CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter contains two main parts. The first presents the theoretical background for the work. The second provides a review of the literature related to previous approaches to correlate the characteristics of the paper production process with resulting outputs, such as the quality of the product. These previous initiatives have also addressed to some degree the issue of developing predictive models of the pulp and paper production process.

2.1 Background:

One of the key objective of environmental management for paper industry is to minimize the resource consumption, pollution load, and environmental risks. Typically, one of two approaches would be emphasized. The first and classical environmental approach would be to treat and control the pollutants formed before they are discharged into the environment (so-called end of the pipe treatment). The second approach, now receiving considerable attention involves improvement of the production processes and systems in order to achieve increased efficiency. This source reduction approach to environmental improvement can also involve the substitution of alternative materials in order to provide environmental improvement. Benefit of these approaches depends on the availability of correct information. Detailed understanding of the relationships and associations between the various types of inputs to the process and their impacts on the characteristics of the streams (both product and non-product) that emerge from the process is necessary in order to be able to predict improvements in output parameters based upon changes in input conditions.

There are many approaches, most based on mathematical relationships that have been used in the past, to uncover non-obvious associations and cause and effect relationships among diversified data sets. One such approach holds promise, based on previous research [3, 4, 5, and 6] of being applicable to the paper input/output relation problem and is explored in this research program. That method is called the multivariate technique and consists of sequential application of factor analysis (FA) and multiple regression analysis (MRA). It has the potential to identify, quantify, and predict the relationship among process and effluent parameters. If this tool is applied properly, the information obtained from the solution is expected to be beneficial in the decision-making regarding production changes for environmental improvement in the paper industry management.

2.1.1 Theory and Philosophy of Factor Analysis (FA):

Factor Analysis (FA) is a data reduction technique. It is a mathematical procedure for removing redundancy from a set of correlated variables and representing the variables with a smaller set of derived variables, or factors (Figure 2.1). The basic function of FA is to summarize data with a minimum loss of information and to identify relationships among the variables. More details of FA theory can be found in references 17-21.

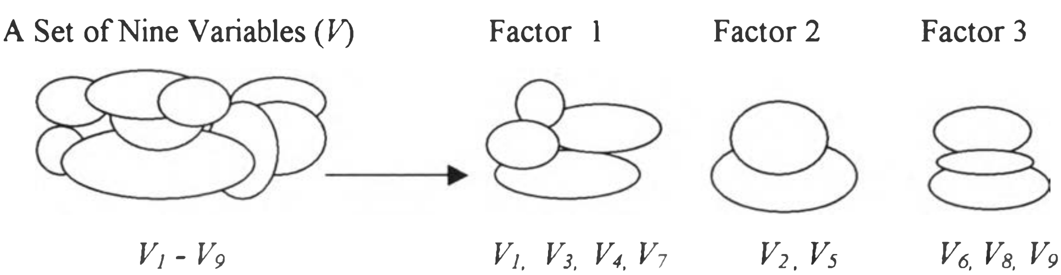


Figure 2.1 A Set of Variables Reduced into Factors

The operating mechanics of FA achieves the smaller number of factors through recognition and rearrangement of the pattern of interrelationship among variables in a dataset-by-dataset transformation via matrix algebra.

The original data for n measurements on p characters can be displayed as a data matrix containing all observations or cases for all variables as follows.

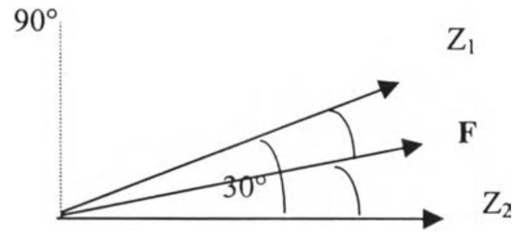
Observation (Case)	Value of Variable			
1	V_{11}	V_{21}	...	V_{p1}
2	V_{12}	V_{22}	...	V_{p2}
:	:	:	...	:
:	:	:	...	:
n	V_{1n}	V_{2n}	...	V_{pn}

This data matrix is then transformed to a correlation matrix. The transformation involves calculation of a standardized variable (Z) for each set of observed variables followed by determination of the covariance between pairs of the standardized variables (Z). This covariance is called the correlation coefficient (r). This approach is described in more detail in Chapter 3. The Correlation Matrix can take the form shown below.

"Correlation Matrix"

Standardized Variable	Z_1	Z_2	...	Z_n
Z_1	r_{11}	r_{12}	...	r_{1n}
Z_2	r_{12}	r_{22}	...	:
:	:	:	...	:
:	:	:	...	:
:	:	:	...	:
Z_n	r_{n1}		...	r_{nn}

The standardized variables (Z_n) can be thought of as vectors. Sequentially, pairs of these variable-vectors are correlated by comparison of the cosine of the angle between them. Those pairs with high correlation can be associated with a new vector drawn to minimize the cosine between the new vector and each member of the variable vector pair. This new vector is called a "Factor" (F) and represents the characteristics of the two variables that are incorporated into it. Later, if other variable vectors are determined to lie in nearby vector spaces (as determined by the size of the cosines) they may be associated into the factor as well. This approach is also discussed in more detail in Chapter 3.



The clusters of these vectors (both the standardized variables and the factors) in the factor space of orthogonal factors provide a way to determine the patterns of variation and association in the matrix. The establishment of the factors is a most important key in the use of FA for this purpose. In reality, there are several dimensions called *n*-dimensions that must be the same as number of factors. However, factors are generally expressed in 2 dimensions. As indicated previously, these factors reflect the pattern of variation of the standardized variable-vectors. Therefore, the factors can be interpreted as the pattern of interrelationships among the variables that make up the data (Figure 2.2).

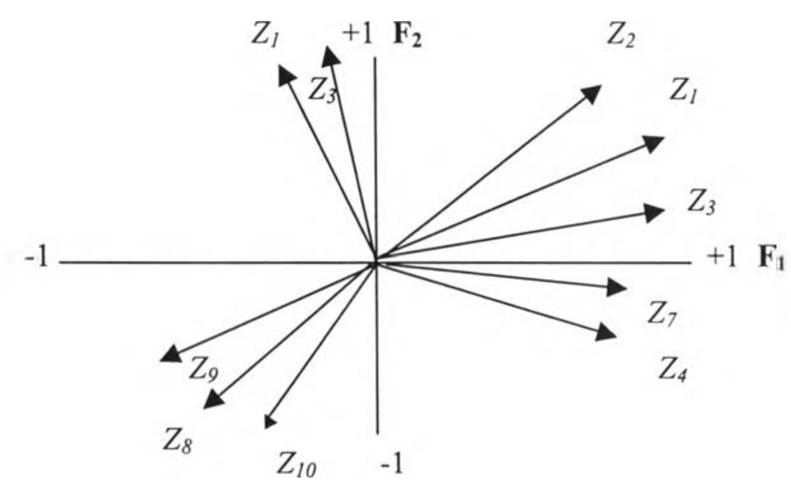


Figure 2.2 Plot Diagram of Ten Variables on the Two Factors

Figure 2.2 displays the plot for the 10 standardized variables on the two orthogonal factors that are unrelated. The smaller the acute angle between any two vectors (or the smaller the cosines), the more highly correlated are those variables (*Z*). The projections of each variable (*Z*) on the factor-axes are called factor loadings (*l*). The "factor loadings" are the correlations between the variables and the factors. As illustrated in Figure 2.2,the first factor F_1 can be defined as an inter-correlated pattern of relationships between variables, Z_1, Z_2, Z_3, Z_4, and Z_7. The second factor F_2 can be defined as an interrelationship between variables, Z_1 and Z_3. It should be noted that variables can be associated with more than one factor. In addition, the other variables Z_8, Z_9, and Z_{10} in this example are not included in the first two factors. This means that subsequent consideration of the matrix is likely to identify additional factors that lie in a different dimension beyond the 2-dimensions that were used to determine F_1 and F_2. This is a normal expectation when using FA.

Through completion of the FA procedure, a factor matrix is obtained with the different percentages of total variance as illustrated below. Note that as defined above, the correlation between a factor and a given standardized variable is termed the "Factor Loading" or "l".

"Factor Matrix"

Standardized Variable	Factors			
	F_1	F_2	...	F_n
Z_1	l_{11}	l_{12}	...	l_{1n}
Z_2	l_{12}	l_{22}	...	:
:	:	:	...	:
:	:	:	...	:
:	:	:	...	:
Z_n	l_{n1}	.	...	l_{nn}

"Factor Loadings"

The form of equation that expresses each common factor (F_1, F_2...F_m) in terms of the contribution of each variable (Z_1, Z_2...Z_m) is shown below. It is obtained by multiplying each variable Z_n with score coefficient (b). The score coefficient is a function of "l" and "r" or the standardized variable Z_n. The details of the calculation of "b" are provided in Chapter 3.

$$F_1 = b_{11} Z_1 + b_{12} Z_2 + \dots\dots\dots + b_{1m} Z_m \qquad 2.1$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$F_m = b_{p1} Z_1 + b_{p2} Z_{12} + \dots\dots\dots + b_{1m} Z_m$$

The form of equation that expresses each variable (Z_1, ... Z_m) in terms of the relationship with common factors (F_1,...F_m) is shown below. It is obtained by function of "*l*" and "*r*" or the standardized variable Z_n. The details of the calculation of "*b*" are provided in Chapter 3.

$$F_1 = b_{11} Z_1 + b_{12} Z_2 + \dots\dots\dots + b_{1m} Z_m \qquad 2.1$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$F_m = b_{p1} Z_1 + b_{p2} Z_{12} + \dots\dots\dots + b_{1m} Z_m$$

The form of equation that expresses each variable ($Z_1, \dots Z_m$) in terms of the relationship with common factors ($F_1, \dots F_m$) is shown below. It is obtained by multiplying each factor with the factor loadings (*l*) that represents the association between particular factors and particular standardized variables.

$$Z_1 = l_{11} F_1 + l_{12} F_2 + \dots\dots\dots + l_{1m} F_m \qquad 2.2$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$Z_m = l_{p1} F_1 + l_{p2} F_m + \dots\dots\dots + l_{1m} F_m$$

Based on the development of these relationships, FA can be applied to the following aspects of data analysis and interpretation.

1. Data reduction and transformation: Data are manipulated from the original largest set of variables to a smaller set of uncorrelated variables that can be used for other subsequent techniques, such as regression analysis, with acceptable loss of variation in the original data.
2. Patterns of interrelationship: FA can be used to separate the relationships among each variable to a separate pattern called the factor structure. Moreover, it makes possible the scoring of the factor for each case, and can be then applied to explore the complex interrelationships among the measurements of observed events.

There are two generally accepted approaches for reduction of data of this type. In addition to FA, another approach called Principle Component Analysis (PCA) is sometimes used. Although both FA and PCA have the same goal, there are two major differences between them as described below.

1. In FA, the variables are expressed as linear combinations of the factors, whereas in PCA the variables are expressed as a few linear functions of the variables.

Moreover, FA is often used as a data-analysis tool in conjunction with other methods such as regression analysis. On the other hand, if the goal is to define a smaller number of variables for input into another analysis, PCA is always used. For the situation of this study with large numbers of variables, and the expectation of a complex set of interrelationships, FA seems to be a better choice for data reduction.

A major goal of this study is to explain the complexities of the interrelationships of many variables involved in papermaking. The anticipated approach to accomplish this depends on the identification of factors (or associations of measurable variables), where these factors are not directly observable. This goal confirms the decision to explore the use of FA rather than PCA, because FA will support an eventual better fit with all of the observable data. Furthermore, the factor scores that emerge from the FA model can be used as new independent variables or predictor variables for further analysis by MRA.

With regard to development of a broader management tool, MRA is needed in order to relate the output of interest in this study (wastewater loads) to the material input.

2.1.2 Theory and Philosophy of Multiple Regression Analysis (MRA) :

In decision-making involving predictive thinking, MRA provides means to develop models that show how one variable is influenced by others. These models involve mathematical curve that explains the behavior of the response variable Y under a range of the set of the predictor variables; $x_1, x_2, \ldots\ldots x_p$. In this case, the response variable Y is the representative of one of the output variables, and the predictor variables are representatives of the various members of the set of input variables. MRA also provides the variability of the responses that can be expected for a particular set of the predictor variables: $x_1, x_2, \ldots\ldots x_p$ [20-22].

The aim of MRA is to develop a model that can use information derived from a set of independent input variables or predictor variables to predict the dependent variable or response variable as accurately as possible (Figure 2.3).

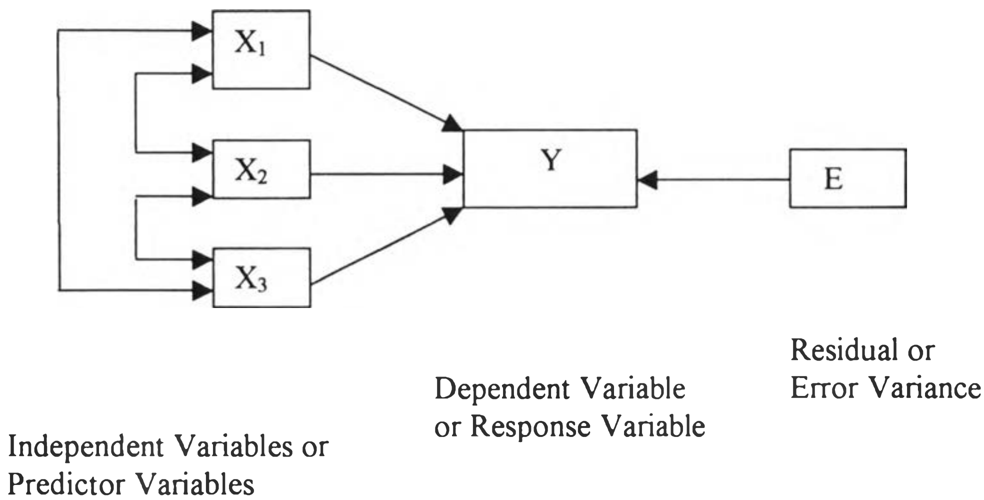


Figure 2.3 A Model of MRA

The more completely MRA can explain the variation in the response variable, the more accurate the prediction will be. However, in practice, there is usually an amount of variance that is unaccounted for by the MRA model. This is termed the residual or the error variance. Further details about MRA can be found in references 13-16 and 22-24.

The concept of MRA is to construct a model that is based on an initial hypothesis of relationships among the variables. The MRA approach continues by testing the closeness of the fit of the predicted response variables with the actual observations and then sequentially modifying the model parameters until the closest fit possible is achieved.

If there are relations between inputs and outputs in the observed data, the MRA model explains those relationships. Furthermore, MRA assesses the accuracy of the description of the relationship and the importance of each of the predictor variables in a relationship as well. Therefore, MRA yields an overall estimate of the variance between the predicted and actual dependent (output) variables (R^2; coefficient of determination), a test of the significance of the variance (F-test), a test of whether each variable is contributing significantly (t-test), and possibly an estimate of each variable's independent effect (β-coefficients), based on three assumptions inherent in MRA: independence, normality and homogeneity of error [23]. The MRA model can include both linear regression (LR) and nonlinear regression (NLR) models.

MRA can be applied to predict values of one or more response variables from a collection of predictor variable values. Further it can be used to assess the effects of changes in the predictor variables on the responses. This last characteristic holds promise of application as a management tool for improvement of environmental outcomes in the papermaking industry.

2.1.3 Design of Work :

This study focuses on an initiative to identify and develop the key understandings necessary to develop an environmental management tool in papermaking that predicts environmental outcomes based upon available process input data. Based on the methods and applicability discussed in the previous section, FA and MRA are used as the mathematical tools to accomplish this goal. Because the papermaking site selected for this study produces two main products, Gypsum liner board (consisting of GF and GB) and Duplex coated board (consisting of DP 450, DP 400, DP 350, DP 310, and DP 270), both FA and MRA models have been constructed for each of the seven sub-products of the industrial paper manufacturing facility. The framework of the components of the whole model for each product is shown below.

The first step of the study is to build an FA model for the material input and utility consumption for each of the product types using a set of input variables for each product. The products of the FA models as significant factors (F_1,.., F_n) are used as the predictor variables to analyze the relationships between each wastewater load parameter: SS, TDS, COD and BOD using MRA for building a predictive model.

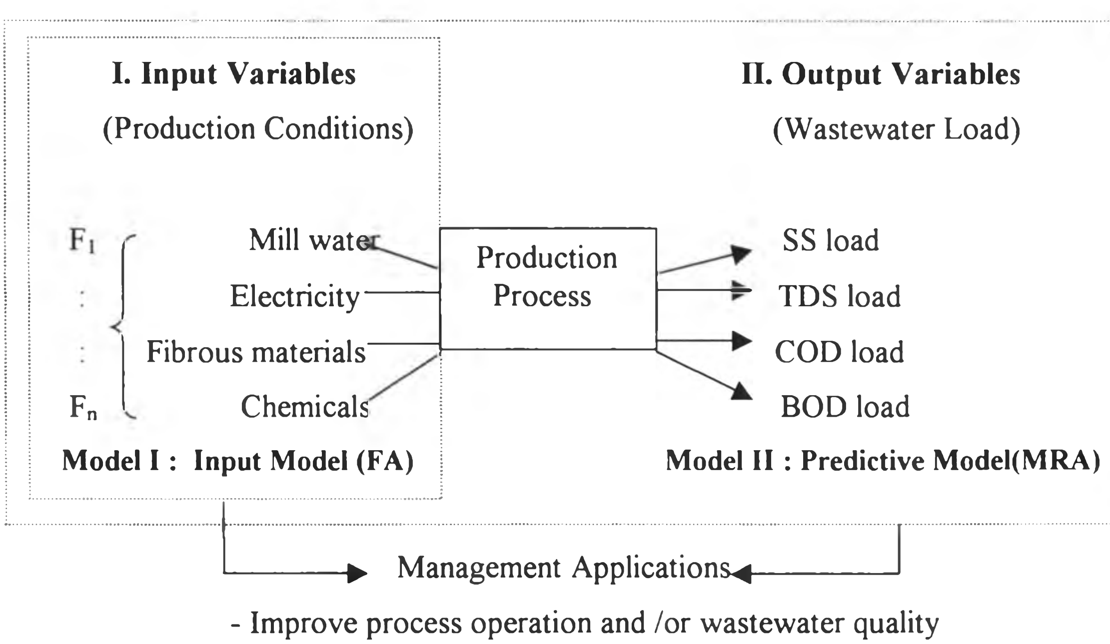


Figure 2.4 Framework of FA and MRA Models for Industrial Papermaking

2.2 Literature Review on Paper production:

In general, the raw materials used in paper manufacturing consist of fibers of various types, water and chemicals, while energy (as electricity and steam) is required throughout [2]. The input fibrous materials may contain virgin fiber and wastepaper or recovered paper obtained from old corrugated containers and old newspaper. Water is used for various applications throughout the process as process water, cooling water and boiler feed water. Chemicals can be classified as product additive materials (such as fillers, sizing and fixing agents, wet strength agents, and coating materials) and as process aids (such as defoaming agents and biocides). Electricity is required for equipment operations, material transportation and ventilation. Steam is used for the heating of water, pulp, air and chemicals to the required processing temperature and in the largest quantity for drying the paper [27]. The details of the papermaking process, the specific industrial paper products from the study site, and the raw materials involved are presented in the following sections.

2.2.1 Process of Paper Production:

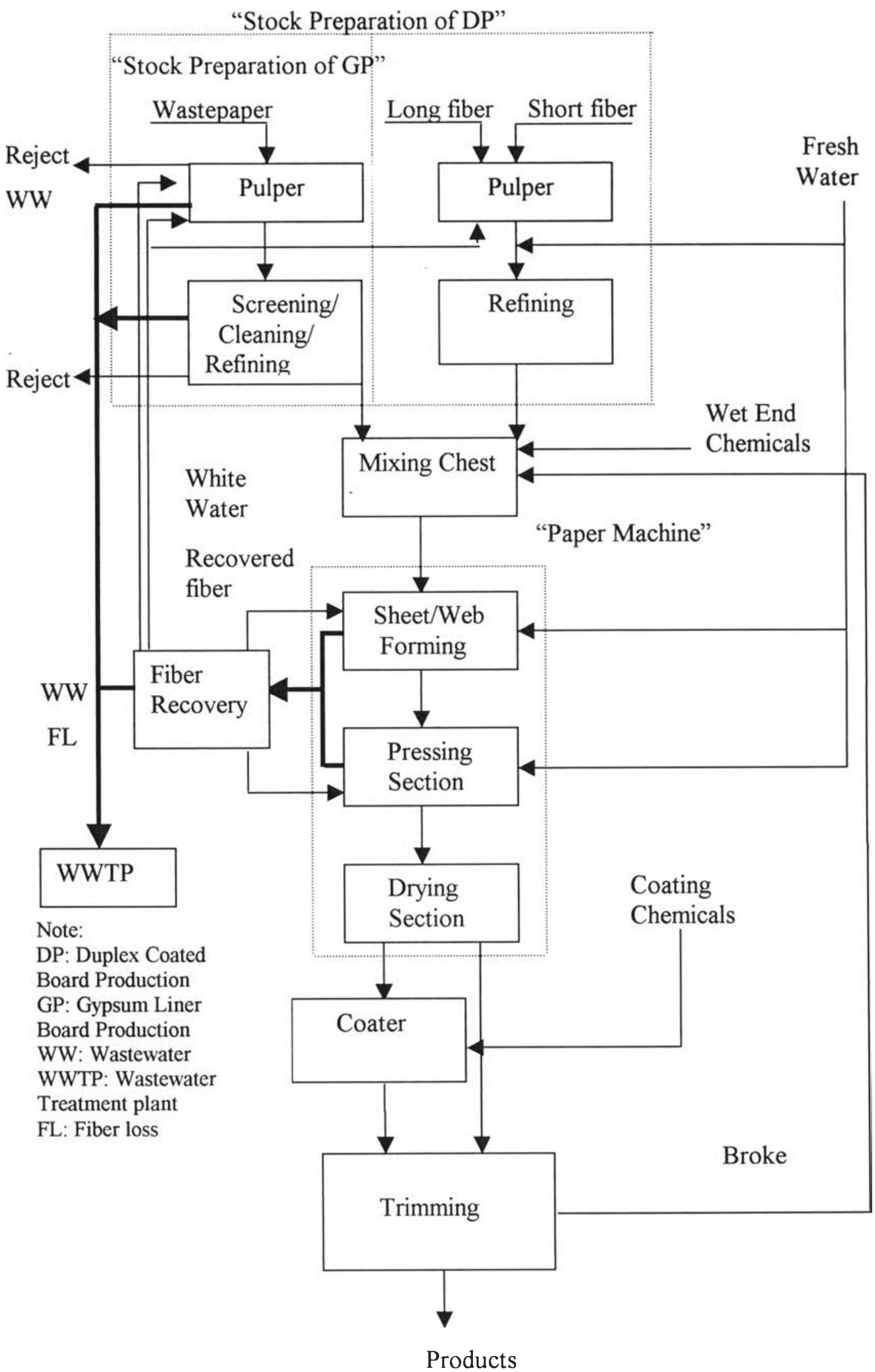
There are three major steps in the papermaking process (Figure 2.5) [1, 28-29].

Step 1: Stock preparation or Material preparation. The fibrous materials are converted into a pulp slurry by re-pulping, cleaning, and refining. Chemicals can be added at the end of this step. At the study site, Duplex coated boards are made from mixtures of both wastepaper and virgin pulp consisting of both long fiber and short fiber. Gypsum linerboards are made from several types of wastepaper.

Step 2: Wet end operations. This step consists of a wire section and a press section. The pulp slurry is delivered to the paper machine and is deposited on a moving wire belt, while wet end chemicals are added for fiber bonding and strength properties. The continuous sheet or web of pulp fibers is then pressed between a series of rollers to remove more water and to compress the web.

Step 3: Dry end operations. In the drying section, the paper fibers begin to bond together upon evaporation of water facilitated by the steam-heated rollers. Coatings can be applied to the paper thereafter in order to improve gloss, color, printing detail and brilliance.

At the study site, industrial paper production is performed by a multilayer (4 ply) paper machine called a Fourdrinier machine.



Source : The production site studied.

Figure 2.5 Flow Diagram of Industrial Paper Production

2.2.2 Product of Industrial Paper:

At the study site, two kinds of industrial paper are produced as described in the following sections [30]:

2.2.2.1 Gypsum linerboard: This product is made entirely from wastepapers consisting of a number of fibrous materials. Gypsum linerboard is used as the facing and backing material in the production of gypsum wallboard for the construction industry. Wall board is used chiefly for interior walls and as a base for plaster.

2.2.2.2 Duplex coated board or folding boxboard: This type of industrial paper is made from a mixture of virgin fiber and fibers from recycled paperboard as well as mixed grades of wastepapers.
The duplex coated board is usually pigment-coated or lined with bleached pulp to enhance its appearance and printing characteristics.

2.2.3 Raw materials:

There are three main kinds of raw materials including the utility consumption [1, 30-31].

2.2.3.1 Wastepaper pulps: These pulps consist of several kinds of fibrous materials from different sources as discussed below [30].

2.2.3.1.1 Old newspaper group: This fibrous material is made from newspapers that have been used in households, and newspapers remaining from sales, including the unprinted newsprint that is white paper. In this study, the material from this group is designated as A_1, A_2, A_9, A_{11}, A_{12}, A_{15}, A_{17}, and A_{18}.

2.2.3.1.2 Old Corrugated Container group: This fibrous material is made from brown boxes that are mostly used in retail stores, factories, and office-buildings. In this study, the materials in this group are designated as A_3, A_4, and A_5 for different sources.

2.2.3.1.3 Mixed Waste (MW): This kind of wastepaper consists of various qualities of paper, predominately obtained from offices and households. In this study, the material from this group is designated as A_{10} and A_{16}.

2.2.3.2 Virgin pulps: These types of pulp consist of both long fiber and short fiber varieties.

2.2.3.2.1 Long fiber: This fiber is made from soft wood such as pine. The length of the fiber is about 3-5 millimeter. In this study, this type of fiber is designated as A_6 and A_{13}.

2.2.3.2.2 Short fiber: This kind of fiber is made from hard wood such as Eucalyptus or bamboo. The length of fiber is about 1- 3 millimeter. In this study, this type of fiber is designated as A_7, A_8, and A_{14}.

2.2.3.3 Chemicals: There are several kinds of chemicals used in the production of industrial paper as described in the following sections [32-35]:

2.2.3.3.1 Wet End Chemicals: These chemicals are used for improving the properties of the pulp slurry before being delivered to paper machine.

2.2 3.3.2 Sizing Agent: This agent (such as emulsion size or emulsifier) is used to control penetration of aqueous liquids into paper, in general water, because cellulose will absorb water from the air and cause the sheet to swell. Thus, sizing agent is needed to improve water resistance of the paper product.

2.2.3.3.3 Alum: This material is used with sizing agent for pH control at 4.5-5.

2.2.3.3.4 Modified Starch: This material is one of the dry strength agents, used to enhance the strength of the individual fibers and of the bonds between the fibers.

2.2.3.3.5 Starch: This material is also one kind of dry strength agent, used to enhance the formation of bonds between the layers.

2.2.3.3.6 Clay: This material is one of the fillers, used to increase the weight, opacity and whiteness of paper, and to reduce the quantity of pulp used.

2.2.3.3.7 Defoamer: This chemical is a processing aid used for eliminating bubbles in pulp slurry.

2.2.3.3.8 Wet Strength Agent: This agent enhances the strength of paper when it is wet.

2.2.3.3.9 Biocide: This material is an anti-biotic agent to control the growth of Microbial organisms, such as slime, that cause contaminants in the paper.

2.2.3.3.10 Dry End Chemicals: These chemicals are coating chemicals used for surface coating and for increasing the strength of the paper's surface and printability, including,

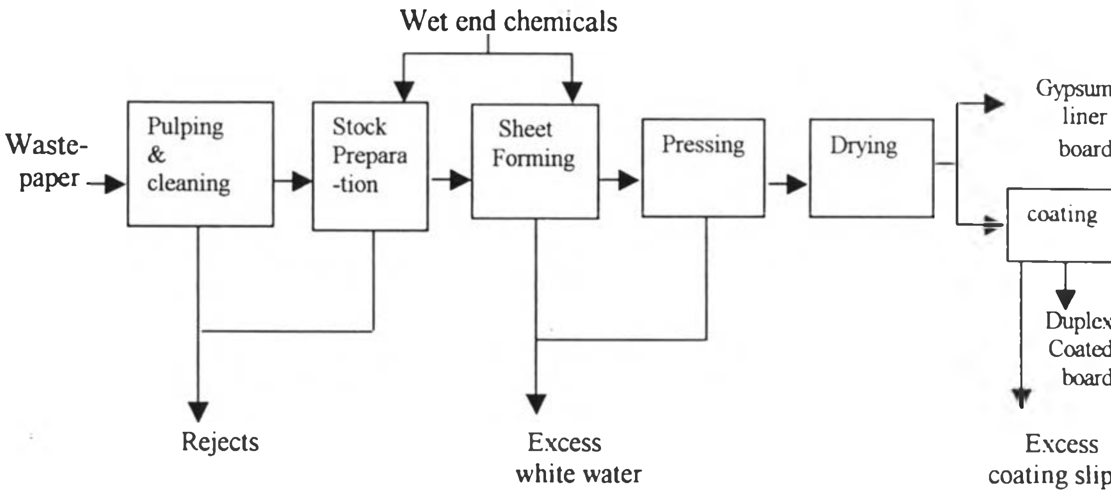
1) pigments; such as calcium carbonate ($CaCO_3$), which is used as a pigment for ink absorption and for increasing the smoothness of the paper's surface, and clay;
2) binder; latex is major binder in coating application;
3) Other chemicals, such as a dispersant, for pigment dispersion.

2.2.4 Utility Consumption: Water and electricity consumption in the production process is considered to be another group of multivariate input data in this study. Steam is not considered as an input variable because it is not related to wastewater generation and is used principally in drying step in paper machine.

2.2.5 Environmental impact from paper production:

Paper production can affect the environment in a number of ways. Pollutants are discharged to water, land, and air and noise is generated from mill operations and transportation of materials in and out the site. However, the major pollutant is the water effluent that varies from mill to mill depending on the degree of whitewater closure, paper grade produced, size of the mill and type of raw material used [35-40]. The discharges consist of dissolved organic and inorganic substances and solid substances such as fibers, fines, and fillers as shown in Figure 2.6 [35-36].

The wastewater generation comes from most steps of papermaking. In stock preparation, part of the effluent that raises concerns about BOD, COD and fine solids is generated by cleaning and refining processes. Some wastewaters with similar characteristics are produced at the wet end operations [38-41]. In the drying section, the water from the web is extracted to the atmosphere by a steam heated drying cylinder. This water constituent is not considered as serious form of environmental pollution. However, the coating operation can cause intermittent water emissions with high BOD, COD and SS loads due to equipment cleaning [42].



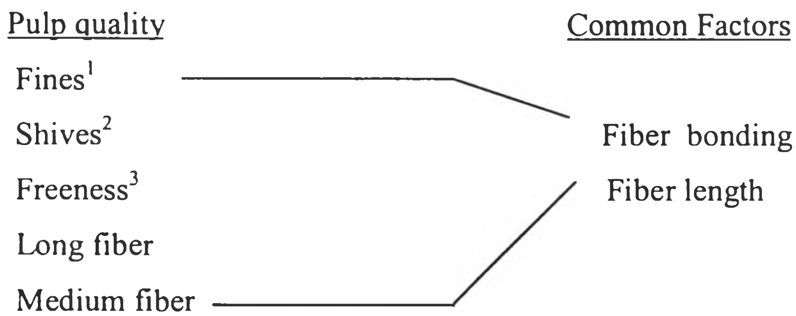
Source : The production site studied.

Figure 2.6 Discharge of Pollutants from a Paper Mill

2.2.6 Modeling of pulp quality and paper quality:

Factor analysis and factor network analysis (consisting of FA and partial least square) for modeling of pulp and paper qualities has been carried out before by Strand, et al. during 1989 - 1998 [3]. The previous work consisted of efforts to related process input variables to characteristics of the paper products. The research did not examine any relationships with environmental outputs as wastewater loads. Examples of their work are discussed below.

In 1989, Strand, Makvist, and Jackson were the first to use factor analysis in analyzing mechanical pulp quality variation in Scandinavian newsprint mills. The result showed that the large number of pulp quality tests can be cut down to a smaller number of "common factors" (Figure 2.5). These factors represented the complex fiber phenomena; fiber bonding and fiber length that ultimately determine pulp quality [3].



Note: 1 and 2 are pulp quality relating to characteristics of suspended solids and 3 is the ability of pulp in water absorption.

Figure 2.7 Common Factor Diagram of Pulp Quality

In 1990, the factor network analysis method, which had been applied successfully in the Scandinavian industry to pulp quality, was applied to paper quality analysis in North America [3]. The work has shown that 15-25 paper quality tests can be reduced to 7 independent "common factors". These factors represented all of the important properties of paper quality. The result from this analysis is called a factor network and led to the development of factor network analysis of paper production, used particularly for newsprint mills, in 1993 [3].

In 1993, Saltin and Strand studied the modeling of the effects of refiner operation on newsprint quality [4]. The result showed that paper quality was determined by seven underlying phenomena: fiber binding, fiber length, fiber number, unbounded surface area, fiber orientation, fiber pressing and light absorption, as shown in Figure 2.8 [4].

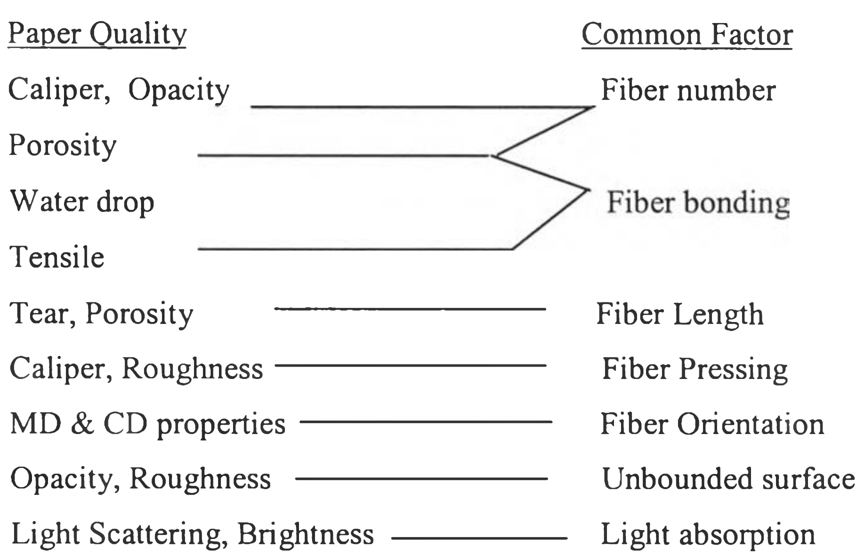


Figure 2.8 Common Factor Diagram of Paper Quality

In 1997, Strand presented an optimization of furnish (a multicomponent mixture, containing several types of pulp, filler and additives) blending for newsprint quality using factor network analysis to build a model for the data over a 6 month-period [6]. This study indicated that factor network analysis could show the interrelationships among the input variables (pulp quality, type of pulp and blends), paper machine operating condition (speed), and the output variables (paper quality). The presence of four common factors could explain 65% of the variation in the variables as shown in Figure 2.9 [6].

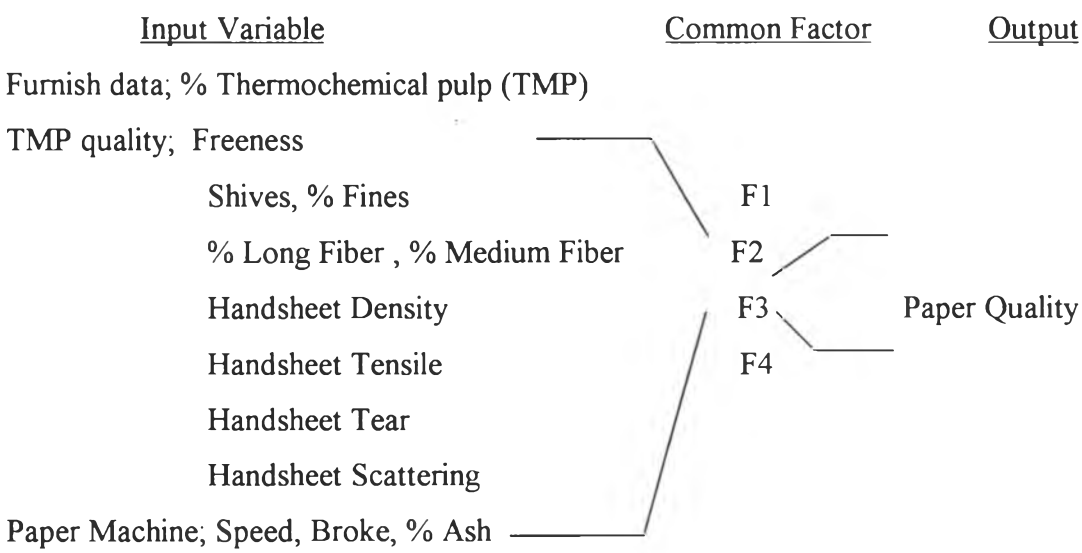


Figure 2.9 Common Factor Diagram of Furnish Blending

In 1998–1999, Strand, et al. studied quality control of mainline refiners at the Pondery Newsprint Co. [5]. They found that the variables used in the control system, namely manipulated variables (such as backflow and infeed dilution), limit variables (such as motor load, specific energy, blowline consistency and dilution flow ratio), and control variables (such as freeness and long fiber content) are related to energy control. Moreover, quality control has reduced the amount of off-spec pulp by over 70% as well as reducing specific energy consumption (kWhr/adt of paper) by 6% [5].

2.2.7 Application of FA and MRA in Other Works :

The earliest work reported in the application of factor analysis was done before 1904 by the psychologist Charles Spearman, who is considered to be the "father" of the method. Spearman's purpose was to provide a mathematical model to represent his theory of general intelligence. According to his theory, "all branches of intellectual activity have in common one fundamental function (or group of functions) whereas the remaining or specific elements of the particular branch of activity seem in every case to be wholly different from those in all other branches (Spearman, 1904) [8]. Spearman's theory of general intelligence and his interest in mathematical models led to the development of his two-factor method.

Much of the early work in FA during the period from 1900 to 1930 was devoted to the application of Spearman's model to a great variety of applied problems and to investigations of conditions under which the model was appropriate. During this period, the general factor model was a single model that was not always adequate to describe the relationships between variables in a set of data.

The multiple-factor theory was developed by Thurstone in 1931, 1938, and 1947. Thurstone was the most prominent of the early modern factor analysis theorists. He had considerable influence on the development of the method from the 1930s to the present. He was responsible for several methods, particularly the concept of simple structure that has been considered as representing an ideal factor-analytic solution.

The early work in FA carried out by the various psychologists tended to be theory oriented, the statistical tests of specific hypotheses concerning the factorial structures of particular sets of variables were not available.

In the mid to late 1950s, the theoretical orientation moved toward what has come to be called exploratory factor analysis. This was encouraged by Thurstone's common-factor theory and principal components, that had not been generally applied before because of the extreme computational labor required.

In the late 1950s and 1960s, FA was used to analyze the apparently complex relationships between variables in a set of data in order to simplify and to interpret the complex relationships between variables. In this period, FA was applied to identify the underlying factor in other fields, education, business and science particularly in chemical analysis, including environmental analysis on a macro scale such as investigations of sources of air pollutant. However, in the case of micro scale applications as in a process operation related to waste generation, there are no reported applications.

In applications where predictions of outcomes are needed, the factor score, which is the result of the FA model, has been used as a predictor variable in some situations, using MRA.

For example, in legal cases involving application of the 1964 Civil Rights Act in the United States, which prohibits discrimination in the workplace based on sex or race, MRA-based predictions have been considered. For women, blacks, and other groups, the courts have ruled that statistics can be used as *prima facie* evidence of discrimination. The complex statistical analyses based on MRA predictions attempt to demonstrate that similarly qualified individuals have not been treated equally [22]. In addition, this mathematical and statistical model has been developed for other applications, such as the relationship between beginning salary and salary progression to various employee characteristics such as seniority, education, and previous work experience. Through this example, one objective is to determine whether sex and race are important predictors of salary. The MRA technique used to build the model is a linear regression analysis, one of the most versatile data analysis procedures. Regression can be used to summarize data as well as to study the relationships among variables.