# APPLICATION OF MULTIVARIATE ANALYSIS TECHNIQUES IN UNDERSTANDING COMPLEX INDUSTRIAL PROCESSES – A PULP MILL EXAMPLE

by

Saket Kumar

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

1999

Program Authorized to Offer Degree: College of Forest Resources

UMI Number: 9952852





Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Doctoral Dissertation

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Saket Kumar

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

Richard R. Gustafson

Reading Committee:

Richard R. Gustafson

David Briggs

Eric C. Turnblom

Date: October 26, 1999

University of Washington

Abstract

# APPLICATION OF MULTIVARIATE ANALYSIS TECHNIQUES IN UNDERSTANDING COMPLEX INDUSTRIAL PROCESSES – A PULP MILL EXAMPLE

by Saket Kumar

Chairperson of the Supervisory Committee:
Chairman, Prof. Richard Gustafson
Department of Paper Science and Engineering

The financial and process benefits of improving the mill fiber line are widely acknowledged. However, process optimization of the fiber line is difficult due to the complex behavior of pulp and paper systems. The dissertation project focused on application of multivariate analysis techniques for understanding and improving fiber line performance. Models for prediction and variability analysis of kappa number and total bleaching cost were developed using data generated by pulping and bleaching operations. The research project led to refinement of earlier methods of data preprocessing and development of algorithmic solution for the data time shifting problem.

Multivariate statistical techniques were used to analyze sources of kappa and bleaching cost variability for Weyerhaeuser Longview mill and Georgia Pacific Ashdown mill. For the Weyerhaeuser Longview mill, factor analysis allowed development of models that successfully predict kappa number out of a continuous digester and $O_2$ delignification stage. The most important cause of kappa variability in the continuous digester was found to be mischarges in alkali. Variations in kappa number can be reduced by 45% in the digester and 40% in the $O_2$ delignification reactor if variables correlating with the important factors are brought under control.

None of the multivariate techniques were successful in predicting K-number for the Georgia Pacific Ashdown mill. The main reason for poor prediction was that the digester was already under tight control as evident from low (6.12%) coefficient of variation of K-number. Processes under tight control appear to generate datasets with minimal correlation structures. Such datasets are not suitable candidates for the purposes of predicting output variables such as K-number.

In the bleaching study, principal component analysis as well as factor analysis models with fourteen upstream variables successfully predicted bleaching cost trend. However, neural networks bleaching cost predictions were poor. Factor analysis and PCA models of the bleaching cost indicated that most of the bleaching cost variability was either due to lignin factor (which represents pulping and washing variables) or due to digester column stability represented by outlet device amperage. A method to compare results from various multivariate methodologies was also developed. The factor model with fourteen variables achieved the highest score on comparison scale for bleaching cost study.

Both the lignin and digester stability factor point at the digester being the major source of bleaching cost variability. It appears that there are variations in pulp lignin content (or some latent variable) that are not measured by the K number test at the Decker, but results in changes in the chlorine requirements at the D/C stage. In this situation, bleaching cost predicted by the model may be used as a soft sensor to manipulate temperature, steam flow in digester to produce pulp with uniform bleach chemical requirements (i.e., consistent latent variable variation). This way cost variability will be reduced, as presumably the variation in lignin content will be minimized.

TABLE OF CONTENTS

List of figures

List of TABLES

## Acknowledgments

I want to thank members of my supervisory committee, Prof. Richard R. Gustafson, Prof. David Briggs, Dr. Eric Turnblom, Dr R James Barbour, and Prof. Christian Grue for their consistent support and valuable input on my research. This thesis would have remained dream without the guidance and input of my advisor Rick Gustafson. I have learnt so much from him and continue to do so. My thanks go to United States Forest Service and Georgia Pacific Corporation for providing financial support for this project. The data used in the project was acquired from Weyerhaeuser Longview mill and Georgia Pacific Ashdown mill. I would like to express my gratitude to Jim Rushton, Senior Consulting Engineer at GP, for his support and encouragement throughout the project.

I want to thank graduate students in paper science department for sharing the roller coaster ride that graduate school is. They have helped in numerous ways at school and off school. My special appreciation goes to Yue Liu, Ravi Managuli for listening to my crazy ideas and still managing to encourage me.

Special gratitude to my parents who have been constant source of support and joy in my life. They have always encouraged me to dream and work for those dreams. They are a big part of whatever I have achieved so far. My brothers and bua were always there when I needed them. Thanks to them for helping me cope up with rough times and achieving my best.

## Dedication

I wish to dedicate this thesis to my parents, who have sacrificed so much to get me in this position. I am very proud to be son of Kuleshwar and Geeta.

# CHAPTER 1: INTRODUCTION

## 1.1 PULPING AND BLEACHING OVERVIEW

Modern pulp and paper mills use wood chips as the basic raw material to produce paper. Inside wood, papermaking fibers are cemented together by an amorphous, highly polymerized substance called lignin. The objective of pulping is to degrade and dissolve away the lignin and leave most of the cellulose and hemicelluloses in the form of intact fibers. The task can be accomplished mechanically, thermally, chemically, or by the combination of these processes. The most prevalent pulping process, called kraft process, involves cooking the wood chips in a solution of sodium hydroxide ($NaOH$), and sodium sulfide ($Na_2S$). The alkaline attack causes fragmentation of the lignin molecules into smaller segments whose sodium salts are soluble in the cooking liquor. "Kraft" is the German word for strong, and kraft pulps produce strong paper products; but the unbleached pulp is characterized by a dark brown color. The dark color of unbleached pulp is attributed to "chromophoric groups" on the lignin. The approach followed to produce white paper, from unbleached pulp, is to completely remove the "chromophoric groups" and residual lignin using bleaching chemicals. Pulp bleaching is achieved through a continuous sequence of process stages using different chemicals such as chlorine dioxide, chlorine, oxygen, and hydrogen peroxide usually with washing between stages. Finally, the bleached pulp is transferred to paper machine to manufacture paper. Mills refer to the set of operations from pulping to paper machine as the *fiber line*. In terms of investment, a 1000 ton per day integrated bleached kraft mill is estimated to cost in excess of 1 billion dollars (typically exceeding one million dollars per worker) [1]. Such a high investment cost coupled with rising raw material costs has necessitated increased impetus on process optimization efforts.

Pulp and paper mills engage in high volume operation. For example, a 1000-ton per day kraft pulp mill uses 2200-ton wood chips per day (assuming a pulp yield of 45% from wood chips). The amount of pulping chemicals charged, heat energy used, and bleaching chemicals used are all based on 2200-ton wood chip consumption. In such a high investment, high volume operation, small process improvements can amount to significant dollar savings for mills. For example, assume a small process improvement results in reduction of kappa number (a measure of residual lignin in pulp) variability by two points. This reduction in kappa variability can lead to an approximately 1% increase in pulp yield (22 tons more pulp per day ~ $10000 in increased profits per day). The decrease in kappa variability will also enable the bleach plant to charge bleach chemicals assuming lower incoming kappa (Figure 1-1). The lower bleach chemical charge per ton of pulp will result in savings in total bleach chemical cost. Additionally, the pulp product will be of higher quality due to reduced pulp strength loss.

It is evident from this discussion that there is room for improvement in a mill fiber line that can lead to monetary and process benefits. However, a better process understanding is a prerequisite for making such an improvement. In other words, a complete understanding of the behavior of the system is essential to modify and optimize an existing process and to develop new ones. At present, the quantitative knowledge of the processes involved in the mill fiber line is generally incomplete and is in a state of continuous evolution [2,3].

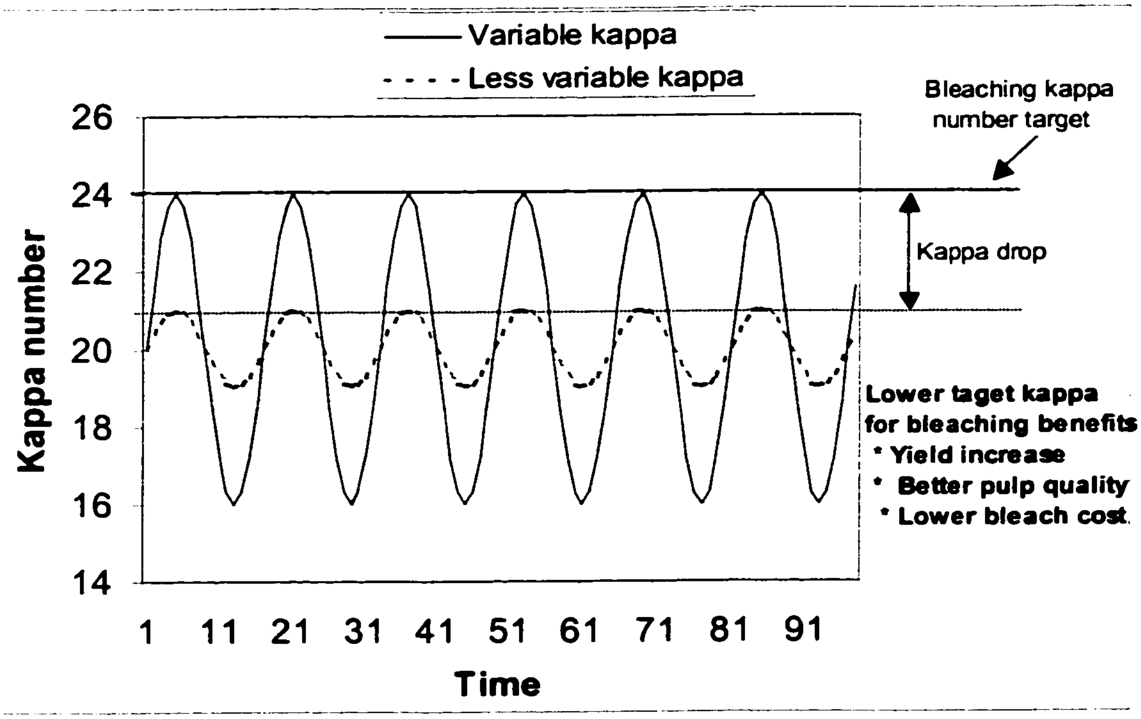



Figure1-1 Benefits of fiber line optimization

## 1.2 PROBLEMS IN ANALYZING PULPING OPERATIONS

There are several reasons for the incomplete understanding of the fiber line operations. Most of the problems in understanding fiber line operations can be attributed to the inherent nature of the pulping process. Pulp and paper mills perform complex physical and chemical unit operations on a daily basis. These operations involve numerous interacting variables, i.e., a change in one input variable affects more than one output variable. Multivariate interactions result in varying degrees of correlations in the data generated by the process. Numerous process control loops present in pulp mills also result in highly correlated data. The process analysis is further complicated by the presence of unknown disturbances such as chip quality variations coming into the system.

A number of other factors make it difficult to understand mill operations. These factors are related to the design of the equipment used. In the case of reactors like the digester and bleaching towers, there are long delay times due to the design of the process equipment. As a result, the process data contains timeshifted variables trends that can't be easily used to model process behavior. Another complicating factor is the presence of non-linear relationships among variables in the dataset generated by the process. Most importantly, it is quite difficult to get a dataset representative of the process as fiber lines run under unsteady states for a significant portion of the time. All these factors in conjunction with the complex nature of the process itself make it difficult to get useful information about the process. Table 1-1 provides a summary of problems in analyzing pulping operations.

Table 1-1. Process and analysis problems while analyzing pulping operations.

| **Process problems** | | **Analysis problems** |
|---|---|---|
| Complex physical and chemical unit operations. | ➡ | Multivariate interactions, correlations, non-linearity. |
| Unknown incoming disturbances e.g., chip quality changes | ➡ | Poor prediction models |
| Numerous process control loops. | ➡ | Highly correlated variables, controlled process. |
| Large dead-time, time lag in the process | ➡ | Time-shifted variables, loss of predictive information. |

## 1.3 PROCESS OPTIMIZATION REQUIREMENTS

Knowledge of the process is a prerequisite to successful design and implementation of fiber line optimization. Many pulp and paper processes are based on quite complex physical and chemical operations involving many interacting variables. With increasing access to computers, there has been development of quantitative process knowledge in the form of mathematical model [2].

### *1.3.1 PROCESS MODELING*

These are two basic ways of modeling process dynamics. The first method is to derive a *mechanistic model (or first principle model)* using laws of physics and chemistry. Mechanistic models are based on mass and energy balances as well as kinetics and thermodynamics of processes. The resulting model usually consists of a set of differential equations. These equations can be numerically integrated to simulate, on a computer, the behavior of the process. The parameters of the differential equation depend on the chosen operating points such as throughput or grade. Process simulations, which are based on these models, are useful for testing control strategies at the design stage. Theoretical pulping models have the advantage of being able, at least in theory, to accurately describe the complex behavior of pulping systems. A disadvantage is that it is difficult to obtain sufficiently reliable parameter estimates for deriving and applying the model. Also such models require substantial computing resources to perform the lengthy numerical integration of simultaneous partial differential equations. For these reasons, empirical models are frequently developed and used.

### *1.3.2 EMPIRICAL MODELS*

Unfortunately, our understanding of the physics and chemistry of many processes such as pulping and bleaching is not sufficient to derive detailed mechanistic models. In this case, a model may be obtained through the analysis of data acquired when the process is operating. This method is referred to as *empirical modeling* and will be dealt with later in the chapter. Experimental data from the process is also used to verify mechanistic models and/or to provide the estimates of unknown process parameters in such mechanistic models.

Empirical models explicitly relate such dependent variables as yield, kappa number, and pulp viscosity to controllable pulping variables such as alkali charge, H-factor, sulfidity, and liquor-to-wood ratio. Hatton [4, 5] succeeded in correlating a large amount of kraft pulping data with simple relationships that relate pulp yield and kappa number to H-factor and effective alkali charge. These equations are applicable to the pulping of thin chips at unspecified and presumably constant values of sulfidity and liquor-to-wood ratio. The form of relationship is

$$Y = A - B \cdot (\log_{10} H) \cdot (Ea^{n1}) \qquad \text{(Eq. 1-1)}$$

$$K = \alpha - \beta \cdot (\log_{10} H) \cdot (Ea^{n2}) \qquad \text{(Eq. 1-2)}$$

where Y is yield, K is kappa number, H is H-factor, EA is effective alkali charge and A, B, $\alpha$, $\beta$, $n_1$, $n_2$ are parameters assumed to be constant for a given species.

The Hatton equations provide a very compact description of the pulping behavior of several species. Their most obvious shortcomings are that they cannot predict the effects of changing sulfidity or liquor-to-wood ratio and that their predictive power for hardwoods is not very great, owing to poorer fits to the original data. Models also have limited range where these can be used.

Tasman [6,7] analyzed several sets of pulping data in the literature with a view of deriving a model that would account for changes in sulfidity as well as in H-factor and alkali charge. He derived the following equations:

$$\log Y = \frac{a - b(EA\log_{10}S/\log_{10}EA)}{H} + I \qquad \text{(Eq. 1-3)}$$

where

$I=(c + d \log_{10}S)/\log_{10}EA)$ for softwoods;

$I=c - d \log_{10}EA$ for hardwoods.

Y is yield, EA is effective alkali charge, S is sulfidity, H is H-factor, and a, b are constant for a particular species.

Similar to Hatton equations, Tasman's equations are useful for estimating tradeoffs between pulping variables and predicting unbleached pulp properties.

Lin et al. [8] reported that the following equations could be used to predict the kappa number of kraft cook from hardwoods from Taiwan and the Ivory Coast:

$$\text{Kappa} =(A'D_0^{0.136})/(Q_0^{1.171}\, H^{0.175}) \qquad \text{(Eq. 1-4)}$$

where $D_0$ is the liquor-to-wood ratio, $Q_0$ is the alkali-to-wood ratio, H is H-factor, and A' is a constant for each wood species.

Lin's equation offers the advantage of being able to predict the effect of varying the liquor-to-wood ratio but doesn't account for sulfidity. The same form of equations may not be applicable to North American species.

Empirical models have their own set of disadvantages. Because they are entirely without theoretical basis they lack generality. A model based on a given set of experimental data is unlikely to be able to predict the outcome of cooks performed under conditions that differ from those used in the particular set of experiments that gave that data. A change in wood species, lignin content, chip size distribution or wood specific gravity is likely to seriously affect the applicability of the model. A theoretical model, on the other hand, can be more readily adapted to new conditions by substituting new values of one or more fundamental parameters such as lignin content, delignification rate constant or diffusivity. Significant improvements in the fiber line using empirical models is not possible due to the lack of generality of these models.

### *1.3.3 MULTIVARIATE ANALYSIS*

Most of the empirical models discussed above involve some variation of linear regression on raw or transformed process data. In the recent past, there has been considerable interest in the use of more sophisticated multivariate analysis techniques to model complex industrial processes. Multivariate techniques such as principal component analysis, factor analysis, and neural networks have been successfully used in pulp and paper industry. Applications cover a wide range of processes including chip refining, continuous and batch digesters, and paper machine operation [9-15].

A number of researchers have used multivariate dimension-reduction techniques such as principal component analysis (PCA) and factor analysis (FA). Arkun and Rigopoulos [9] have used PCA to compress and filter data from on-line sensors in paper machines. Using PCA they identified the most significant features of cross machine direction profile while filtering out the random noise. The filtered data was used for predictive control in a closed-loop system. Strand [10] has used factor analysis to analyze the effects of raw material variation on mechanical pulp properties. He has also modeled the behavior in newsprint wood refiners [11, 12, 13]. The application of factor analysis has mostly been limited to mechanical pulping.

Neural network is a non-linear multivariate analysis technique. These networks have also been used to discover hidden patterns inside data generated by mill operations. Rudd used [14] a neural network to predict the mat consistency and soda loss of the pulp leaving a bleached washer. The neural network used was based on 14 variables available in an existing control system. While working with the washer networks, Rudd found that a single network to predict consistency, density, and soda losses was not the most effective approach. The network tried to average the accuracy with which it could predict these variables. As a result, he used two networks which predicted different output variables using many common input variables. Additionally, he found that training networks for various operating conditions, such as pulp grade changes could result into much higher accuracy.

Dumont, et al. [15] examined a neural network developed using data from an industrial chip refiner. The research investigated the feasibility of using a feed-forward neural network as an alternative to mathematical modeling of complex processes. Outputs predicted by the network compared favorably with industrial refiner data. In addition, it was shown that the network structure could be modified to optimize refiner operation and product quality. Khorasani [16] proposed a backpropagation neural network-based controller to replace the self-tuning regulator (STR) for closed-loop control of specific energy in the refining process. Miyanishi [17] applied artificial neural networks to the diagnosis of paper web breaks in a commercial newsprint paper machine. He proposed a three-stage multilayer neural network and backpropagation method to extract essential causes of paper web breaks.

## 1.4 DISSERTATION OBJECTIVE

It is evident from previous discussions that multivariate tools are effective in analyzing paper machines as well as smaller unit operations (e.g., washing, refining) in the pulp mill. If multivariate techniques were used to analyze a pulp mill, an improved understanding of the kraft fiber line could be expected along the lines successes achieved in similar efforts in other areas of pulping and papermaking.

A pulp mill fiber line is a complex network of storage tanks and unit operations. Upsets due to shutdowns, slowdowns, rate, and species changes in one process tend to propagate through the network and hence influence the operation of other unit operations. A complete study of a mill fiber line, with an objective of understanding and optimization, has not been not been done before. Multivariate techniques, not first principle or empirical models, appear to be a well suited tool to analyze this complex system. Thus, the dissertation approach involved use of a spectrum or combination of multivariate techniques to study the kraft pulping process.

The main objective of the dissertation was to investigate the use of multivariate statistics for analyzing and optimizing pulp mill fiber line. The efficacy of different multivariate techniques, in analyzing the fiber line, was also compared.

# CHAPTER 2: MULTIVARIATE ANALYSIS METHODS

Pulp mill operations generate datasets with hundreds of measurements. However, relatively few events may be occurring despite the number of measurements being large. The data from these measurements must therefore be mapped into meaningful descriptions of event(s). m The multivariate techniques principal component analysis, factor analysis, and neural networks were used in our project to describe fiber line operations. An overview of multivariate techniques is presented before our project approach and results can be discussed. The goal of all multivariate analysis techniques is to predict output variables (called response variables) using input variables to the process (called explanatory variables) using as few independent characteristics as possible. The problem is not trivial because explanatory variables as well as response variables are intercorrelated. A detailed discussion of the multivariate techniques is presented in the following sections.

## 2.1 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) has been extensively used to reduce the number of explanatory variables. About a century ago, the idea of PCA appeared in psychology and some social science fields [18]. Today, a great deal of research is still being done in the general area of PCA and PCA is being used in engineering. PCA has been extensively applied in many disciplines, including chemistry, biology, meteorology, and chemical engineering. In chemical engineering, it has been used in data visualization, correlation and prediction, quality control, sensor calibration, and processes monitoring [19, 20, 21]. Wise and co-workers [22] give a theoretical basis for the use of PCA for monitoring processes. They find that under most circumstances the PCA model will span the same space as a linear state-space model.

Conventional PCA can be viewed as a technique for linearly mapping multidimensional data onto lower dimensions with minimal loss of information [23]. Let $X = [x_1, x_2, ..., x_n]$ be a normalized *n* dimensional data set. PCA is a method to transform X into lower dimensional latent structures. The set X would typically contain the measurements of a process. In many real-world processes, e.g., pulp mill data, the measurements are correlated; that is, knowledge of a subset of measurements defines the rest to some degree. With PCA, one can develop a linear model that explains the maximum variance of the X data for a given model complexity.

Linear PCA involves the orthogonal decomposition of X along directions that explain the maximum variation of the data. The largest eigenvector $t_1$ of the covariance matrix of X is usually called the first principal component loading. The eigenvector $t_1$ gives the direction of the first principal component. The projection of an original data point onto the $t_1$ eigenvector defines a point in the subspace. Projections of all original data points in the subspace on first eigenvector, constitute principal component scores along the dimension $p_1$. The second principal component is defined by the second largest eigenvector and so on. If the variables in X are correlated, as often is the case, most of the variation in the data set X after calculating *m* principal components with $m << n$ will have been explained. Mathematically, X can be written as

$$X = p_1 t_1' + p_2 t_2' + .... + p_m t_m' + E \qquad \text{(Eq. 2-1)}$$

where X is the normalized data set of process measurements or variables. $t_i$ (i=1,2,..., m) are eigenvectors of the covariance matrix of X. $p_i$ (i=1,2,..., m) are *m* dimensions in the subspace also called principal components. E is the residual data matrix.

More generally, we can write:

$$X = PT^{T} + E \qquad \text{(Eq. 2-2)}$$

where P is defined as principal component scores, and T is defined as principal component loadings or eigenvector of the covariance matrix of X.

By summarizing the information in X using the new variables ($p_i$, i=1, ..., m), one has reduced the dimensionality of the space from *n* to *m*. In geometrical terms this is equivalent to approximating the n-dimensional observation space by the projections of the observations down onto a much smaller m-dimensional space. Selecting *m* is very important in the PCA calculation. Ideally *m* is chosen such that there is no significant process information left in the residual E. Rather E should represent random error, and adding one more principal component would only result in fitting some of this random error. There are several ways for selecting *m*; one can proceed until the percent of the variation explained by adding an additional principal component is small. A better procedure is to use cross validation [24, 25] whereby one holds back a certain fraction of the observations (say 1/3) and then performs a PCA analysis on the remaining data. Following the PCA analysis the Square Prediction Error (SPE) for those observations is computed.

$$SPE = \| X - PT' \|^2 \qquad \text{(Eq. 2-3)}$$

This calculation is repeated until every observation has been left out once. The optimal order of the principal components model, *m*, is taken as that order minimizing the sum of the SPE values from the data used for the model development and the testing data.

## 2.2 FACTOR ANALYSIS

Another data reduction technique that is often used is factor analysis (FA). Although very similar to PCA, FA is different from PCA in one important aspect. FA assumes that the variability of a variable has two parts [26]. The first part, called common variability, is influenced by other variables in the data set. The second part of variability is independent of the effect of other variables (i.e., it is random). So, in a sense FA may be better at describing systems such as the pulping fiber line where not all the variations of the fiber line come from variable interactions alone.

Mathematically speaking factor analysis assumes that the variance of a variable can be broken down into two additive parts. The fundamental partition is one between that portion of the variance that a variable shares with other variables, its *common variance*, and that portion which is not shared, its *unique variance*. The factor analysis model can be described as

$$X = f_1 t_1' + f_2 t_2' + \dots + f_m t_m' + U \qquad \text{(Eq. 2-4)}$$

where X is normalized data set of process measurements or variables. $t_i$ (i=1,2,..., m) are eigenvectors of the modified covariance matrix of X. $f_i$ (i=1,2,..., m) are $m$ dimensions in the subspace also called factors. U is called the uniqueness matrix. The U matrix represents the *unique portion of the variance* of variables.

The factor analysis model for ith observation of a single variable j would be

$$X_{ij} = a_{ij}F_{1i} + a_{j2}F_{2i} + \dots + a_{jm}F_{mi} + U_{ji} \qquad \text{(Eq. 2-5)}$$

Where $X_{ij}$ =normalized score of the ith observation on variable j.

$a_{jk}$ = regression weight of the kth factor for predicting the jth variable.

$F_{ki}$ = score of ith observation on kth factor.

$U_{ji}$= uniqueness score for the observation on variable j.

The contribution of each factor to the variation in $X_j$ is given by the square of its regression weight for predicting that variable. The sum of the squares of these factor loadings (called *communality*, symbol $h_j^2$) is the proportion of the variance of variable j that is accounted for by the set of factors. Mathematically speaking,

$$Communality = h_j^2 = \sum_{k=1}^{m} a_{jk}^2 \qquad \text{(Eq. 2-6)}$$

$$Uniqueness = u_j^2 = 1 - h_j^2 \qquad \text{(Eq. 2-7)}$$

The total variance of normalized variable j follows from the above equation and is given by

$$\sigma_j^2 = 1 = u_j^2 + h_j^2 \qquad \text{(Eq. 2-8)}$$

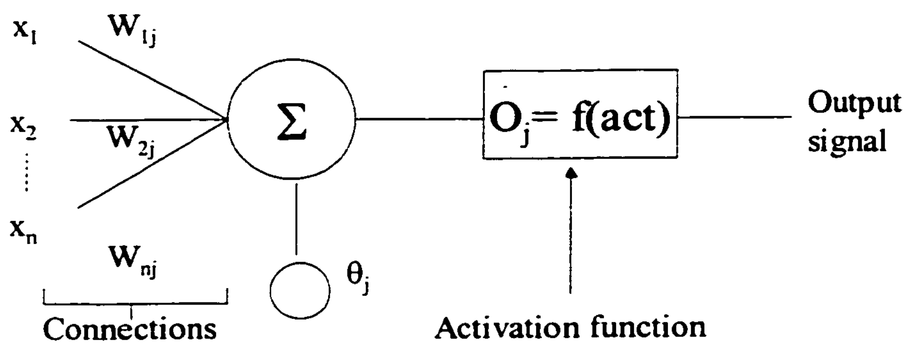
There are two problems associated with factor analysis. The first problem is finding out the optimum number of factors. The second problem is related to the reduction of correlation matrix (also known as estimating communalities of the variables). The decision about the optimal number of factors is based in part on estimates of common variance, and estimates of the communalities (common variance) depend in part on the number of factors one chooses to retain. In other words, these two problems are closely related and thus difficult to separate. The frequently used solution is to make an arbitrary decision about one feature, either number of factors or communality, and to obtain the solution of other by direct computation. In our project communality was estimated and number of factors was obtained by direct computation. The main reason for this choice was unavailability of any information about the number of independent events (i.e., number of factors) occurring in the process data.

Another important characteristic of factor analysis is factor rotation. Factor rotation can be defined as a process of finding an optimal reference frame to clarify the meaning of the factors. A number of computational algorithms such as quartimax and varimax have been developed that rotate the factors to simple structure without the intervention of personal bias. Each of these procedures involves finding a rotation that maximizes the variance in factor loadings across rows of the factor matrix. The criterion of meaningfulness of factors is, ultimately, a necessary condition for an adequate factor solution, but it is far from sufficient, unless the meaning derives from an understanding of the process, e.g., fiber line operation in our project. Note that the communalities of the variable remains unchanged by the process of rotation.

## 2.3 NEURAL NETWORKS

Both PCA and FA are good at describing a dataset containing linear interactions [23, 26]. To investigate non-linearities present in pulping and bleaching processes a non-linear multivariate technique such as neural network analysis is desirable. For many years linear modeling has been the commonly used technique in most modeling domains. Linear models have well-known optimization strategies. Where the linear approximation was not valid (which was frequently the case) the models suffered accordingly. Artificial neural networks have shown usefulness for making of non-linear models. Neural networks grew out of research in artificial intelligence; specifically, attempts to mimic the fault-tolerance and capacity to learn of biological neural systems by modeling the low-level structure of the brain [27]. These networks have been used to solve problems of prediction, classification or control in areas as diverse as finance, medicine, engineering, geology, and physics [28]. This sweeping success can be attributed to the capability of neural networks to model extremely complex non-linear functions. Neural networks are relatively easy to use as they *learn by example*. The user gathers representative data, and then invokes *training algorithms*, which assist neural networks in automatically learning the structure of the data. The user does need to have heuristic knowledge of the selection and preparation of data. For the selection an appropriate neural network, and interpretation of results, the level of user knowledge needed to successfully apply neural networks is much lower than would be the case using (for example) more traditional statistical methods.

Artificial neural networks can achieve some remarkable results using a simplified model of biological brain. Artificial Neural networks are mathematical systems that are comprised of a number of "processing units" that are linked via weighted interconnections. A processing unit is essentially an equation, which is often referred to as a "transfer or activation function" (Fig. 3.1). A processing unit takes weighted signals from other neurons, possibly combines them, transforms them and outputs a numeric result.




where $x_1, x_2, \ldots, x_n$ are input signals to the neural network (NN).

$W_{1j}, w_{2j}, \ldots, w_{nj}$ are connection weights inside the NN which are adjusted during NN training.

$\theta_j$ is the bias factor that can be adjusted during NN training process.

f(act) is the activation function which transforms set of input signals into output signal.

Figure 2-1. Details of an artificial neuron.

For any useful network, there must be inputs (which carry the values of input variables of the process) and outputs (which form predictions, or control signals). There must also be hidden neurons, which play an internal role in the network. The input, hidden, and output neurons need to be connected together. The layers have similar characteristics and execute their transfer simultaneously. A simple network has a *feedforward* structure [28]: signals flow from inputs, forwards through any hidden units, eventually reaching the output units. A typical feedforward network is shown in the Figure 2-2.

The behavior of neural networks, how they map input data to output data, is influenced primarily by the transfer functions of neurons, how they are interconnected and the weights of those interconnections. The Figure 2-3 shows three commonly used transfer functions, i.e., linear, nonlinear, and semi-linear (sigmoid). Typically, an architecture or structure of a neural network is established and one of a variety of mathematical algorithms are used to determine what the weights of the interconnections should be used to maximize the accuracy of the outputs produced. Neural networks are "trained", meaning they use previous examples to learn the relationships between the input variables and the predicted variables by setting these weights. Once these relationships are established, the neural network can be presented with new input variables and it will generate predictions.

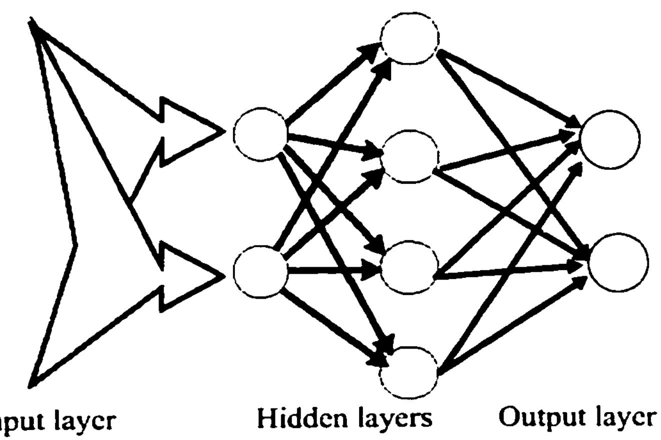



Figure 2-2. Neural network with input, hidden, and output layers

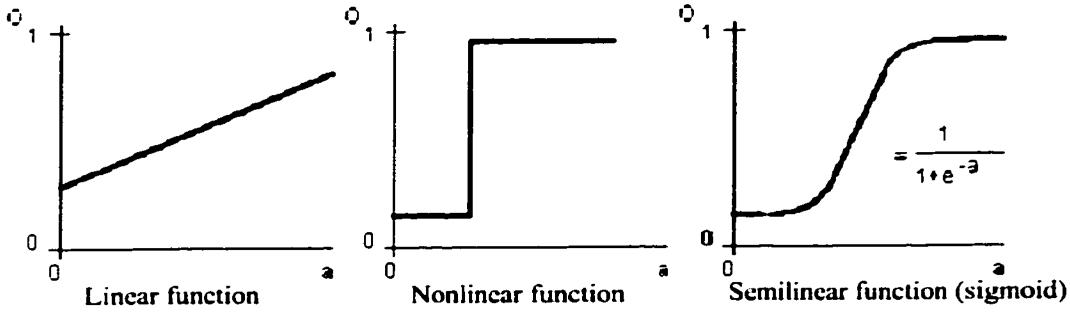



Figure 2-3. Commonly used transfer or activation function in neural networks.

A number of nonlinear dynamic models have been built by applying neural network techniques for chemical industries. Bhat et al. [29] used backpropagation technique to train a feedforward neural network. Schweiger [30] and Rudd used neural networks to predict and control paper machine parameters.

Similar work was done for an industrial chip refiner by Dumont, et al. [15]. The outputs of neural networks developed by Dumont compared favorably with industrial refiner data. For closed-loop control of specific energy in the refining process, Khorasani [16] proposed a backpropagation neural network-based controller to replace the self-tuning regulator (STR). In a commercial newsprint paper machine, Miyanishi [17] applied artificial neural networks to the diagnosis of paper web breaks. He extracted the essential causes of paper web breaks by using a three-stage neural network and backpropagation method.

## 2.4 TIME SERIES ANALYSIS

Time series (TS) is commonly used to investigate the dynamic relationship among variables when data are collected over a regular time interval or are time dependent

[31]. Time series provided insights into timelags and deadtime problems that are present in the dataset acquired from a pulp mill.

Mathematical model creation using time series is an extremely involved process. Much exploratory data analysis must take place before modeling is even attempted. This exploratory data analysis begins by looking at the probabilistic structure of the data set. If the probabilistic structure of dataset is unaffected by a shift in the time origin or a data series looks the 'same' at whatever point in time the observations are started and it shows similar behavior throughout its duration, then the time series is said to be stationary. Alternatively, a time series is said to be stationary if the mean and standard deviation of successive series do not change significantly with time. Time series have characteristics, which cannot be measured by the mean and standard deviation. For example, two kappa number time series (associated with two production runs) can have similar mean and standard deviation but have different short-term and long-term variations. Time Series procedures include the following:

1. AUTO CORRELATION: This is a method for comparing a sequence to itself to determine the correlation between successive measurements. It determines the degree of continuity of the data, if it is cyclic or periodic, i.e., if it repeats itself, if a trend is present or if the data are random. This is done by passing the sequence over itself and determining the goodness of fit at successive positions. This goodness of fit is the measure of similarity and dissimilarity.

2. CROSS CORRELATION: This procedure compares one sequence with another to identify and locate positions of high correlations between the two sequences. The techniques involved are slightly different to the autocorrelation procedure. Cross-correlations for positive and negative lags must be calculated as the entire length of one series must be moved past the other and the position of the theoretical lags zero is not known. The successive comparisons are called match positions and not lags.

For continuous processes where dynamics are important, as in pulp and paper mills, time series analysis techniques are particularly suited to analyze process variability. The process data routinely collected can be used to monitor control performance and identify the source of product quality variations. Croteau [32] et al. reviewed time series techniques and procedures with examples from various mill applications. Applications range from correlating the active alkali change to a continuous kraft digester with the pulp's kappa number in the blowline to the air/solids ratio with the carbon monoxide concentration in recovery-furnace flue gas.

Allison [33] et al. used closed-loop time series identification to develop the process model of chip level in a Kamyr digester. Based on model predictions, a generalized predictive controller (GPC) was designed. A significant reduction in chip level and P-no. (covariance matrix) variability was observed after GPC installation. Ritala [33] used time series analysis to develop a process analysis system. This system identified causes of quality variations by accommodating irregular signals as well as long process dead times.

# CHAPTER 3: PROJECT APPROACH

Multivariate data analysis techniques were used in this research project to provide a concise mathematical description (model) of the fiber line. Models of the fiber line were developed to improve the process as well as to study the sources of product variability. The project approach consisted of following four steps

1. Preprocessing of raw data acquired from fiber line operations.
2. Analysis of preprocessed data using multivariate techniques to predict important process output variables.
3. Interpretation of analysis results from step 2.
4. Comparison of the results from various multivariate methodologies.

Project steps are described in detail in the following paragraphs.

## 3.1 RAW DATA PREPROCESSING

There are always problems with real world data. Data problems are situations that prevent efficient use of data analysis tools or that result in generating unacceptable results. It is reasonable to either rectify the data problems ahead of time or recognize the effects of data problems on the results. Data problems in pulp mill data can be caused by the changes in process characteristics and operating conditions, as well as in the data collection process itself. For example, pulp mill process equipment such as continuous digesters, bleach towers and storage vessels often have large timelags and deadtimes as these hold large volumes of chips/pulp and liquor. In the context of data analysis, a great deal of predictive information is lost unless the deadtime present in the raw data set is accounted for.

Data from process operations undergoes various transformations and changes before they are acquired for analysis. A necessary condition to get any useful result from a data analysis involves elimination of effects of transformations and changes implicit in the raw data. Data preprocessing consists of all the actions taken before the actual data analysis process started. It can be defined as a transformation T that transforms the raw mill data vectors $X_{ik}$, to a set of new data vectors $Y_{ij}$

$$Y_{ij} = T\,(X_{ik}) \qquad \text{(Eq. 3.1)}$$

such that: (i) $Y_{ij}$ preserves the "valuable predictive information" in $X_{ik}$,

(ii) $Y_{ij}$ is more useful than $X_{ik}$.

In the above relation:

$i = 1, \ldots n$ where $n$ = number of observations,

$j = 1, \ldots m$ where $m$ = number of variables after preprocessing,

$k = 1, \ldots l$ where $l$ = number of variables before preprocessing,

and in general, $m < l$.

Data preprocessing is performed to achieve a number of objectives. In addition to solving data problems, such as corrupt, noisy data or irrelevant data in the data sets, it helps in learning more about the nature of the data. Some of the data problems in the pulp mill study were related to the size of the dataset, data filtering, and choice of variables for predicting output variables. These problems are discussed in detail in the following sections.

### *3.1.1 SIZE OF DATASET*

The size of a dataset acquired for analysis from a pulp mill and bleach plant is quite important. A comprehensive analysis of a large volume of data was difficult as the dataset represented a variety of process upsets and disturbances as well as a number of different production runs. Additionally, the dataset may have contained a number of production runs with different operational campaigns. Efforts were made to reduce the size of the dataset so that only few important campaigns could be modeled. Production runs, based on constant chip meter RPM or production rate, were chosen as representative of pulp mill operations. Similarly, constant production runs were chosen for the bleach plant. The time duration for pulp mill and bleach plant datasets were similar as the research project focused on determining upstream variables (i.e., pulp mill variables) that were affecting important output variables in the bleach plant.

### *3.1.2 IRRELEVANT DATA*

A large number of irrelevant variables were present in the preliminary stages of the multivariate analysis. The number of variables in the preliminary analysis could be up to 150 for pulp mill and over 200 variables for bleach plant. Extraction of meaningful variables from a large set of variables was not a trivial task. The main goal of eliminating irrelevant data was to narrow the search space in the data analysis. An important criterion in removing irrelevant variables was the presence or absence of predictive information about important output variables such as kappa number or bleaching cost in the variables. Some of the steps taken to remove irrelevant variables from the analysis were the following:

(i) Variables related to mechanical equipment such as mixers, pumps etc. were removed from the analysis. These variables did not have any significant, direct impact on process operations.

(ii) Variables whose trend did not change significantly in a production run were removed. Variables that were removed had a coefficient of variation (COV) of less than 3%.

(iii) Variables were removed from the analysis based on the change of accuracy in the developed model. Variables with the least predictive ability for kappa number or bleach cost were removed from the analysis.

### *3.1.3 DATA FILTERING*

The raw dataset from the mill contained noisy trends for process variables. Since the study was focused on long-term variations of important output variables, data filtering was required. Some of the most common filtering techniques are the following:

(i) Time domain filtering, where the mean or median of the measured data in a window of predetermined size is taken.

(ii) Frequency domain filtering, where data are transformed via Fourier analysis and high frequency contributions are eliminated from the data.

(iii) Time-frequency domain filtering, where the measured data are transformed simultaneously in the time and frequency domain.

For our study, time domain filtering was used with a five-hour window. In other words, five-hour moving average of variables was used in the multivariate analysis.

### *3.1.4 TIMESHIFTING IN RAW DATA*

The fact that large quantities of process data are readily available leads one to consider using these data to improve process operations. However, there exist several problems inherent in the pulping and bleaching processes that serve as obstacles in using such data. Pulp and paper processes often have large timelags and deadtimes. Process equipment such as continuous digesters, bleach towers and storage vessels hold large volumes of chips/pulp and liquor, which may be conveyed as a plug ("plug flow") through the equipment. Additionally, large storage tanks (high-density tanks) often separate process units (pulp mill, bleach plant, paper machine, etc.). These storage tanks isolate one area from upsets in other areas. The resulting timelags may vary from minutes to several hours, depending on the production rate and other processing variables. When analyzing pulp and paper process data, a great deal of predictive information is lost unless the deadtime present in the raw data set is accounted for.

The loss of predictive information in kappa number trend with respect to total bleaching cost trend is shown in Table 3-1. Theoretically total bleaching cost should be strongly correlated with kappa number, as the cost of bleaching pulp is directly proportional to the amount of lignin in the pulp (*kappa number*). However, the correlation between kappa number and total bleaching cost for raw data is quite low (-0.36). In other words, it is difficult to predict total bleaching cost from kappa if the data are not conditioned. The low correlation is due to time shifting of total bleaching cost with respect to kappa number. As the effects of time shifting are progressively removed, the correlation between kappa number and total bleaching cost increases (Table 3-1, Figure 3-1). Figure 3-2 illustrates approximate retention times in the Georgia Pacific Ashdown pulp mill fiber line at a chip meter speed of 12 rpm (900 TPD). Looking at this figure, one can easily understand why time shifting of process data was required before any sensible analysis could be performed.

**Time-synchronization**

The raw data set implicitly contains the dynamics of the digester, diffusers, high-density tanks, and bleaching towers. In time synchronization, the dynamic effects of process equipment and storage towers, on individual variables, are eliminated from the raw data. The timeshifted dataset contains only correlations among variables, not auto correlation of each variable (which represents the process dynamics).

For time synchronizing correlated variables such as kappa number at the end of cooking and kappa number after washing, the cross-correlation function of time series analysis can be used. However most of the variables in pulping are not correlated in a similar manner. For time synchronizing all variables an algorithm was developed and implemented in the VBA (Visual Basic Application) language that comes with

Table 3-1. Effect of timelag on correlation coefficient between kappa number and total bleaching cost.

| | *Correlation with Kappa number (filtered)* |
|---|---|
| Total Bleach Cost (*filtered*) | -0.360 |
| Total Bleach Cost (timeshift, 30 hr) | 0.184 |
| Total Bleach Cost (timeshift, 60 hr) | 0.749 |
| Total Bleach Cost (timeshift, 90 hr) | 0.510 |

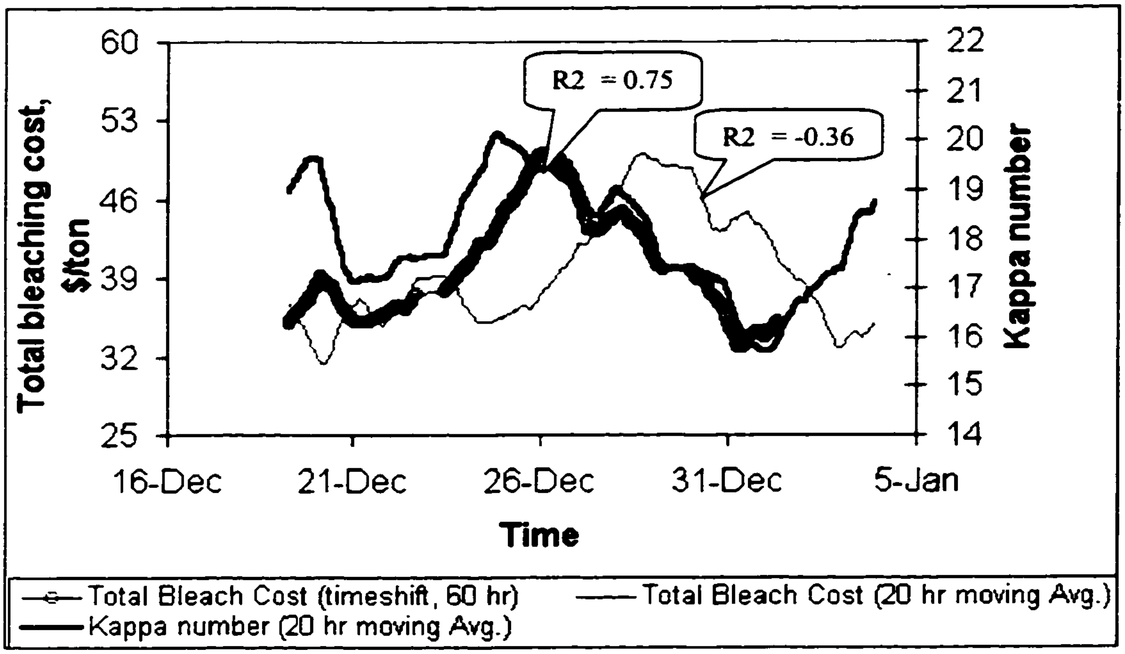



Figure 3-1. Effects of timeshifting on correlation between kappa number and total bleaching cost.

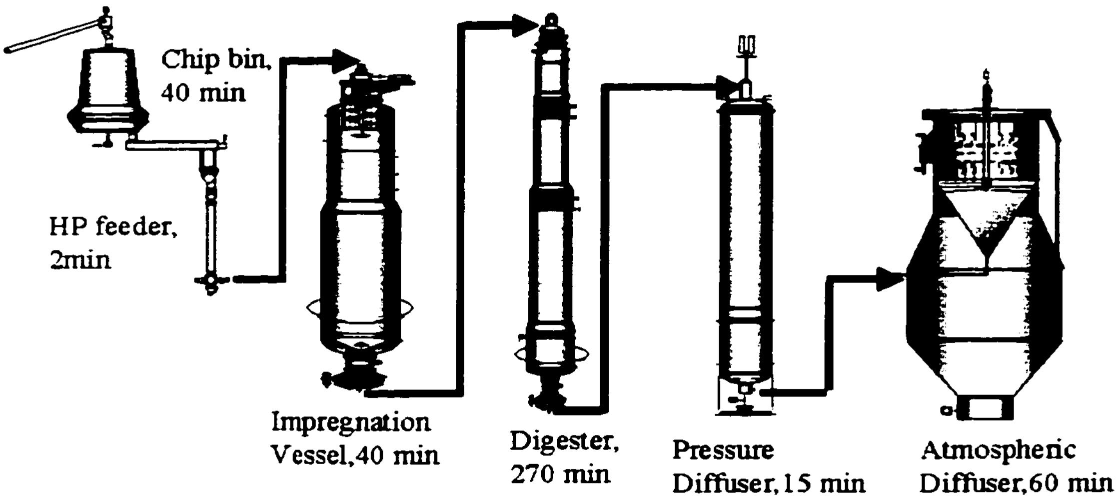



Figure 3-2. Transport delays in Ashdown pulp mill at 12 chip meter rpm.

the MS-Excel program. In the algorithm mill, the fiber line is divided into different areas based on retention time of pulp in those areas. Essentially there were two types of mill areas. In the first type of mill area, pulp retention time was directly related to the production rate, e.g., continuous cooking section and bleach plant towers. In the continuous cooking section, chip meter RPM determines the time pulp sample spends in that area (actual time is related to design volumes of the pulping equipment as well as chip column compaction in the digester). Similarly, bleach plant production determines the time pulp spends in bleach plant towers.

In the second type of mill area, pulp retention time was variable and could not be directly from fiber line production rate. High-density (HD) storage vessels are an example of this type of mill area. Different area types in Ashdown fiber line are shown in Figure 3-3.

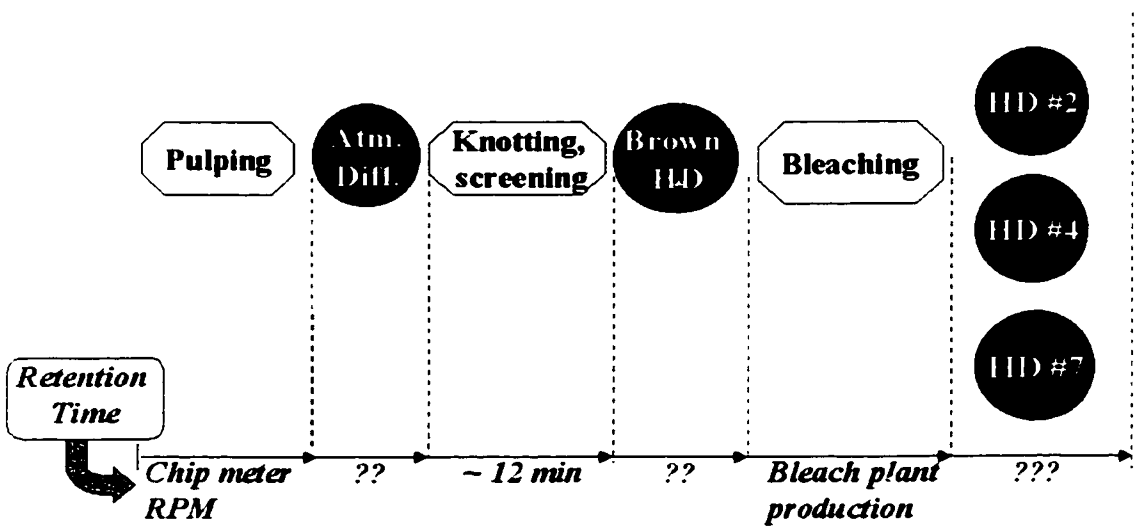



Figure 3-3. Different area types (based on pulp retention time) in Ashdown fiber line. (* *circle represent variable time areas.*)

Finding retention time in high density (HD) towers was the most difficult aspect of the time synchronization calculation. In the case of HD towers, the retention time is determined by the entry and withdrawal rate of pulp, which can be highly variable. The method for calculation of retention time towers is described below.

When a sample enters a HD tower the level of the tank in tons is recorded. Assuming that the pulp moves as a plug through the tower, the retention of the pulp sample is determined by the rate the plug moves downwards, as determined by the tag value of the flow meter just after the HD. Mathematically speaking, the retention time would be given by the difference of the limits of the following integral.

$$M_o = \int_{t_0}^{t_r} F_t dt \approx F_{avg}(t_r - t_0) \qquad \text{(Eq. 3-1)}$$

where,

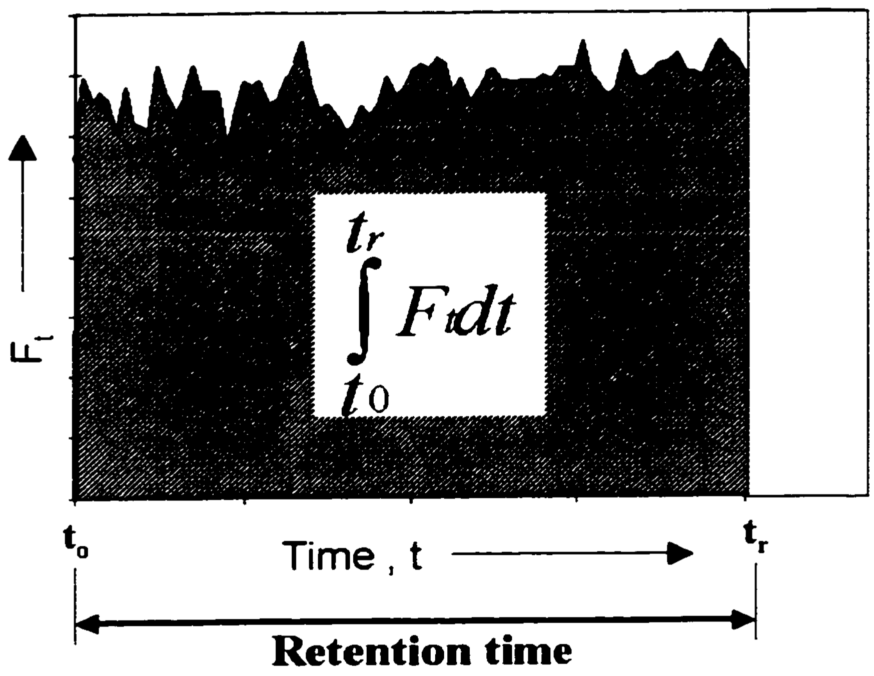
$M_0$ is the amount of pulp in the HD tower when a pulp sample comes into the tower, in tons.

$F_t$ is the rate at which pulp is drawn off the tower, in tons per hour.

$t_0$ represents the time pulp sample comes into the HD tower.

$t_r$ represents the time when this pulp sample leaves the HD tank.

The retention time for the pulp sample coming into HD tower at $t_0$ hour is $(t_r-t_o)$ hours. For calculating retention time, one needs to know either $t_r$ or $F_{avg}$. Both $t_r$ and $F_{avg}$ are functionally related as $F_{avg}$ depends on the pulp flow profile between time periods $t_r$ and $t_o$. Since both $t_r$ and $F_{avg}$ are unknown, the algorithm calculates the approximate part of the equation by assuming a value for $t_r$ and then iterating to reach the solution (detailed iteration algorithm is presented in Appendix A). The retention time in a tank is a variable




* Note that retention time in HD
changes as $F_t$(machine pull) profile changes.

Figure 3-4. Retention calculation for a particular pulp sample

quantity as it depends on the flow rate profile of flow into and out of tank, which changes over the period of time (Figure 3-4). All time-synchronization steps discussed so far pertain only to a particular pulp sample. In order to time-synchronize a large dataset, several retention time calculations were performed. These calculations were quite computationally intensive. The exact time (varying from twenty minutes to several hours on a Pentium 300 MHz machine) and number of iterations depended on the number of observations in the dataset.

## 3.2 MULTIVARIATE ANALYSIS OF PREPROCESSED DATA

In the dataset generated by pulp mill operations, there may be hundreds of measurements. Although the number of measurements is large, only few events may be driving the pulping process. In the second step of the process analysis, data from process measurements were mapped into meaningful descriptions of event(s) occurring in the pulp mill. The multivariate techniques principal component analysis, factor analysis, and neural networks were used to describe process characteristics of the pulping process. These techniques were also used for predicting important output variables such as kappa number and bleaching cost. The goal of using all multivariate analysis techniques was to predict important output variables (called response variables) as function of input variables to the process (called explanatory variables) using as few independent characteristics as possible. The problem was not trivial because explanatory variables as well as response variables were correlated among themselves. A strong correlation structure between input and output variables is also needed to mathematically represent pulp mill operations.

The goal of pulp mill operations is to produce pulp of the best available quality with minimum costs. However, quality variations do occur due to several types of disturbances coming into the pulp mill. For digester, quality variations are measured in terms of kappa variations (i.e., deviation of kappa number from its mean value).

## 3.2 INTERPRETATION OF ANALYSIS RESULTS

Results from the multivariate models were interpreted in the third step of the project. Predictions of important output variables from multivariate models were analyzed to see if these made sense from a process operations point of view. In addition to this, the underlying patterns in the dataset were identified from the variability analysis. The main goal of variability analysis is to be able to understand, and therefore reduce, variability of important variables such as kappa number. The variations in output variables such as kappa number can be split into two important components. The first component is variation in kappa attributable to other process variables. The second is random variations around mean kappa number that are not attributable to process variables. Controlling the variability of input variables that covary with kappa can reduce variations in kappa number.

## 3.3 COMPARISON OF METHODOLOGIES

A comparison of results from the application of multivariate techniques was done in the final step of the dissertation project. The performance of each of the multivariate analysis techniques in understanding and optimizing mill fiber line was compared. A basis to compare the utility of multivariate tools was developed. The evaluation was based on several criteria. These criteria were data preprocessing ability, ease of use, and accuracy of prediction of important output variables and amount of useful information obtained through the use of methodology. The ability of a methodology to provide predictive information about important output variables like kappa number and total bleaching cost was of considerable value from a process operations point of view. An ability to improve process understanding (eventually leading to optimization) was another important criterion. When combined together these criteria helped in assessing the utility of data analysis techniques.

# CHAPTER 4: RESULTS OF KAPPA ANALYSIS

Two case studies were performed based on the project approach described in Chapter 3. The first case study focused on the pulping portion of the fiber line. The pulp digester is a major unit operation in the pulp mill. Its proper control is very important to pulp production in the entire fiber line. The main objective of digester operation is to produce the same quality pulp from differing chips as they pass through the digester. A decrease in pulp quality variations results in economic fiber line operation.

Kappa number is the main quality control variable in digester operation. It is a measure of residual lignin in the pulp. A pulp sample with uniform kappa number distribution represents uniform quality pulp. All pulp mills strive for tight kappa number control. In many cases, however, kappa number is not tightly controlled and the sources of kappa variability are unknown. In the first case study of the dissertation, kappa number prediction models were developed using multivariate techniques. Prediction models were used to analyze sources of kappa variability in two pulp mills.

Weyerhaeuser Longview mill data was used to develop models for predicting kappa number of pulp out of the continuous digester and oxygen delignification reactor. Prediction models were developed using factor analysis. Results from the Longview analysis indicated the presence of deterministic relationships *(for kappa number prediction)* among process variables. Sources of digester kappa variability were also found. The Longview study was followed by a detailed study which used the dataset acquired from Georgia Pacific's Ashdown mill. In this study, the multivariate techniques principal component analysis, factor analysis, and neural network analysis were used to develop models for K-number (equivalent of kappa number at Ashdown mill) prediction. Details of Weyerhaeuser Longview mill and Georgia Pacific mill study are presented in the following paragraphs.

## 4.1 WEYERHAEUSER KAPPA STUDY

The Weyerhaeuser Longview pulp mill uses softwoods, Douglas fir, Hemlock/Pine mix and hardwoods to produce bleached paperboard and fine paper products. It has a capacity of about 1250 tons per day (TPD). The mill fiber line has a Kamyr continuous digester with an extended modified continuous cooking (EMCC) section. The digester also has capability of running in losolids mode which like EMCC cooking, can produce pulp with a very low lignin concentration and high fiber strength. The pulp, cooked in the continuous digester, is washed in a pressure diffuser and then stored in one of two blow tanks with capacity of 175 tons each. From the blow tanks pulp goes to knotting and screening sections. The screened pulp is washed in brownstock washers before it is sent to an oxygen delignification stage. After $O_2$ delignification, pulp is washed once again before it is stored in brown stock high-density storage tanks.

The digester throughput or production rate is determined by chip meter revolutions per minute (rpm). For the Longview mill, a digester chip meter rpm of 9.5 corresponds to a production rate of 1431 air-dried tons of pulp per day (ADST/D). White liquor is used to transport chips, impregnate chips, and is added to the digester at various points in the cooking process. White liquor is distributed throughout the pulping system in the following approximate proportions:

- 60% - Chip Feed System (chip chute, high pressure feeder, top well of impregnation vessel)
- 10%- Bottom Circulation System

- 10%- Modified Cook System
- 20%- Wash Circulation System

After being used in the digester, white liquor is colored with the lignin and other solids from the wood chips, and becomes black liquor. Pulp is sent to the bleach plant section and finally to paper machine.

Kappa number is used to determine the extent of lignin present in the pulp after cooking. The data from Longview pulp mill was used in the first phase of kappa number case study. Factor analysis was used to see if there was a deterministic process present in the acquired mill data that caused kappa number vary. The study focused on the development and application of factor analysis models to analyze kappa variability for the Kamyr continuous digester and oxygen delignification reactors.

### *4.1.1 DATA PREPROCESSING*

A dataset, consisting of 93 variables and 750 observations representing three days of production, was acquired from the Longview mill. For the analyzed dataset, the mean kappa number was 24.04 with a standard deviation of 4.1 units. The coefficient of variation of kappa was 16.9%. For the oxygen delignification reactor, post $O_2$ kappa number had a mean value of 11.74 with standard deviation 0.79 units. The coefficient of variation of post $O_2$ kappa was 6.77%.

Douglas fir production runs at the Longview mill were chosen for the kappa number study as it had the highest kappa variability (COV of 16.9 %). Several variables were removed from the original dataset in the data preprocessing stage. Most of the mechanical control variables such as differential pressures (DP) of digester screens, DP of pressure diffuser were removed from the analysis. A number of flows around digester, which contained redundant information, were also dropped in the preprocessing stage. Other variables with low variability (coefficient of variation less than 2%) were also removed from the analysis. In the end, forty-one variables were selected for the digester kappa number study whereas thirty-two variables were used for $O_2$ kappa study. Datasets for the digester and $O_2$ stage were time synchronized using time delays shown in Table 4-1. The conditioned and time synchronized datasets were used for prediction and variability analysis of kappa number. The results of the kappa variability study for Weyerhaeuser Longview mill are discussed in detail in the following sections.

Table 4-1.Time delays in the Longview mill fiber line.

| S. No. | Section | Time, in minutes |
|---|---|---|
| 1 | Impregnation vessel | 55 |
| 2 | Digester, trim section | 60 |
| 3 | Digester, extraction section | 48 |
| 4 | Digester, MC section | 66 |
| 5 | Digester, Wash section | 103 |
| 6 | Digester, bottom section | 41 |
| 7 | Pressure diffuser | 10 |
| 8 | Blow tanks | 150 |
| 9 | Screen room | 10 |
| 10 | Brown stock press | 10 |
| 11 | O2 reactor | 60 |
| 12 | Post O2 press | 10 |

### *4.1.2 MULTIVARIATE ANALYSIS*

The pulp mill was divided into two sections to study them separately. These sections were the digester and oxygen delignification reactor. Two separate models were developed for these sections. The data analysis flow sheet is shown in Figure 4-1. Time synchronized and conditioned data, representing 3-4 days of fir production, was analyzed using factor analysis. For development of the model, the conditioned data were partitioned into two separate parts. One part was used to analyze the data and build a factor model, while the second part was used to validate the model. In other words, the resulting model was applied to the second part of data to test the accuracy and robustness of the model. The exact division of a dataset into model development and model validation datasets was based on two criteria. "60% of *data for development and 40% for validation*," being the first criterion. Approximately sixty percent of data is used for model development and the remaining dataset for model validation. The exact division of dataset is based on the extent of kappa variation, which was the second criterion. Since multivariate models are databased models, the dataset with the largest variation were used for model development. The partitioning information for the two sections is summarized in Table 4-2.

Table 4-2. Partitioning of data for factor models.

| **Section** | **Total observations** | **Observations for model development** | **Observations for model validation** |
|---|---|---|---|
| **Digester** | 600 | 400 | 200 |
| **$O_2$ delignification** | 330 | 230 | 100 |

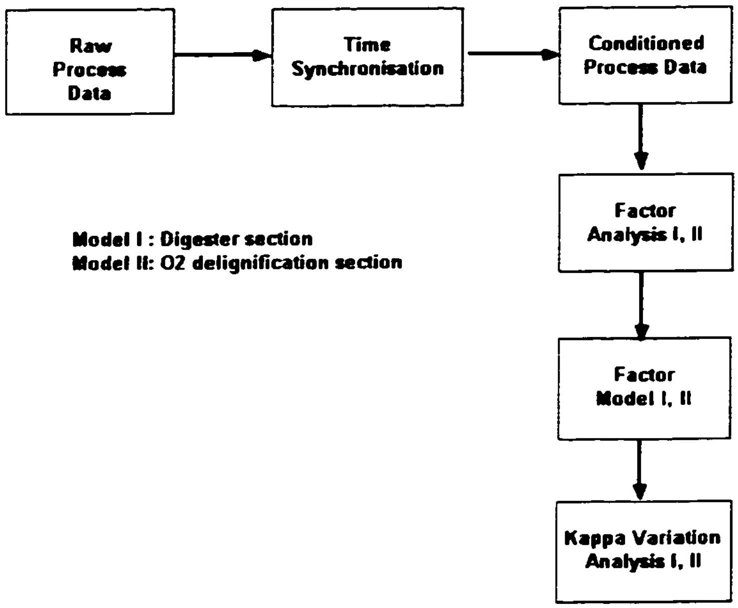



Figure 4-1. Flow sheet of Longview mill kappa prediction and variability analysis.

The time synchronized and conditioned data could now be analyzed. The presence of non-linear relationships was determined by looking at xy plots. Because of the limited data range in the single grade, the non-linear relationships were found to be negligible and thus not used. In addition, the presence of cross interrelationships among the process variables was determined. Several significant cross interrelationships were detected and included in the model. For example, Pre $O_2$ stage kappa number strongly correlated with digester kappa number; therefore, only pre $O_2$ kappa was included in the analysis for the $O_2$ delignification reactor.

In the next step, several variables were eliminated from the analysis for the sake of parsimony. In this step, it was determined that a large number of the variables were not significant as these had low coefficient of variation and low average correlation coefficients. Approximately one half of the available process variables were removed from the analysis for pulping and $O_2$ delignification. The remaining 41 variables (digester) and 32 variables ($O_2$ stage) were used in factor analysis.

**Kamyr digester**

All factor analysis modeling was done using FactNet (a statistical analysis software developed by Pacific Simulation based in Moscow, ID). A factor model using forty-one variables was created for the digester. The resulting accuracy for the factor model for all observations is shown in Figure 4-2 and Figure 4-3. The R-square value for the plot in Figure 4-2 is approximately 70%. The model predicted the kappa number reasonably well for the data used in the model development stage as well as model validation stage.

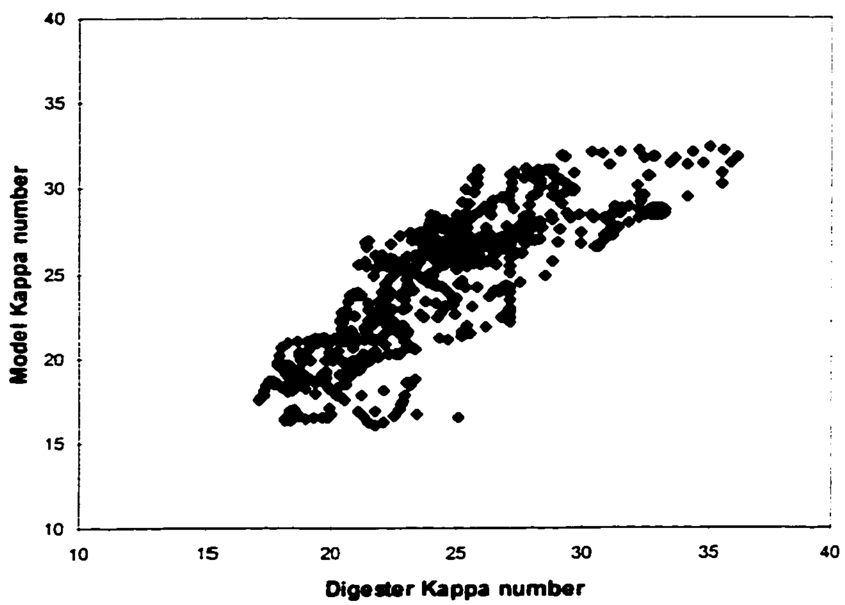



Figure 4-2. Factor analysis kappa number predictions for the digester (development and validation dataset).

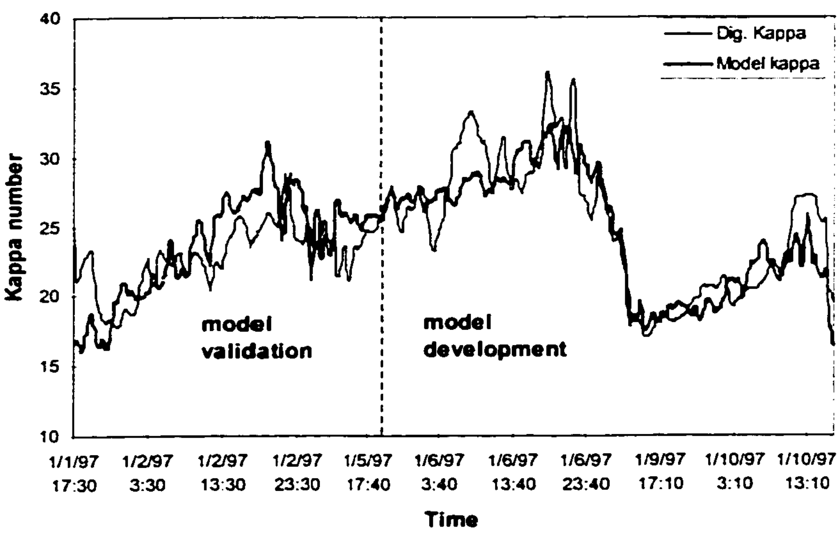



Figure 4-3. Digester kappa number trend for factor analysis predictions.

The model predicted kappa number for the data not used in the model development (more than 200 observations) as shown in Figure 4-3. As can be seen in Figure 4-3, the model tracked the behavior of the measured kappa over the entire time period. The model doesn't track spikes (large but brief excursions) in kappa number due to process upsets, which have been filtered out in the data conditioning stages.

**$O_2$ Delignification reactor**

A factor model of the $O_2$ reactor, using thirty-two variables, was created. The resulting accuracy for the factor model for all observations is shown in Figure 4-4. The R-square value for plot in Figure 4-4 is approximately 70%. The model predicted the kappa number well for the data used in the model development stage as well as model validation stage. The model predicted the kappa number for the data not used in the model development (more than 100 observations) as shown in Figure 4-5.

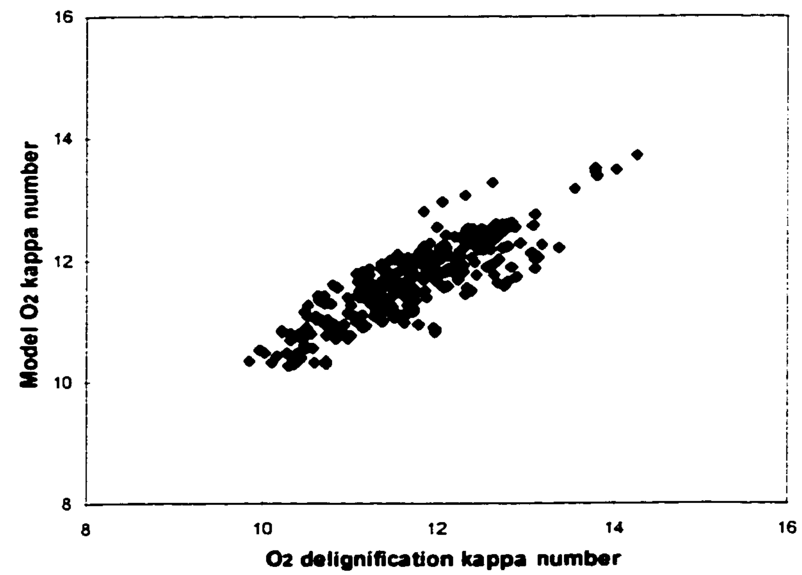



Figure 4-4. Factor analysis kappa number predictions for the $O_2$ reactor (development and validation dataset).

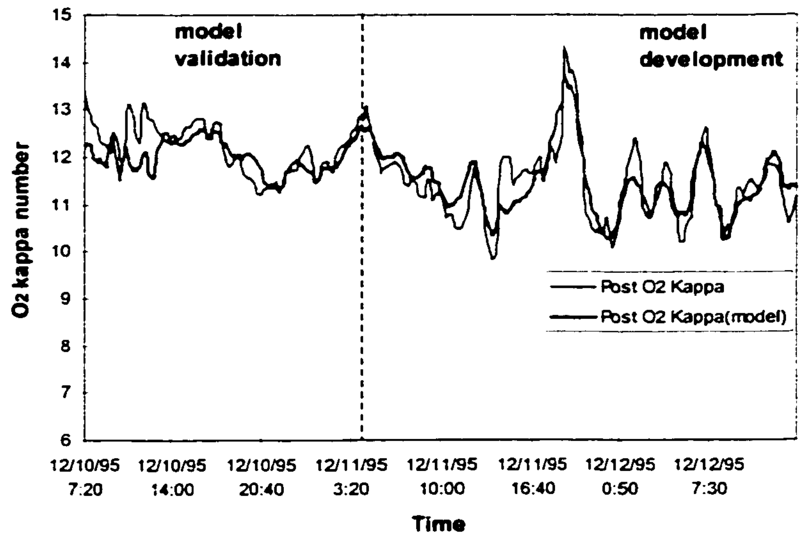



Figure 4-5. $O_2$ reactor kappa number trend for factor analysis predictions.

### *4.1.3 INTERPRETATION OF RESULTS*

**Kamyr digester**

After modeling the kappa number, the underlying patterns in the dataset were identified from the variability analysis. The variability analysis indicated that there were four major patterns underlying the process variables used in the analysis. The patterns are shown in Table 4-3 in the order of importance in the factor analysis model.

Table 4-3. Common factors and primary patterns present in digester data from mill.

| *Factor* | *Primary pattern* |
|---|---|
| Residual Alkali Factor | As MCC residual alkali increases, kappa number decreases. |
| Chip factor | As chip moisture and bulk density increase, kappa number increases. |
| Wash Factor | As wash white liquor, wash circulation increases, kappa number increases. |
| Heating control factor | As BC temperature, Upper extraction screen temperature increase, kappa number decreases. |

The factor analysis showed that 45% of the variation in kappa number was due to variations in residual alkali, chip, wash, and heating control factors. These factors are not true mechanistic parameters as used in first principle modeling. Instead the factors are statistical representations of mechanistic parameters calculated by the factor network.

As shown in Table 4-3, MCC residual alkali increases lead to kappa number decreases. Kappa variations will track liquor alkalinity if other process variables were causing kappa variability, i.e., higher residual alkali will correlate with pulp having higher kappa numbers. This was not the case, indicating that kappa deviations were a result of variation in alkali to wood ratio.

Chip factor is highly correlated with chip bulk density and chip moisture. An increase in chip bulk density results in a kappa increase. This is explainable as the digester control at Longview is based on the volumetric feed rate of chips (i.e., on chip meter speed). For the same chip meter speed, higher bulk density chips will result in more chip weight through the digester at a chemical charge similar to that of lower bulk density chips. As a result, the chemical to wood ratio decreases causing higher kappa pulp. Similarly, higher moisture chips will result in more water going into the digester and thus diluting incoming white liquor. The liquor to wood ratio goes up resulting in higher kappa number.

The heating control factor showed that an increase in BC temperature and upper extraction screen temperature result in a decrease in kappa. This result reinforces the importance of good temperature control.

Factor analysis shows that by controlling variables correlated with residual alkali factor, chip factor, wash factor, and heating control factor, the variations in kappa number can be reduced by 45%. The remaining 55% variations in kappa number was partly due to random variations and partly due to variables not included in this analysis such as chip quality variables. In terms of mill operations, a reduction in kappa number variability by 45% can be quite significant. Such a reduction in the case of the Longview mill would bring kappa coefficient of variation to under 10% resulting in improved product uniformity for unit operations downstream from pulping.

**$O_2$ Delignification reactor**

After factor analysis modeling the $O_2$ kappa number, the underlying patterns in the dataset were identified to do the variability analysis. The analysis indicated that there are three patterns underlying the process variables used in the analysis that affect kappa number after $O_2$ stage. The patterns are shown in Table 4-4 in the order of importance in the model. Sixty percent of the post $O_2$ kappa variation was either random or due to variables not included in the analysis. The factor analysis indicates that 40% of the variations in kappa number after the $O_2$ stage were contributed to by the variations in digester factor, stage factor, and pulp-flow factor. Pre $O_2$ stage kappa is strongly related to post $O_2$ kappa (Table 4-4). In other words, the variability of kappa out of the digester was a major source of variation in kappa after the $O_2$ reactor. It appears that the feed-forward control on the $O_2$ reactor is unable to handle all the

Table 4-4. Common factors and primary patterns present in $O_2$ delignification stage data from mill.

| *Factor* | *Primary pattern* |
|---|---|
| Digester factor | As Pre $O_2$ kappa increases, kappa number increases. |
| Stage factor | As $O_2$ reactor zone temperature increases, kappa number decreases. |
| Pulp flow factor | This factor is strongly correlated to differential pressures of screens and knotters. |

incoming kappa variability. The absence of a correlation between caustic concentration and post $O_2$ kappa suggest caustic charges in $O_2$ reactor is not having much effect in controlling kappa variability. Only $O_2$ temperature seems to have some effect on kappa variability. The stage factor indicates the importance of temperature control in the $O_2$ reactor. The pulp flow factor suggests the influence of screens and knotter operations in oxygen delignification. The pulp flow factor could be suggesting effects of pulp uniformity on kappa variability.

The factor analysis shows that by controlling variables correlated with the digester factor, stage factor, pulp-flow factor, the variations in kappa number could be reduced by 40%. The remaining 60% of the variations in kappa number was partly due to random variations and partly due to variables not included in this analysis.

## 4.2 GEORGIA PACIFIC K-NUMBER STUDY

Results from the Weyerhaeuser study showed that kappa variations could be related to the input variables in the data. These results led to a larger study involving detailed analyses of the data acquired from the Georgia Pacific (GP) Ashdown mill. The objectives of study were to test the effectiveness of all multivariate techniques in predicting K-number and to give insight into kappa number variability.

The pulp mill at Ashdown uses southern pine to produce bleached kraft pulp. With a capacity of about 1000 TPD, the fiber line has a Kvaerner continuous digester system followed by a five-stage bleach plant (See Figure 4-6). The washed pulp coming out of the digester goes into a pressure diffuser followed by two-stage atmospheric diffusion washing. The atmospheric diffuser sits at the top of a high-density tank. Brownstock pulp is then stored in a high-density tank before it is bleached. At the time of study, the bleaching sequence was D/CEoD1E2D2 with 50% target substitution in the D/C stage.

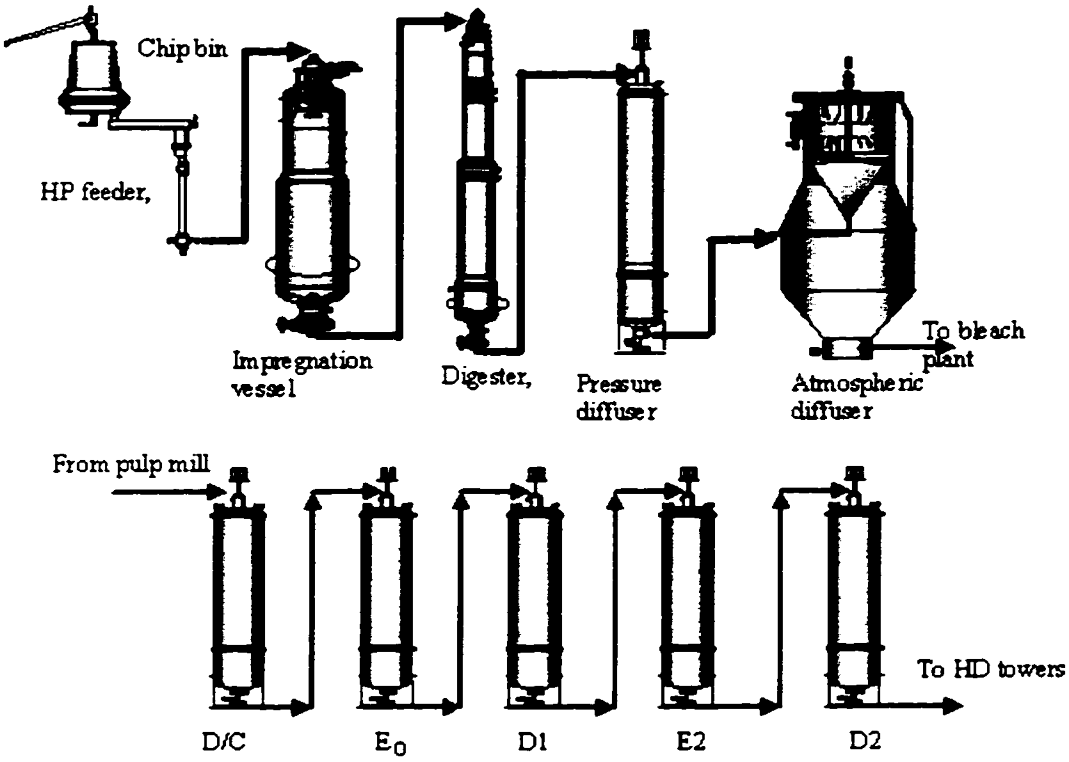



Figure 4-6. Fiber line at Georgia Pacific Ashdown mill.

### *4.2.1 DATA PREPROCESSING*

The pulping data, consisting of 137 variables and 5000 observations representing three months of production, was acquired from the mill databases. For the acquired dataset, digester K-number average was 18.46 with a standard deviation of 1.42 units. The coefficient of variation of K-number was 7.7%.

The acquired pulp mill data consisted of multiple production runs with several startups and shutdowns interspersed between them. The focus of the K-number study was to analyze only steady state operations of the fiberline. To represent steady operations, three production runs were selected from the available dataset. Each of these datasets represented different production runs and operational strategies. Combined together, three pulping data sets represented approximately one month of mill production. In the next step of data preprocessing, a number of variables were removed from the dataset. Most of the removed variables were lower level control variables such as levels in tanks, silos or current flowing to pumps and mixers. A number of variables that were not very important from K-number modeling perspectives (e.g., reject flow out of high pressure feeder) were also removed. Other variables were removed from dataset based on their lower process importance and low variability (coefficient of variation less than 2%). In all eighty variables were removed from the pulping datasets. Data from the remaining fifty-seven variables were carefully edited to remove outliers, incorrect observations, and blank rows. The resulting dataset was then time synchronized using method described in Chapter 3. Each time-stamped set of observations was time synchronized separately to account for variable time delay present in the mill dataset. Table 4-5 presents representative time delays in different areas of the fiberline at a chip meter rpm of thirteen or pulp production rate of 975 TPD. Cleaned and time synchronized datasets were filtered in

Table 4-5. Time delays in different areas of pulp mill fiber line at chip meter rpm 13.

| S. No. | Section | Time delay, in min |
|---|---|---|
| 1 | To chip silo | 5 |
| 2 | Hula bin | 35 |
| 3 | Chip feeders | 2 |
| 4 | IV | 25 |
| 5 | Co-current cook zone | 107 |
| 6 | Wash zone | 100 |
| 7 | Pressure Diffuser | 15 |
| 8 | Atm. Diffuser | 76 |
| 9 | Knotters | 5 |
| 10 | Screens | 5 |
| 11 | Decker | 2 |
| 12 | UBLch HDTower | 60 |

Table 4-6. Datasets for K-number analysis (BCH is bottom circulation heater).

| Dataset | Number of variables | Number of observations | K-number | | | Important operational strategy change |
|---|---|---|---|---|---|---|
| | | | **Average** | **Stdev** | **COV** | |
| **Digester 1** | 57 | 373 | 18.65 | 1.40 | 7.49 | BCH installed |
| **Digester 2** | 57 | 270 | 18.44 | 1.31 | 7.09 | Chip meter RPM 13 |
| **Digester 3** | 57 | 183 | 18.11 | 1.05 | 5.77 | Chip meter RPM 13.5 |
| **Composite Digester set** | 57 | 725 | 18.54 | 1.14 | 6.12 | All of the above |

the next step of preprocessing. A five-hour moving average filter was used to filter time-synchronized datasets. At this stage there were three preprocessed datasets representing three production runs in the pulp mill. These three datasets had different operational strategies. A fourth dataset containing all three datasets was also created. It represents a combination of operational strategies. Details of preprocessed datasets are presented in Table 4-6. These datasets were used for prediction and variability analysis of K- number. Results of the multivariate analysis of all four pulping datasets were quite similar. Only results from the composite pulping dataset are presented in the following paragraphs.

### *4.2.2 MULTIVARIATE ANALYSIS*

The preprocessed composite pulping dataset was used to develop multivariate models using three techniques; factor analysis, principal component analysis, and neural network analysis. Fifty-seven pulping variables were used in all three multivariate models for the purposes of K-number prediction and variability analysis. In addition, a smaller set of thirty-five pulping variables was also used to develop another set of multivariate models. The smaller set of variables was used to see if K-number prediction improved with a simpler variable structure. The thirty-five variables were a subset of original list of fifty-seven variables. The smaller list of variables was obtained by removing strongly controlled variables (e.g., temperature around digester) and variables containing redundant information (e.g., flows at the bottom of the digester). Variables were removed from the analysis based on the change of accuracy of the developed model. The accuracy of the model was based on its predictive ability for K-number. A list of the developed models with their codes and descriptions is presented in Table 4-7. Tables 4-8, 4-9, and 4-10 present detailed list of variables used in different multivariate models for the K-number case study.

Table 4-7. Model code and description for K-number multivariate models.

| S.No. | Model code | Model description |
|---|---|---|
| 1 | FA_57 | Factor analysis model with fifty-seven variables. |
| 2 | FA_35 | Factor analysis model with thirty-five variables. |
| 3 | PCA_57 | Principal component analysis model with fifty-seven variables. |
| 4 | PCA_35 | Principal component analysis model with thirty-five variables. |
| 5 | NN_57 | Neural network analysis model with fifty-seven variables. |
| 6 | NN_35 | Neural network analysis model with thirty-five variables. |

Table 4-8. Pre-digester variables used in K-number case study.

| S.No. | Tag | Code | Description | model_57 | model_35 |
|---|---|---|---|---|---|
| 1 | 20-FC003 . | wl_flo | FV-3 WHITE LIQUOR FLOW | Y | Y |
| 2 | 20-FC008 . | fedr_prg | FV-3A HP FEEDER PURGE FLOW | Y | Y |
| 3 | 20-FC009 . | Reswlflo | FV-3B RES WHITE LIQ FLW | Y | |
| 4 | 20-FC027 . | iv_sluce | FV-61 IV SLUICE FLOW | Y | |
| 5 | 20-FI012 . | top_circ | F-6 TOP CIRC FLOW | Y | Y |
| 6 | 20-FI013 . | mkWL_flo | F-7 MAKE-UP LIQUOR FLOW | Y | Y |
| 7 | 20-FI022 . | chut_fl | F-24 CHIP CHUTE CIRC FLOW | Y | |
| 8 | 20-LC041 . | L10_lvl | L-10 DIG CHIP GAMMA LVL | Y | Y |
| 9 | 20-LI042 . | Srgelvl | DIG CHIP SG LVL CALC | Y | Y |
| 10 | 20-SC005 | CM_rpm | CHIP METER SPEED | Y | |
| 11 | 20-TI157 . | wl_temp | 2PM WHITE LIQUOR TEMP | Y | Y |
| 12 | K2-CAUST .PE | caus_stn | #2 MILL CAUSTIC STRENGTH | Y | |
| 13 | K2-SULF .PE | Wlsulf | #2 PULP WHITE LIQ SULFIDIY | Y | Y |
| 14 | K2-WLSULF.PE | wl_sulf | WHITE LIQUOR SULF TEST | Y | |

Table 4-9. Digester variables used in K-number case study.

| S.No. | Tag | Code | Description | model_57 | model_35 |
|---|---|---|---|---|---|
| 1 | 20-CHPLIQ. | C_liq_df | DIG CHIP-LIQ LVL DIFF CALC | Y | Y |
| 2 | 20-DPI071. | od_Dp | DP-11 DIG OUTLET DEV DP | Y | Y |
| 3 | 20-DPI072. | extA_dp | DP-16A EXTRACT SCRN DP | Y | Y |
| 4 | 20-DPI073. | exB_dp | DP-16B EXTRACT SCRN DP | Y | Y |
| 5 | 20-DPI125. | wash_dp | WASH SCREEN DIFF PRESS | Y | Y |
| 6 | 20-FC002A. | extA_fl | F-16A EXTRACT FLOW UPPER | Y | Y |
| 7 | 20-FC002B. | extB_fl | F-16B EXTRACT FLOW LOWER | Y | Y |
| 8 | 20-FC016 . | blo_flo | FV-12B DIG 'B' BLOW FLW | Y | Y |
| 9 | 20-FC017 . | cblo_flo | FV-13 COLD BLOW FLOW | Y | Y |
| 10 | 20-FC019 . | cnt_wash | FV-18 COUNTER WASH FLOW | Y | Y |
| 11 | 20-FC152 . | botc_byp | HV-31 BOT CIRC FLOW BYPASS | Y | Y |
| 12 | 20-FI001 . | dig160 | F-1 160# STM TO DIG FLOW | Y | Y |
| 13 | 20-FI135 . | flsh2_fl | DIGESTER DILUTION FACTOR | Y | Y |
| 14 | 20-HC070 .CO | blo_dilu | HV-87 BLOW LINE DILUTION | Y | |
| 15 | 20-II369 . | od_amps | DIG OUTLET DEVICE AMPS | Y | Y |
| 16 | 20-TC480 . | botcheat | BOT CIRC HTR DISC TEMP | Y | Y |
| 17 | 20-TI075 . | digstemp | 160# STM TO DIG TEMP | Y | Y |
| 18 | 20-TI085A. | t12_tmp | T-12 BLOW LINE TEMP | Y | |
| 19 | 20-TI085B. | blo_tmp | T-12B BLOW LINE TEMP | Y | |
| 20 | 20-TI089 . | wahC_tmp | T-20C WASH CIRC HTR IN TMP | Y | |
| 21 | 20-TI092 . | botC_tmp | T-60 BOTTOM CIRC TEMP | Y | |
| 22 | 20-TI128 . | topS_tmp | TOP SEP INLET TEMP | Y | |
| 23 | 20-TI156 . | blowtemp | BLOW LINE TEMP | Y | |
| 24 | 20-WASHRT.PE | wshcircR | 2PM WASH CIRC RATIO | Y | |
| 25 | K2-DIKNUM.PE | Lo_knum | KAMYR DIG. LOWER K NUMBER | Y | Y |
| 26 | K2-SOLAVG.PE | wbl_soli | #2 WBL % SOLIDS 8 HR AVG | Y | |
| 27 | K2-UPKNUM.PE | up_knum | KAMYR DIG. UPPER K NUMBER | Y | |
| 28 | K2-WBLRES.PE | wbl_res | #2 MILL WBL RESIDUAL | Y | Y |

Table 4-10. Post-digester variables used in K-number case study.

| S.No. | Tag | Code | Description | model_57 | model_35 |
|---|---|---|---|---|---|
| 1 | 20-AI242 . | was_cdc | WASH HTR 160# COND CONDUCT | Y | Y |
| 2 | 20-CC415 . | stk_PD | STK TO PRESS DIFF CONSIST | Y | |
| 3 | 20-FC138 . | stage1wsh | FV-27 1ST STAGE WASH FLW | Y | Y |
| 4 | 20-FC144 . | stage2wsh | FV-37 2ND STAGE WASH FLW | Y | Y |
| 5 | 20-FC414 . | dil_liq | DILUTION LIQ FLW | Y | Y |
| 6 | 20-FC423 . | was_PD | WASH LIQ TO PD FLW | Y | |
| 7 | 20-FI021 . | Washcirc | F-20 WASH CIRC FLOW | Y | Y |
| 8 | 20-HC096A. | wscr_bf | WASH SCREEN BACKFLUSH FLOW | Y | Y |
| 9 | 20-PC352 . | hb_steam | PD DISCHARGE PRESS | Y | |
| 10 | Atm. diffuser | atm_dfsr | TIME DELAY IN ATMOSPHERIC DIFFUSER | Y | |
| 11 | Brown HD | brownHD | BROWN HD TIME DELAY | Y | |
| 12 | K2-BSKNUM.PE | Bsknum | #2 MILL BROWN STK K NUMBER | Y | Y |
| 13 | K2-DKCOND.MI | dek_cond | #2 DECKER CONDUCTIVITY | Y | Y |
| 14 | K2-DWCOND.MI | DifW_cnd | DIFFUSION WASHER CONDUCT. | Y | Y |
| 15 | K2-DWCONS.MI | DifWcons | DIFFUSION WASHER CONSIST. | Y | |

**Factor analysis**

The time synchronized and preprocessed pulping data was analyzed using factor analysis. All factor analysis modeling was done using FactNet. The conditioned data with, 725 observations, was partitioned into two separate parts. The data partitioning was based on "60% of *data for development and 40% for validation*" criterion. The first 425 observations were used in building factor model while remaining 300 observations was used to validate the model. The model development dataset had coefficient of variation slightly higher than that of model validation data.

The presence of significant non-linear relationships was determined by looking at xy plots. For the purposes of factor analysis, K-number is assumed to be linearly related to input variables. The factor model (FA_57), with fifty-seven variables, was developed for K-number prediction. FA_57 model, with eight factors, represented pulping operations as inferred from dataset generated by the process. Details of factor model are shown in Table 4-11. The accuracy of the resulting factor analysis model for all observations is shown in Figure 4-7 and Figure 4-8. The coefficient of correlation (R-square) values for the plot in Figure 4-8 is approximately 49% (33% for model validation data). It is evident from Figures 4-7 and 4-8 that the factor model does a poor job of predicting K-number in the validation dataset. Poor K-number prediction can be explained in terms of low variability of actual K-number. The digester is under a tight control as evident from low (6.12%) coefficient of variation of K-number. With such low coefficient of variation, most of K-number variability probably came from random factors such a testing variations or from unknown factors such as wood quality variability. As information about testing variations and wood quality were not included in the analysis, factor model couldn't predict K-number variability or the test was not variable enough.

In the next step, the input variable space was reduced to thirty-five variables in order to produce a simplified representation of the pulping process. The variable elimination process has been described earlier in the data preprocessing section. Tables 4-8, 4-9, 4-10 show all thirty-five variables used to develop the second factor model (FA_35) for K-number case study. Details of factor model FA_35 are shown in Table 4-11. The K-number predictions of this model, built from five factors, are shown in Figure 4-7, Figure 4-9. The R-square value for the plot in Figure 4-9 is approximately 38% (27% for model validation data). This correlation value is lower than correlation value in case of *fifty-seven-variable factor model*. The correlation coefficient for K-number predictions is low in both cases. Both factor models were unsuccessful when trying to predict K-number variability around the mean.

Table 4-11. Factor analysis models of K-number for GP mill.

| Model → | FA_57 | FA_35 |
|---|---|---|
| Number of factors | 8 | 5 |
| Intercept | 18.63 | 18.63 |
| Factor 1 coeff. | -0.31 | -0.31 |
| Factor 2 coeff. | -0.31 | -0.44 |
| Factor 3 coeff. | 0.21 | 0.23 |
| Factor 4 coeff. | -0.39 | -0.26 |
| Factor 5 coeff. | -0.07 | 0.32 |
| Factor 6 coeff. | -0.23 | 0.00 |
| Factor 7 coeff. | -0.34 | 0.00 |
| Factor 8 coeff. | 0.25 | 0.00 |
| | | |
| $R^2$ (all data) | 49 | 38 |
| $R_v^2$ (validation data) | 33 | 27 |
| Error standard deviation | 0.84 | 0.99 |

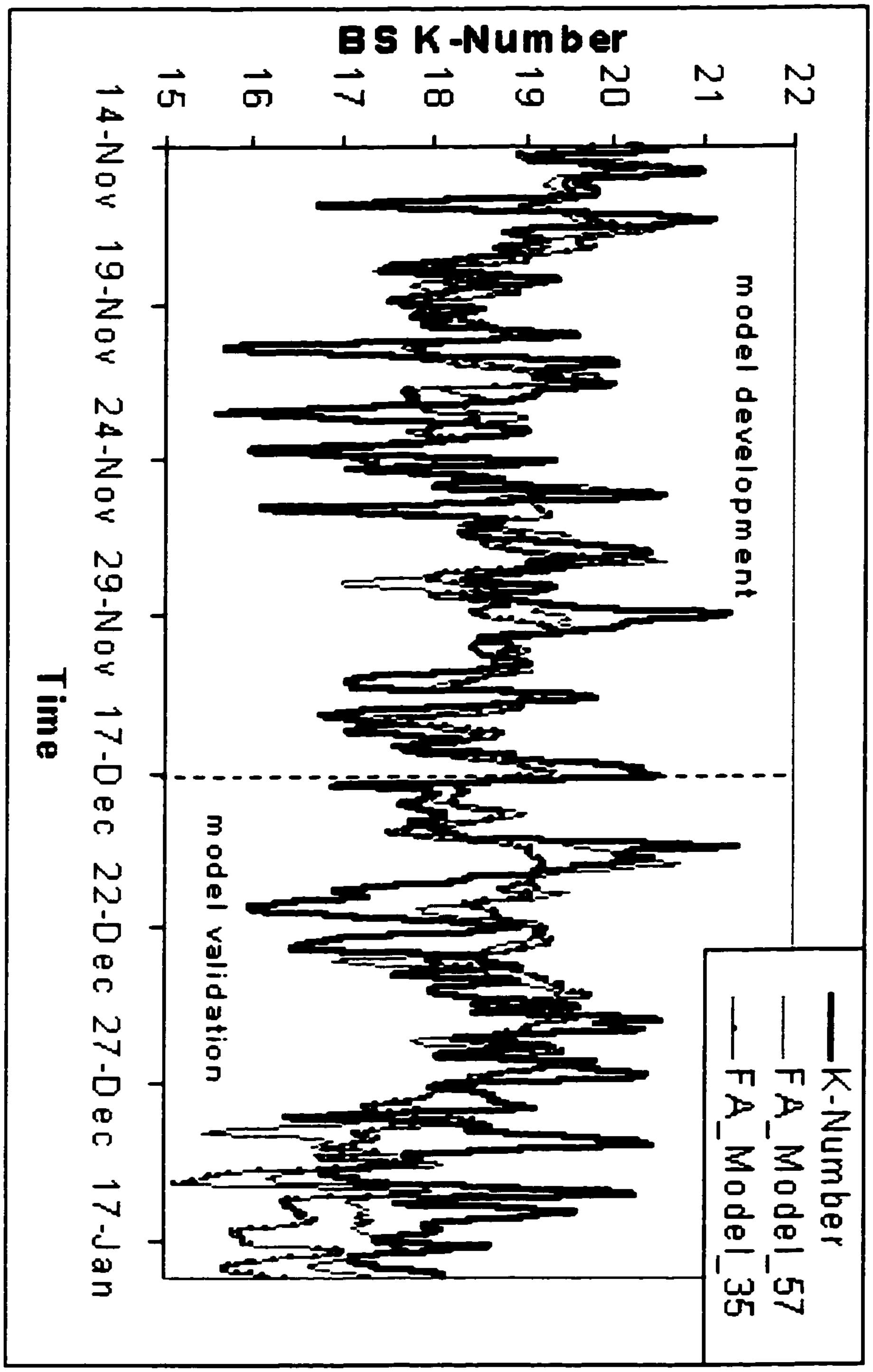



Figure 4-7. Factor analysis predictions of brown stock K-number (BSKNUM).

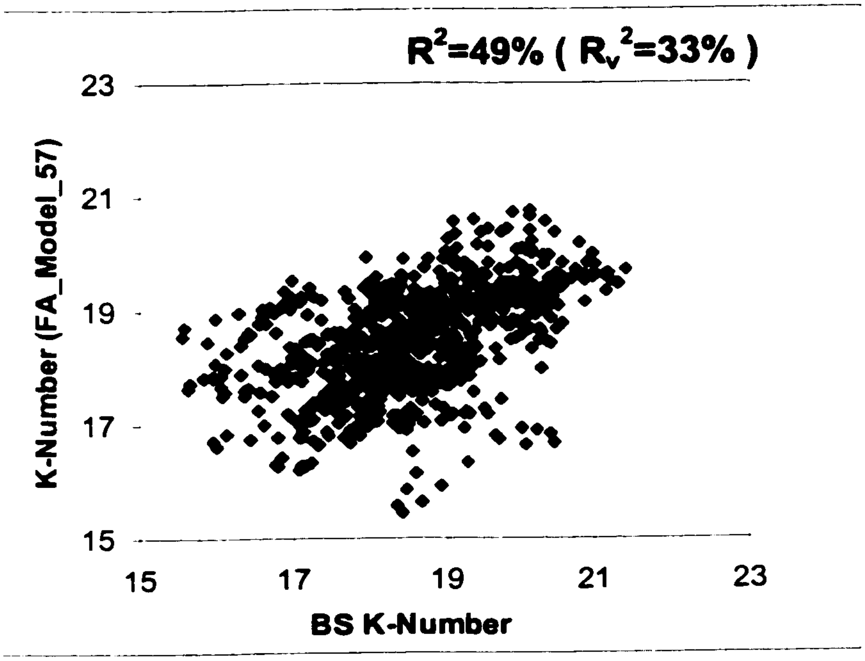



Figure 4-8.Correlation between factor analysis model (FA_57) predictions and actual brown stock (BS) K-number for continuous digester.

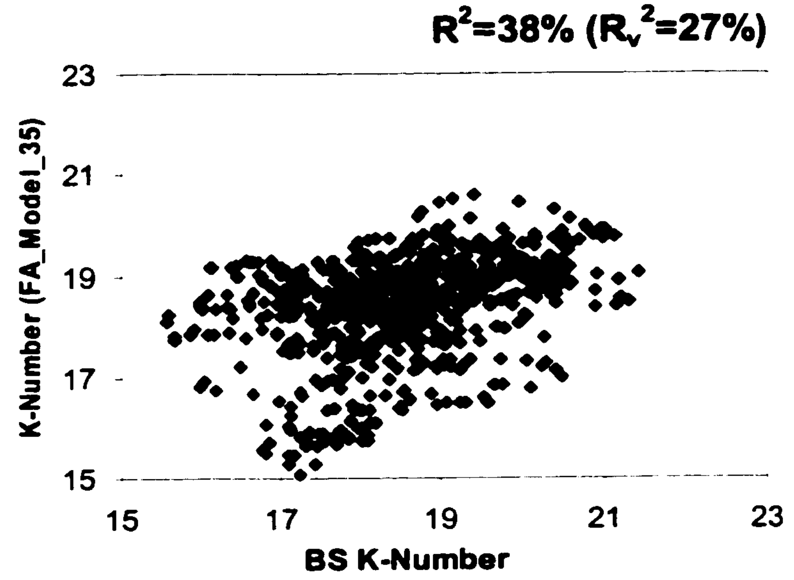



Figure 4-9.Correlation between factor analysis model (FA_35) predictions and actual brown stock (BS) K-number for continuous digester.

**Principal component analysis**

Factor analysis models assume that variations in K-number can be split into two components; (1) variation in K-number attributable to other process variables, and (2) random variations around the mean K-number that are not attributable to process variables. Principal component analysis models assume that variations in K-number are entirely due to other process variables. In the next phase of the study, PCA models were developed for K-number prediction using the pulping dataset.

Statictica, a software package developed by Statsoft Inc., was used for principal component modeling. The time synchronized and preprocessed pulping data was analyzed using principal component analysis. The data partitioning for principal component analysis was the same as that used for factor analysis. The first part with

425 observations was used to analyze the data and build a principal component model, while the second part with 300 observations was used to validate the model

For the purposes of principal component analysis (PCA), K-number was assumed to be a linear related to input variables. The principal component model (PCA_57), with fifty-seven variables with eight components, was developed for K-number prediction. PCA_57 model. Details of PCA model are shown in Table 4-12. The accuracy of the resulting PCA model for all observations is shown in

Figure 4-10 and Figure 4-11. The coefficient of correlation (R-square) value for the plot in Figure 4-11 is approximately 40% (21% for model validation data). It can be seen that much of the variability comes from brief excursions of K-number which the model is unable to simulate. Similar to the factor analysis model, the principal component model fails to predict K-number in the validation dataset. The poor K-number prediction for the PCA model can be explained in terms of the low variability of actual K-number. The digester is under a tight control as evident from low (6.12%) coefficient of variation of K-number. Most of the variability in K-number is either due to testing variations or due to variables such as wood quality. As the information about testing variations and wood quality was not included in the PCA analysis, it was impossible to predict K-number using PCA models. The next step in the PCA was to reduce input variable space to thirty-five variables to facilitate simplified representation of the pulping process. The variable elimination process has been described earlier in the data preprocessing section. Tables 4-8, 4-9, and 4-10 show all thirty-five variables used to develop the second principal component model (PCA_35). Details of principal component model PCA_35 are shown in Table 4-12. The K-number predictions of this model, built from five components, are shown in Figure 4-10 and Figure 4-12. The R-square value for the plot in is approximately 32% (18% for model validation data). The correlation value for PCA_57 model was higher than that for the PCA_35. In fact, correlation value for K-number predictions is low in both cases. None of the PCA models were successful in predicting K-number.

Table 4-12. Principal component analysis models of K-number for GP mill.

| **Model →** | **FA_57** | **FA_35** |
|---|---|---|
| **Number of components** | 8 | 5 |
| **Component 1 coeff.** | -0.313 | 0.290 |
| **Component 2 coeff.** | -0.078 | -0.263 |
| **Component 3 coeff.** | 0.260 | 0.144 |
| **Component 4 coeff.** | -0.157 | 0.072 |
| **Component 5 coeff.** | -0.021 | -0.127 |
| **Component 6 coeff.** | -0.077 | 0.000 |
| **Component 7 coeff.** | -0.172 | 0.000 |
| **Component 8 coeff.** | -0.189 | 0.000 |
| | | |
| **$R^2$ (all data)** | 38 | 27 |
| **$R_v^2$ (validation data)** | 32 | 18 |

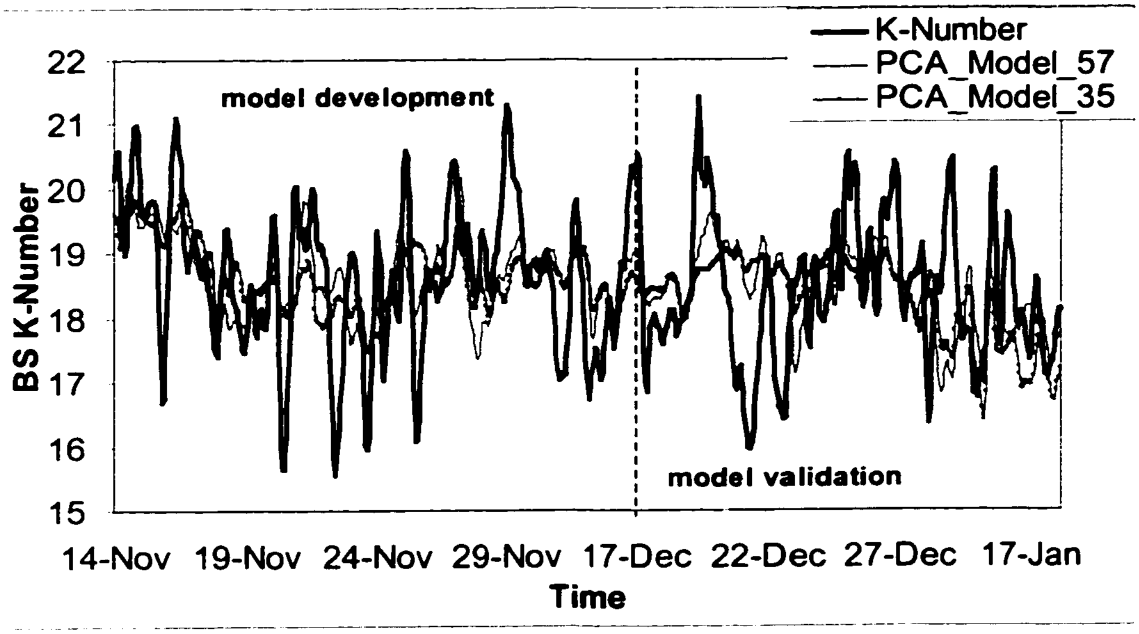



Figure 4-10. Principal component analysis predictions of brown stock K-number (BSKNUM).

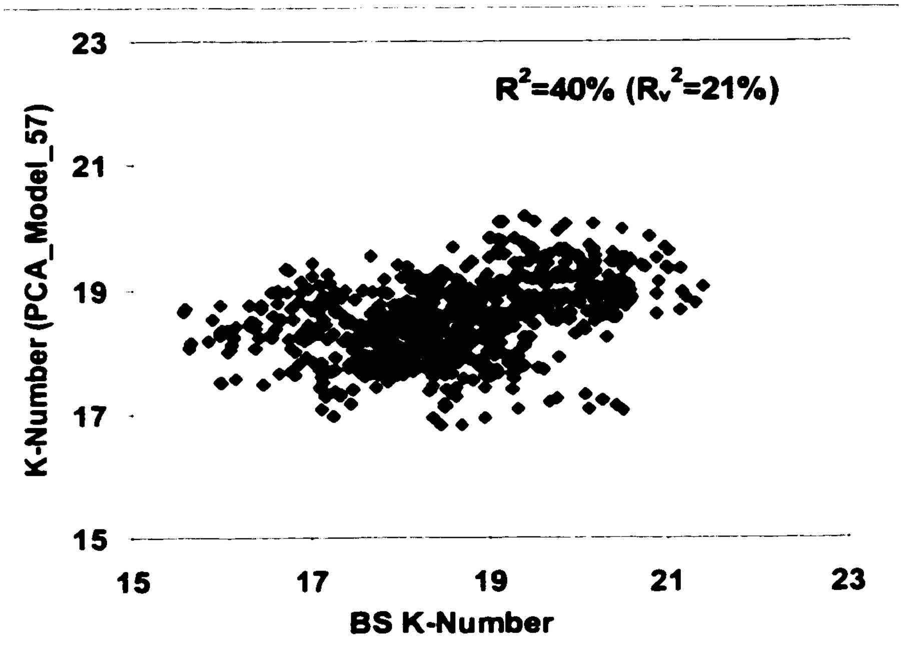



Figure 4-11.Correlation between principal component analysis model (PCA_57) predictions and actual brown stock (BS) K-number for continuous digester.

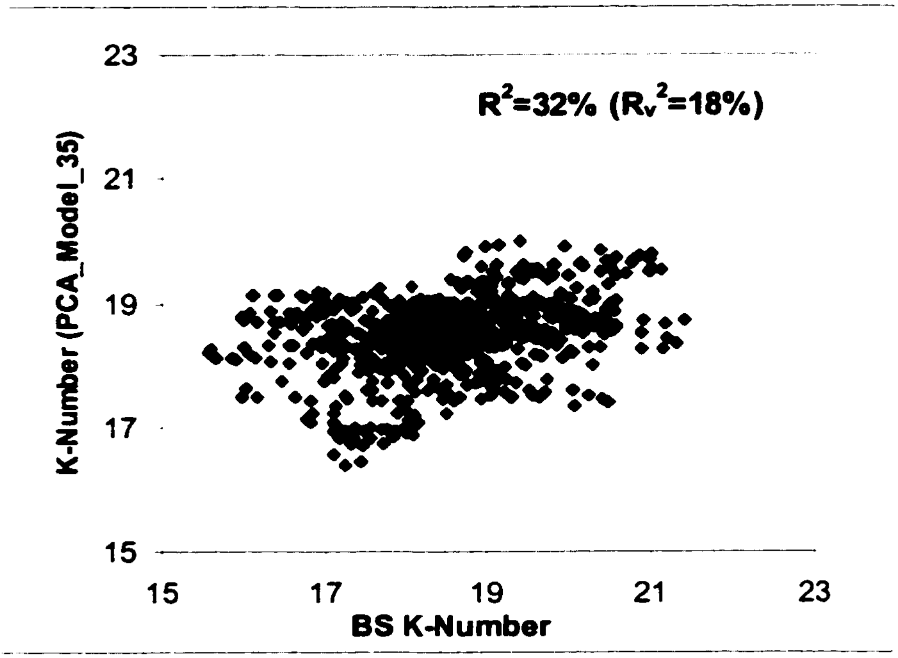



Figure 4-12.Correlation between principal component analysis model (PCA_35) predictions and actual brown stock (BS) K-number for continuous digester.

**Neural network analysis**

The failure of FA and PCA to predict K-number motivated the use of neural networks, a non-linear multivariate technique. Both PCA and FA models assume a dataset containing only linear interactions among variables. In order to investigate the presence of non-linearities in the pulping dataset a non-linear multivariate technique, neural network, was used.

The time synchronized and preprocessed data, representing about one month of pine production at the mill were analyzed using neural network (NN) analysis. Process Insights, a neural network software developed by Pavilion Technologies based at Austin, TX, was used for neural network modeling. The conditioned data was partitioned into two separate parts. The split in the dataset for model development and validation for NN is different from PCA and FA models. More data was required for training neural networks. However, the exact division into model development and validation parts was done by looking at K-number trend and choosing the portion with higher K-number variability. First 547 observations were used to build a NN model, while remaining 176 observations were used to validate the model. The model validation was done to test the accuracy and robustness of the developed NN model.

A NN analysis model was developed using fifty-seven variables for K-number prediction. The sensitivity list of variables in the NN_57 model is presented in Table 4-13. This list shows the twenty most important variables for predicting K-number. The accuracy for the resulting NN model for all observations is shown in Figure 4-13 and Figure 4-14. The R-square value for the plot in Figure 4-15 is approximately 64% (50% for model validation data). Similar to the PCA and FA models, the NN model didn't predict K-number well as evident from low correlation coefficient for model validation dataset. The poor K-number prediction can again be explained in terms of the tightly controlled process K-number (coefficient of variation = 6.12 %). However, the behavior of the NN model in model development stage is quite different from that of the FA and PCA models. NN model better approximates the raw K-number trend in comparison with FA, PCA models. The reason for this is that NN, being a nonlinear modeling technique, does a great job of fitting a model to unknown variations (testing variations and chip quality) of K-number in the model development dataset. The NN_57 model falls in the validation stage as testing and chip quality variations of K-number, in this dataset, are different from that in development dataset (by the very definition of being random).

Similar to FA and PCA modeling, a NN with smaller input variable space was developed using thirty-five variables. Variables were eliminated as described earlier in the data preprocessing section. The list of thirty variables used in the NN_35 K-number case study is presented in Tables 4-4, 4-5, 4-6. The sensitivity list of NN

Table 4-13. Sensitivity list of neural network models of K-number.

| | NN_57 | | | | NN_35 | |
|---|---|---|---|---|---|---|
| Rank # | Input Variable | Sensitivity Value | | Rank # | Input Variable | Sensitivity Value |
| 1 | WHITE LIQUOR SULF TEST | -0.368 | | 1 | FV-37 2ND STAGE WASH FLW | 0.582 |
| 2 | #2 DECKER CONDUCTIVITY | 0.347 | | 2 | FV-3A HP FEEDER PURGE FLOW | -0.512 |
| 3 | DIGESTER DILUTION FACTOR | -0.254 | | 3 | #2 DECKER CONDUCTIVITY | 0.495 |
| 4 | DILUTION LIQ FLW | 0.248 | | 4 | HV-31 BOT CIRC FLOW BYPASS | -0.486 |
| 5 | FV-13 COLD BLOW FLOW | -0.242 | | 5 | DIG OUTLET DEVICE AMPS | -0.422 |
| 6 | BLOW LINE TEMP | 0.228 | | 6 | DIG CHIP-LIQ LVL DIFF CALC | -0.428 |
| 7 | Brown HD | 0.209 | | 7 | DIGESTER DILUTION FACTOR | -0.314 |
| 8 | T-20C WASH CIRC HTR IN TMP | -0.217 | | 8 | L-10 DIG CHIP GAMMA LVL | 0.191 |
| 9 | #2 WBL % SOLIDS 8 HR AVG | 0.203 | | 9 | DP-16B EXTRACT SCRN DP | -0.250 |
| 10 | HV-31 BOT CIRC FLOW BYPASS | 0.196 | | 10 | F-16A EXTRACT FLOW UPPER | 0.302 |
| 11 | DP-16A EXTRACT SCRN DP | -0.196 | | 11 | WASH HTR 160# COND CONDUCT | -0.220 |
| 12 | DIG CHIP-LIQ LVL DIFF CALC | -0.194 | | 12 | #2 PULP WHITE LIQ SULFIDIY | -0.094 |
| 13 | DIG CHIP SG LVL CALC | -0.184 | | 13 | DIFFUSION WASHER CONDUCT. | 0.139 |
| 14 | WASH HTR 160# COND CONDUCT | -0.179 | | 14 | #2 MILL WBL RESIDUAL | -0.006 |
| 15 | #2 PULP WHITE LIQ SULFIDIY | 0.174 | | 15 | DP-16A EXTRACT SCRN DP | -0.141 |
| 16 | FV-3A HP FEEDER PURGE FLOW | -0.169 | | 16 | DIG CHIP SG LVL CALC | -0.230 |
| 17 | DIG OUTLET DEVICE AMPS | -0.163 | | 17 | DILUTION LIQ FLW | 0.242 |
| 18 | TOP SEP INLET TEMP | 0.160 | | 18 | F-6 TOP CIRC FLOW | -0.200 |
| 19 | #2 MILL WBL RESIDUAL | 0.155 | | 19 | FV-13 COLD BLOW FLOW | -0.150 |
| 20 | F-1 160# STM TO DIG FLOW | 0.148 | | 20 | WASH SCREEN BACKFLUSH FLOW | 0.126 |

model NN_35 is shown in Table 4-9. The K-number predictions of this model are shown in Figure 4-13 and Figure 4-15. The R-square value for the plot in the Figure 4-15 is approximately 70% (20% for model validation data). This correlation value is lower than the correlation value for the *NN_57 model.* In fact, the correlation value for K-number predictions for validation dataset is quite low in both cases.

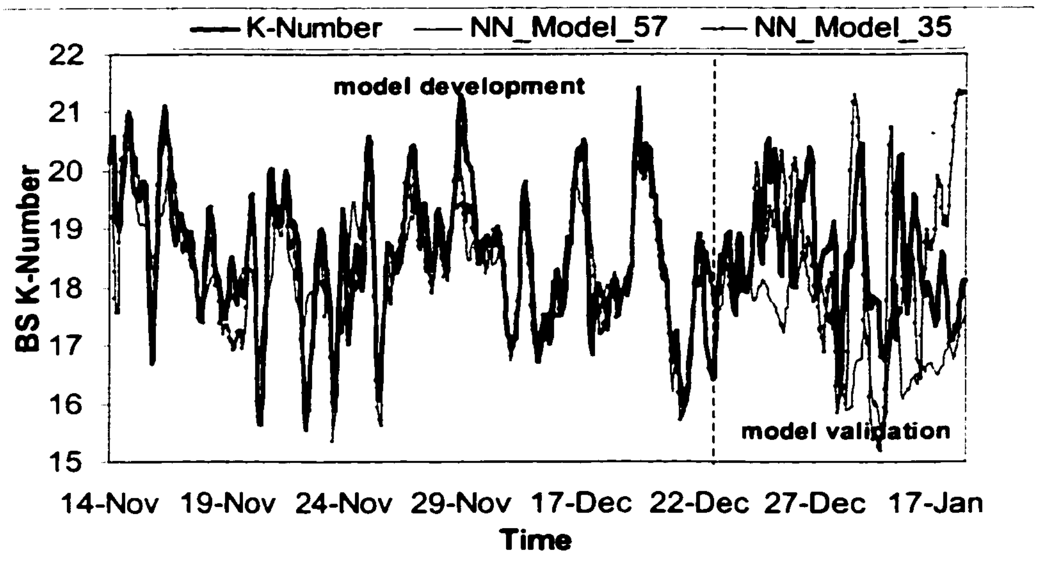



Figure 4-13. Neural network prediction of brown stock K-number (BSKNUM).

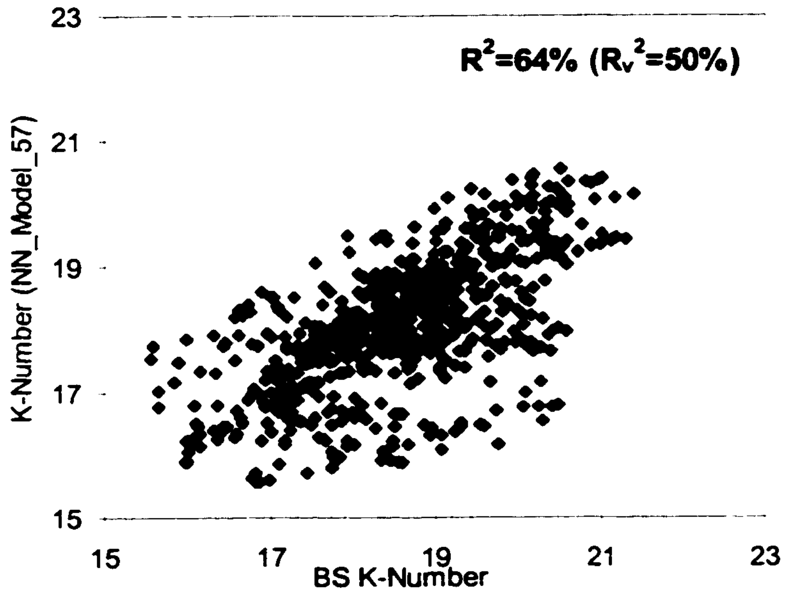



Figure 4-14.Correlation between neural network model (PCA_57) predictions and actual brown stock (BS) K-number for continuous digester.

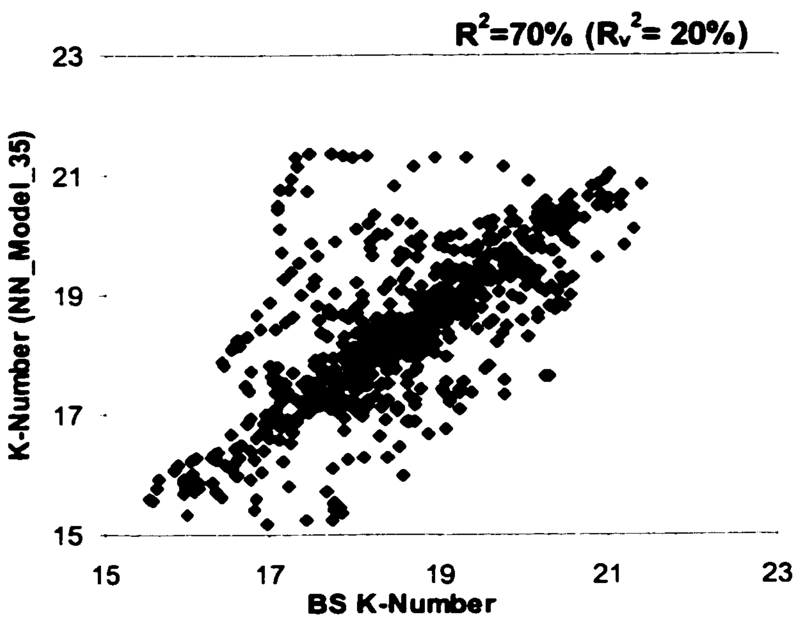



Figure 4-15.Correlation between neural network model (PCA_35) predictions and actual brown stock (BS) K-number for continuous digester.

### *4.2.3 INTERPRETATION OF THE RESULTS*

All three multivariate techniques were unable to model K-number prediction using the dataset acquired from the Ashdown mill as evident from low correlation coefficients for model validation datasets. Only the neural network could fit K-number variability in the model development dataset. Models developed using data acquired from the process failed to capture inherent processes driving the pulping at the GP mill. Some of the reasons for poor performance of multivariate models could be lack of predictive correlation structure in the preprocessed dataset or absence of important process information such as chip quality variables.

The goal of the dissertation was to understand the pulping process from the data generated from process operations. The pulping process happens to be strongly controlled and poorly understood. Additionally, the application of various control strategies masks the behavior of the process (as inferred from the generated dataset) partly or sometimes completely. The relationships, which are expected from general pulp and paper theory, are only partly valid or quite different for the controlled process (Figure 4.16). As a result, the analysis of the data, which comes out of controlled process, doesn't give the understanding of the process. An interesting situation arises when important output variables such as kappa number are strongly controlled by process operations. The dataset generated by a strongly controlled process has an insignificant correlation structure. Insignificant correlation structures are not suitable candidates for developing data based models for the purposes of predicting output variables such as kappa number or total bleaching cost. This happens to be the case in GP pulping dataset.

The poor modeling ability of the pulping dataset can also be explained if one looks carefully at Ashdown mill *modus operandi*. The mill operates by setting a constant production rate (using constant chip meter RPM). Every other input variable is ratioed to the chip meter RPM (pulp mill production rate). In other words, alkali charge is fixed and operators try to keep temperature around digester within a narrow band (coefficient of variation very less than 2% for project dataset). The K-number of pulp coming out of the digester also has a low coefficient of variation. As a result the pulping dataset has a poor correlation structure. The poor correlation structure is further evident from low correlation values among variables in the correlation matrix. The poor correlation structure leads to poor predictive models as evident from the multivariate results.

The K-number case study didn't include any information about testing variations and chip quality variations, as no pertinent variables were available in the mill databases. In other words, the correlation structure of pulping dataset had unknown disturbances such as testing and chip quality variations implicit in it. There was no way of isolating unknown disturbances implicit in the correlation structure. This could also have led to poor prediction models.

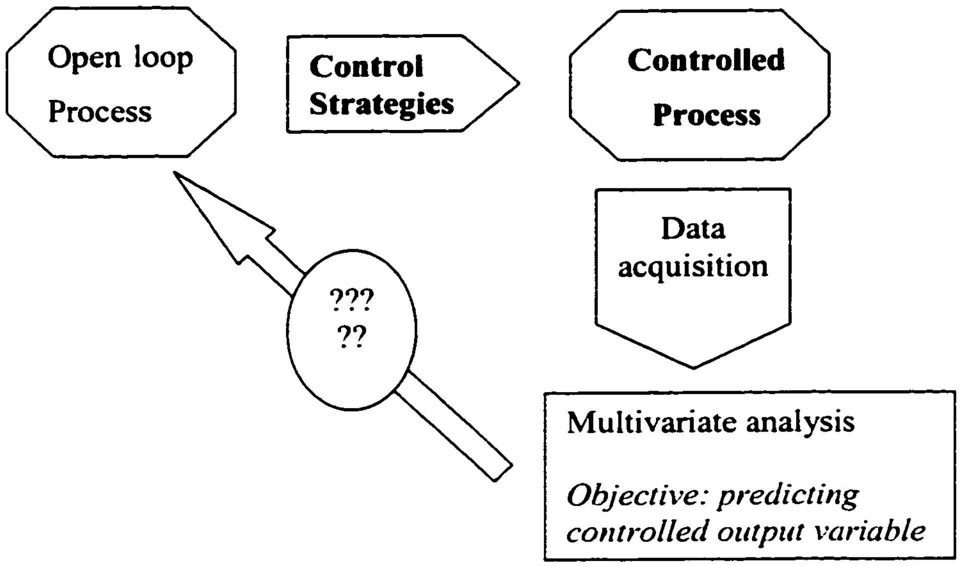



Figure 4-16. Process control strategies change/mask the behavior of process as inferred from acquired dataset.

## 4.4 CONCLUSIONS OF KAPPA STUDY

Two sets of studies were done to investigate kappa number prediction and variability using data from a pulp mill. The first study used data from the Weyerhaeuser Longview mill to predict kappa number out of the digester and $O_2$ delignification reactor. In the Longview study, factor analysis allowed development of models that successfully predict kappa number out of a continuous digester and $O_2$ delignification stage. The most important cause of kappa variability in case of the continuous digester was found to be mischarges in alkali. The major source of kappa variability in the $O_2$ delignification reactor is variability of the pulp kappa out of the digester. Better feedforward digester control might lead to improved $O_2$ reactor performance. Factor analysis results can be used in control of kappa number as they point out factors and in turn the variables, which if controlled, will reduce kappa number variability. Variations in kappa number can be reduced by 45% in case of the digester and 40% in case of the $O_2$ delignification reactor if variables correlating with the important factors are brought under control.

In the second study, the multivariate techniques principal component analysis, factor analysis, and neural networks were used to develop K-number prediction models for the Georgia Pacific Ashdown mill. None of these models were successful in predicting K-number. The main reason for poor prediction was that the digester was already under tight control as evident from low (6.12%) coefficient of variation of K-number. The dataset generated by such a strongly controlled process doesn't have significant correlation structure, which is necessary to develop predictive models. In other words, datasets with insignificant correlation structures are not suitable candidates for the purposes of predicting output variables such as K-number. Another reason for poor prediction models could be the absence of important variables such as chip quality.

# CHAPTER 5: RESULTS OF COST ANALYSIS

Two case studies were performed based on the project approach described in Chapter 3. The first case study focused on the pulping portion of the fiber line. The second case study focused on total bleaching cost. The objective of the bleaching process is to achieve the desired pulp brightness at a desired production rate with minimum expenditure in bleaching chemicals and energy. Bleaching also strives to maintain pulp strength while meeting environmental constraints. The process is subject to several disturbances including incoming K-number variations, washing losses, long delays, channeling, and improper mixing. Manual control therefore commonly uses chemicals dosed in excess of the actual demand. Using improved models of bleaching, the excess may be diminished, leading to a decrease in variability of the quality variables (mostly brightness), and a decrease in total cost of bleaching chemicals and improved environmental performance.

Pulp brightness is the main quality control variable in the bleaching operation. Mill operators control pulp brightness within specifications by changing chemical charges in different bleaching stages. The total cost of bleaching chemicals is what determines the economics of bleach plant operations. The second case study in the research involved developing models, using multivariate techniques, for total bleaching cost in the bleach plant.

The total bleaching cost study was done in two phases. The bleaching stage contributing most to total bleaching cost variability was determined in the first phase of the project. In the second phase of the study, upstream variables in the fiberline (i.e., variables from pulping and washing sections) were used in predictive modeling of bleaching cost. Principal component analysis, factor analysis, and neural network analysis were used in second phase to develop models for bleaching cost prediction.

## 5.1 BLEACH PLANT AT GEORGIA PACIFIC MILL

The pulp mill at Ashdown uses southern pine to produce bleached kraft pulp. With a capacity of about 1000 TPD, the fiber line has a Kvaerner continuous digester system followed by five-stage bleach plant (See Figure 5-1. Bleach plant at Georgia Pacific Ashdown mill.). The washed pulp coming out of pulping and washing areas is stored in brown stock high-density tank. At the time of study, the bleaching sequence was D/CEoD1E2D2 with 50% target substitution in the D/C stage. Various bleach plant operating parameters are shown in Table 5-1.

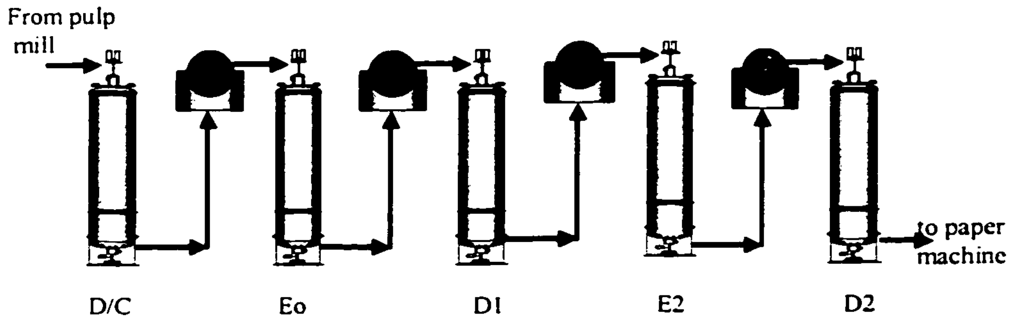



Figure 5-1. Bleach plant at Georgia Pacific Ashdown mill.

Table 5-1. Bleach plant operating parameters for bleaching stages

| D/C stage | | Min. | Target | Max. |
|---|---|---|---|---|
| | Cl2 brightness | 34 | 36 | 38 |
| | Vat Resid. | | trace | |
| | % ClO2 sub. | 10 | 50 | 100 |
| EO stage | | | | |
| | Vat pH | 9.5 | 10.0 | 10.4 |
| | Steam Mixer Temp. | | 160 F | |
| D1 stage | | | | |
| | D1 brightness | 82 | 84 | 86 |
| | Vat pH | 3.5 | 3.7 | 3.9 |
| | D1 tube temp. | 145 | 150 | 155 |
| E2 stage | | | | |
| | Vat pH | 9.5 | 10.0 | 10.4 |
| | Steam mixer Temp. | | 170 F | |
| D2 stage | | | | |
| | Final brightness | 88.5 | 89.5 | 90.5 |
| | Final pH | 3.7 | 4.0 | 4.2 |
| | Vat pH | 3.5 | 3.7 | 3.9 |
| | Vat Residual | | trace – 0.02 | |
| | Steam Mixer Temp. | | 185 F | |

## 5.2 TOTAL BLEACHING COST STUDY: PHASE I

The bleaching stage contributing most to total bleaching cost variability was determined in the first phase of the project. In this phase, factor analysis models were developed to see if a deterministic relationship existed among input variables for predicting total bleaching cost.

### *5.2.1 DATA PREPROCESSING*

The bleaching data, consisting of more than 160 variables and 5000 observations representing three months of production, was acquired from the mill. For the acquired dataset, total bleaching cost had a mean value of 40.29 $/ton with standard deviation 6.25 $/ton. The coefficient of variation (COV) of total bleaching cost was 15.32%. The pulp mill would like to see total bleaching cost COV to be under 10%.

The acquired bleach plant data consisted of multiple production runs with several startups and shutdowns interspersed between them. The focus of the bleaching cost study was to analyze only steady state operations of the fiberline. To represent steady operations, five production runs were selected from the available dataset. Combined together these production runs consisted of 1170 observations. In the next step of data preprocessing, a number of variables were removed from the dataset. Most of the removed variables were lower level control variables such as current flowing to pumps and mixers. A number of variables, which contained redundant process information, e.g., process flows, were also removed. Other variables were removed from dataset based on their lower process importance and low variability (coefficient of variation less than 2%). Approximately, 65% of available process tags (variables) from the bleach plant were removed from the analysis. The trend from remaining fifty-five variables were carefully edited to remove outliers, incorrect observations, and blank rows. The resulting dataset was then time synchronized using method described in Chapter 3. Each time-stamped set of observations was time synchronized separately to account for variable time delay present in the mill dataset. Table 5-2 presents representative time delays in different areas of the fiberline at a chip meter rpm of thirteen. Cleaned and time synchronized datasets were filtered in the next step of preprocessing. A five-hour moving average filter was used to filter time-synchronized datasets. Results of the first phase of the total bleaching cost study are presented in the following section.

Table 5-2. Time delays in different areas of bleach plant at chip meter rpm 13.

| Sr. No. | Section | Time delay, in min |
|---|---|---|
| 1 | #2 Cl2 | 25 |
| 2 | E1 Pre tube | 20 |
| 3 | E1 Tower | 46 |
| 4 | D1 Pre tube | 33 |
| 5 | D1 Tower | 70 |
| 6 | E2 Pre tube | 20 |
| 7 | E2 Tower | 46 |
| 8 | D2 Pre tube | 33 |
| 9 | D2 Tower | 92 |

### *5.2.2 MULTIVARIATE ANALYSIS*

After going through time synchronization and other preprocessing steps, a dataset with 1170 observations was chosen for factor analysis. For model development, the conditioned data were partitioned into two separate parts. One part was used to analyze the data and build a factor model, while the second part was used to validate the model. One part (consisting of 690 observations) was used to analyze the data and build a factor model, while the second part (remaining 480 observations) was used to validate the model. The division of the dataset into model development and model validation datasets was based on the 60/40 criteria, i.e., "60% of *data for development and 40% for validation*". The last 60% of the dataset was chosen for model development as it had larger bleaching cost variation.

A factor model using fifty-five variables was created for the bleaching section. The resulting accuracy for the factor model for all observations is shown in Figure 5-2 and Figure 5-3. The coefficient of correlation ($R^2$) value for plot in Figure 5-2 is approximately 72% (69% for validation dataset). The correlation coefficient for the validation dataset is denoted by $R_v^2$. The model predicted the total bleaching cost well for the data used in the model development stage as well as for the data used for model validation. The model predicted the total bleaching cost for the data not used in the model development (more than 470 observations) as shown in Figure 5-3.

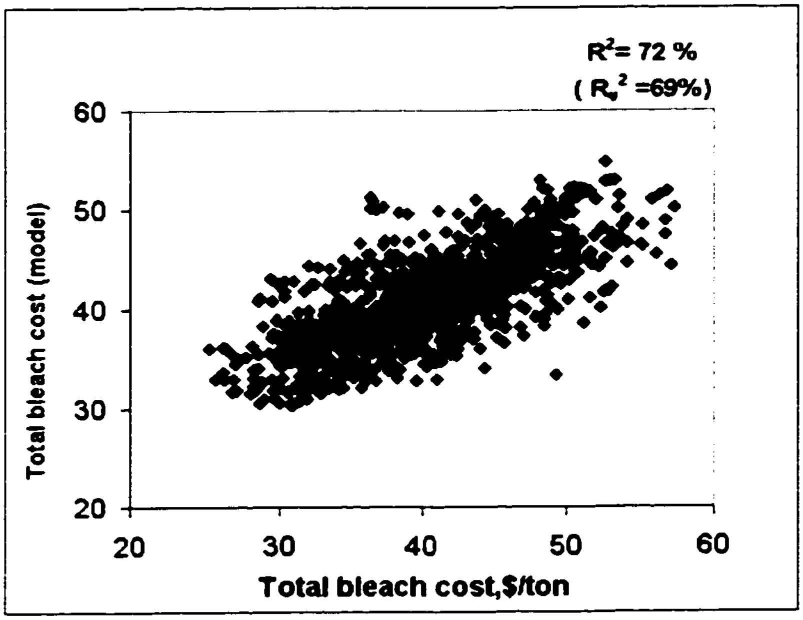



Figure 5-2. Correlation between factor analysis model predictions and actual total bleaching cost at Ashdown mill.

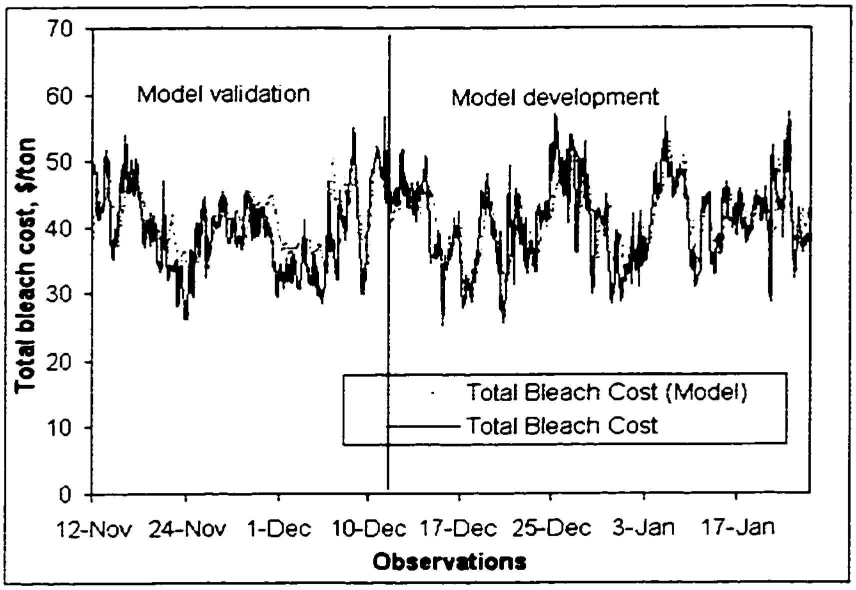



Figure 5-3. Factor analysis prediction of total bleaching cost at Ashdown mill

### *5.2.3 INTERPRETATION OF RESULTS*

After modeling the total bleaching cost, the underlying patterns in the dataset were identified by variability analysis. The results indicated that there were three patterns underlying the process variables used in the analysis that affect the total bleaching cost. The patterns are shown in Table 5-3 in the order of importance in model. The factor analysis indicates that 65% of the variations in total bleaching cost can be attributed to variations in the Dioxide factor and the First Stage Chlorine factor. As evident from the factors, a large portion of the total bleaching cost variability can be attributed to the first stage of bleaching.

Table 5-3. Common factors and primary patterns present in data from bleaching section of mill.

| ***Factor*** | ***Primary pattern*** |
|---|---|
| Dioxide factor | Strong positive correlation with chlorine dioxide flow to first stage mixers and decker kappa factor. As dioxide factor increases total bleaching cost increases. |
| First stage chlorine factor | Strong positive correlation with chlorine flow to first stage. |
| pH control factor | This factor is strongly correlated to D/ C stage stock pH and vat pH (some relation with first stage $Cl_2$ and $ClO_2$ charge). |

## 5.3 TOTAL BLEACHING COST STUDY: PHASE II

In the second phase of the bleaching study, upstream variables in the fiberline were used in the predictive modeling of bleaching cost. Bleach plant upstream variables refer to variables from pulping and washing sections. Results from the first phase of bleaching cost study indicated that most of the variability in total bleaching cost came from the first stage of bleaching, i.e., D/C stage. The second phase of bleaching study focused only on developing multivariate models of D/C bleaching cost using upstream variables.

### *5.3.1 DATA PREPROCESSING*

The bleaching data, consisting of 160 variables and 5000 observations representing three months of production, was acquired from the mill databases. For the acquired dataset, total bleaching cost had a mean value of 40.29 $/ton with standard deviation 6.25 $/ton. The coefficient of variation of total bleaching cost was 15.32%.

The acquired pulp mill data consisted of multiple production runs with several startups and shutdowns interspersed between them. To represent steady operations, three production runs were selected from the available dataset. Combined together, the three bleaching data sets represented approximately one month of mill production. In the next step of data preprocessing, a number of variables were removed from the dataset as described in Section 5.2.1. In all ninety-seven variables were removed from chosen bleaching datasets. Trends from remaining sixty-three variables were carefully edited to remove outliers, incorrect observations, and blank rows. The resulting dataset was then time synchronized using method described in Chapter 3. Each time-stamped set of observations was time synchronized separately to account for variable time delay present in the mill dataset. Table 5-4 presents representative time delays in different areas in the mill fiberline at a chip meter rpm of thirteen (equivalent to production rate of 975 tons per day). Cleaned and time synchronized datasets were filtered in the next step of preprocessing. A five-hour moving average filter was used to filter time-synchronized datasets. At this stage there were three preprocessed datasets representing three production runs in the pulp mill (Table 5-5). These three datasets had different operational strategies. A composite dataset containing all three datasets was created. It represented all operational strategies following during model development period.

The composite dataset with sixty-three upstream variables and 825 observations (approximately one month of mill production) was used to find deterministic relationship between upstream variables and D/C stage bleaching cost (*D/C stage bleaching cost will be referred to as bleaching cost hereafter*). Results from the bleaching cost study are presented in the following discussion.

Table 5-4. Time delays in Georgia Pacific fiberline at chip meter rpm 13.

| S. No. | Section | Time delay, in min |
|---|---|---|
| 1 | To chip silo | 5 |
| 2 | Hula bin | 35 |
| 3 | Chip feeders | 2 |
| 4 | IV | 25 |
| 5 | Co-current cook zone | 107 |
| 6 | Wash zone | 100 |
| 7 | Pressure Diffuser | 15 |
| 8 | Atm. Diffuser | 76 |
| 9 | Knotters | 5 |
| 10 | Screens | 5 |
| 11 | Decker | 2 |
| 12 | UBLch HDTower | 60 |
| 13 | #2 Cl2 | 25 |
| 14 | E1 Pre tube | 20 |
| 15 | E1 Tower | 46 |
| 16 | D1 Pre tube | 33 |
| 17 | D1 Tower | 70 |
| 18 | E2 Pre tube | 20 |
| 19 | E2 Tower | 46 |
| 20 | D2 Pre tube | 33 |
| 21 | D2 Tower | 92 |

Table 5-5. Datasets for bleaching cost analysis.

| Dataset | Number of variables | Number of observations | Bleach cost Average | Bleach cost Stdev | Bleach cost COV | Important operational strategy change |
|---|---|---|---|---|---|---|
| Bleach 1 | 63 | 344 | 24.49 | 4.60 | 19.57 | BCH installed |
| Bleach 2 | 63 | 285 | 25.61 | 6.49 | 26.35 | Chip meter RPM 13 |
| Bleach 3 | 63 | 186 | 23.14 | 2.63 | 11.35 | Chip meter RPM 13.5 |
| Composite bleach set | 63 | 825 | 25.05 | 6.35 | 25.30 | All of the above |

### *5.3.2 MULTIVARIATE MODELS*

The preprocessed composite bleaching dataset was used to develop multivariate models using three techniques; factor analysis, principal component analysis, and neural network analysis. Sixty-three variables, upstream from bleaching, were used in all three multivariate models for the purposes of bleaching cost prediction and variability analysis. In addition to this, a number of smaller subsets of sixty-three upstream variables were used to develop other multivariate models. The smaller sets of variables were used to see if bleaching cost prediction improved with simpler variable structure in the models. Both process knowledge as well as process variability were important factors in the variable elimination/selection step of model building. Following steps were taken to obtain smaller sets of upstream variables.

1. Removing strongly controlled variables (COV < 4%). Thirty-eight variables remained in the dataset.

2. Eliminating variables with least average correlation with other variables. Twenty-one variables remained in the dataset.

3. Retaining only variables with high COV and process importance in the model. Fourteen variables remained in the dataset.

A list of developed models with their codes and descriptions is presented in Table 5-6. Tables 5-7, 5-8, 5-9 present detailed list of variables used in the different multivariate models. All the variables excluded from the analysis were mathematically insignificant from model accuracy point view.

Table 5-6. Model code and description for bleaching cost multivariate models.

| S.No. | Model code | Model description |
|---|---|---|
| 1 | FA_63 | Factor analysis model with sixty-three variables. |
| 2 | FA_38 | Factor analysis model with thirty-eight variables. |
| 3 | FA_21 | Factor analysis model with twenty-one variables. |
| 4 | FA_14 | Factor analysis model with fourteen variables. |
| 5 | PCA_63 | Principal component analysis model with sixty-three variables. |
| 6 | PCA_38 | Principal component analysis model with thirty-eight variables. |
| 7 | PCA_21 | Principal component analysis model with twenty-one variables. |
| 8 | PCA_14 | Principal component analysis model with fourteen variables. |
| 9 | NN_63 | Neural network model with sixty-three variables. |
| 10 | NN_38 | Neural network model with thirty-eight variables. |
| 11 | NN_21 | Neural network model with twenty-one variables. |
| 12 | NN_14 | Neural network model with fourteen variables. |

Table 5-7. Pre-digester variables used in bleaching cost study.

| Sr. No. | Tags | Description | FA_63 | FA_38 | FA_21 | FA_14 |
|---|---|---|---|---|---|---|
| 1 | 20-FC003 . | FV-3 WHITE LIQUOR FLOW | y | | | |
| 2 | 20-FC008 . | FV-3A HP FEEDER PURGE FLOW | y | y | y | |
| 3 | 20-FC027 . | FV-61 IV SLUICE FLOW | y | | | |
| 4 | 20-FI011 . | F-5 CHIP CHUTE RELIEF FLOW | y | | | |
| 5 | 20-FI012 . | F-6 TOP CIRC FLOW | y | | | |
| 6 | 20-FI013 . | F-7 MAKE-UP LIQUOR FLOW | y | | | |
| 7 | 20-FI022 . | F-24 CHIP CHUTE CIRC FLOW | y | y | y | |
| 8 | 20-LC484 . | CHIP METER CHUTE LEVEL | y | | | |
| 9 | 20-PC352 . | 60# STM TO HB | y | | | |
| 10 | 20-SC005 | CHIP METER SPEED | y | | | |
| 11 | 20-TI157 . | 2PM WHITE LIQUOR TEMP | y | y | | |
| 12 | K2-SULF.PE | #2 pulp white liq sulfidiy | y | y | y | y |

Table 5-8. Digester variables used in bleaching cost study.

| Sr. No. | Tags | Description | FA_63 | FA_38 | FA_21 | FA_14 |
|---|---|---|---|---|---|---|
| 1 | 20-CHPLIQ. | DIG CHIP-LIQ LVL DIFF CALC | y | y | y | y |
| 2 | 20-DPI071. | DP-11 DIG OUTLET DEV DP | y | y | | |
| 3 | 20-DPI072. | DP-16A EXTRACT SCRN DP | y | y | y | y |
| 4 | 20-DPI073. | DP-16B EXTRACT SCRN DP | y | y | y | y |
| 5 | 20-FC002A. | F-16A EXTRACT FLOW UPPER | y | y | | |
| 6 | 20-FC002B. | F-16B EXTRACT FLOW LOWER | y | | | |
| 7 | 20-FC009 . | FV-3B RES WHITE LIQ FLW | y | | | |
| 8 | 20-FC016 . | FV-12B DIG 'B' BLOW FLW | y | y | | |
| 9 | 20-FC017 . | FV-13 COLD BLOW FLOW | y | y | | |
| 10 | 20-FC152 . | HV-31 BOT CIRC FLOW BYPASS | y | y | | |
| 11 | 20-FI001 . | F-1 160# STM TO DIG FLOW | y | | | |
| 12 | 20-FI021 . | F-20 WASH CIRC FLOW | y | | | |
| 13 | 20-FI135 . | F-24 FILT/#2 FLSH TNK FLW | y | y | | |
| 14 | 20-HC058 . | HV-11 KAMYR PAD AIR PRESS | y | y | | |
| 15 | 20-HC070 .CO | HV-87 BLOW LINE DILUTION | y | y | | |
| 16 | 20-II369 . | DIG OUTLET DEVICE AMPS | y | y | y | y |
| 17 | 20-LC045 . | L-15 DIG LIQUOR LVL CNTRL | y | | | |
| 18 | 20-LC046 . | LV-16 1A FLASH TANK LVL | y | y | y | |
| 19 | 20-LC114 . | LV-16B #1B FLASHTANK LVL | y | y | y | |
| 20 | 20-LI042 . | DIG CHIP SG LVL CALC | y | y | y | |
| 21 | 20-PC049 . | P-1 160# STM TO DIG PRES | y | y | | |
| 22 | 20-PC062 . | PV-17 2 FLSH TNK REL PRESS | y | | | |
| 23 | 20-TC480 . | BOT CIRC HTR DISC TEMP | y | y | | |
| 24 | 20-TI075 . | 160# STM TO DIG TEMP | y | y | y | |
| 25 | 20-TI085B. | T-12B BLOW LINE TEMP | y | | | |
| 26 | 20-TI089 . | T-20C WASH CIRC HTR IN TMP | y | | | |
| 27 | 20-TI092 . | T-60 BOTTOM CIRC TEMP | y | y | | |
| 28 | 20-TI128 . | TOP SEP INLET TEMP | y | y | | |
| 29 | 20-TI156 . | BLOW LINE TEMP | y | y | | |
| 30 | K2-DIKNUM.PE | KAMYR DIG. LOWER K NUMBER | y | y | y | y |
| 31 | K2-SOLAVG.PE | #2 WBL % SOLIDS 8 HR AVG | y | y | | |
| 32 | K2-UPKNUM.PE | KAMYR DIG. UPPER K NUMBER | y | y | y | y |
| 33 | K2-WBLRES.PE | #2 MILL WBL RESIDUAL | y | | | |

Table 5-9. Post-digester variables used in bleaching case study.

| Sr. No. | Tags | Description | FA_63 | FA_38 | FA_21 | FA_14 |
|---|---|---|---|---|---|---|
| 1 | 20-AI242 . | WASH HTR 160# COND CONDUCT | y | y | y | |
| 2 | 20-FC019 . | FV-18 COUNTER WASH FLOW | y | y | y | y |
| 3 | 20-FC138 . | FV-27 1ST STAGE WASH FLW | y | y | y | y |
| 4 | 20-FC144 . | FV-37 2ND STAGE WASH FLW | y | y | y | y |
| 5 | 20-FC423 . | WASH LIQ TO PD FLW | | y | | |
| 6 | 20-WASHRT.PE | 2PM WASH CIRC RATIO | y | | | |
| 7 | 33-CD$ .PE | 2L CD STAGE TOTAL COST | y | y | y | y |
| 8 | 33-FC006 .PE | #2 BLCH PLNT ACT RATE T/DY | y | | | |
| 9 | Atm. diffuser | Atmospheric diffuser retention time | y | | | |
| 10 | Brown HD | Brown HD retention time | y | y | y | y |
| 11 | K2-BSCONS.MI | #2 BROWNSTOCK CONSISTENCY | y | y | | |
| 12 | K2-BSKNUM.PE | #2 MILL BROWN STK K NUMBER | y | | y | y |
| 13 | K2-DKCOND.MI | #2 DECKER CONDUCTIVITY | y | y | y | y |
| 14 | K2-DWCOND.MI | DIFFUSION WASHER CONDUCT. | y | | | |
| 15 | K2-DWCONS.MI | DIFFUSION WASHER CONSIST. | Y | | | |
| 16 | K2-SPKCNT.MI | #2 PM BS SPECK COUNT | Y | | | |
| 17 | K2-WHDTON.PE | 2PM WASH HD LEVEL IN TONS | Y | | | |
| 18 | 20-CC415 . | STK TO PRESS DIFF CONSIST | Y | y | | |

### *5.3.3 FACTOR ANALYSIS*

**Modeling results**

The time synchronized and preprocessed data, representing about one month of southern pine production at the mill, were analyzed using factor analysis. The bleaching data was partitioned into two separate parts. One part, consisting of 425 observations, was used to analyze the data and build a factor model, while the other part with 400 observations was used to validate the model. The initial division of the dataset into model development and model validation datasets was based on 60/40 criteria, i.e., "60% of *data for development and 40% for validation*". The 60/40 criterion was, however, modified in case of bleaching study as a deterministic model could be built using smaller number of observations (only 425 observations). A smaller dataset for model building also facilitated using a larger dataset for testing the accuracy and robustness of the developed model. The last 425 observations of the preprocessed dataset were used for model building whereas first 400 observations were used for model validation.

For the purposes of factor analysis, K-number is assumed to be a linearly related to input variables. A factor analysis model (FA_63) was developed using sixty-three variables (listed in Tables 5-6,7,8) for bleach cost prediction. The factor model, with eight factors, represented pulping and washing operations upstream from the D/C bleaching stage. The accuracy of the resulting factor analysis model for all observations is shown in Figure 5-4 and Figure 5-5. The coefficient of correlation (R-square) value for plot in Figure 5-5 is approximately 88% (74% for model validation data). It quite evident that FA_63 model successfully predicts bleaching cost. The model predicted bleaching cost for the data not used in the model development (more than 400 observations) as shown in Figure 5-5.

The sixty-three variables factor model (FA_63) was quite successful at predicting the bleaching cost, but interpretation of eight factors in terms of sixty-three variables is difficult. Thus, a number of other factor models with smaller number variables were developed (details in Table 5-9). The first row in Table 5-9 shows the number of factors in different factor models. Factor coefficients represent the quantitative importance of different factors in a model. Variables in the smaller factor models were carefully chosen from sixty-three variables used in FA_63 model to retain

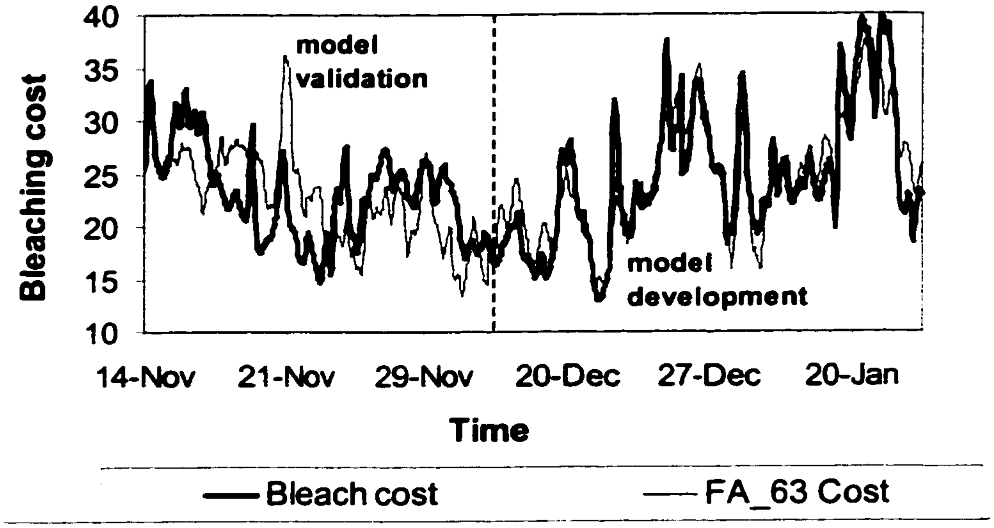



Figure 5-4. Factor analysis prediction of bleaching cost for factor model with sixty-three upstream variables.

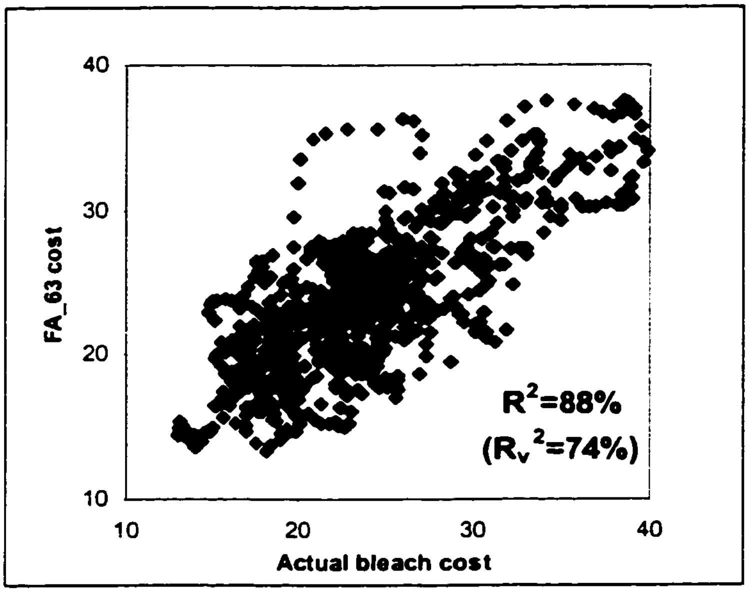



Figure 5-5.Correlation between factor analysis model predictions and actual bleaching cost for factor model with sixty-three upstream variables.

Table 5-10. Factor analysis models of bleaching cost for GP mill.

| **Model →** | **FA_63** | **FA_38** | **FA_21** | **FA_14** |
|---|---|---|---|---|
| **Number of factors** | 8 | 5 | 3 | 2 |
| **Intercept** | 25.05 | 25.08 | 25.08 | 25.05 |
| **Factor 1 coeff.** | 1.55 | 3.15 | 3.60 | 5.02 |
| **Factor 2 coeff.** | 3.55 | 2.07 | 1.36 | -0.75 |
| **Factor 3 coeff.** | 3.32 | 3.17 | 2.96 | |
| **Factor 4 coeff.** | 0.19 | -1.15 | | |
| **Factor 5 coeff.** | 1.00 | 1.76 | | |
| **Factor 6 coeff.** | 0.58 | | | |
| **Factor 7 coeff.** | -0.18 | | | |
| **Factor 8 coeff.** | -2.11 | | | |

important predictive information. The variable selection process for these models is described in the previous section (section 5.3.2). In all, three smaller factor models for predicting bleaching cost were developed. These models used 38, 21, 14 upstream variables respectively (listed in Tables 5-6,7,8) to predict bleaching cost. For all models, data partitioning into the model development and model validation sets was similar to that of FA_63 model.

Using factor analyses, three prediction models (with five, three, and two factors) were developed to represent pulping and washing operations. The accuracy of the resulting factor analysis models for all observations is shown in Figures 5-6, 5-7, 5-8. The coefficients of correlation (R-square) for the three smaller models and FA_63 model are presented in Table 5-11. It is quite clear from the Table 5-11 that all three smaller factor models as well as FA_63 model were able to predict bleach cost successfully. However, for simpler representation of mill operations, the model with smallest number of variables (FA_14) was chosen for next stage of the analysis.

Table 5-11. Correlation coefficients of various bleach cost factor models.

| Factor Model | Number of variables | Number of factors | $R^2$, all data | $R_v^2$, validation data | Squared residual average |
|---|---|---|---|---|---|
| FA_63 | 63 | 8 | 0.88 | 0.74 | 21.58 |
| FA_38 | 38 | 5 | 0.83 | 0.74 | 23.04 |
| FA_21 | 21 | 3 | 0.74 | 0.70 | 13.81 |
| FA_14 | 14 | 2 | 0.75 | 0.74 | 10.41 |

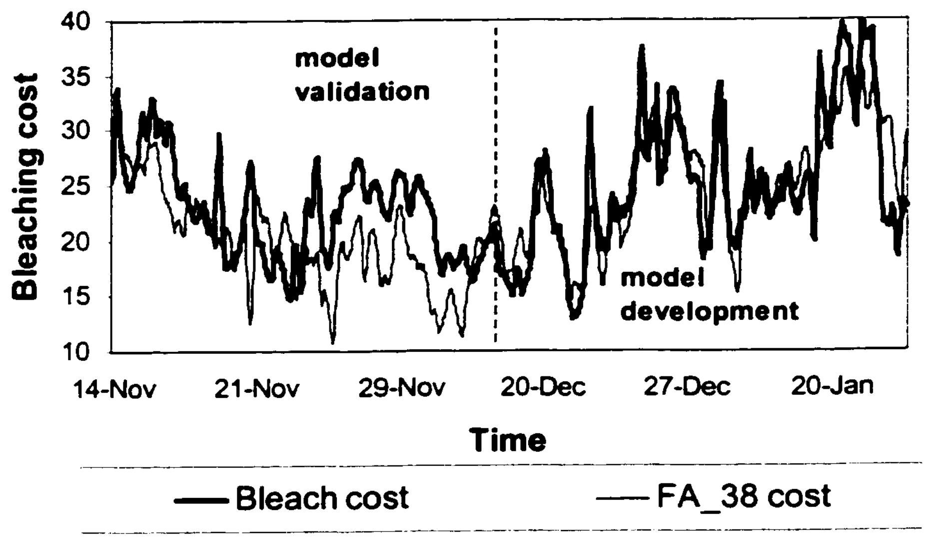



Figure 5-6. Factor analysis prediction of bleaching cost for FA_38 model.

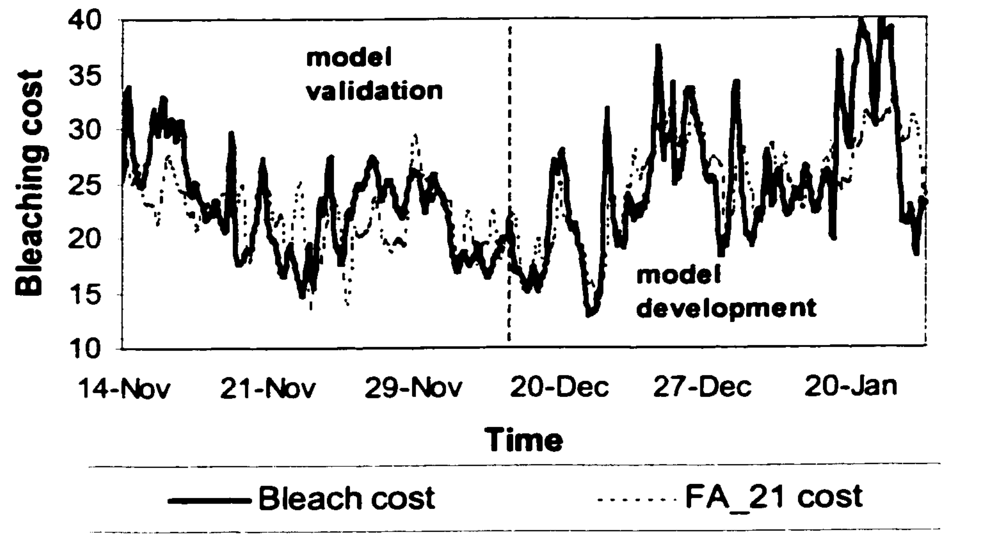



Figure 5-7. Factor analysis prediction of bleaching cost for FA_21 model.

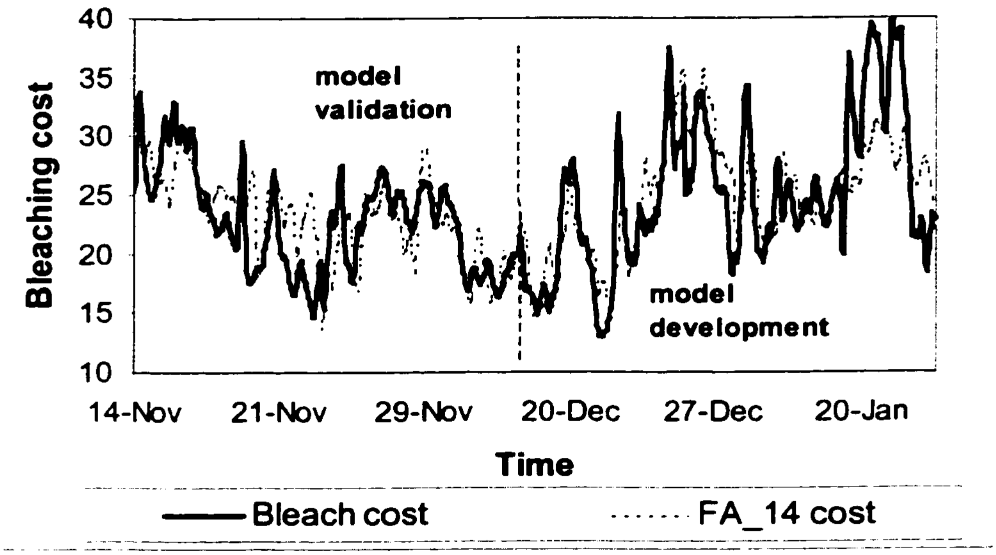



Figure 5-8. Factor analysis prediction of bleaching cost for FA_14 model.

**Interpretation of factor model**

After modeling the bleaching cost, the underlying patterns in the dataset were identified from the variability analysis. The variability analysis indicated that there were approximately two patterns underlying the upstream process variables used in the bleaching cost analysis. The patterns are shown in Table 5-12 in order of importance in the model.

The factor analysis indicated that 65% of the variations in bleaching cost were contributed by variations in lignin factor and digester stability factors. These factors were not true mechanistic parameters as used in first principle modeling. Instead the factors were statistical representations of mechanistic parameters calculated by the factor analysis.

The lignin factor can be thought of as a latent variable that combines the effects of pulping as well as washing. As shown in Table 5-12, digester K-number increases lead to increase in bleach cost. This follows from the fact that higher K-number pulp

Table 5-12. Common factors and primary patterns present in digester data from mill.

| *Factor* | *Primary pattern* |
|---|---|
| Lignin Factor | As digester K-number increases, bleaching cost increases.<br>As atmospheric diffusion flow rate increases, bleaching cost decreases. |
| Digester stability factor | Related to outlet device amperage and counter wash flow rate at the bottom of the digester. |

has higher lignin content thus requiring more bleaching chemical in the D/C stage. Increased bleach chemical usage results in increased bleach cost. The lignin factor is also strongly correlated to the wash water flows to first and second stages of atmospheric diffusion washing. As the flow rates increase, pulp is washed better which decreases bleach cost. Digester stability factor is the other pattern present in the prediction dataset. This factor is strongly correlated to outlet device (OD) amperage, which is the amount of current drawn by the outlet device. OD amperage is a sign of column stability inside the digester. It is related to brown stock consistency and is also a function of K-number. Stable OD amperage is a prerequisite for good digester operation. Variation in OD signifies unsteady state operation and result in varying bleach cost.

Factor analysis shows that by controlling variables correlated with lignin factor and digester stability factor, the variations in bleaching cost can be reduced by 65%. The remaining 35% variations in bleach cost was partly due to random variations and partly due to variables not included in this analysis such as chip quality variables.

### *5.3.4 PRINCIPAL COMPONENT ANALYSIS*

**Modeling results**

The time synchronized and preprocessed pulping data, representing about one month of pine production at the mill, were analyzed using principal component analysis (PCA). Data partitioning in PCA modeling was identical to that in factor analysis modeling. A PCA analysis model was developed using sixty-three variables (listed in Tables 5-6,7,8) for bleach cost prediction. A multivariate model with eight principal components was chosen to represent the set of upstream variables predicting bleach cost. The accuracy for the resulting PCA model for all observations is shown in Figure 5-9. It is evident that the PCA model does not predict bleach cost for the first seventy observations. The PCA_63 model bleach cost trend does track actual bleach cost trend from observation 70 onwards. The coefficient of correlation for bleach cost model predictions is 35% for the validation dataset (excluding the first seventy observations). This correlation value is very low if these 70 observations are included.

The question arises what is happening in first seventy observations that leads to poor prediction for bleach cost in PCA_63 model. As mentioned earlier, a composite dataset from three different productions runs was used for multivariate analysis (Table 5-4). The first production run in the dataset contained an important shift in operational strategy. A bottom circulation heater (BCH) was installed which lead to change in distribution of heat energy going into pulping system. As a result of BCH installation, steam consumption at the top of the digester decreased as black liquor in the bottom circulation was heated indirectly in the BCH. This resulted in a big change in bottom circulation temperature (after first seventy observations) as shown in Figure 5-10.

There is an assumption about common variability of variables in PCA. The PCA model assumes that all variability of a variable can be attributed to other variables. PCA_63 model was developed using observations 400-825. As per PCA assumption, all variability of bottom circulation temperature (BC Temp) was functionally related to bleaching cost. As a result when BC Temp changed significantly in the model validation dataset, bleaching cost prediction changed significantly as well. This is not the case with the FA model as FA assumes that only part of variability of a variable is attributed to other variables.

Similar to factor analysis, a number of other PCA models were developed with smaller number of variables for better interpretation of PCA models. Variables in the smaller PCA models were carefully chosen from sixty-three variables used in the first

PCA model to retain important predictive information (details in Section 5.3.2). In all, three smaller PCA models for predicting bleaching cost were developed. These models used 38, 21, 14 upstream variables (listed in Tables 5-7, 8, 9) to predict bleaching cost. The variable selection process for these models is described in Section 5.3.2. For all models, data partitioning was same as in case of PCA_63 model.

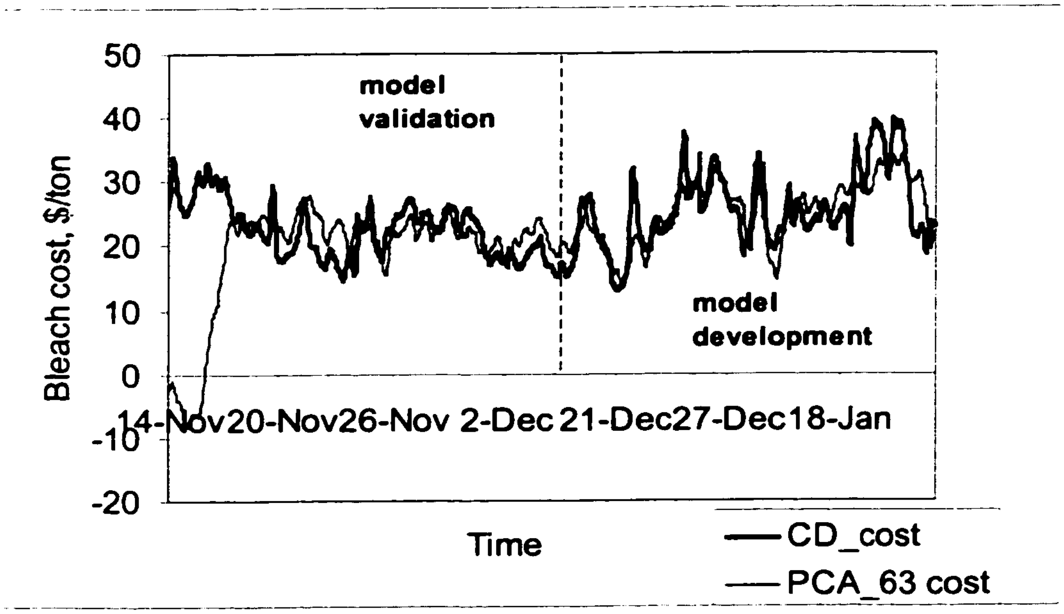



Figure 5-9. Principal component analysis predictions of bleaching cost for PCA_63 model with sixty-three upstream variables.

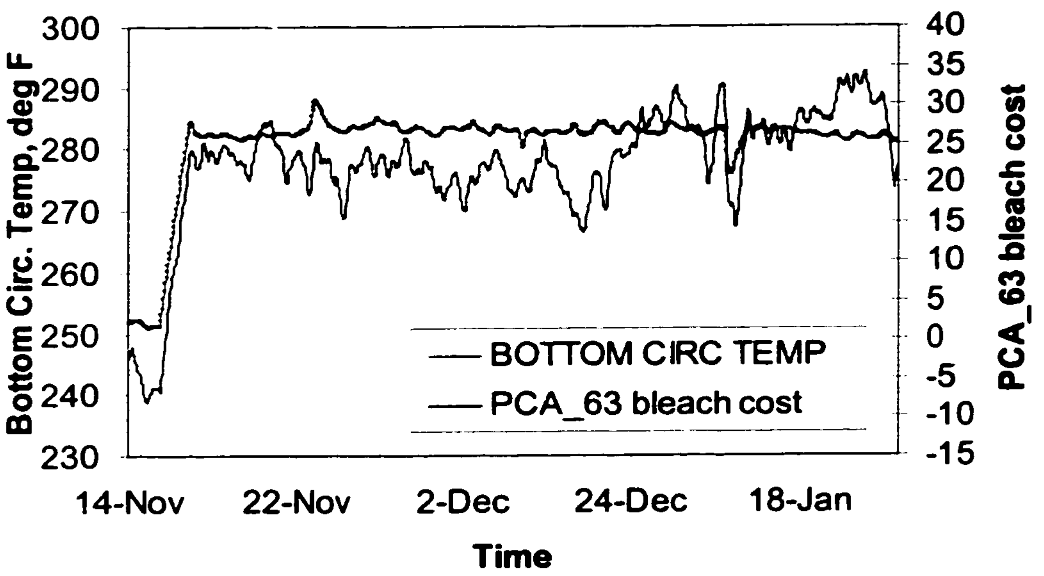



Figure 5-10. Bottom circulation temperature and PCA_63 model predictions of bleach cost.

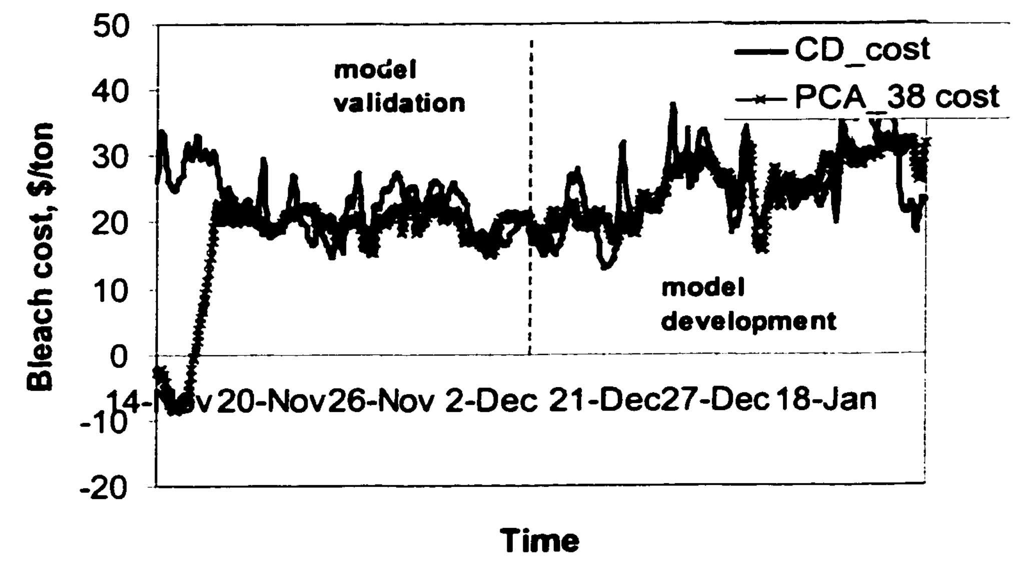



Figure 5-11. Principal component analysis predictions of bleaching cost for PCA_38

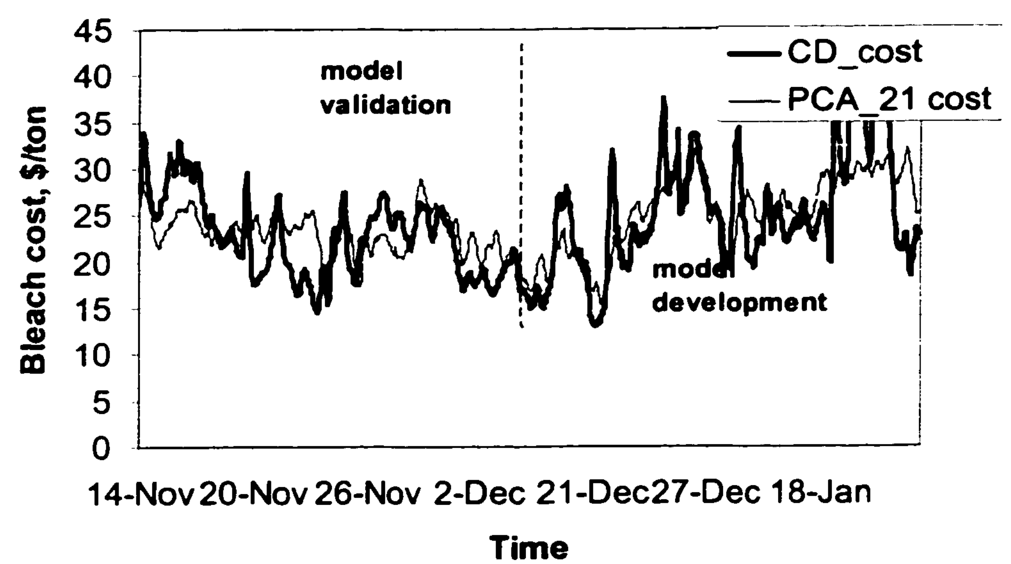



Figure 5-12. Principal component analysis predictions of bleaching cost for PCA_21

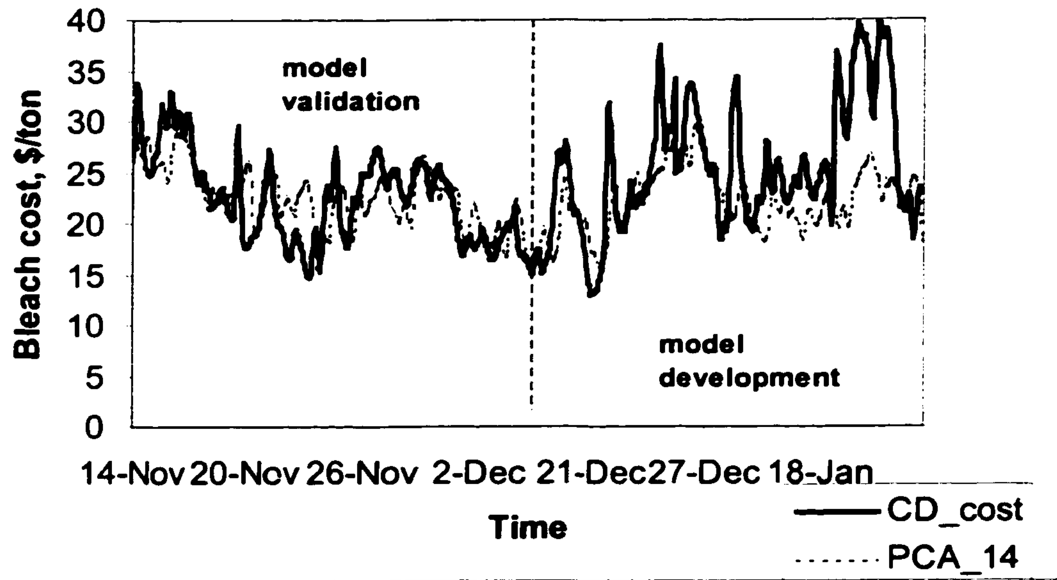



Figure 5-13. Principal component analysis predictions of bleaching cost for PCA_14

Details of the PCA models are presented in Table 5-13. The accuracy of the resulting principal component analysis models for all observations is shown in Figures 5-11, 5-12, and 5-13. The coefficients of correlation (R-square) for the three smaller models and PCA_63 model are presented in Table 5-13. Bleaching cost predictions from model PCA_38 could not predict first seventy observations for the reasons similar to PCA_63 model. Principal component analysis models PCA_21 and PCA_14 were able to predict bleach cost successfully for all the observations (Table 5-14). These models performed well for the first seventy observations as well. The main reason for improved cost prediction of these smaller models is exclusion of temperature variables for PCA_21, PCA_14 models. For simpler representation of mill operations, the model with the smallest number of variables (PCA_14) was chosen for next stage of the analysis.

Table 5-13. Principal component analysis models of bleaching cost for GP mill.

| Model → | PCA_63 | PCA_38 | PCA_21 | PCA_14 |
|---|---|---|---|---|
| Component 1 | 0.41 | 0.52 | 0.15 | 0.43 |
| Component 2 | -0.22 | -0.11 | 0.52 | -0.54 |
| Component 3 | 0.53 | 0.12 | -0.47 | |
| Component 4 | -0.02 | -0.39 | | |
| Component 5 | -0.15 | 0.27 | | |
| Component 6 | 0.06 | | | |
| Component 7 | 0.29 | | | |
| Component 8 | 0.12 | | | |

Table 5-14. Correlation coefficients of various bleach cost PCA models.

| | Number of variables | Number of components | R2, all data | R2, for validation data | Squared residual average |
|---|---|---|---|---|---|
| PCA_63 | 63 | 8 | 0.76* | 0.34* | 14.08* |
| PCA_38 | 38 | 5 | 0.71* | 0.44* | 13.35* |
| PCA_21 | 21 | 3 | 0.68 | 0.50 | 16.33 |
| PCA_14 | 14 | 2 | 0.68 | 0.73 | 13.66 |

* refers to values excluding first seventy observations

**Interpretation of PCA models**

After modeling the bleaching cost, the underlying patterns in the dataset were identified for the sources of variability. There were two patterns underlying the upstream process variables used in the bleaching cost analysis. First pattern (Lignin component) was strongly correlated to digester K-number and wash flow rates to atmospheric diffusion washers. In essence, first pattern represented the importance of cooking and washing in determining the bleaching cost of pulp. The second pattern in PCA_14 dataset, DigStability component is related to digester column stability as it was strongly correlated to outlet device current consumption and differential pressures of extraction screens. Column stability is a very important factor as a stable column represents steady state operation. If the digester is operating under unsteady state condition pulping is performed under non-uniform cooking conditions, thus resulting into varying pulp-bleaching cost. PCA_14 has essentially same variable structure when compared to FA model FA_14. However exact weights of different variables on factors/components are different. This is to be expected from differing assumptions about common variability of variables when developing FA and PCA models.

### *5.3.5 NEURAL NETWORK ANALYSIS*

Both PCA and FA models assume a dataset contain only linear interactions among variables. In order to investigate the presence of non-linearities present in bleaching cost dataset a non-linear multivariate technique, neural network, was used. The time synchronized and preprocessed data, representing about one month of pine production at the mill, were then analyzed using neural network (NN) analysis. The conditioned data was partitioned into three separate parts. The first part, with 520 observations, was used to train and build a NN model, while the second part with 200 observations was used to validate the model. Another part consisting of 105 observations was used as a testing dataset for the NN model. Testing refers to a process of preventing the neural network from over fitting the dataset. The model validation was done to test the accuracy and robustness of the NN model. The split in the dataset for model development and validation for NN is different from PCA, FA models as more data are required for training neural networks. However, the exact division of data into model development, testing, and validation parts was done by looking at bleaching cost trend and choosing the portion with higher bleaching cost variability.

Similar to factor analysis and principal component analysis, several NN models were developed using sets of upstream variables. In all four NN models were developed using 63, 38, 21, and 14 upstream variables. The training and testing results of all these NN models are presented in Table 5-15. Table 5-15 shows that neural network with sixty-three variables, NN_63, has no predictive capability for bleaching cost. In fact most of the neural network models didn't predict bleaching cost. Only the neural network model (NN_14) with fourteen variables had some reasonable prediction of bleaching cost as shown in Figure 5-20. This could be because a neural network with a large number of input variables tries to model complex relationships among all the input variables. The actual predictive structure for bleaching cost is masked by noise and disturbances that neural networks (with large number of input variables) try to model. It is noticeable that even for the NN_14 model the coefficient of correlation ($R^2$ value) for the validation dataset was lower than $R^2$ values for principal component analysis and factor analysis models using same variables. A non-linear model of upstream variables to predict bleaching cost doesn't seem to work. In contrast, both PCA, FA models with 14 upstream variables were able to predict bleaching cost quite well.

It appears that linear modeling techniques such as FA, PCA are quite effective in ignoring noise, disturbances present in the dataset while managing to capture predictive structures present in the process dataset. The case is different for neural networks which try to model all variations present in the dataset including noise. As a result NN do a poor job of predicting bleaching cost. If the number of variables is large, then more noise and disturbances are present in the dataset. The NN will try to model the noise while ignoring important predictive structure. When the number of variables in NN is small, there is less noise present in the dataset. In such cases, NN models do a relatively better job of predicting bleaching cost as shown by higher coefficient of correlation for the validation dataset (Table 5-15).

Table 5-15. Training and validation results of the neural network models for bleaching cost.

| | Observations | | | | R2 value | |
|---|---|---|---|---|---|---|
| Model | train | test | validation | Squared residual average | all data | Validation |
| NN_63 | 520 | 105 | 1-200 | 30.76 | 0.58 | 0.01 |
| NN_38 | 520 | 105 | 1-200 | 21.38 | 0.50 | 0.31 |
| NN_21 | 520 | 105 | 1-200 | 21.20 | 0.41 | 0.33 |
| NN_14 | 520 | 105 | 1-200 | 16.09 | 0.51 | 0.48 |

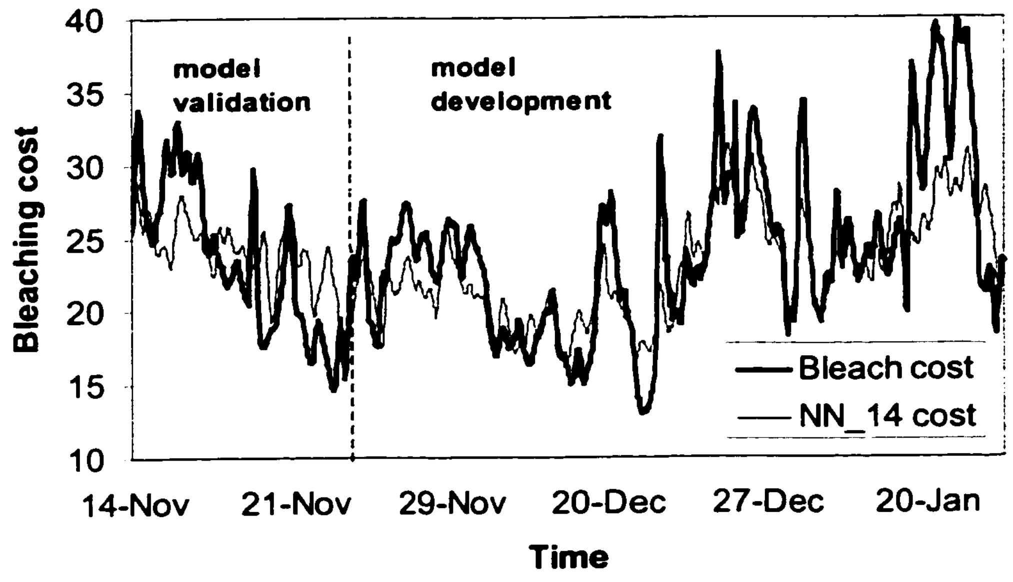



Figure 5-14. Neural network prediction of bleach cost using fourteen variables.

### *5.3.6 COMPARISON OF RESULTS USING DIFFERENT METHODOLOGIES*

In the previous sections, three multivariate methodologies were used to develop bleaching cost prediction models using the pulping dataset. The comparison of these methodologies was necessary to see which one of the multivariate techniques was better at modeling process characteristics. A basis for comparing the utility of multivariate techniques was developed. Each of the multivariate methodology was given a score on the following criteria

1. Accuracy of bleaching prediction in the validation dataset.

2. Ease of use.

3. Effort to build model.

4. Robustness to data preprocessing.

5. Explanation of process behavior.

Each criterion was assigned a arbitrary weight as per its importance, and a score on a scale of five (1-5) was given to each category for each multivariate model. The overall score of a multivariate methodology was its weighted average of scores in each of the five categories.

The overall scores of all models are given in Table 5-16. Table 5-17 shows average squared residuals for different models' validation datasets along with their 95% confidence interval. F_14 model has the best overall score of 4.2, followed by PCA_14 with score of 3.6. NN model with fourteen variables (NN_14) has the lowest score of 3.2. The score for "accuracy' is based on the magnitude of the average residuals (Table 5-17). As it is evident from the Table 5-17, PCA and FA models with fourteen variables give the best prediction of bleaching cost.

In terms of model building efforts, it was easier to build models with larger number of variables. Selecting variables for models with smaller number of upstream variables was quite difficult. Scores for "ease of use" were assigned based on personal experience. PCA models were fairly easy to use, but NN models were hard to operate.

PCA as well as FA models with smaller number of variables were easier to use and understand.

"Robust to data processing" refers to the ability of being able to construct models without sophisticated data preprocessing. PCA were very sensitive to the data preprocessing steps. FA models were slightly better than PCA models in this respect. Using NN didn't need detailed data preprocessing when compared to PCA and FA models. Choice of dataset, however, was quite important for NN compared to required for PCA and FA models. PCA and FA were better at extracting predictive relationships from noisy dataset. NN was not so flexible. As a result, using different datasets for NN yielded vastly different results. "Explanation of process behavior" refers to the ability to reconcile model interpretation to process knowledge. For the upstream dataset it was easiest for FA models followed by PCA models. NN models were black box models, as a result the explanation of process behavior was rather difficult.

Table 5-16 Overall scores of comparison of multivariate models for bleach cost prediction.

| Weight | 4.0 | 2.0 | 3.0 | 2.0 | 3.0 | 14.0 |
|---|---|---|---|---|---|---|
| ***Model*** | ***Accuracy*** | ***Ease of use.*** | ***Effort to build model.*** | ***Robustness to data preprocessing.*** | ***Explanation of process behavior.*** | ***Overall*** |
| FA_14 | 4.0 | 4.0 | 5.0 | 4.0 | 4.0 | **4.2** |
| PCA_14 | 4.0 | 4.0 | 3.0 | 3.0 | 4.0 | **3.6** |
| NN_14 | 3.0 | 2.0 | 5.0 | 4.0 | 2.0 | 3.2 |

Table 5-17 Squared residual average in bleaching cost prediction for validation datasets of different models.

| **Model** | **Squared residual average** | **95% C.I.** |
|---|---|---|
| **FA_14** | 10.41 | 10.41±1.54 |
| **PCA_14** | 9.20 | 9.20±1.34 |
| **NN_14** | 16.09 | 16.09±2.46 |

## 5.4 FIBERLINE VARIABILITY: K-NUMBER AND BLEACHING COST

The dissertation, so far, has focused on studying bleaching cost variability and K-number variability for the Georgia Pacific mill. Since the fiber line consists of a pulp mill followed by the bleach plant, there should be a link between kappa number variations and bleaching cost variations. The exact quantitative effect of K-number variations on the bleaching cost variability is determined in this section.

Bleaching cost predictions for all observations for PCA_14 model is shown in the Figure 5-18. The coefficient of correlation ($R^2$) is 68% for this plot. Figure 5-19 shows bleaching cost predictions for PCA_14 when K-number variables (bsknum, lo_knum) are excluded from the model. The coefficient of correlation ($R^2$) is 45% for bleaching cost predictions. It is evident that bleaching cost variability, as predicted by model, is reduced by 23% when information about K-number variations is not included in the analysis. It was evident from the discussions in Chapter 4 that K-number variability can't be predicted reliably for the Georgia Pacific mill digester. As a result not much can be done about controlling a significant portion of bleaching cost variations as predicted by the PCA_14 model. Controlling other sources of bleaching cost variability, i.e., washing variables and DigStability factor, can lead, however, to reduction in the cost variations and thus economic operations of the bleach plant.

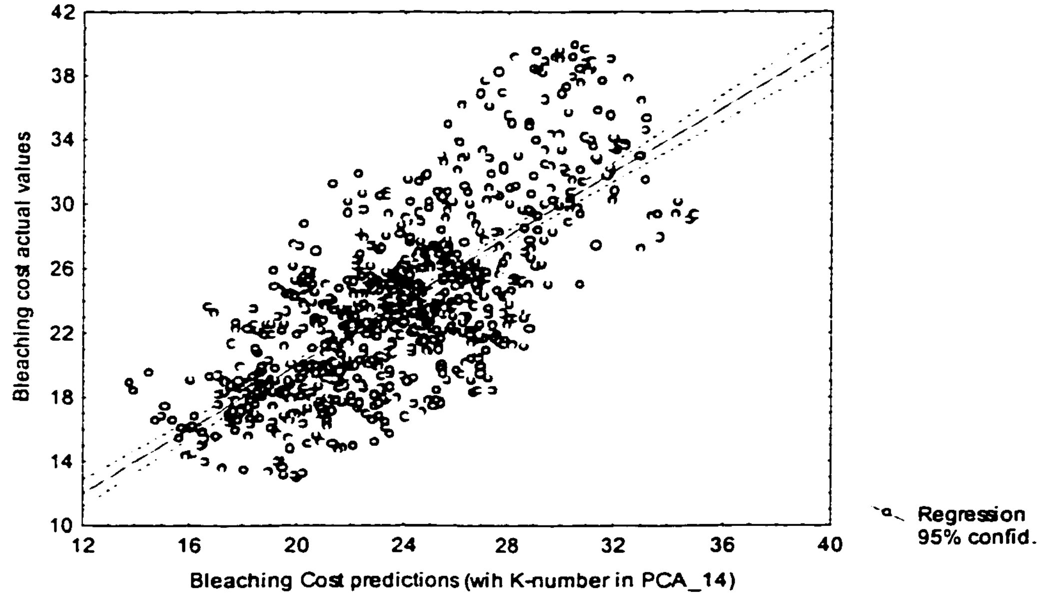



Figure 5-15. Predicted vs. Observed Values for bleaching cost (with K-number in model).

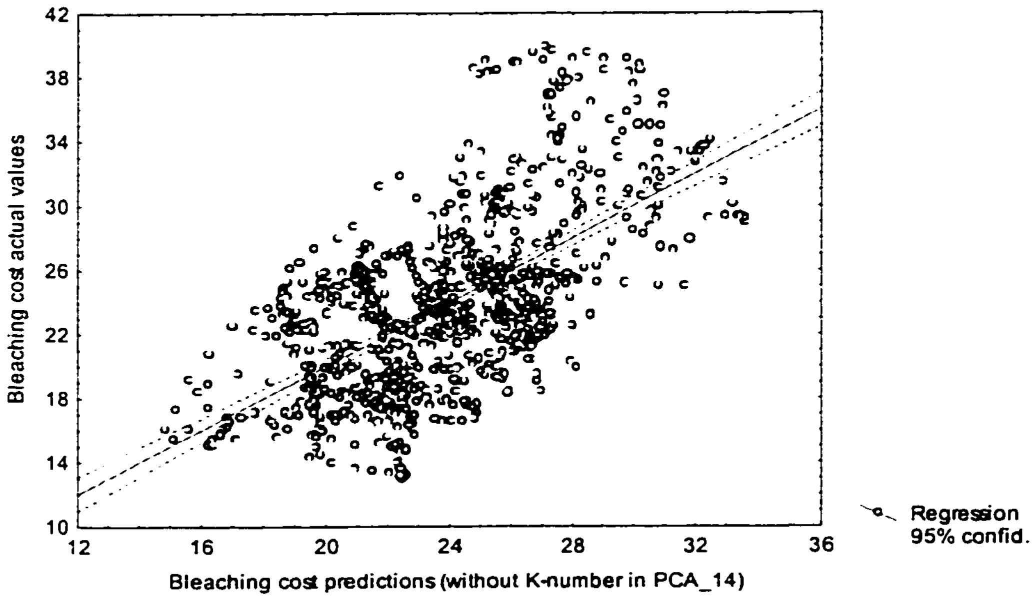



Figure 5-16. Predicted vs. Observed Values for bleaching cost (without K-number in model).

## 5.5 CONCLUSIONS OF BLEACHING COST STUDY

The bleaching cost study was done in two phases. The first phase determined the bleaching stage that was contributing most to total bleaching cost variability. Results from the first phase indicated that most of the variability in total bleaching cost came from the first stage of bleaching, i.e., D/C stage. In the second phase of the bleaching study, upstream variables in the fiberline were used to develop predictive models of bleaching cost. A number of models were developed using principal component analysis, factor analysis, and neural network analysis.

For principal component analysis as well as factor analysis, models with fourteen variables successfully predicted the bleaching cost trend. Neural networks bleaching cost predictions were poor. Factor analysis and PCA models of the bleaching cost indicated that most of the bleaching cost variability was either due to lignin factor (which represents pulping and washing variables) or due to digester column stability represented by outlet device amperage. A method to compare results from various multivariate methodologies was also developed. The factor model with fourteen variables achieved the highest score of all the models.

# CHAPTER 6: CONCLUSIONS

The financial and process benefits of improving the mill fiber line are widely acknowledged. However, process optimization of the fiber line is difficult due to the complex behavior of pulp and paper systems. The dissertation project focused on the application of multivariate analysis techniques for understanding and improving fiber line performance. More specifically, the research applied different methods for prediction and variability analysis of kappa number and total bleaching cost. These models were developed using data generated by pulping and bleaching operations. The research project led to refinement of earlier methods of data preprocessing. A number of concepts relating to the analysis of data generated by mill operations were proposed. Also a method of comparing results from different multivariate techniques was proposed. Detailed conclusions are described in the following sections.

## 6.1 DATA PREPROCESSING

Large quantities of process data are readily available in pulp mills for improving process operations. However, there exist several problems inherent in the pulping and bleaching processes that serve as obstacles in using such data. A number of steps to eliminate effects of transformations and changes implicit in the raw process data were proposed in the dissertation.

An algorithmic solution for time shifting problems present in the mill dataset was developed. Pulp tracking led to models with a more realistic correlation structure and better prediction of kappa number and bleaching cost.

## 6.2 KAPPA STUDY CONCLUSIONS

Two studies were done for kappa number prediction using data from a pulp mill. In the first study, the Weyerhaeuser Longview mill data was used to predict kappa number out of digester and $O_2$ delignification reactor. In the Longview study, factor analysis allowed development of models that successfully predict kappa number out of a continuous digester and $O_2$ delignification stage. The most important cause of kappa variability in case of the continuous digester was found to be mischarges in alkali. The major source of kappa variability in the $O_2$ delignification reactor is variability of the pulp out of the digester. Better digester control can lead to improved $O_2$ reactor performance. Factor analysis results can be used to reduce kappa variability as they point out factors and in turn the variables that cause kappa variability. Variations in kappa number can be reduced by 45% in case of the digester and 40% in case of the O2 delignification reactor if variables correlating with the important factors are brought under control.

In the second study, principal component analysis, factor analysis, and neural network were used to develop K-number prediction models for the Georgia Pacific Ashdown mill. None of these models were successful in predicting K-number. The main reason for poor prediction was that the digester was already under tight control as evident from the low (6.12%) coefficient of variation of K-number. It appears that the dataset generated by a strongly controlled process doesn't have significant correlation structure, which is necessary to develop predictive models.

## 6.3 BLEACHING COST CONCLUSIONS

The bleaching cost study was done in two phases. The first phase determined the bleaching stage which was contributing most to total bleaching cost variability. Results from the first phase of the bleaching cost study indicated that most of the variability in total bleaching cost came from the first stage of bleaching. In the second phase of the bleaching study, upstream variables in the fiberline were used in the predictive modeling of bleaching cost. A number of models were developed using principal component analysis, factor analysis, and neural network analysis for predicting bleaching cost.

For principal component analysis as well as factor analysis, models with fourteen upstream variables successfully predicted bleaching cost trend. However, in case of neural networks bleaching cost predictions however were poor. Factor analysis and PCA models of the bleaching cost indicated that most of the bleaching cost variability was either due to lignin factor (which represents pulping and washing variables) or due to digester column stability represented by outlet device amperage. A method to compare results from various multivariate methodologies was also developed. The factor model with fourteen variables achieved the highest score on a comparison scale for the bleaching cost study.

## 6.4 WHAT ALL THIS MEANS TO MILL?

It is evident from K-number and bleach plant case studies that upstream variables for bleaching have a linear predictive structure for bleaching cost. However, two important components of the bleaching cost predictive model are the lignin factor and digester stability factor. Both of these factors point at the digester as being major source of bleaching cost variability. At the same time, the K-number case study shows that digester is operating smoothly as evident by small coefficient of variation of quality variable, K-number. It seems impossible to predict or correlate digester K-number with input variables to the pulping process for the Georgia Pacific mill. The question arises, "what mill can do to reduce cost variability?"

It appears that either something is happening in the digester that is not picked up by the K number test or some unknown disturbances (such as chip quality) are passing through the digester causing bleach cost to vary. If the first hypothesis is correct, then K number determined at the end of digester is not sensitive enough to assess variations in lignin content. One can say that there are variations in pulp lignin content that are not measured by the K number test, but are picked up by changes in the chlorine requirements at the D/C stage. In this situation, the bleaching cost predicted by the model may be used as a soft sensor to manipulate temperature, steam flow in digester to produce pulp with uniform quality (i.e., consistent latent variable variation). This way cost variability will be reduced, as presumably the variation in bleach cost will be minimized although lignin content may stay constant.

In case of unknown disturbances such as chip quality variations passing through the digester, there are no other choices than installing systems to monitor chip quality which will enable better digester control in terms of low bleaching cost variability.

## Bibliography


1. Smook, Gary A., Handbook for pulp and paper technologists, Angus Wilde Publications , p2-3(1992).

2. Brewster, D.B., Pulp and Paper manufacture, TAPPI PRESS, Atlanta, 1993, p 33

3. Fadum, Ole-Kristian, Tappi Journal, "The process information renaissance," 79(10):69(1996).

4. Hatton, J.V., Tappi 56(7):96-100(1973)

5. Hatton, J.V., Tappi 59(8):48-50(1976)

6. Tasman, J.E., Tappi 64(3):175-176(1981)

7. Tasman, J.E., Trans. Tech. Sect. CPPA 62(2):TR29(1980)

8. Lin, C.P., Mao, M.Y. and Jane, C.Y., Tappi 61(2):72 (1978)

9. Arkun, Y.; Rigopoulos, A.; "Principal Components Analysis in Estimation and Control of Paper Machines", European Symposium on Computer-Aided Process Engineering-6: Supplement to Computers & Computer Engineering [Volume 20, Part B] (Computers & Chemical Engineering): S1059-1064 (May 29, 1996; Elsevier Science Ltd)

10. Strand, B.C., “Analyzing the Effects of Raw Material Variation on Mechanical Pulp Properties Using Integrated Factor Networks,” CPPA Conference, Montreal, January, 1993.

11. Strand, B.C., Saltin, J.F., TAPPI 1992 Pulping Conference Proceedings, TAPPI PRESS, Atlanta, GA.

12. Strand, B.C., "Applications of Factor Network Analysis in the Pulp and Paper Industry", TAPPI, 1996

13. Saltin, Johan and Strand, B. C., "Modeling the Effects of Refiner Operation on Newsprint Quality," Mechanical Pulping Conference, Oslo, Norway, 1993

14. Rudd, J. B.; DeGroot, D. L., "Advanced Supervisory Control of Brownstock Washers Using a Neural Optimization Controller: Mill Trial", TAPPI/ISA PUPID Process Control Conf. (Lake Buena Vista) Proc.: 119-126 (March 3-7, 1991).

15. Dumont, G. A.; Tessier, P.; Yu, Q.; "Modeling a Wood-Chip Refiner Using Artificial Neural Networks ", Tappi Journal 78, no. 6: 167-174 (June 1995).

16. Khorasani, Kooi, S. B. L.; K.; " Control of Wood-Chip Refiner Using Neural Networks "Tappi J. 75, no. 6: 156-162 (June 1992)

17. Miyanishi, Takanori, "Neural networks in web break diagnosis of a paper machine", Proceedings of the 1998 65th Pulp and Paper Research Conference: 132-137 (1998; Japan TAPPI).

18. Pearson, K., "On Lines and Planes of Closest Fit to Systems of Points in Space," Philos. Mag., 2, 559-572, (1901).

19. Stephanopoulos, G. N., and H. Guterman (1989). "Pattern Recognition in Fermentation Processes,". In ACS Meeting, Miami Beach.

20. Kresta, J., T. E. Marlin, and J. F. MacGregor (Nov., 1990). "Choosing Inferential Variables Using Projection to Latent Structures ( PLS ) with Application to Multicomponent Distillation,". In Proceedings of AIChE Annual Meeting, Chicago IL.

21. Carey, W. P., and S. S. Yee, "Calibration of Nonlinear Solid-state Senor Arrays Using Multivariate Regression Techniques," Sensors and Actuators B., 9, 113- 122, (1992).

22. Wise, B. M., N. L. Ricker, D. J. Veltkamp, and B. R. Kowalski, "A Theoretical Basis for the Use of Principal Component Model for Monitoring Multivariate Processes," Process Control and Quality, 1(1), 41-45, (1992).

23. T. Joliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.

24. Stone, M., "Cross-validation: a Review," Math. Operations forsch. Statist., Ser. Statistics, 9(1), (1978).

25. Wold, S., "Cross-validatory Estimation of the Number of Principal Components in Factor and Principal Component Analysis," Technometrics, 20, 397- 405, (1978).

26. E Frank and J. H. Friedman, A statistical view of some chemometrics regression tools," Technometrics,35, pp. 109-35(1993)

27. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing.

28. Patterson, D. (1996). Artificial Neural Networks. Singapore: Prentice Hall.

29. Bhat, N., and T. J. McAvoy (Nov. 12-16, 1990). "Information Retrieval from a Stripped Backpropagation Network,". In Presented at AIChE National Meeting, Chicago.

30. Schweiger, C. A.; Rudd, J. B., "Prediction and Control of Paper-Machine Parameters Using Adaptive Technologies in Process Modeling ", Tappi Journal 77, no. 11: 201-208 (November 1994).

31. Alwan, L.C., and H.V. Roberts, "Time-series modeling for statistical process control," Journal of Business and economic statistics, 6(1), 87-95, 1988.

32. Croteau, A. P., Nobleza, G. C..; Roche, A. A.; "Analysis Techniques: Practical Tool for Mill-Wide Quality Improvements",Pulp Paper Can. 91, no. 7: 76-81 (July 1990)

33. Allison, B. J. ; Cheetham, W. J.; Dumont, G. A.; Novak, L. H.,, "Adaptive-Predictive Control of Kamyr Digester Chip Level ", AIChE J. 36, no. 7: 1075-1086 (July 1990)

34. Ritala, R., "New Tool for Powerful Process Analysis", Pap. Puu 75, no. 7: 488-490 (Sept. 24, 1993)

# APPENDIX A: Pulp tracking algorithm

## VISUAL BASIC CODE FOR PULP TRACKING PROGRAM.

```
'-----------------------------------------------------------------------------------

' author: Saket Kumar :University of washington
'              phone: 206 543 8142 (email : saket@u.washington.edu)
' Author would like to thank pulp mill and technical dept. engineers
' for extending their full support and cooperation in this project. Thanks
' georgia pacific, Ashdown operations
'-----------------------------------------------------------------------------------
   Public calcSheet As Object
   Public numVar As Integer
   Public msg As String
   Public NewLine As String    ' New-line.
   Public Const ObsInterval As Integer = 60    ' interval in min between two
observations
   Public Const maxSyncTime As Integer = 2600 ' time in minutes
   Public Const one_day As Integer = 1440 'minutes in a day
   Public probTag As String
   Public mySheet As Object
   Public probTime As Date
   Public adjProbTime As Date

Sub Auto_Open()
   ThisWorkbook.Sheets("Pulp tracking").Select
End Sub

Sub Auto_close()
    ActiveWorkbook.Close (False)
End Sub


'following functions handles time spent in HD towers
```

```
Function RetnTime(sTagname As String, flowRateTag As String, sTime As Date)
  Dim eTime As Date
  Dim eLTime As Date
  Dim sLTime As Date
  Dim deadTime As Integer ' dead time in minutes
  Dim avgValveOpenTime  As Integer ' for flow from #2BP to HDs, in minutes
  Dim tankLevel ' don't dimension otherwise lose precision
  Dim machineFlow ' don't dimension otherwise lose precision
  Dim avgFlowInGPM
  Dim avgConsistency
  Dim mySheet As Object
  Dim pulpToMachine As Long
  Dim machineFlow62
  Dim machineFlow63
  Dim machineFlow64
  Dim tagArea 'As Integer
  Dim convFac

   avgValveOpenTime = 20

  ' what will happen if this value is non-numeric use value ten minutes after
  sLTime = adjustTime(sTime, -avgValveOpenTime / 2)
  eLTime = adjustTime(sTime, avgValveOpenTime / 2)
  tankLevel = VarToInt(TagCalAvgVal(sTagname, sLTime, eLTime))
  '-----------------------------------------------------------------------------
  ' This code handles the conversion of TPD flow to TPH in case of bleach tag
  convFac = 1

  If Not (Left(sTagname, 8) = "30-WI023") Then
  tagArea = Left(flowRateTag, 2)
  If tagArea = 33 Then
    convFac = 24 ' as bleach plant rate is in TPD
  End If
  End If
  '--------------------------------------------------------------------------------
  i = 1
  deadTime = 1
  Do
  start = Timer

    eTime = adjustTime(sTime, deadTime)
    'what happens if this value is non-numeric
```

```
    ' handles the atm diffuser case where there is no flowrate.PE tag
    If Left(sTagname, 8) = "30-WI023" Then '"30-WI023."
       avgFlowInGPM = VarToInt(TagCalAvgVal("30-FI001 .", sTime, eTime))
       avgConsistency = VarToInt(TagCalAvgVal("30-CC025 .", sTime, eTime))
       machineFlow = avgFlowInGPM * (8.34 * 60 / 2000) * (avgConsistency / 100) ' flow out to machine in TPH
  ElseIf sTagname = "36-LI050 ." Then  ' deals with HD tower #6 (three pulp lines out of HD#7)
    MsgBox "pulp came out HD#7"
    machineFlow62 = VarToInt(TagCalAvgVal("36-FI058 .PE", sTime, eTime)) / convFac
    machineFlow63 = VarToInt(TagCalAvgVal("36-FI052 .PE", sTime, eTime)) / convFac
    machineFlow64 = VarToInt(TagCalAvgVal("36-FI137 .PE", sTime, eTime)) / convFac
    machineFlow = machineFlow62 + machineFlow63 + machineFlow64
    Else
       ' for HD #2 , HD #6
        machineFlow  =  VarToInt(TagCalAvgVal(flowRateTag,  sTime,  eTime))  /
convFac
    End If

    pulpToMachine = (machineFlow * deadTime / 60)
    pulpInTank = tankLevel - pulpToMachine

  'determines dead time step size based on the pulp left in
   If pulpInTank > 30 Then
       deadTime = deadTime + 100
   ElseIf pulpInTank <= 30 And pulpInTank > 10 Then
         deadTime = deadTime + 10
    ElseIf pulpInTank <= 10 Then
      deadTime = deadTime + 1
    End If
   i = i + 1 ' counter to prevent infinte looping
    If i > 2000 Then
       MsgBox "Please check flow rate" & flowRateTag & " profile."
       Exit Do
    End If
  Loop Until pulpInTank < 0.1
  RetnTime = deadTime
End Function


Function timeSpentInTank(sTagname As String, flowRateTag As String, eTime As Date)
  Dim avgConsistency
  Dim avgFlowInGPM
  Dim avgValveOpenTime  As Integer ' for flow from #2BP to HDs, in minutes
  Dim convFac As Integer
  Dim deadTime As Integer ' dead time in minutes
  Dim eLTime As Date
  Dim loopsTime As Date
```

```
Dim machineFlow ' don't dimension otherwise lose precision
Dim machineFlow62
Dim machineFlow63
Dim machineFlow64
Dim pulpToMachine As Long
Dim sLTime As Date
Dim sTime As Date
Dim tagArea 'As Integer
Dim tankLevel ' don't dimension otherwise lose precision

avgValveOpenTime = 20
deadTime = 10 'guessed deadtime
Do
 start = Timer
 loopsTime = adjustTime(eTime, -deadTime)
 sLTime = adjustTime(loopsTime, -avgValveOpenTime / 2)
 eLTime = adjustTime(loopsTime, avgValveOpenTime / 2)
tankLevel = VarToInt(TagCalAvgVal(sTagname, sLTime, eLTime)) ' pulp in HD when pulp sample came in

 ' what will happen if this value is non-numeric use value ten minutes after
 '-----------------------------------------------------------------------------
 ' This code handles the conversion of TPD flow to TPH in case of bleach tag
 convFac = 1
 If Not (Left(sTagname, 8) = "30-WI023") Then
  tagArea = Left(flowRateTag, 2)
  If tagArea = 33 Then
    convFac = 24 ' as bleach plant production rate is in TPD not TPH
  End If
 End If
 '-----------------------------------------------------------------------------

 'what happens if this value is non-numeric
 'handles the atm diffuser case where there is no flowrate.PE tag
 If Left(sTagname, 8) = "30-WI023" Then '"30-WI023."
    avgFlowInGPM = VarToInt(TagCalAvgVal("30-FI001 .", loopsTime,
eTime))
    avgConsistency = VarToInt(TagCalAvgVal("30-CC025 .", loopsTime,
eTime))
  machineFlow = avgFlowInGPM * (8.34 * 60 / 2000) * (avgConsistency / 100) ' flow out to knotters in TPH
 ElseIf sTagname = "36-LI050 ." Then ' deals with HD tower #6 (three pulp lines out of HD#7)
    ' MsgBox "pulp came out HD#7"
  machineFlow62 = VarToInt(TagCalAvgVal("36-FI058 .PE", loopsTime, eTime)) / convFac
    machineFlow63 = VarToInt(TagCalAvgVal("36-FI052 .PE", loopsTime, eTime)) / convFac
  machineFlow64 = VarToInt(TagCalAvgVal("36-FI137 .PE", loopsTime, eTime)) / convFac
    machineFlow = machineFlow62 + machineFlow63 + machineFlow64
```

```
    Else
      ' for HD #2 , HD #6
        machineFlow = VarToInt(TagCalAvgVal(flowRateTag, loopsTime, eTime)) / convFac
    End If
  pulpToMachine = (machineFlow * deadTime / 60)
  pulpInTank = tankLevel - pulpToMachine

  'determines dead time step size based on the pulp left in tank goto
  If pulpInTank > 30 Then
     deadTime = deadTime + 20
  ElseIf pulpInTank <= 30 And pulpInTank > 10 Then
       deadTime = deadTime + 10
  ElseIf pulpInTank <= 10 Then
      deadTime = deadTime + 1
  End If

 Loop Until pulpInTank < 0.1
 timeSpentInTank = deadTime

End Function
```

```
Function timeSpentSmlHD(sTagname As String, flowRateTag As String, eTime As Date)
   Dim sTime As Date
   Dim loopsTime As Date
   Dim bestStartTime1 As Date
   Dim bestStartTime2 As Date
   Dim bestStartTime3 As Date
   Dim bestStartTime4 As Date
   Dim calcDeadTime As Integer ' dead time in minutes
   Dim tankLevel
   Dim machineFlow
   Dim mySheet As Object
   Dim pulpToMachine As Long
   Dim guess As Integer
   Dim guessRetnTime ' As Integer
   Dim errTime
   Dim timeStep
   Dim errSign As Boolean
   Dim c As Object

  'ActiveWorkbook.Save
   'Application.ScreenUpdating = False

   errSign = True: i = 1 ' Initialize variables.
   Do
     loopsTime = eTime - i * 30 / 60 / 24
     calcDeadTime = RetnTime(sTagname, flowRateTag, loopsTime)
     pulpOutTime = adjustTime(loopsTime, calcDeadTime)
     errTime = (eTime - pulpOutTime) * 24 * 60
     If errTime > 0 Then
       errSign = False
      bestStartTime1 = loopsTime + 30 / 60 / 24 ' when error changes sign
     End If
      i = i + 1
   Loop Until errSign = False 'Or errTime = 0    ' Exit outer loop immediately.

   If Abs(errTime) < 1 Then
       timeSpentSmlHD = calcDeadTime
       Application.ScreenUpdating = True
      Exit Function
    End If
```

```
errSign = True: i = 0 ' Initialize variables.
Do
 loopsTime = bestStartTime1 - i * 15 / 60 / 24
 calcDeadTime = RetnTime(sTagname, flowRateTag, loopsTime)
 pulpOutTime = adjustTime(loopsTime, calcDeadTime)
 errTime = (eTime - pulpOutTime) * 24 * 60
 If errTime > 0 Then
  errSign = False
  bestStartTime2 = loopsTime + 15 / 60 / 24 ' when error changes sign
 End If
  i = i + 1
Loop Until errSign = False    ' Exit outer loop immediately.

If Abs(errTime) < 1 Then
    timeSpentSmlHD = calcDeadTime
    Application.ScreenUpdating = True
   Exit Function
 End If

errSign = True: i = 0 ' Initialize variables.
Do
 loopsTime = bestStartTime2 - i * 5 / 60 / 24
 calcDeadTime = RetnTime(sTagname, flowRateTag, loopsTime)
 pulpOutTime = adjustTime(loopsTime, calcDeadTime)
 errTime = (eTime - pulpOutTime) * 24 * 60
 If errTime > 0 Then
  errSign = False
  bestStartTime3 = loopsTime + 5 / 60 / 24 ' when error changes sign
 End If
  i = i + 1
Loop Until errSign = False 'Or errTime = 0    ' Exit outer loop immediately.

If Abs(errTime) < 1 Then
    timeSpentSmlHD = calcDeadTime
    Application.ScreenUpdating = True
   Exit Function
 End If


 errSign = True: i = 0 ' Initialize variables.
 Do
 loopsTime = bestStartTime3 - i * 1 / 60 / 24
 calcDeadTime = RetnTime(sTagname, flowRateTag, loopsTime)
```

```
    pulpOutTime = adjustTime(loopsTime, calcDeadTime)
    errTime = (eTime - pulpOutTime) * 24 * 60
    If errTime > 0 Then
     errSign = False
     bestStartTime4 = loopsTime + 1 / 60 / 24 ' when error changes sign
    End If
    i = i + 1
  Loop Until errSign = False 'Or Abs(errTime) < 1  ' Exit outer loop immediately.

     If Abs(errTime) < 1 Then
      timeSpentSmlHD = calcDeadTime
      Application.ScreenUpdating = True
     Exit Function
   End If
   timeSpentSmlHD = RetnTime(sTagname, flowRateTag, bestStartTime4)
'  Application.ScreenUpdating = True
```

**End Function**

```
'==============================================================================
' following subroutine finds the retention time in various sections for problem 1 as
' well as problem 2

Sub GetRetentionTime()
    Application.ScreenUpdating = False
    NewLine = Chr(13) + Chr(10)
    Set mySheet = Sheets("Pulp tracking")
    clearEntry ' erases old retention values from tracking sheet
   mySheet.Activate
   Call TrackPulp(mySheet.Range("prob1Cell"),mySheet.Range("prob1RetnTime"), mySheet.Range("prob1hd"))
   Call TrackPulp(mySheet.Range("prob2Cell"), mySheet.Range("prob2RetnTime"), mySheet.Range("prob2hd"))
    Range("f20").Select
    Application.ScreenUpdating = True
    msg = " Pulp tracker has found retention time in" & NewLine
    msg = msg + " various areas. Please check if these " & NewLine
    msg = msg + " look OK and then click on topTenMacro to " & NewLine
    msg = msg + "get top ten variables that have changed." & NewLine
    MsgBox msg, 48, " Pulp tracking Info"
End Sub


' lists top ten changed variables in pulping/bleaching/ k2 sections
Sub Display_Top_Ten_Variable()
   Dim hrColnNum As Integer
   Dim DBox As Object
   Dim ComboList As Object
   NewLine = Chr(13) + Chr(10)
   Set DBox = ThisWorkbook.DialogSheets("ComboDlg")
   Set ComboList = DBox.DropDowns("ComboList")

'  Clear the list
  If ComboList.ListCount <> 0 Then ComboList.RemoveItem Index:=1, Count:=ComboList.ListCount

'  Fill the list
   ComboList.AddItem Text:="1 Hour"
   ComboList.AddItem Text:="4 Hour"
   ComboList.AddItem Text:="12 Hour"
   ComboList.AddItem Text:="24 Hour"
   ComboList.AddItem Text:="7 days"
   ComboList.AddItem Text:="20 days"

'  Display the dialog
```

```
DBoxOK = DBox.Show

' If not canceled, show the selection
Number = ComboList.ListIndex  ' Initialize variable.
Select Case Number  ' Evaluate Number.
   Case 1 ' 1 hour difff.
      hrColnNum = 8
      Mystring = "1 hr. "
   Case 2 ' 4 hour difff.
      hrColnNum = 11
      Mystring = "4 hr. "
   Case 3 ' 12 hour difff.
      hrColnNum = 14
      Mystring = "12 hr. "
   Case 4 '24 hour difff.
      hrColnNum = 17
      Mystring = "24 hr. "
   Case 5 ' 7 day difff.
      hrColnNum = 20
      Mystring = "7 day "
   Case 6 '20 day difff.
      hrColnNum = 23
      Mystring = "20 day "
   Case Else   ' Other values.
       Mystring = " "
End Select
msg = " Please be patient as excel gets values from PI" & NewLine
msg = msg + " Average program time : 5-9 min(depends on computer speed)" & NewLine
MsgBox msg, 48, " Program information"

'updates column top
Set mySheet = Sheets("Pulp tracking")
mySheet.Range("G35") = Mystring + " before"
mySheet.Range("H35") = Mystring + " after"

mySheet.Range("G49") = Mystring + " before"
mySheet.Range("H49") = Mystring + " after"

mySheet.Range("G65") = Mystring + " before"
mySheet.Range("H65") = Mystring + " after"

mySheet.Range("G79") = Mystring + " before"
mySheet.Range("H79") = Mystring + " after"
```

```
mySheet.Range("G95") = Mystring + " before"
mySheet.Range("H95") = Mystring + " after"

mySheet.Range("G109") = Mystring + " before"
mySheet.Range("H109") = Mystring + " after"

 Application.ScreenUpdating = False
' calculate the variable information based on retention time
Sheets("Pulp tracking").Select
Range("A131:AB900").Select
Selection.Calculate

'problem 1 top ten
    Call GetTopTags(Sheets("Pulp tracking").Range("keyBleachInfoCell"), hrColnNum, 36) ' (data rownum, hour coln num, display position)
    Call GetTopTags(Sheets("Pulp tracking").Range("keyPulpInfoCell"), hrColnNum, 66) ' (data rownum, hour coln num, display position)
    Call GetTopTags(Sheets("Pulp tracking").Range("keyK2InfoCell"), hrColnNum, 96) ' (data rownum, hour coln num, display position)

' problem 2 top ten
Call GetTopTags(Sheets("Pulp tracking").Range("keyBleachInfoTwo"), hrColnNum, 50) ' (data rownum, hour coln num, display position)
Call GetTopTags(Sheets("Pulp tracking").Range("keyPulpInfoTwo"), hrColnNum, 80) ' (data rownum, hour coln num, display position)
Call GetTopTags(Sheets("Pulp tracking").Range("keyK2InfoTwo"), hrColnNum, 110) ' (data rownum, hour coln num, display position)
 Application.ScreenUpdating = True
End Sub

' prints the report
Sub PrintReport()
 Sheets("Pulp tracking").Activate
 Application.DisplayAlerts = False
 ActiveSheet.PageSetup.CenterHeader = "Pulp Tracking Report"
 ActiveWindow.SelectedSheets.PrintOut From:=2, To:=5, Copies:=1
 Application.DisplayAlerts = True
End Sub

'=====================================================================

Private Sub clearEntry()
 Sheets("Pulp tracking").Select
 ' clears the retention time cells and hd tower numbers
 Range("C16").Select
 Selection.ClearContents
 Range("C22").Select
 Selection.ClearContents
```

```
    Range("F20:F26").Select
    Selection.ClearContents
    Range("G20:G26").Select
    Selection.ClearContents
    Range("F15:F16").Select
    Selection.ClearContents
    Range("G15:G16").Select
    Selection.ClearContents

    ' clears the topt ten lists
    Range("B36:K45").Select
    Selection.ClearContents
    Range("B50:K60").Select
    Selection.ClearContents
    Range("B66:K76").Select
    Selection.ClearContents
    Range("B80:K90").Select
    Selection.ClearContents
    Range("B96:K106").Select
    Selection.ClearContents
    Range("B110:O120").Select
    Selection.ClearContents
    Range("B114").Select
    Range("C8").Select
End Sub

Private Sub TrackPulp(ProblnfoCell As Object, RetnInfoCell As Object, hdCell As Object)
    Dim sTagname As String
    Dim syncTime As Integer
    Dim myVal ' As Variant
    Dim tagCell As Object
    Dim tagArea As Integer
    Dim flowRateTag As String
    Dim twoDaysAgo As Date
    Dim whenPulpLeftHDBot As Date
    Dim whenPulpLeftPineHDBot As Date
    Dim whenPulpLeftAtmDifBot As Date
    Dim whenPulpLeftSilo As Date
    Dim whenPulpEntersHD As Date
    Dim whenPulpEntersPineHD As Date
    Dim whenPulpEntersAtmDif As Date
    Dim hdTowerNum As Integer
    Dim timeInMachineArea As Integer
```

```
   Dim timeInPineHD As Integer
   Dim timeInAtmDif As Integer
   Dim avgCMRpm As Integer
   Dim avgBleachProdn As Integer
   Dim varAreaTime As Integer
   Dim totPulpingTime As Integer
   Dim totBleachingTime As Integer

  ' Application.ScreenUpdating = False
    NewLine = Chr(13) + Chr(10)
   ActiveWorkbook.Save
   Set mySheet = Sheets("Pulp tracking")
   probTag = mySheet.Range("C8").Value
   probTime = ProbInfoCell.Value

   'find average bleach prodn., avg. CM RPM  for last two days
    twoDaysAgo = adjustTime(probTime, -one_day * 2)
    avgCMRpm = VarToInt(TagCalAvgVal("20-SC005 .", twoDaysAgo, probTime))
    avgBleachProdn = VarToInt(TagCalAvgVal("33-FC006 .PE", twoDaysAgo,
probTime))

   'enter and change CM_rpm and Unbleached production values in sync sheets
   'get time spent in each of the sections
   Sheets("DigSyncInfo").Select
   Range("CM_rpm").Value = Fix(avgCMRpm)
   ActiveSheet.Calculate
   totPulpingTime = -(ActiveSheet.Range("f6").Value)

   Sheets("BleSyncInfo").Select
   Range("UB_production").Value = Fix(avgBleachProdn)
   ActiveSheet.Calculate
   totBleachingTime = -(ActiveSheet.Range("f6").Value)

  'find HD time using paper machine tag in question << which HD pulp came from
  'need to know paper machine for real as that will determine the HD tower and fiber
line
     tagArea = whichArea(probTag) ' area where problem arose
    Select Case tagArea ' based on where problem arose we determine if need to forecast/backcast

       Case Is = 27, 41, 42, 43, 44, 48, 49, 61, 62, 63, 64

'____________________________________________________________________________
__
```

```
'if problem arose in papermachine, stock prep, or pulp dryer then
'we need to backcast all the variables
'===================================================================
'following if statements determine which HD pulp sample came from

        If tagArea = 61 Or tagArea = 41 Or tagArea = 42 Then ' papermachine 61
        area>>HD#2
      hdTowerNum = 2
      sTagname = "36-LI009 ."
      flowRateTag = "36-FI008 .PE"
    ElseIf tagArea = 27 Then ' pulp dryer pulp comes from HD#6
      hdTowerNum = 6
      sTagname = "36-LI032 ."
      flowRateTag = "36-FI054 .PE"
      ' MsgBox "Problem in PD section"
    ElseIf tagArea = 64 Or tagArea = 63 Then ' machine 64 pulp comes from
HD#7
      hdTowerNum = 7
      sTagname = "36-LI050 ."
      flowRateTag = "36-FI137 .PE"
    ElseIf tagArea = 48 Or tagArea = 49 Then ' machine 63 pulp comes from
HD#7
      hdTowerNum = 7
      sTagname = "36-LI050 ."
      flowRateTag = "36-FI052 .PE"
    ElseIf tagArea = 43 Or tagArea = 44 Or tagArea = 62 Then' machine 62 pulp >> HD#7
      hdTowerNum = findTower(probTime, probTag) ' find where pulp came
from
      If hdTowerNum = 7 Then
        sTagname = "36-LI050 ." ' for HD#7
        flowRateTag = "36-FI058 .PE"
      ElseIf hdTowerNum = 4 Then
        sTagname = "36-LI016 ." ' for HD#4
        flowRateTag = "36-FI015 .PE"
      End If
    End If
    '==============================================================
    timeInMachineArea = timeToMachine(probTag) ' from HD bottom to problem tag point
    whenPulpLeftHDBot = adjustTime(probTime, -timeInMachineArea)
    ' time spent in HD depends on tank level when sample came in and machine pull afterthat
    timeInHD = timeSpentInTank(sTagname, flowRateTag, whenPulpLeftHDBot)

   ' following codes are common to all tags if problem is in machine
   whenPulpLeftPineHDBot = adjustTime(whenPulpLeftHDBot, -(timeInHD + totBleachingTime))
```

```
timeInPineHD = timeSpentSmlHD("33-LI346B.", "33-FC006.PE", whenPulpLeftPineHDBot)'
timeUBlchdHD
 Sheets("DigSyncInfo").Select
  Application.Goto Reference:="knot_scrn"
  totScrngTime = Application.Sum(Selection)

  whenPulpLeftAtmDifBot = adjustTime(whenPulpLeftPineHDBot, -(timeInPineHD + totScrngTime))
timeInAtmDif = timeSpentSmlHD("30-WI023 .", flowRateTag, whenPulpLeftAtmDifBot)
 '-----------------------------------------------------------------
 ' update digester sheet for changes in atm diffuser retention time
 Sheets("DigSyncInfo").Select
 Range("d21").Value = Fix(timeInAtmDif)
 ActiveSheet.Calculate
 totPulpingTime = -(ActiveSheet.Range("f6").Value) ' updates the total pulping
time
 '-----------------------------------------------------------------

' time at for chip sample at silo
whenPulpLeftSilo = adjustTime(whenPulpLeftAtmDifBot, -(timeInAtmDif + totPulpingTime))

'now we have tracked that sample all the way back to chip silo for problem in machine
 ' display important retention times in different zone and different HD
 ' with a possibility of changing them changing the sheet calculation

 RetnInfoCell.Offset(-5, 0).Value = avgCMRpm
 RetnInfoCell.Offset(-4, 0).Value = avgBleachProdn
 hdCell.Value = hdTowerNum

 RetnInfoCell.Value = timeInMachineArea
 RetnInfoCell.Offset(1, 0).Value = timeInHD
 RetnInfoCell.Offset(2, 0).Value = totBleachingTime
 RetnInfoCell.Offset(3, 0).Value = timeInPineHD
 RetnInfoCell.Offset(4, 0).Value = totScrngTime
 RetnInfoCell.Offset(5, 0).Value = timeInAtmDif
 RetnInfoCell.Offset(6, 0).Value = totPulpingTime
 '_________________________________________________________________

Case Is = 33
'__________________________________________________________________
 ' if problem arose in bleachplant then need to forecast papermachine variable if
 ' possible and backcast into variables of pulping section

 ' MsgBox "the problem is in bleaching section"
 varAreaTime = -varSyncTime(probTag)
 '=================================================================
```

```
''''''''''''''
'forecasting ''
''''''''''''''
whenPulpEntersHD = adjustTime(probTime, varAreaTime) ' time adjusted to reflect when
    ' pulp enters in one of the bleached HD tower
'==========================================================================
''''''''''''''
'backcasting ''
''''''''''''''
whenPulpLeftPineHDBot = adjustTime(probTime, -(totBleachingTime - varAreaTime))
timeInPineHD = timeSpentSmlHD("33-LI346B.", "33-FC006.PE", whenPulpLeftPineHDBot) ' time UBlchdHD
Sheets("DigSyncInfo").Select
Application.Goto Reference:="knot_scrn"
totScrngTime = Application.Sum(Selection)
whenPulpLeftAtmDifBot = adjustTime(whenPulpLeftPineHDBot, -(timeInPineHD + totScrngTime))
timeInAtmDif = timeSpentSmlHD("30-WI023 .", flowRateTag, whenPulpLeftAtmDifBot)

'-----------------------------------------------------------------
' update digester sheet for changes in atm diffuser retention time
Sheets("DigSyncInfo").Select
Range("d21").Value = Fix(timeInAtmDif)
ActiveSheet.Calculate
totPulpingTime = -(ActiveSheet.Range("f6").Value) ' updates the total pulping
time
'-----------------------------------------------------------------
whenPulpLeftSilo = adjustTime(whenPulpLeftAtmDifBot, -(timeInAtmDif + totPulpingTime))

' now we have tracked that sample all the way back to chip silo
' display important retention times in different zone and different HD
' with a possibility of changing them changing the sheet calculation

RetnInfoCell.Offset(-5, 0).Value = avgCMRpm
RetnInfoCell.Offset(-4, 0).Value = avgBleachProdn
hdCell.Value = "Not Known"

RetnInfoCell.Value = timeInMachineArea
RetnInfoCell.Offset(1, 0).Value = 0
RetnInfoCell.Offset(2, 0).Value = totBleachingTime
RetnInfoCell.Offset(3, 0).Value = timeInPineHD
RetnInfoCell.Offset(4, 0).Value = totScrngTime
RetnInfoCell.Offset(5, 0).Value = timeInAtmDif
RetnInfoCell.Offset(6, 0).Value = totPulpingTime
'_______________________________________________________________

Case Is = 20
```

```
'_____________________________________________________________
            ' if problem arose in bleachplant then need to forecast papermachine &
            bleaching 'variables if' possible and backcast variables of pulping
            section

'     MsgBox "the problem is in pulping section"
        varAreaTime = -varSyncTime(probTag)

'=================================================================

            ''''''''''''''
        'forecasting "
        ''''''''''''''
        whenPulpEntersAtmDif = adjustTime(probTime, varAreaTime) ' time adjusted to reflect when
timeInAtmDif = RetnTime("30-WI023 .", flowRateTag, whenPulpEntersAtmDif) ' please note that
        ' function Retn time ignores the flowRateTag when sTagName is 30-WI023
        Sheets("DigSyncInfo").Select
        Application.Goto Reference:="knot_scrn"
        totScrngTime = Application.Sum(Selection)
whenPulpEntersPineHD = adjustTime(whenPulpEntersAtmDif, (timeInAtmDif + totScrngTime))
timeInPineHD = RetnTime("33-LI346B.", "33-FC006.PE", whenPulpEntersPineHD) 'time in UBlchdHD
whenPulpEntersHD = adjustTime(whenPulpEntersPineHD, (totBleachingTime + timeInPineHD))
        ' pulp enters in one of the bleached HD tower

'=================================================================

        ''''''''''''''
        'backcasting "
        ''''''''''''''

        '---------------------------------------------------------------
        ' update digester sheet for changes in atm diffuser retention time
        Sheets("DigSyncInfo").Select
        Range("d21").Value = Fix(timeInAtmDif)
        ActiveSheet.Calculate
        totPulpingTime = -(ActiveSheet.Range("f6").Value) ' updates the total pulping
time
        '---------------------------------------------------------------
        whenPulpLeftSilo = adjustTime(probTime, -(totPulpingTime - varAreaTime))

        ' now we have tracked that sample all the way back to chip silo

        ' display important retention times in different zone and different HD
        ' with a possibility of changing them changing the sheet calculation
```

```
    RetnInfoCell.Offset(-5, 0).Value = avgCMRpm
    RetnInfoCell.Offset(-4, 0).Value = avgBleachProdn
    hdCell.Value = "Not Known"

    RetnInfoCell.Value = timeInMachineArea
    RetnInfoCell.Offset(1, 0).Value = 0
    RetnInfoCell.Offset(2, 0).Value = totBleachingTime
    RetnInfoCell.Offset(3, 0).Value = timeInPineHD
    RetnInfoCell.Offset(4, 0).Value = totScrngTime
    RetnInfoCell.Offset(5, 0).Value = timeInAtmDif
    RetnInfoCell.Offset(6, 0).Value = totPulpingTime
    '_____________________________________________

End Select


' display important retention times in different zone and different HD
' with a possibility of changing them changing the sheet calculation

'mySheet.Range("f15").Value = avgCMRpm
'mySheet.Range("f16").Value = avgBleachProdn
'mySheet.Range("c16").Value = hdTowerNum

' mySheet.Range("Prob1RetnTime").Value = timeInMachineArea
'mySheet.Range("Prob1RetnTime").Offset(1, 0).Value = timeInHD
' mySheet.Range("Prob1RetnTime").Offset(2, 0).Value = totBleachingTime
' mySheet.Range("Prob1RetnTime").Offset(3, 0).Value = timeInPineHD
' mySheet.Range("Prob1RetnTime").Offset(4, 0).Value = totScrngTime
' mySheet.Range("Prob1RetnTime").Offset(5, 0).Value = timeInAtmDif
' mySheet.Range("Prob1RetnTime").Offset(6, 0).Value = totPulpingTime


Sheets("Pulp tracking").Select
Range("B5:N30").Select
Selection.Calculate
Range("c13").Select

Sheets("DigSyncInfo").Select
ActiveSheet.Calculate
Sheets("BleSyncInfo").Select
ActiveSheet.Calculate
Sheets("K2SyncInfo").Select
```

```
   ActiveSheet.Calculate
   Sheets("PMcSyncInfo").Select
   ActiveSheet.Calculate
   Sheets("HDtSyncInfo").Select
   ActiveSheet.Calculate
   Sheets("Pulp tracking").Activate
  ' Application.ScreenUpdating = True

 End Sub

'finds out which area tag came from [in case of K2 tags it looks for
' area information in sheet "K2SyncInfo" column "N"
Private Function whichArea(sTagname As String) As Integer
   Dim c  As Object
   Dim probArea 'As String

   probArea = Left(sTagname, 2) ' area where problem arose

    Select Case probArea  ' tag area.
      Case Is = "K2" ' pulping/ bleaching manual tests
      Set c = Sheets("K2SyncInfo").Range("K12")
      counter = 0
      Do While (c.Value <> sTagname)  ' Inner Loop.
        counter = counter + 1   ' Increment Counter.
        If counter > 70 Then ' as there are not more than 70 K2 variables
          Exit Do
        End If
        Set c = c.Offset(1, 0)
       Loop
       If c.Value = sTagname Then
         whichArea = c.Offset(0, 3).Value
       Else
        MsgBox "No such tag in my database" ' 48
        whichArea = 0
      End If
     Case Else
      whichArea = CInt(Left(sTagname, 2)) ' if problem is not in K2 area then it is in
area
                        ' given by first two entries of tag
      End Select
End Function

Private Function findTower(probTime As Date, probTag As String)
```

```
Dim eTime As Date
Dim avgValveOpenTime  As Integer ' for flow from #2BP to HDs, in minutes
Dim deadTime As Integer ' dead time in minutes
Dim eLTime As Date
Dim loopsTime As Date
Dim whenPulpLeftHDBot As Date
Dim machineFlow ' don't dimension otherwise lose precision
Dim machineFlow624
Dim machineFlow627
Dim machineFlow637
Dim machineFlow647
Dim pulpToMachine4 As Long
Dim pulpToMachine7 As Long
Dim sLTime As Date
Dim sTime As Date
Dim tagArea 'As Integer
Dim tankLevelHD7 ' don't dimension otherwise lose precision
Dim tankLevelHD4 ' don't dimension otherwise lose precision
Dim timeInMachineArea As Integer

timeInMachineArea = timeToMachine(probTag) ' from HD bottom to problem tag
point
whenPulpLeftHDBot = adjustTime(probTime, -timeInMachineArea)
eTime = whenPulpLeftHDBot

avgValveOpenTime = 20
deadTime = 10 'guessed deadtime
Do
  start = Timer
  loopsTime = adjustTime(eTime, -deadTime)
  sLTime = adjustTime(loopsTime, -avgValveOpenTime / 2)
  eLTime = adjustTime(loopsTime, avgValveOpenTime / 2)

  ' calculates pulp in HD 4
  tankLevelHD4 = VarToInt(TagCalAvgVal("36-LI016 .", sLTime, eLTime)) ' pulp
in HD
  machineFlow624   =   VarToInt(TagCalAvgVal("36-FI015   .PE",   loopsTime,
eTime))
  pulpToMachine4 = (machineFlow624 * deadTime / 60)
  pulpInTank4 = tankLevelHD4 - pulpToMachine4


  ' calculates pulp in HD 7
```

```
    tankLevelHD7 = VarToInt(TagCalAvgVal("36-LI050 .", sLTime, eLTime)) ' pulp
in HD
    machineFlow627    =    VarToInt(TagCalAvgVal("36-FI058    .PE",    loopsTime,
eTime))
    machineFlow637    =    VarToInt(TagCalAvgVal("36-FI052    .PE",    loopsTime,
eTime))
    machineFlow647    =    VarToInt(TagCalAvgVal("36-FI137    .PE",    loopsTime,
eTime))
    machineFlow7 = machineFlow627 + machineFlow637 + machineFlow647
    pulpToMachine7 = (machineFlow7 * deadTime / 60)
    pulpInTank7 = tankLevelHD7 - pulpToMachine7

   If pulpInTank4 >= pulpInTank7 Then
     pulpInTank = pulpInTank4
   Else
     pulpInTank = pulpInTank7
   End If

    'determines dead time step size based on the pulp left in tank goto
    If pulpInTank > 30 Then ' if machine draw is zero (machine is down)
       'MsgBox " pulp in tank is : " & pulpInTank
       deadTime = deadTime + 20
    ElseIf pulpInTank <= 30 And pulpInTank > 10 Then
        deadTime = deadTime + 10
    ElseIf pulpInTank <= 10 Then
       deadTime = deadTime + 1
    End If

  Loop Until pulpInTank7 + pulpInTank4 < 0.2

  ' calculates which HD send more pulp to # 62 machine HD#7 or HD#4
  sTime = adjustTime(eTime, -deadTime)
  machineFlow624 = VarToInt(TagCalAvgVal("36-FI015 .PE", sTime, eTime))
  machineFlow647 = VarToInt(TagCalAvgVal("36-FI137 .PE", sTime, eTime))
  If machineFlow624 >= machineFlow647 Then
     findTower = 4
  Else
     findTower = 7
  End If
 ' MsgBox "our pulp came from HD#" & findTower
End Function

' following functions finds out how much time is spent by pulp from HD bottom to
```

```
' the point where problem is noted
Private Function timeToMachine(sTagname As String) ' As Integer
   Dim tagString As String
   Dim tagTable As Range

      tagString = sTagname
     ' first two digits of tag and find out whether tag is from bleach or digest
     ' and then look for appropriate sheet
      tagArea = Left(tagString, 2)

      Select Case tagArea   ' tag area.
        Case Is = ("27") 'Pulp dryer"
          Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
              tagString = "PD" + Left(sTagname, 2)
       Case Is = "41", "42", "61" ' papermachine
         Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM61"
       Case Is = "43", "44", "62" ' papermachine
         Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM62"
       Case Is = "48", "49" ', "63" ' papermachine
         Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM63"
       Case Is = "63", "64" ' papermachine
         Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM64"
      Case Else
          timeToMachine = 0 ' no machine time incase problem arose in bleaching
plant or before that
         Exit Function
     End Select
    '=======================================================================
   timeToMachine = -1 * Application.VLookup(tagString, tagTable, 2, False) ' gets time lag value from table
    '=======================================================================

End Function


Function totSyncTime(sTagname As String)
  Dim varAreaTime As Integer
  Dim curCell As Object

  Set mySheet = Sheets("Pulp tracking")
  Set curCell = mySheet.Range("prob1RetnTime")
```

```
timeInHD = curCell.Offset(1, 0).Value
totBleachingTime = curCell.Offset(2, 0).Value
timeInPineHD = curCell.Offset(3, 0).Value
totScrngTime = curCell.Offset(4, 0).Value
timeInAtmDif = curCell.Offset(5, 0).Value
totPulpingTime = curCell.Offset(6, 0).Value

varAreaTime = -varSyncTime(sTagname)
tagArea = whichArea(sTagname) 'Left(sTagname, 2) ' where is problem

Select Case tagArea   ' tag area.
     Case Is = "64", "63", "62", "61", "49", "48", "44", "43", "42", "41", "27" '"36"

         ' paper machine 64;  have to shift tags all the way back
                 totSyncTime = varAreaTime + timeInHD + totBleachingTime + timeInPineHD +
                 totScrngTime + timeInAtmDif + totPulpingTime
     Case Is = "33" ' Bleaching
                 totSyncTime = (totBleachingTime - varAreaTime) + timeInPineHD + totScrngTime +
                 timeInAtmDif + totPulpingTime
     Case Is = "30" ' screening
         totSyncTime = (totScrngTime - varAreaTime) + timeInAtmDif +
totPulpingTime
     Case Is = "20", "19" ' pulping
         totSyncTime = (totPulpingTime - varAreaTime)
     Case Else
        totSyncTime = 0
   End Select
End Function

Function varSyncTime(sTagname As String)
  Dim tagString As String
  Dim tagTable As Range

    tagString = sTagname
   ' first two digits of tag and find out whether tag is from bleach or digest
   ' and then look for appropriate sheet
    tagArea = Left(tagString, 2)

    Select Case tagArea   ' tag area.
      Case Is = "19", "20" ' chip supply area
         Set tagTable =
ThisWorkbook.Sheets("DigSyncInfo").Range("dig_sync_table")
   Case Is = "33" ' bleach plant
       Set tagTable = ThisWorkbook.Sheets("BleSyncInfo").Range("ble_sync_table")
      Case Is = ("27") 'Pulp dryer"
```

```
            Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PD" + Left(sTagname, 2)
    Case Is = "K2" ' pulping tests
        Set tagTable = ThisWorkbook.Sheets("K2SyncInfo").Range("k2_sync_table")
    Case Is = "11" ' retention towers
        Set tagTable = ThisWorkbook.Sheets("HDtSyncInfo").Range("hd_sync_table")
    Case Is = "41", "42", "61" ' papermachine
        Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM61"
    Case Is = "43", "44", "62" ' papermachine
            Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM62"
    Case Is = "48", "49" ', "63" ' papermachine
        Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM63"
    Case Is = "63", "64" ' papermachine
        Set tagTable = ThisWorkbook.Sheets("PMcSyncInfo").Range("pmc_sync_table")
            tagString = "PM64"
    Case Else
        MsgBox "No sync value available for -->> " & tagString
        Exit Function
      ' miscellaneous
    End Select
'==========================================================================
varSyncTime = Application.VLookup(tagString, tagTable, 2, False) ' gets time lag value from table
'==========================================================================

End Function

Function pulpK2SyncTime(sTagname As String)
  Dim varAreaTime As Integer
  Dim mySheet As Object

  Set mySheet = Sheets("Pulp tracking")
  totPulpingTime = mySheet.Range("prob1RetnTime").Offset(6, 0).Value
varAreaTime = -varSyncTime(sTagname) ' time from the end of pulping where variable was recorded
  pulpK2SyncTime = (totPulpingTime - varAreaTime)

End Function

Function bleachK2SyncTime(sTagname As String)
  Dim varAreaTime As Integer
  Dim mySheet As Object
  Dim curCell As Object

  Set mySheet = Sheets("Pulp tracking")
```

```
  Set curCell = mySheet.Range("prob1RetnTime")
  timeInHD = curCell.Offset(1, 0).Value
  totBleachingTime = curCell.Offset(2, 0).Value
  timeInPineHD = curCell.Offset(3, 0).Value
  totScrngTime = curCell.Offset(4, 0).Value
  timeInAtmDif = curCell.Offset(5, 0).Value
  totPulpingTime = curCell.Offset(6, 0).Value

  varAreaTime = -varSyncTime(sTagname) ' time from the end of bleaching where variable was recorded
          bleachK2SyncTime = (totBleachingTime - varAreaTime) + timeInPineHD + totScrngTime +
          timeInAtmDif + totPulpingTime

End Function

Function totSyncTime2(sTagname As String)
  Dim varAreaTime As Integer
  Dim curCell As Object

  Set mySheet = Sheets("Pulp tracking")
  Set curCell = mySheet.Range("prob2RetnTime")

  timeInHD = curCell.Offset(1, 0).Value
  totBleachingTime = curCell.Offset(2, 0).Value
  timeInPineHD = curCell.Offset(3, 0).Value
  totScrngTime = curCell.Offset(4, 0).Value
  timeInAtmDif = curCell.Offset(5, 0).Value
  totPulpingTime = curCell.Offset(6, 0).Value

  varAreaTime = -varSyncTime(sTagname)
  tagArea = whichArea(sTagname) 'Left(sTagname, 2) ' where is problem

  Select Case tagArea   ' tag area.
      Case Is = "64", "63", "62", "61", "49", "48", "44", "43", "42", "41", "27" '"36"
                  ' paper machine 64;  have to shift tags all the way back
            totSyncTime2 = varAreaTime + timeInHD + totBleachingTime + timeInPineHD + totScrngTime +
            timeInAtmDif + totPulpingTime
      Case Is = "33" ' Bleaching
                totSyncTime2 = (totBleachingTime - varAreaTime) + timeInPineHD + totScrngTime +
                timeInAtmDif + totPulpingTime
      Case Is = "30" ' screening
      totSyncTime2 = (totScrngTime - varAreaTime) + timeInAtmDif + totPulpingTime
      Case Is = "20", "19" ' pulping
          totSyncTime2 = (totPulpingTime - varAreaTime)
      Case Else
         totSyncTime2 = 0
```

```
      End Select

End Function

Function pulpK2SyncTime2(sTagname As String)
   Dim varAreaTime As Integer
   Dim mySheet As Object

   Set mySheet = Sheets("Pulp tracking")
   totPulpingTime = mySheet.Range("prob2RetnTime").Offset(6, 0).Value
   varAreaTime = -varSyncTime(sTagname) ' time from the end of pulping where variable was recorded
   pulpK2SyncTime2 = (totPulpingTime - varAreaTime)
End Function


Function bleachK2SyncTime2(sTagname As String)
   Dim varAreaTime As Integer
   Dim mySheet As Object
   Dim curCell As Object

   Set mySheet = Sheets("Pulp tracking")
   Set curCell = mySheet.Range("prob2RetnTime")
   timeInHD = curCell.Offset(1, 0).Value
   totBleachingTime = curCell.Offset(2, 0).Value
   timeInPineHD = curCell.Offset(3, 0).Value
   totScrngTime = curCell.Offset(4, 0).Value
   timeInAtmDif = curCell.Offset(5, 0).Value
   totPulpingTime = curCell.Offset(6, 0).Value
  ' MsgBox "bleachingtime :" & totBleachingTime
   varAreaTime = -varSyncTime(sTagname) ' time from the end of bleaching where variable was recorded
bleachK2SyncTime2 = (totBleachingTime - varAreaTime) + timeInPineHD + totScrngTime + timeInAtmDif + totPulpingTime

End Function


'-----------------------------------------------------------------------------------
' PI FUNCTIONS CALLED BY VBA
'==============================================================

' returns the PIArcVal based on a tagname and time
Function TagArcVal(sTagname As String, arcTime As Date) As Variant
    Dim sTime As String     'start Timestamp
    sTime = Format((arcTime), "dd-mmm-yy h:mm") '<<< PI timeformat
    TagArcVal = Application.Run("PIArcVal", sTagname, sTime, 0, "ASHPI")
```

```
End Function

' returns the PICalcVal based on a tagname and time
Function TagCalAvgVal(sTagname As String, sTime As Date, eTime As Date) As Variant
    Dim avgStime As String
    Dim avgEtime As String

   avgStime = (Format((sTime), "dd-mmm-yy h:mm")) '<<< PI timeformat
   avgEtime = (Format((eTime), "dd-mmm-yy h:mm")) '<<< PI timeformat
  TagCalAvgVal = Application.Run("PICalcVal", sTagname, avgStime, avgEtime, "average", 1, 0, "ASHPI")
End Function

' returns the PI calculates average value based on a tagname and stime and interval in
' minutes (between start and end times
Function CalAvgVal(sTagname As String, sTime As Date, difTime As Integer) As Variant
    Dim avgStime As String
    Dim avgEtime As String
    Dim temp As String
   avgStime = (Format((sTime), "dd-mmm-yy h:mm")) '<<< PI timeformat
  avgEtime = Format(adjustTime(sTime, (difTime)), "dd-mmm-yy h:mm") '<<< PI timeformat
   If difTime < 0 Then
     temp = avgEtime
     avgEtime = avgStime
     avgStime = temp
  End If
  CalAvgVal = Application.Run("PICalcVal", sTagname, avgStime, avgEtime, "average", 1, 0, "ASHPI")
End Function
'------------------------------------------------------------------------------


' this function shifts the probtime by the varSyncTime (in minutes)
' and returns time stamp in PI format, i.e., as string [note that negative
' varSyncTime means timeshifting backwards]
Function adjustTime(probTime As Date, varSyncTime As Integer) As Date
  Const one_hour As Date = 1 / 24
  Dim addedTime As Date

  addedTime = probTime + one_hour * varSyncTime / 60
  adjustTime = Format((addedTime), "dd-mmm-yy h:mm") '<<< PI timeformat
End Function


'------------------------------------------------------------------------------------
' converts variant data type into integer by writing and then reading from worksheet
```

```
Private Function VarToInt(inVar As Variant) 'As Integer
     Set calcSheet = ActiveSheet
      calcSheet.Range("IV9000") = (inVar)
      VarToInt = Val(calcSheet.Range("IV9000").Value)
      calcSheet.Range("IV9000").Clear
End Function

' converts variant data type into string by writing and then reading from worksheet
Private Function VarToString(inVar As Variant) As String
   Set calcSheet = ActiveSheet
      calcSheet.Range("IV9000") = (inVar)
      VarToString = calcSheet.Range("IV9000").Value
      calcSheet.Range("IV9000").Clear
End Function
'--------------------------------------------------------------------------------------------
' following gets the first data row number, first hour column number(column number
for
 'which data is to be sorted(e.g., column K for four hour avg. difference))
' and displays top ten variables that have changed in writeHere row (starting in
column B)

Sub GetTopTags(keyPulpInfoCell As Object, hrColnNum As Integer, writeHere As Integer)
    Dim numVar As Integer
    Dim array1() ' Tags
    Dim array2() 'tag descriptors
    Dim array3() As Long '% change in variable
    Dim array4() 'time when sample was there
    Dim array5() 'value of variable when our problem sample was there << based on array4
    Dim array6() 'avg. four hour before
    Dim array7() 'avg. four hour after
    Dim array8() 'tag units
    Dim Key_cell As Object, CurrCell As Object, FirstCell As Object

    keyCellRowNum = keyPulpInfoCell.Row
    Set mySheet = Sheets("Pulp tracking")
    Set FirstCell = mySheet.Range("cc1000")
    Set Key_cell = mySheet.Cells(keyCellRowNum, 1)

    mySheet.Select
   ' counts number of observations
    Key_cell.Activate
    anchor_cell = ActiveCell.Address
    ActiveCell.End(xlDown).Select
```

```
    Range(anchor_cell, ActiveCell).Select
    numVar = Application.Count(Selection)

    ReDim array1(1 To numVar)
    ReDim array2(1 To numVar)
    ReDim array3(1 To numVar)
    ReDim array4(1 To numVar)
    ReDim array5(1 To numVar)
    ReDim array6(1 To numVar)
    ReDim array7(1 To numVar)
    ReDim array8(1 To numVar)

  'get range from worksheet to the VBA array
   For i = 1 To numVar 'First To Last
     Set CurrCell = Key_cell.Offset(i - 1, hrColnNum - 1)
     If i > 32 Then

    End If
      If Not Application.IsNonText(CurrCell) Or Not
Application.IsNonText(CurrCell.Offset(0, 1)) Then  'Or (CurrCell.Value = 0 And
CurrCell.Offset(0, 1).Value = 0) Then 'Or Not IsNumeric(CurrCell.Value) Then ' = 0
Or TypeName(CurrCell) Then
        array3(i) = 0
      ElseIf (CurrCell.Value = 0) Then
        array3(i) = 0
      Else  ' calculates absolute percentage change
     array3(i) = 100 * Abs((CurrCell.Value - CurrCell.Offset(0, 1).Value) / CurrCell.Value)
     End If

      array1(i) = Key_cell.Offset(i - 1, 1).Value ' Tags
      array2(i) = Key_cell.Offset(i - 1, 2).Value 'tag descriptors
      array4(i) = Key_cell.Offset(i - 1, 4).Value  'time when sample was there
      array5(i) = Key_cell.Offset(i - 1, 5).Value 'value of variable when our problem
sample was there << based on array4
      array6(i) = Key_cell.Offset(i - 1, hrColnNum - 1).Value 'avg. four hour before
      array7(i) = Key_cell.Offset(i - 1, hrColnNum).Value 'avg. four hour after
      array8(i) = Key_cell.Offset(i - 1, 3).Value 'tag units
  Next i

   First = LBound(array3)
   Last = UBound(array3)

'  Transfer array to worksheet
```

```
  For i = First To Last
     FirstCell.Offset(i - 1, 1).Value = array1(i)
     FirstCell.Offset(i - 1, 2).Value = array2(i)
     FirstCell.Offset(i - 1, 3).Value = array3(i)
     FirstCell.Offset(i - 1, 4).Value = array4(i)
     FirstCell.Offset(i - 1, 5).Value = array5(i)
     FirstCell.Offset(i - 1, 6).Value = array6(i)
     FirstCell.Offset(i - 1, 7).Value = array7(i)
     FirstCell.Offset(i - 1, 8).Value = array8(i)
  Next i

'  Sort the worksheet range
   FirstCell.CurrentRegion.Sort Key1:=FirstCell.Offset(0, 3), Order1:=xlDescending,
Orientation:=xlTopToBottom


'  Present top 10 values
  Set newCell = mySheet.Cells(writeHere, 1) 'write top ten variables that have
changed
   For i = 1 To 10
     newCell.Offset(0, 1).Value = FirstCell.Offset(i - 1, 1).Value
     newCell.Offset(0, 2).Value = FirstCell.Offset(i - 1, 2).Value
     newCell.Offset(0, 3).Value = FirstCell.Offset(i - 1, 3).Value
     newCell.Offset(0, 4).Value = FirstCell.Offset(i - 1, 4).Value
     newCell.Offset(0, 5).Value = FirstCell.Offset(i - 1, 5).Value
     newCell.Offset(0, 6).Value = FirstCell.Offset(i - 1, 6).Value
     newCell.Offset(0, 7).Value = FirstCell.Offset(i - 1, 7).Value
     newCell.Offset(0, 8).Value = FirstCell.Offset(i - 1, 8).Value

    Set newCell = newCell.Offset(1, 0) ' go to next row
   Next i

   ' clears the worksheet range that contains sorted data
    Range(FirstCell.Address, FirstCell.Offset(numVar + 1, 9)).Clear
 Range("c36").Select
End Sub

Sub ShowComboBox()
Dim DBox As Object
Dim ComboList As Object

   Set DBox = ThisWorkbook.DialogSheets("ComboDlg")
   Set ComboList = DBox.DropDowns("ComboList")
```

```
' Clear the list
  If ComboList.ListCount <> 0 Then ComboList.RemoveItem Index:=1,
Count:=ComboList.ListCount

' Fill the list
  ComboList.AddItem Text:="1 Hour"
  ComboList.AddItem Text:="4 Hour"
  ComboList.AddItem Text:="12 Hour"
  ComboList.AddItem Text:="24 Hour"
  ComboList.AddItem Text:="7 days"
  ComboList.AddItem Text:="20 days"

' Display the dialog
  DBoxOK = DBox.Show

' If not canceled, show the selection
  Number = ComboList.ListIndex  ' Initialize variable.
  Select Case Number  ' Evaluate Number.
    Case 1 ' 1 hour difff.
      hrColnNum = 8
      Mystring = "1 hr. "
    Case 2 ' 4 hour difff.
      hrColnNum = 11
      Mystring = "4 hr. "
    Case 3 ' 12 hour difff.
      hrColnNum = 14
      Mystring = "12 hr. "
    Case 4 '24 hour difff.
      hrColnNum = 17
      Mystring = "24 hr. "
    Case 5 ' 7 day difff.
      hrColnNum = 20
      Mystring = "7 day "
    Case 6 '20 day difff.
      hrColnNum = 23
      Mystring = "20 day "
    Case Else   ' Other values.
       Mystring = " "
  End Select
   MsgBox Mystring & " column : " & hrColnNum
   Set mySheet = Sheets("Pulp tracking")
  mySheet.Range("G35") = Mystring + " before"
```

```
    mySheet.Range("H35") = Mystring + " after"
    mySheet.Range("G49") = Mystring + " before"
    mySheet.Range("H49") = Mystring + " after"

    mySheet.Range("G65") = Mystring + " before"
    mySheet.Range("H65") = Mystring + " after"

    mySheet.Range("G79") = Mystring + " before"
    mySheet.Range("H79") = Mystring + " after"

    mySheet.Range("G95") = Mystring + " before"
    mySheet.Range("H95") = Mystring + " after"

    mySheet.Range("G109") = Mystring + " before"
    mySheet.Range("H109") = Mystring + " after"
End Sub
```

# VITA

*SAKET KUMAR*

University of Washington, Ph.D., 1999

University of Roorkee, B.E., 1994