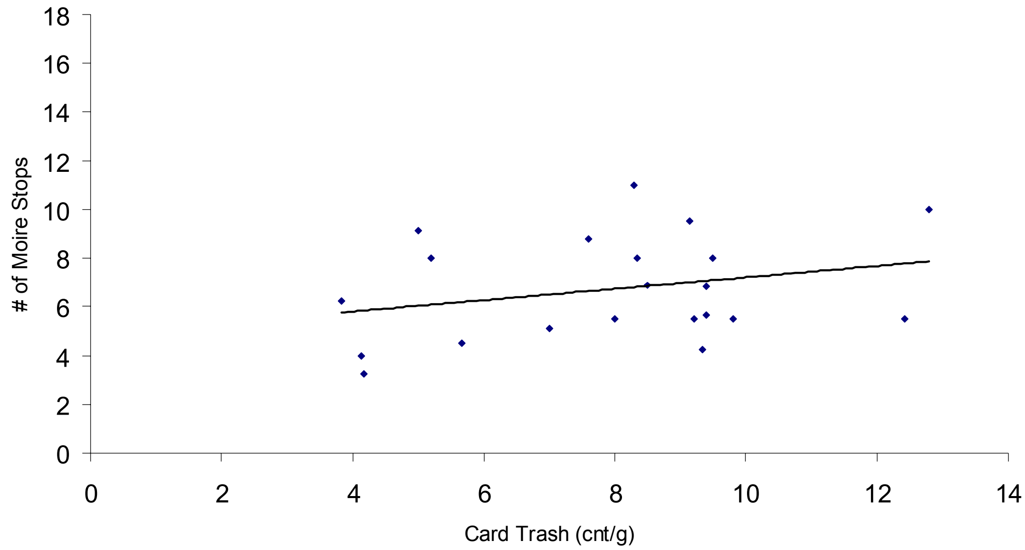


Figure 5.30: Moiré stops vs. card trash with the lag

5.5.2 Level II Analysis

In order to perform analysis on complex relationships, multiple regression in SAS JMP 6 was used. SAS Enterprise Miner contains many data management tools as well as a variety of data mining tools, which were not needed for the analysis. SAS Enterprise Miner is also prepared to handle terabytes of data and the combined dataset for this analysis is very small. Multiple regressions are useful to look at several data elements at the same time to see which best predict a response variable. The same combined data sets, the original and 3 day lag from the Excel analysis, were used for this analysis. Table 5.2 shows the number of data points in each dataset with and without the lag taken from yarn style A.

Table 5.2: Data points for dataset with and without the lag

Datasets	Data Points with NO Lag	Data Points with the Lag
Online Spinning AFIS Card HVI	22	12
Online Spinning AFIS Card	24	21
Online Spinning HVI	96	83
HVI Offline Spinning	29	23

As can be seen, when trying to find any relationships between the datasets, online spinning and HVI provides the largest number of data points for analysis. The problem when using the other combinations is a very limited number of data points. With so few points, the accuracy of analysis is questionable. For this reason, the analysis was performed on the combined online spinning and HVI data for two yarn styles, with and without the lag.

The yarn properties investigated were quality stops/1000lb and moiré stops/1000lb. HVI micronaire, length, uniformity, strength, Rd, plus b, trash % area, and trash leaf were all used as potential predictors for yarn properties. The purpose of these analyses was to explore the data to find any relationships between fiber properties and both yarn quality stops/1000lb and moiré stops/1000lb.

5.5.3 Discussion

What is seen from the different Excel charts is that sometimes slight trends can be seen, and sometimes not. The problem is the data is inconsistent for the different stages of processing. This makes it difficult to determine which parameters affect the yarn performance. Understanding what affects yarn properties can be seen from this equation:

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \cdots a_nx_n$$

In the equation, Y is some yarn or spinning parameter (e.g. quality stops or moiré stops); x is some measured variable; a_0 is the intercept and a_1, a_2, a_3, etc is the slope for each corresponding x value. When the slope is zero then there is no effect of x on y. When plotting $Y=a_0 + ax_1$, there may not be a visible effect or relationship because the other elements are not taken into account. For example, yarn strength depends on several elements such as fiber strength, fiber fineness, fiber length, and extension. It is dangerous to assume that a strong fiber equals a strong yarn. The assumption is only true if the other parameters did not change. Studying one variable to see if there is an effect on the yarn may not be of much use. The benefit of data mining is all elements are taken into account and compared to evaluate which have the most affect on the target variable, which in this case would be some yarn or spinning parameter (Y).

In order to test this assumption, stepwise regression was applied to the combined data set with and without a lag of 3 days. This analysis determined what percent of the variation of quality stops and moiré stops could be predicted by some sliver or fiber properties x. Relationships were found for each experiment, but not necessarily the same when comparing no lag to a 3 day lag time. It is difficult to evaluate the precision of the results since there was limited amount of data and the lag time was estimated.

Multiple regression takes into account several elements x in order to predict some variable y. This allows for experimentation with complex data sets. The Level I analysis was limited in this feature, as specific x and y elements were tested against each other with the assumption they were independent. For this reason, multiple regression should prove to provide more accurate results. A problem with this analysis was the estimation of lag time. Since the three days was estimated, it is not known whether or not it was correct. Creating the lag times was difficult due to the varied number of measurement times for each data set.

Table 5.3 shows the frequency of collection/storage times for each of the data sets used for the analysis. As can be seen, the frequencies vary for each dataset. This resulted in different numbers of measurements for different dates, making it difficult to compare between the datasets. The online spinning data is constantly updated in the software, since it can be seen as instantaneous, however the shift averages are only stored twice a day. For HVI, AFIS, and offline spinning the frequency is the collection and storage of the data.

Table 5.3: Collection/Storage frequencies of the data sets

Data Set	Frequency
HVI	Each laydown (anywhere from one to several days)
AFIS from Cards	Once a week
Online Spinning	Twice daily
Offline Spinning	Once a week

When data mining, it would be beneficial for there to be consistent frequencies of data collection/storage at each collection point. When comparing the different elements, results were inconclusive due to these inconsistencies. Online spinning data was collected twice daily, which means this dataset contained the most data points. The AFIS data from carding was only collected once a week, which means there was only one data point for every seven from the online spinning.

Data mining could prove to be very successful in finding unknown relationships and trends within cotton spinning data. It would be beneficial to collect data every day from each collection point over a set time period, such as six months. This would provide enough data to analyze so that the results would prove useful and accurate.

For more precise results, it would require more accurately estimating the lag times. The purpose of cotton spinning is to mix the cotton as much as possible, making it very hard to track throughout the process. A way of tracking the cotton would be to code each can and track the placements for each process. This would provide an understanding of where each can is at all times.

There continues to be a excess of data in yarn manufacturing, however the data collection systems were not designed with the intentions of applying data mining techniques. An opportunity does exist with the current data collection systems for predictive analysis. As can be seen from the data analysis, results are provided even with a limited amount of data. Improving on areas such as inconsistent collection frequencies and lag times should only improve the value of the analysis.

6 Conclusions and Recommendations for Future Studies

The focus of this research was to explore the applicability of data mining in yarn manufacturing. The methodology involved interviews with textile spinning companies as well as a case study with one cotton open end spinning plant that met the specifications. Data was collected from several different processes and data collection systems. The following section provides a summary of the results from each research objective.

6.1 Conclusions

1. Data collection requirements for textile spinning vary according to the different plants, machine vendors, and software vendors. Machine and software vendors set up their equipment and software to collect certain data elements. These cannot be changed and can only be used to the abilities set by the vendors. Different plants use different machine and software vendors, therefore collecting different data.
2. There are several data quality issues that exist in textile spinning. Data that is automatically downloaded is typically not a problem, unless there is a machine or software error. Data that is input manually may introduce operator error. Not all data is accessible or can be exported to a database such as Excel. Data exported to Excel leads to formatting issues, missing values, duplicate values, and incorrect values. Plus these files are separate files that need to be combined. Since data is recorded from different machines and data

systems, the measurement times are different. This is a quality concern when comparing data from different processes.

3. The data elements needed for monitoring and controlling product quality are those set by the machine and software vendors. These capabilities and settings are predetermined and cannot be changed. The software is preset to measure and record certain data elements. A list of the data elements limited to the case study plant can be found in Appendix B. It is up to the discretion of each plant to utilize the capabilities of the different data monitoring systems.
4. There is little or no exchange of data between the different data analysis systems. The data system vendors did not set up the software to communicate with other software packages from other vendors. Each machine and software vendor holds their information proprietary, with no sharing.
5. When investigating the relationship between process performance and final product quality, the results were inconclusive. It was difficult to draw any concrete conclusions from the data since the collection times were different and the amount of visible data was inadequate. Yarn performance is affected by several elements. It is difficult to draw any conclusions from comparisons of just one or two elements, without taking into account all the elements.
6. There is potential applicability of data mining in cotton yarn spinning. However, limited data collection systems and infrequent collection times between the different systems allows for difficult analysis. Relationships can

be seen from the different analyses. However, it is unknown how precise the results are due to limited amount of data and an estimated lag time.

6.2 Future Studies

For future studies, it might prove beneficial to collect data at the same frequencies across the spinning process. Collecting this data every day for a given time period of three to six months would provide an sufficient amount of data to be used for data mining. This would provide more accurate results considering the increase in number of data values for each date. Data mining software, such as Enterprise Miner or even Jmp, is designed to handle very large amounts of data. The techniques would prove more successful with a greater number of data points.

It might be useful to designate a person be part of the data collection process at a plant. By doing this, one could better understand the capabilities of the different collection systems as well as limitations. A person could also manually collect data from the machine panels at different process stages at a regular interval. This would provide a greater amount of data, frequencies could be controlled, and the data elements could be chosen. With a larger dataset, more data mining techniques could be applied. For this study, regression was the only technique used due to the limited number of data points. It would be interesting to see results from several other techniques.

It would be interesting to determine a means of tracking the cotton throughout the manufacturing process. This would give more exact lag times, which would result in a more effective analysis. It is difficult to estimate lag times for a spinning plant, especially one that changes yarn counts frequently. For this reason, conducting this type of research with more than one plant would be valuable for comparisons. Some plants change their yarn counts more frequently than others, making these difficult to study and estimate the lags. It might be more useful to focus on a plant that is limited to only a few yarn counts. Different plants also have different data collection systems and vary in the amount of online and offline data they collect. Studying several plants would provide very interesting comparisons.

References

Agosta, L. (2004). The future of data mining – Predictive analytics. *DMReview*. Retrieved March 25, 2005, from http://www.dmreview.com/article_sub.cfm?articleID=1007209

Barco. http://www.barco.com/textiles/

Beal, B. (2005). Application vendors to dig into data mining. *SearchCRM.com*. Retrieved September 1, 2005, from http://searchcio.techtarget.com/originalContent/0,289142,sid19_gci1047637,00.html

Berry, M. & Linoff, G. (1997). *Data mining techniques for marketing, sales, and customer support.* New York: John Wiley & Sons.

Berry, M. & Linoff, G. (2000). *Mastering data mining. The art and science of customer relationship management.* New York: John Wiley & Sons.

Black, A. (2005). Customer intelligence survey review: How are companies responding to the demand for more meaningful customer insights. *SAS Data Mining Technology Conference*. October 24-25, 2005.

Braha, D. (2001). *Data mining for design and manufacturing.* Norwell, MA.: Kluwer Academic Publishers

Coppock, D. (2003). Data mining and modeling: Application of data mining techniques. *DMReview*. Retrieved March 25, 2005, from http://www.dmreview.com/article_sub.cfm?articleID=7191

Cotton Incorporated. http://www.cottoninc.com/

Data Mining. *Megaputer*. Retrieved April 7, 2005, from http://www.megaputer.com/dm/dm101.php3

Data mining industry: Overview and future. (2005). *Intertangent Technology Directory*. Retrieved April 1, 2005, from http://www.intertangent.com/023346/Articles_and_News/1473.html

Dubois, L. Data quality in the integrated enterprise. *TDWI*. Retrieved November 3, 2005, from http://www.tdwi.org/Publications/display.aspx?id=6457&t=y

Henderson, D. (1998). Overcoming data integration problems. *DM direct newsletter*. Retrieved February 8, 2006, from http://www.dmreview.com/article_sub.cfm?articleId=924

Hodge, G. & Oxenham, W. (2002). Information engineering: Textile industry's value-adding key to effective decision-making. *National textile center annual report*. Retrieved October 15, 2005, from http://www.ntcresearch.org/pdf-rpts/AnRp02/S99-NS10-A2.pdf

International Organization for Standardization. (1998). *ISO 10782-1: Definitions and attributes of data elements for control and monitoring of textile processes.* First Edition.

Kdnuggets: Polls: Data Mining Techniques. (2005). Retrieved April 5, 2005, from http://www.kdnuggets.com/polls/2005/data_mining_techniques.htm

Kuonen, D. (2005). Is data mining for gold "statistical déjà vu"? *Statoo Consulting*. Retrieved April 1, 2005, from http://www.crm2day.com/library/EEpZElkAkyJMPQLSal.php

Leavitt, N. (2002). Industry trends: Data mining for the corporate masses? *Leavitt Communications, Inc.* Retrieved April 1, 2005, from http://www.leavcom.com/ieee_may02.htm

McCreight, D., Feil, R., Booterbaugh, J. and Backe, E. (1997). Short staple yarn manufacturing. *Carolina Academic Press.*

Moss, L. T. (2003). Defining data mining. *BusinessIntelligence.com.* Retrieved March 25, 2005, from http://www.businessintelligence.com/ex/asp/code.64/xe/article.htm

Nisbet, R. (2004). How to choose a data mining suite. *DM Direct Special Report*. Retrieved February 29, from http://www.dmreview.com/article_sub.cfm?articleID=1000465

Schertel, S. L. (2002). Data mining and its potential use in textiles: A spinning mill. *Dissertation, North Carolina State University*

Shearer, C. (2004). Anticipating consumer behavior with analytics. *SPSS Inc.* Retrieved March 16, 2005, from http://www.crm2day.com/library/EEplpkZllyASiURqyN.php

Smith, G. (2005). The road to data mining: A WWF case study on business intelligence in the non-profit sector. Information Technology World Wildlife Fund, Washington, DC. *SAS Data Mining Technology Conference*. October 24-25, 2005.

Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Education, Inc.

Thearling, K. (2004). An introduction to data mining. *Data warehouse knowledge base.* Retrieved March 10, 2005, from http://www.thearling.com/text/dmwhite/dmwhite.htm

The business intelligence and data warehousing glossary. Retrieved January 30, 2006, from http://www.sdgcomputing.com/glossary.htm#DIKSpectrum

Two Crows Corporation. (1999). *Introduction to Data Mining and Knowledge Discovery.* (3rd ed.) Retrieved April 18, 2005, from http://www.twocrows.com/intro-dm.pdf

United States Department of Agriculture. (2005). *Cotton classification: Understanding the data.* Retrieved February 18, 2006, from http://www.ams.usda.gov/cotton/Word%20documents/Understanding%20the%20Data%20%20Revised%20April%202005.doc

Ye, Nong. (2003). *The handbook of data mining*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Appendices

Appendix A: USDA definitions for HVI properties (adapted from USDA)

- **Gin Code Number** – Composed of five digits. The first two digits denote the Classing Office, and the last three digits identify the gin. Code assigned by local Classing Office.

- **Gin Bale Number** – Seven digit bale numbers assigned by the gin.

- **Permanent Bale Identification (PBI) tag** – A bar-coded identification tag placed on a bale, consisting of the gin code number and gin bale number.

- **Date Classed** – The date the bale was classed in the Classing Office. The format is YYYYMMDD.

- **Module, Trailer, or Single Bale** – One-digit code that indicates whether the sample was outturned as a single bale or as a bale that was module/trailer averaged. (Single bale = 0, Module = 1, Trailer = 2)

- **Module/Trailer Number** – A five-digit number assigned at the gin.

- **Bales in Module/Trailer** – A two-digit number identifies the number of bales in the module/trailer that were released with the module average calculation.

- **Official Color Grade** – Color refers to the gradations of grayness and yellowness in the cotton. Upland is determined by instrument based measurements and Pima is determined by a human classer.

Upland Color Grades	**Pima Color Grades**
11, 21, 31, 41, 51, 61, 71, 81 12, 22, 32, 42, 52, 62, 82 13, 23, 33, 43, 53, 63, 83 24, 34, 44, 54, 84 25, 35, 85	1, 2, 3, 4, 5, 6, 7

Special Condition Codes for Upland Cotton	
96	Mixture of Upland and Pima
97	Fire Damaged
98	Water Damaged

Special Condition Codes for Pima Cotton	
93	Mixture of Pima and Upland
94	Fire Damaged
95	Water Damaged

- **Fiber Staple Length (32^{nds} of an inch)** – Classification instruments measure length in hundredths of an inch. Length is reported on the classification record in both 32nds and 100ths of an inch. Length measurements are converted to 32nds of an inch as shown below:

Upland Length Conversion Chart			
Length (32nds)	**Length (Inches)**	**Length (32nds)**	**Length (Inches)**
24	0.79 & shorter	36	1.11 – 1.13
26	0.80 – 0.85	37	1.14 – 1.17
28	0.86 – 0.89	38	1.18 – 1.20
29	0.90 – 0.92	39	1.21 – 1.23
30	0.93 – 0.95	40	1.24 – 1.26
31	0.96 – 0.98	41	1.27 – 1.29
32	0.99 – 1.01	42	1.30 – 1.32
33	1.02 – 1.04	43	1.33 – 1.35
34	1.05 – 1.07	44 & +	1.36 & +
35	1.08 – 1.10		

American Pima Length Conversion Chart	
Length (32nds)	**Length (Inches)**
40	1.20 and lower
42	1.21 – 1.25
44	1.26 – 1.31
46	1.32 – 1.36
48	1.37 – 1.42
50	1.43 – 1.47
52	1.48 & +

- **Micronaire** – Cotton's resistance to air flow per unit mass is measured to determine micronaire. Micronaire is a measure of the cotton's fineness. Micronaire and maturity are highly correlated within a cotton variety.

- **Strength (grams/tex)** – The fiber strength measurement is made by clamping and breaking a bundle of fibers with a 1/8-inch spacing between the clamp jaws. Results are reported in terms of grams per tex to the nearest tenth. A tex unit is equal to the weight in grams of 1,000 meters of fiber. Therefore, the strength reported is the force in grams required to break a bundle of fibers one tex unit in size. The following table shows some general descriptions of strength measurements in grams per tex.

Fiber Strength Table	
Descriptive Designation	**Strength (grams per tex)**
Weak	23.0 & below
Intermediate	24.0 – 25.0
Average	26.0 – 28.0
Strong	29.0 – 30.0
Very Strong	31.0 & above

- **Leaf Grade** – Leaf refers to small particles of the cotton plant's leaf, which remain in the lint after the ginning process. Upland leaf grades are determined by the classer and are identified as numbers 1 through 7, all represented by physical standards. Upland leaf grade 8 (Below Grade) is used to identify samples having more leaf than leaf grade 7. American Pima leaf grades are also determined by the classer and are identified as numbers 1 through 6, all represented by physical standards, and leaf grade 7 (Below Grade), which is used to describe samples having more leaf than leaf grade 6.

- **Extraneous Matter** – Extraneous matter is any substance in the cotton other than fiber or leaf. Examples of extraneous matter are bark, grass, spindle twist, seed coat fragments, dust, and oil. The kind of extraneous matter, and an indication of the amount (light or heavy), are noted by the classer on the classification record. The amount of extraneous matter in the cotton is reported as level 1 or level 2, with level 2 indicating the heavier contamination. The code numbers identifying the presence and level of extraneous matter in a sample are as follows:

Extraneous Matter	
01	Prep Level 1
02	Prep Level 2
11	Bark Level 1
12	Bark Level 2
21	Grass Level 1
22	Grass Level 2

31	Seed Coat Fragments Level 1
32	Seed Coat Fragments Level 2
41	Oil Level 1
42	Oil Level 2
51	Spindle Twist Level 1
52	Spindle Twist Level 2
61	Other Level 1
62	Other Level 2

- **Remarks** – The instrument assigns the remarks code 75 where applicable. Classers identify other special conditions that may cause processing problems and lower yarn quality. The following remarks codes identify special condition cotton:

75	Other Side Two or More Color Grades and/or Color Groups or One Color Grade and One Color Group Higher
76	Reginned
77	Repacked
78	Redder Than Normal (Pima)
92	Pima Ginned on Saw Gin

- **Instrument Color Grade** – Official color grade for Upland cotton. Each color grade is subdivided to denote differences within a color grade. This information is reported as a two-digit Color Grade and a single-digit color quadrant.

- **Color Quadrant** – This three-digit number is derived by locating on the color diagram the intersection of the Rd and +b readings. Not used for American Pima color.

- **Color Rd** – Grayness that indicates how light or dark the sample is.

- **Color +b** – Yellowness that indicates how much yellow color is in the sample.

- **Non-Lint Content (Trash Percent Surface)** – The two-digit trash code reported on the classification record is the percent of the sample surface covered by trash particles as determined by image analysis. For example, a reading of 04 indicates that trash particles cover 0.4 percent of the sample surface. Trash particles include extraneous matter such as grass, bark, etc., but these particles cannot be distinguished one from another by this measurement. Therefore, the classer will continue to designate samples containing extraneous matter particles.

- **Fiber Length (100^{ths} of an inch)** – see above

- **Length Uniformity Index (percent)** – A three-digit number that is a measure of the degree of uniformity of the fibers in a sample to the nearest tenth (the decimal is not displayed). The descriptive terms listed below may be helpful in explaining the measurement results.

Descriptive Designation	Length Uniformity
Very Low	Below 76.5
Low	76.5 – 79.4
Average	79.5 – 82.4
High	82.5 – 85.4
Very High	Above 85.4

- **Upland or Pima** – One-digit code. (1 = Upland, 2 = Pima)

- **Record Type** – One-digit code. (0 = Original, 1 = Review, 2 = Rework)

- **Record Status** – One-digit code indicating whether or not the manual classing information has been corrected. (0 = Not a correction, 1 = Correction)

- **CCC Loan Premiums and Discounts** – This five-digit code gives the CCC loan premium and discount points for Upland cotton. The physical loan price for Pima cotton is shown in cents per pound.

Appendix B: Case Study Data Elements for Monitoring/Controlling Product Quality

	Machine Panel	Online Data	Offline Data
Bale Laydown	• Shift • Advance • Height • Remaining Runtim • Efficiency • Malfunctions • Downtime	• Length • Uniformity • Strength • Elongation • Rd • Plus b • Micronaire • Trash Leaf • Color-Q • Classer Grade • Maturity Ratio • Maturity Percent • Extraneous Matter • Remark 1 & 2 • Ship Bale Number • Mill Bale Number • Trash Count • Sugar • Trash % Area • Consume Date • Mark • Contract • Gin Bale Number • Gin ID • Warehouse • Bin Number • Position • Crop Year • Mill	N/A

	Machine Panel	Online Data	Offline Data
Mixers	• Shift • Machine Settings • Chamber Malfunctions • Fan Speed	N/A	N/A

	Machine Panel	Online Data	Offline Data
Cleaners	• Shift • Machine Settings • Air Pressure • Roller RPMs	N/A	N/A

	Machine Panel	Online Data	Offline Data
Carding	• Shift • Machine Settings • Production • Production Speed • Delivery Speed • Efficiency • Sliver Breaks • Stops • Downtime • Can Changes • CV% • Fineness • Web Thickness • Draft • Spectrogram	N/A	• Test Date • Machine • Comment(Mat or Sliver) • Nep Size • Neps • Length (w) • Length (w) CV • Upper Quartile Length (w) • Short Fiber Content (w) • Length (n) • Length (n) CV • Short Fiber Content (n) • Length 5% • Total Count • Trash Size • Dust • Trash • Visible Foreign Matter • SCN (um) • SCN Count • Fineness • IFC • Maturity Ratio • Count • Count CV • Evenness • Weights

	Machine Panel	Online Data	Offline Data
Drawing	• Shift • Delivery Speed • Efficiency • Autoleveling • Production • Can Capacity • Sliver Weight • Stops • A% Stop Limit • Length • Spectrogram	• Date • Machine • Sliver Break Input Stops • Sliver Break Output Stops • Sliver Monitor Stops • Autoleveling Stops • Lap Up Creel Stops • Draft Roller Lap Stops • Web Funnel Jam Stops • Sliver Jam Coiler Stops • RQM A% Stops • RQM CV% Stops • No Empty Can Stops • Can Truck Missing Stops • RQM Spectrogram Stops • Thick Place Stops • Total Stops • Efficiency • Production	• Date • Machine • Count • Count CV • Evenness • Weights

	Machine Panel	Online Data	Offline Data
Spinning	• Shift • Yarn Count • Rotor Speeds • Length • Yarn Breaks • Production • Twist • Draft	• Start Date • Machine • Style • Count • Production • Total Stop Time • Quality Stops • Quality Stop Time • Production Stops • Production Stop Time • Doffings • Doffing Time • Red Lights • Red Light Time • Short Thick Stops • Long Thick Stops • Thin Stops • Moiré Stops • Thin Sliver Stops • Thick Sliver Stops • Foreign Fiber Stops • Piecer Attempts • IPI: Neps/km • IPI: Thins/km • IPI: Thicks/km • IPI: CV • Machine Stop Time • Doffings/1000rhr • QualityStops/1000rhr	• Test Date • Machine • Count • Count CV • Single End • Single End CV • Elongation • Elongation CV • CVm • Thin • Thick • Hairiness • Neps • CV by Length • Work to Break • Skein

Appendix C: Collection Points and Data Elements from Plant Interviews

Online Testing:
None

Offline Testing:
HVI

Tests:
- Fiber Length
- Fiber Strength
- Mic
- Plus b
- Uniformity
- Trash
- etc.

Online Testing:
Proprietary System

Tests:
CV, Draft

Offline Testing:
AFIS, USTER,
Weights

Tests:
- Neps
- Length
- SFC
- Trash
- Maturity Ratio
- etc.

Online Testing:
Proprietary System

Tests:
CV, Thick Places, etc.

Offline Testing:
AFIS, USTER,
Weights

Tests:
- Neps
- Length
- SFC
- Trash
- Maturity Ratio
- etc.

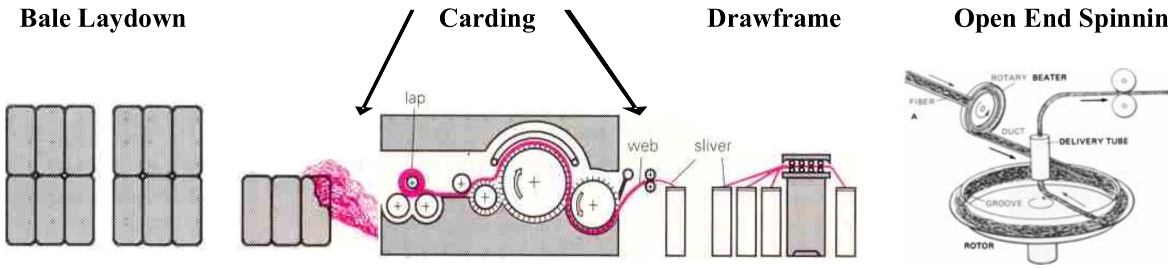
Online Testing:
Proprietary System

Tests:
Yarn Breaks,
Yarn Efficiencies

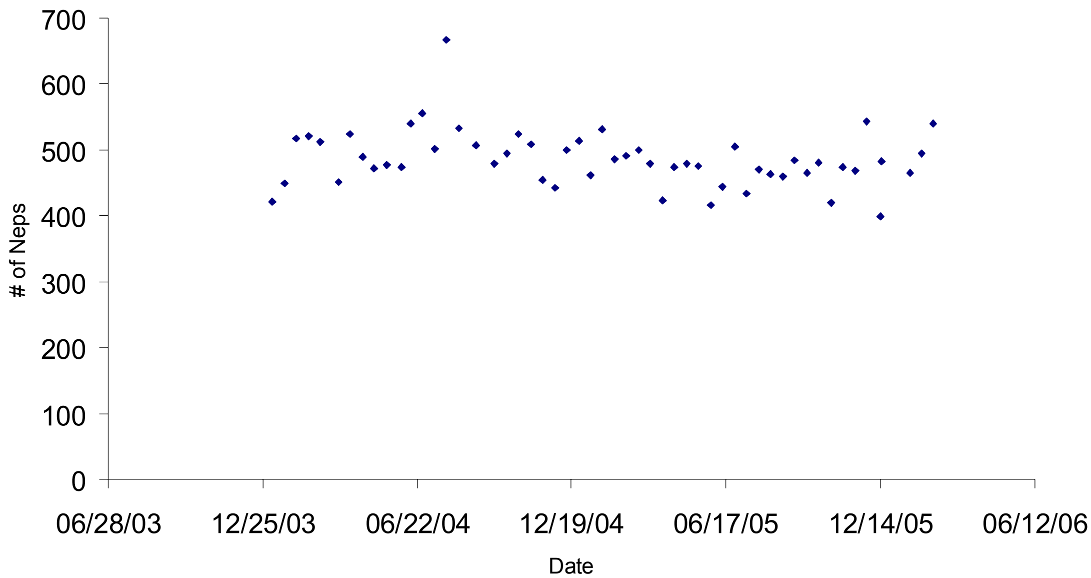
Offline Testing:
Evenness, Tensile,
Count & Skein Break

Tests:
- Strength
- Count
- Twist
- Hairiness
- Regularity
- etc.

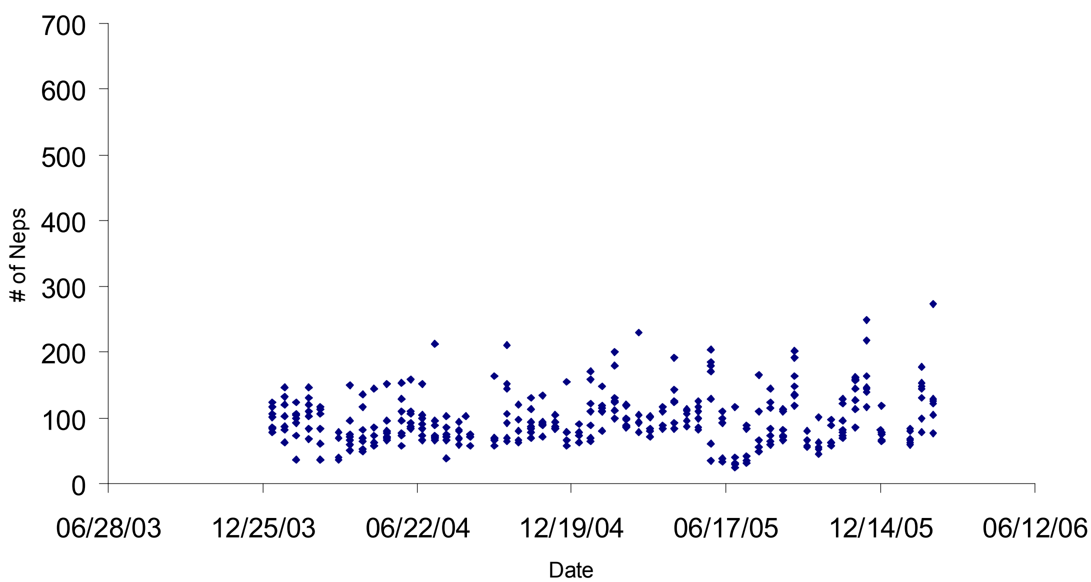
Bale Laydown **Carding** **Drawframe** **Open End Spinning**



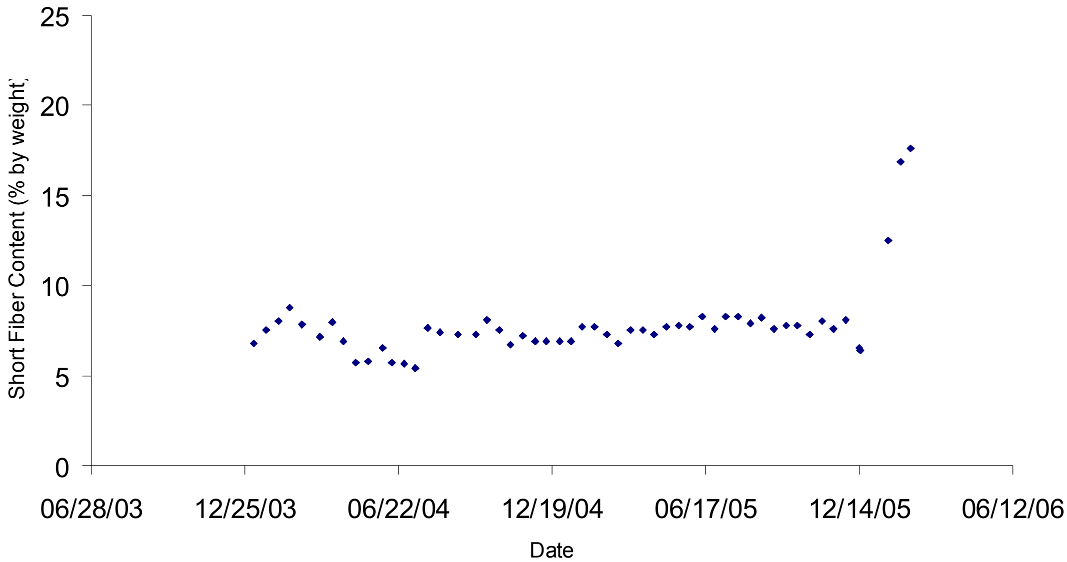
Appendix D: Charts from Level I Analysis



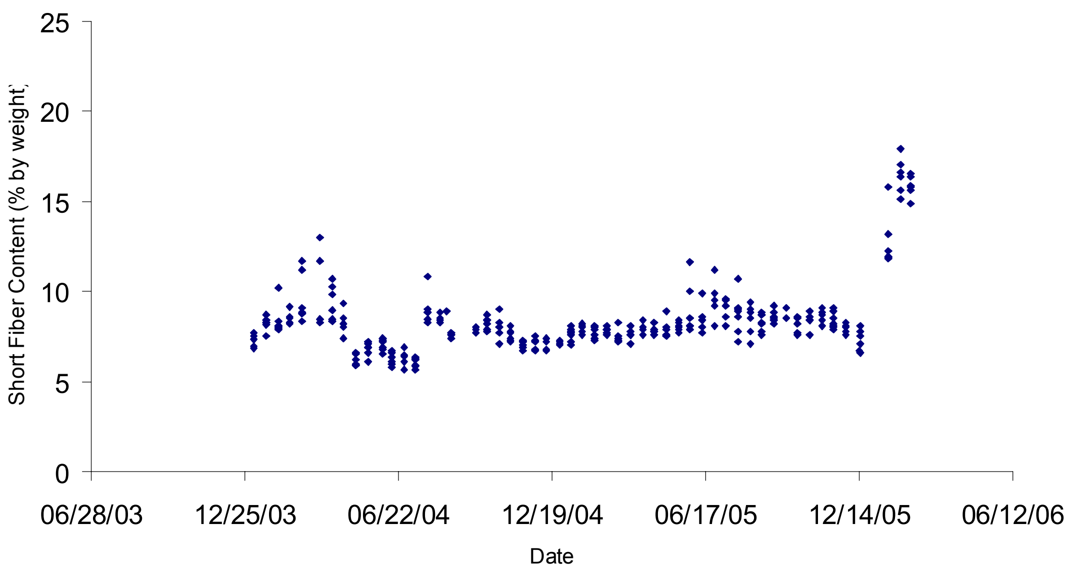
Mat Neps for Card Line 2



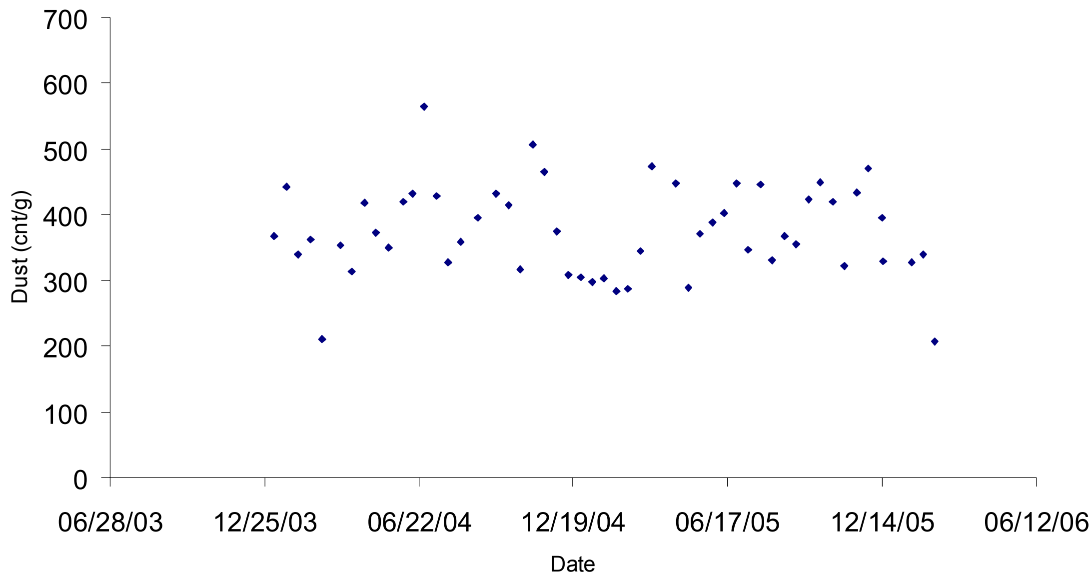
Sliver Neps for Card Line 2



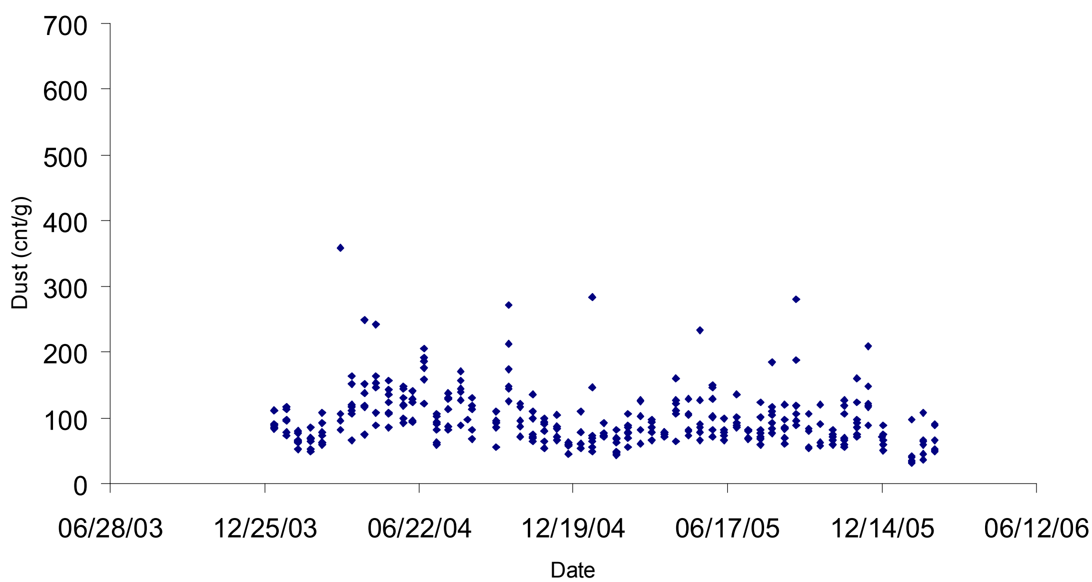
Mat Short Fiber Content (% by weight) for Card Line 2



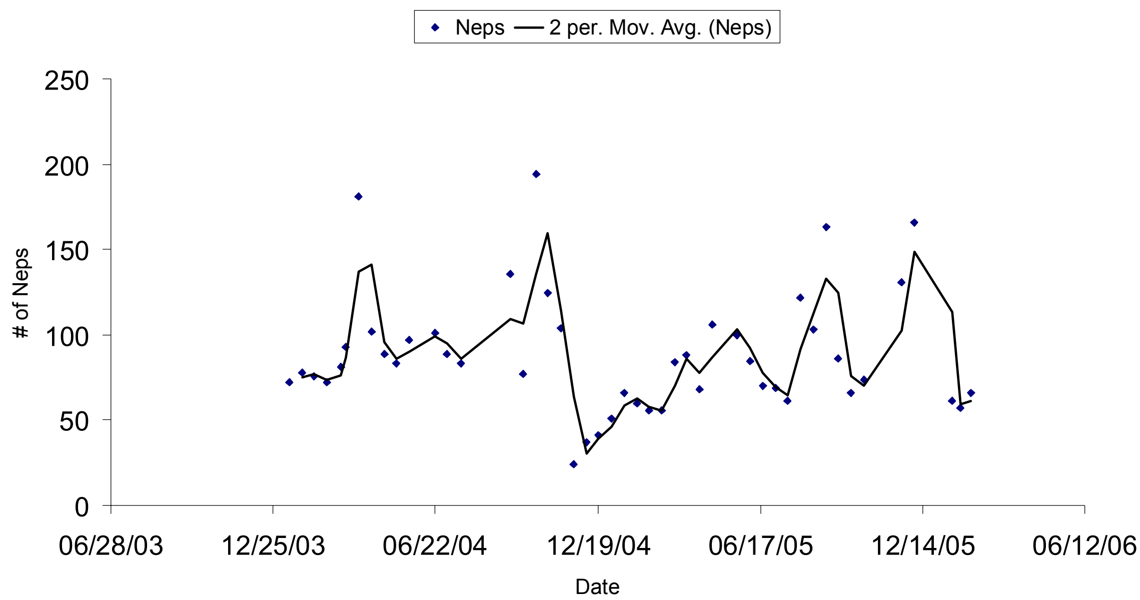
Sliver Short Fiber Content (% by weight) for Card Line 2



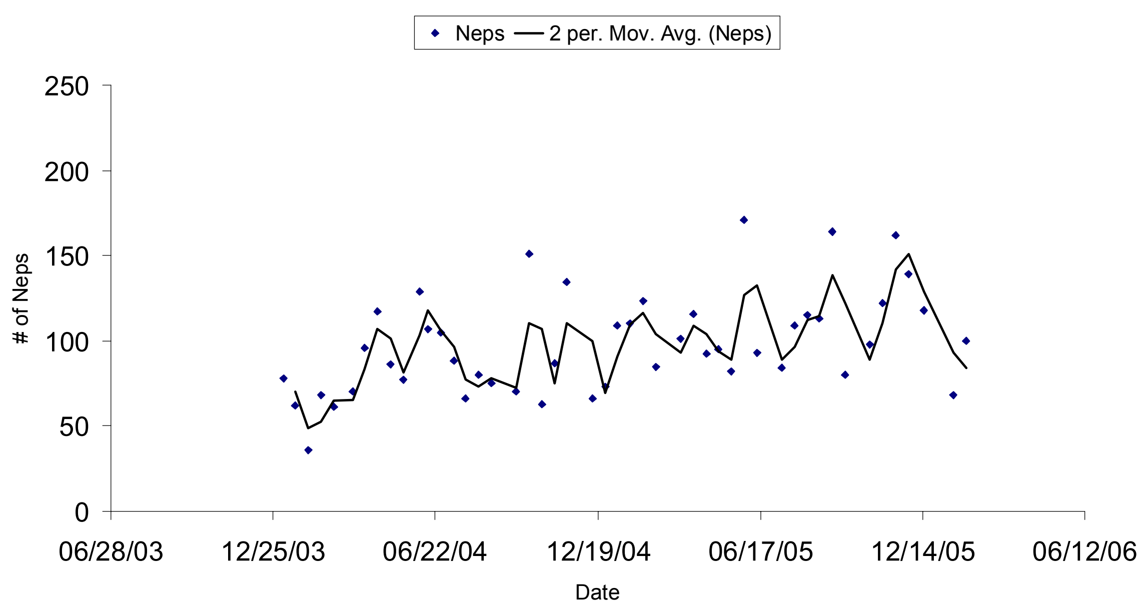
Mat Dust (cnt/g) for Card Line 2



Sliver Dust (cnt/g) for Card Line 2

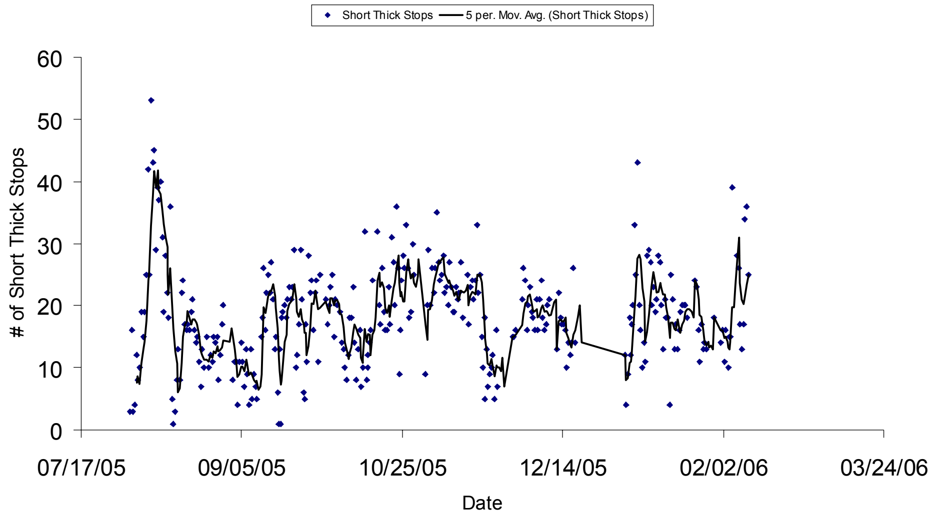


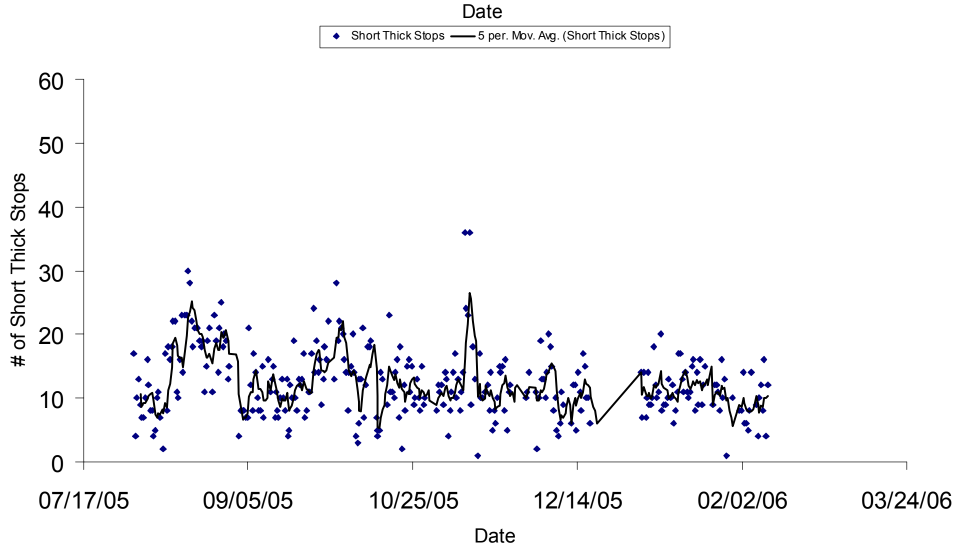
Neps for a Card in Line 1

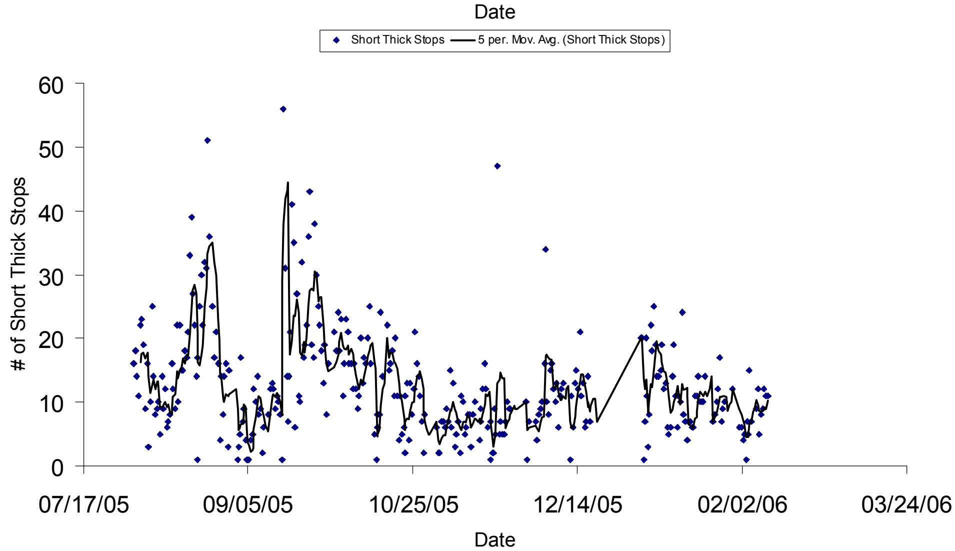


Neps for a Card in Line 2

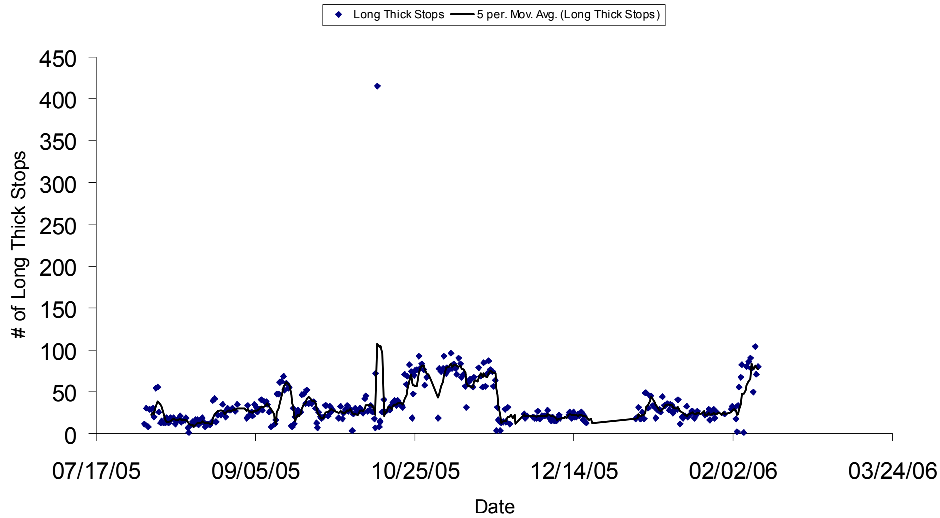
Short Thick Stops for 3 Different Spinning Frames

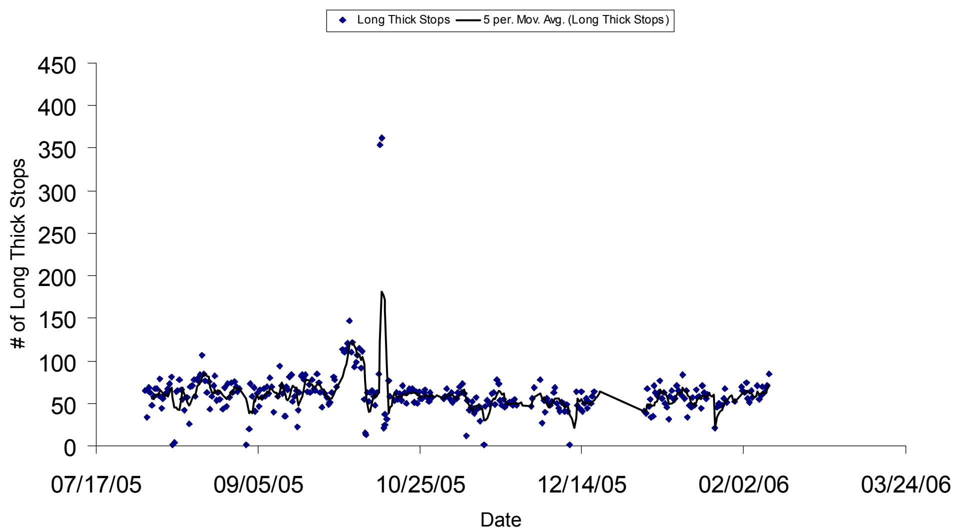


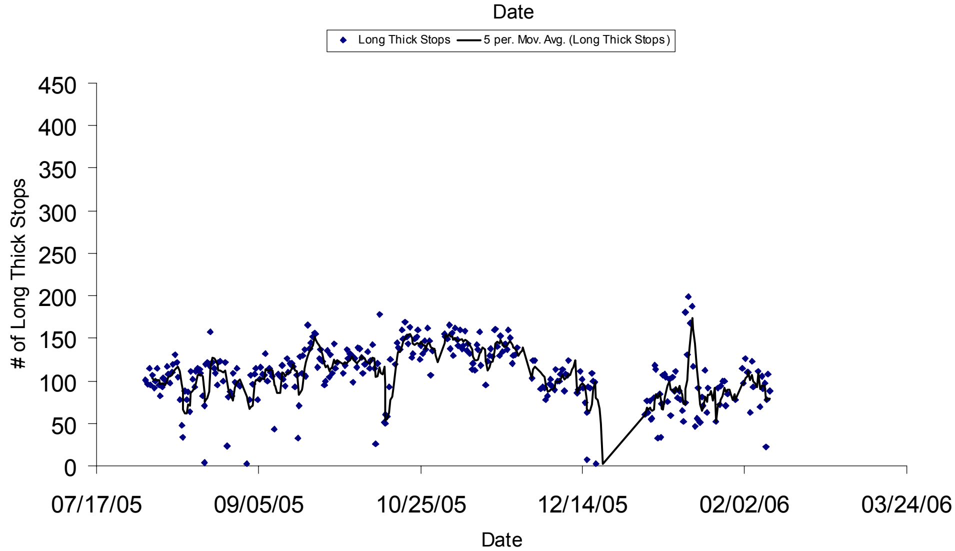




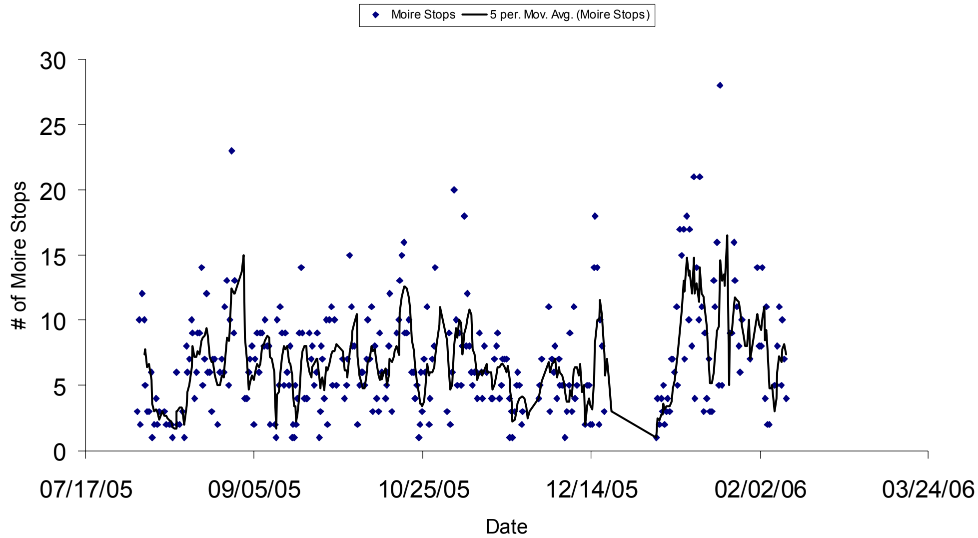
Long Thick Stops for 3 Different Spinning Frames

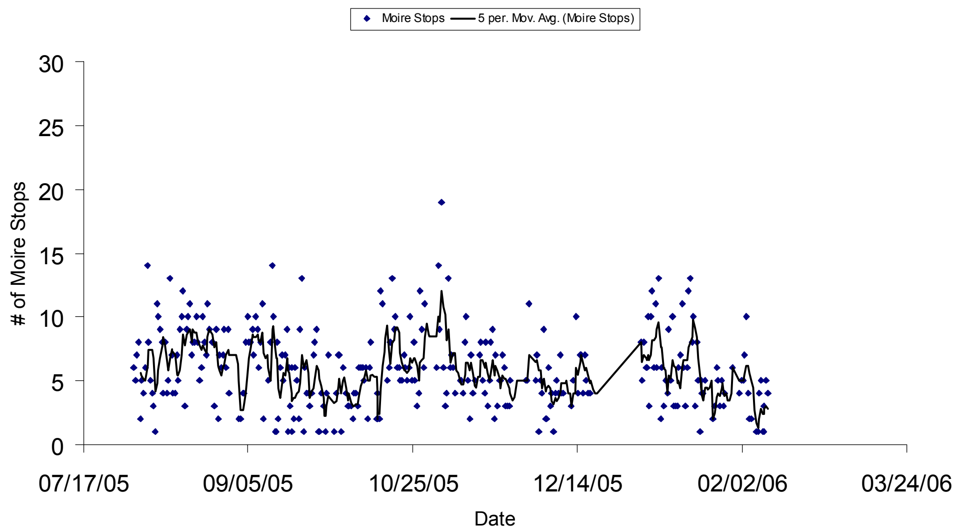


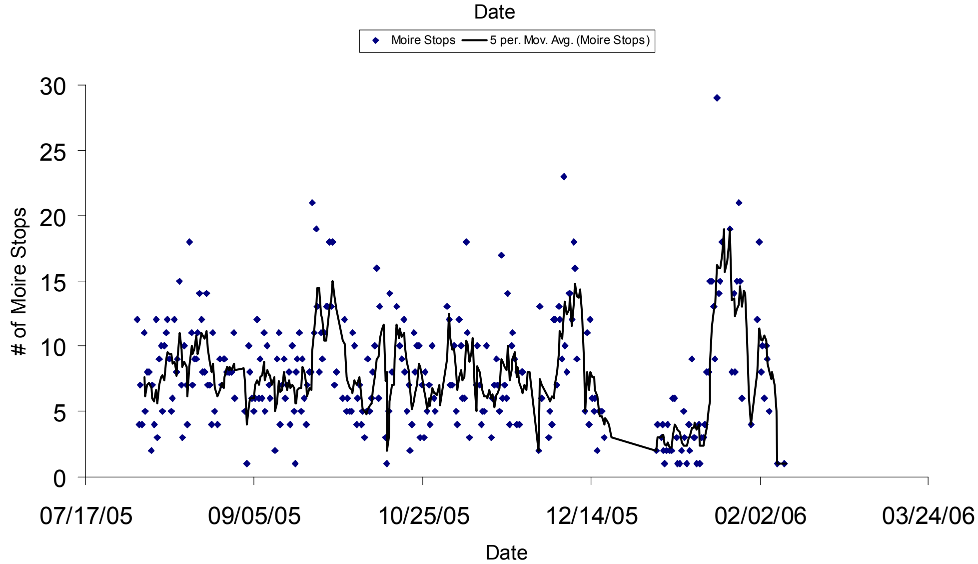




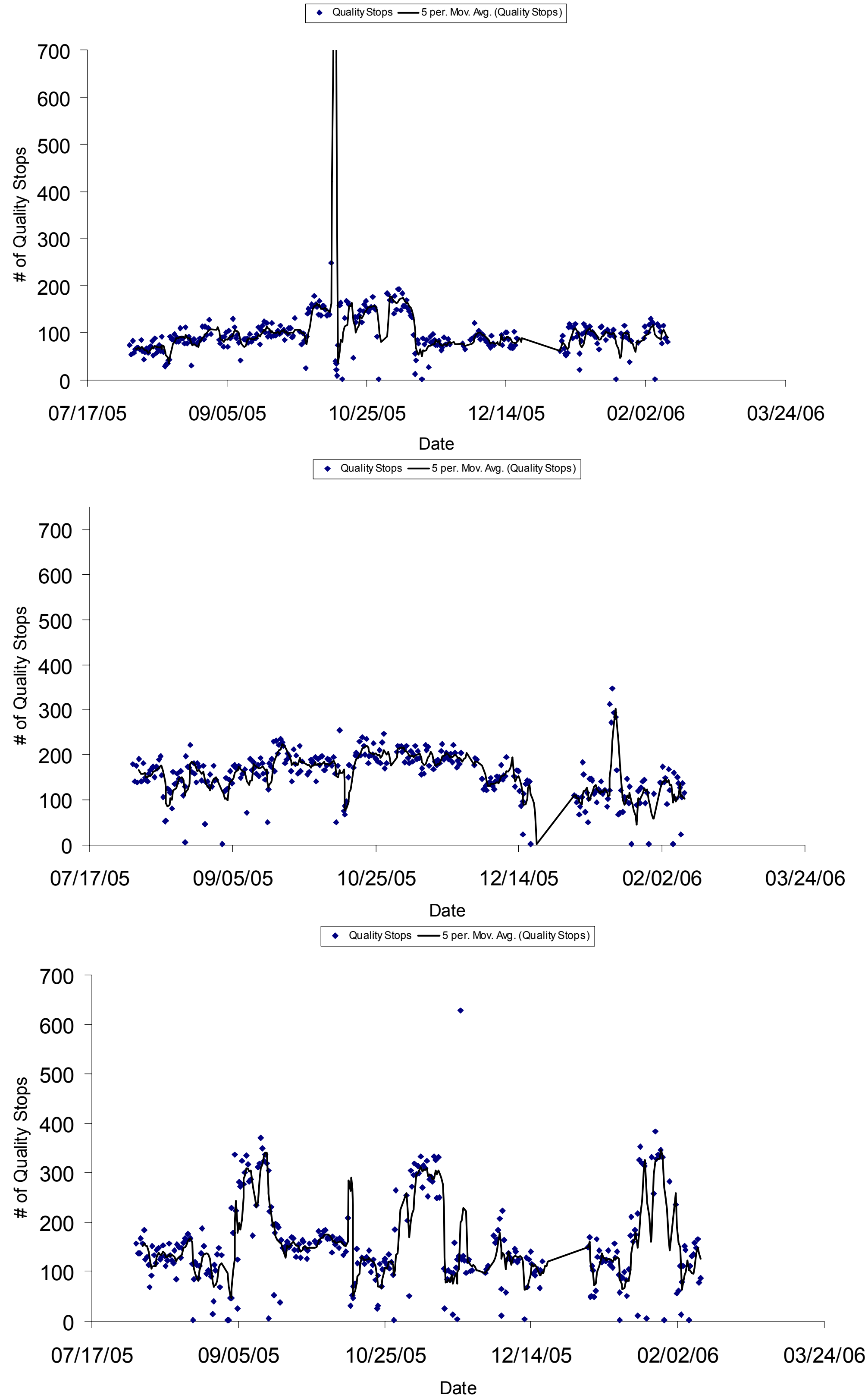
Moiré Stops for 3 Spinning Frames

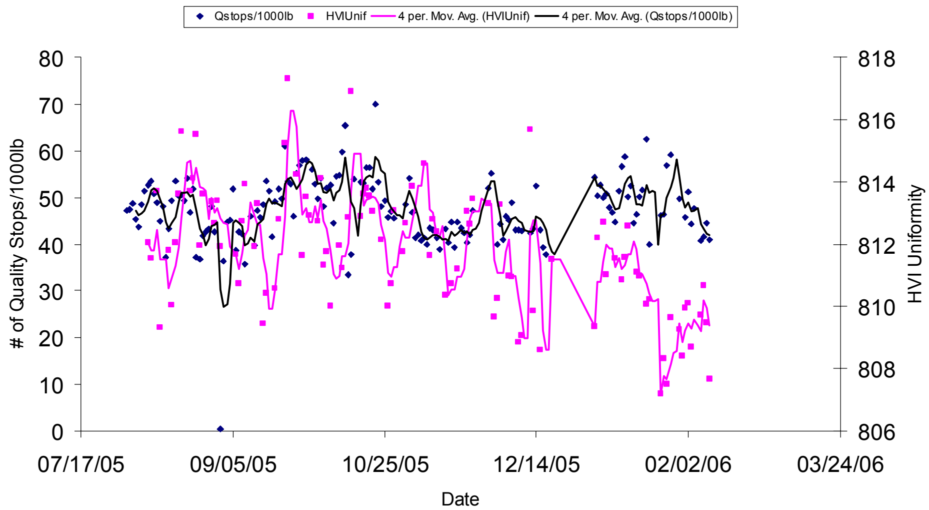




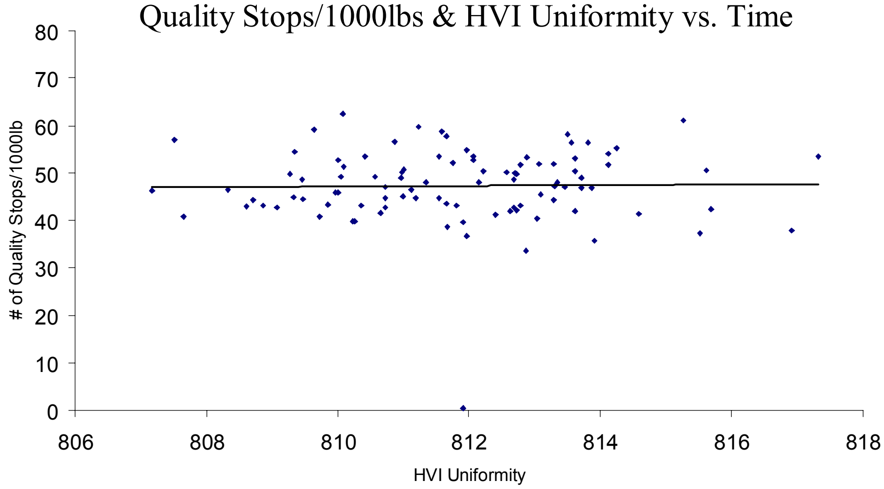


Quality Stops for 3 Spinning Frames

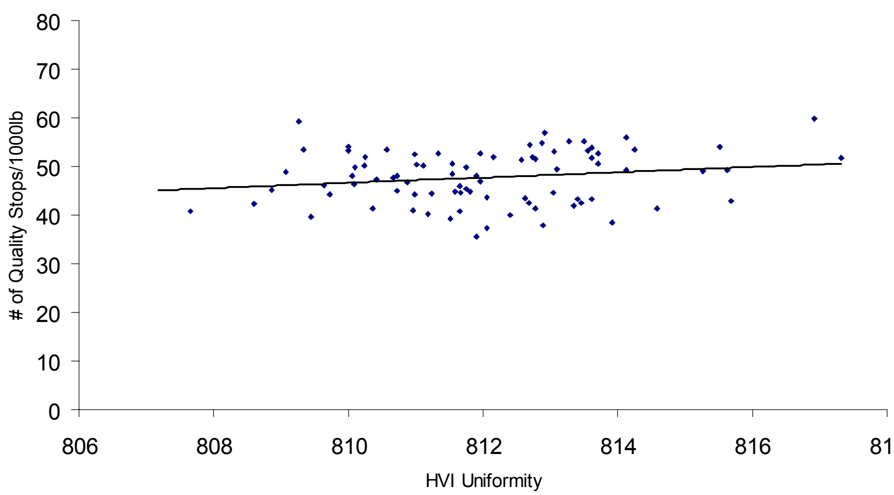




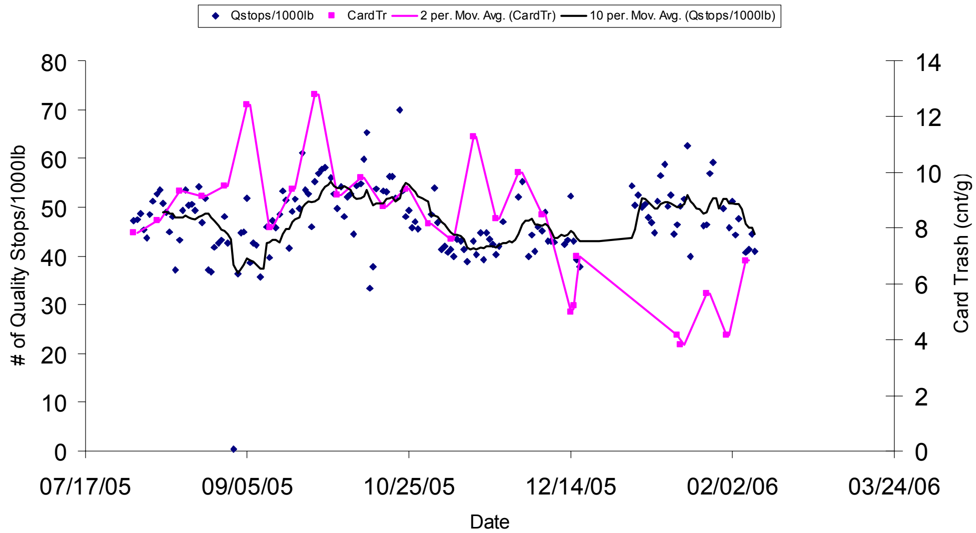
Quality Stops/1000lbs & HVI Uniformity vs. Time



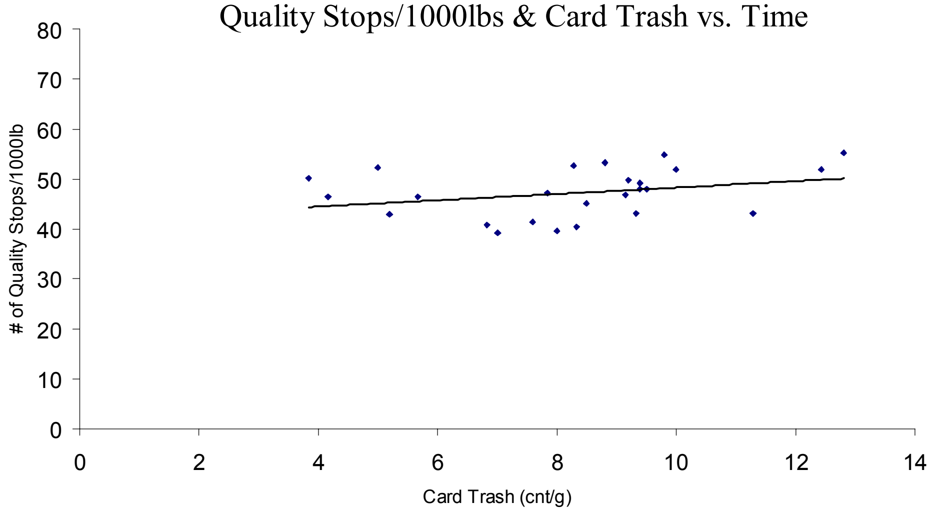
Quality Stops/1000lbs vs. HVI Uniformity



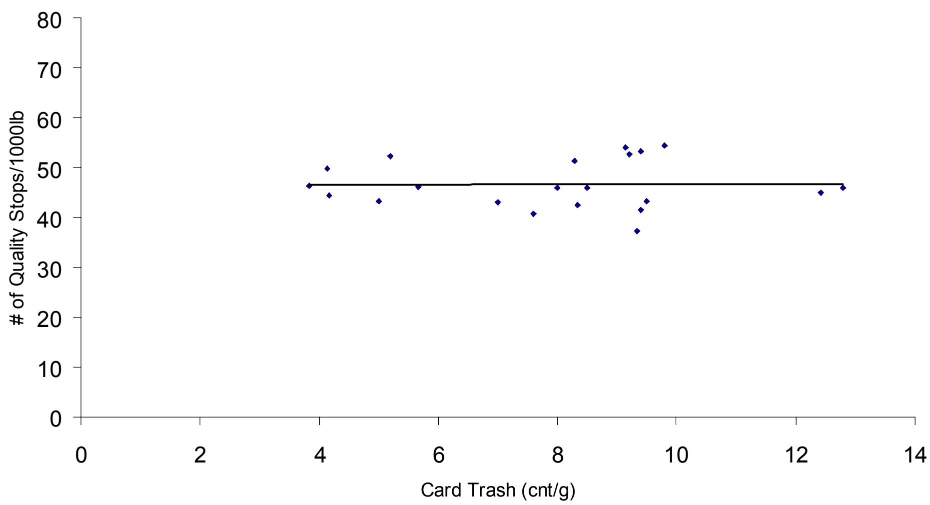
Quality Stops/1000 lbs vs. HVI Uniformity with 3 Day Lag time



Quality Stops/1000lbs & Card Trash vs. Time



Quality Stops/1000lbs vs. Card Trash



Quality Stops/1000 lbs vs. Card Trash with 3 Day Lag time