CHAPTER 4

SITE, DATA COLLECTION AND PRE-ANALYSIS

This chapter contains three major parts; discussion about surveying production sites to identify the one that would provide the most applicable data, collection of the data and pre-analysis analysis of the data for two main product type categories of industrial paper manufacturing, specifically, gypsum liner board and duplex coated board.

4.1 Site Surveying:

In order to find the most appropriate site to be studied, various characteristics of process and product were considered at three possible sites. The optimum site for study would be one that produced a variety of products, one that was relatively uncomplicated in terms of numbers of production lines, and one that had available detailed records of production inputs and outputs.

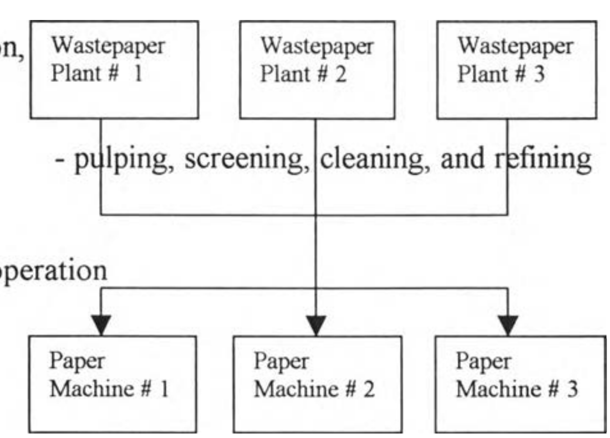
4.1.1 Characteristic of Process Operation:

Three industrial papermaking sites (consisting of site A, site B, and site C) were surveyed. They exhibit markedly different process configurations within their own operations as below.

Site A :

Step I. Stock preparation,

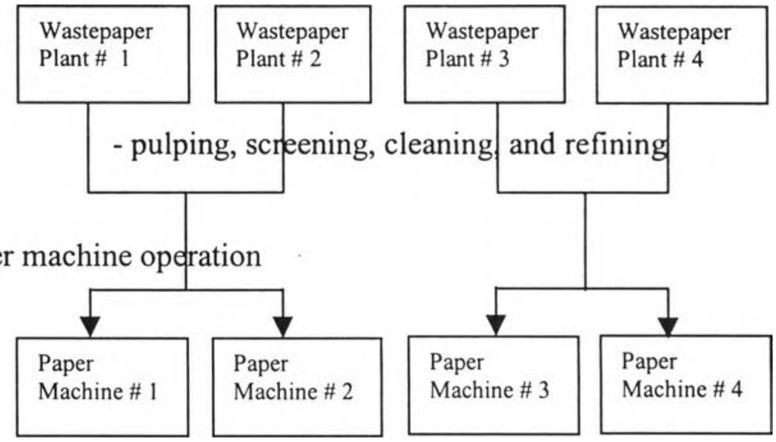
Step II. Paper machine operation



- sheet forming, pressing and drying

Site B :

Step I. Stock preparation,

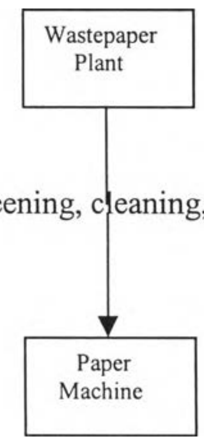


Step II. Paper machine operation

- sheet forming, pressing and drying

Site C :

Step I. Stock preparation,



- pulping, screening, cleaning, and refining

Step II. Paper machine operation

- sheet forming, pressing and drying

At the first two sites, papermaking is performed by several paper machines that are connected into the same system as a combined operation. At the last site, industrial paper is made by one paper machine operating as a single system. With regard to the values of the input and output variables, the process operation from a single system provides accurate data for industrial paper production without complications from alternative product lines being operated at the same time. The same operations from a combined system could provide valid data, but there would be uncertainty about overlap with outputs from other product lines in assessing the relationships with wastewater loadings for example. Moreover, the input and output data from the production process was available at the last site. Based upon the likelihood of the quality of the production data as well as the readily available data, the most appropriate site for this study was deemed to be the production site of the last paper mill. Although the site studied is based upon a single system, it is possible to apply the same modeling approach to a combined system for other sites. If the input and output through several paper machines are considered and can be measured the same as a single paper machine. In particular, the records of data for both input and output variables are also available and are of good quality.

At the site studied, there is a single paper machine that produces multi-grade papers generating two main product types. The first type of product is Gypsum liner board, which is produced in two forms: gypsum face liner (GF) and gypsum back liner (GB). The other product type is duplex coated board (DP) that is produced in a range of five different basis weight categories: DP 450, DP 400, DP 350, DP 310, and DP 270. Although there are several grades of the products, the input and output variable data resulting from their production can be accurately measured in the course of the daily operation. The sequence of production typically starts from the machine shut down condition (sd) → DP → GB → GF, however, occasionally this sequence starts from sd → DP → GF and back again. The duration of time for producing each of the individual products is different depending on demand. During the period of the study, typical production runs were about 15 days for gypsum liner board and 5 days for duplex coated board.

4.1.2 Characteristics of Products:

Although both gypsum liner board and the duplex coated board products are produced on the same multi-layer paper machine called a Fourdrinier machine, the number of layers of paper they contain and their eventual applications are different. Gypsum liner board consisting of three layers of paper is used in the production of gypsum wall board which has eventual application in construction as interior walls. Duplex coated board consists of four layers of paper and is used as the material in packaging, such as white coated boxes for dry goods, or for poster board. Both types of products are made from several types of wastepaper. However, DP also contains some kinds of virgin pulp in order to provide the quality desired in the final product. While DP production includes a coating operation in order to improve the smoothness of the surface for printability, GF and GB do not require coating. Therefore, the production of DP utilizes the coater stage of the paper machine during production, while GF and GB production by-passes the coater. This illustrates the versatility of the paper machine and indicates why there may be differences in the package of outputs from the processes for the different types of paper products even though they use the same machinery.

4.2 Data Collection and Pre-analysis:

Data collection and pre-analysis were conducted in two phases at the site studied. The approach was based upon the available data from daily operations of the individual process steps at the facility. The major categories of these steps include the wastepaper plant, the chemical preparation section, the papermaking itself, and the wastewater from these steps. The first phase of the data was designated for model building. The data set covered a period of 14 months (424 days) from January 2001 to February 2002 because of the change of some production conditions after this period. The second phase of the data was designated for model validation. This data set covered a period of 7 months (214 days) from March 2002 to September 2002. The data sets included data for production of both Gypsum liner board and Duplex coated board. The details of data collection and pre-analysis are discussed in the following sections.

4.2.1 Data Used for Model Building:

The raw process data for both input and output variables were obtained from each of the process steps of the industrial papermaking for all products in a 14 month period of the first phase of data collection (Table 4.1).

From Table 4.1, it can be seen that the total quantity of data obtained from the industrial papermaking operations during the 14 months represent 424 unique data sets or cases that consist of 327 complete data, and 97 incomplete data. Each case represents the data collected for a single day's production. A complete data means that all input and output variables are obtained for the same operation day. An incomplete data means that some variables such as wastewater load parameters or paper grade input information were not measured and identified. Most often this is because the paper machine was shut down and there were no analysis results of wastewater quality. This may have occurred because of holidays, weekends, or equipment repair. The complete data can be further classified into valid data and inappropriate data. Inappropriate data sets refer to data that was collected on days when there was a transition between one type of product and another. On such days, the input data and the output data in terms of wastewater loadings cannot be directly assigned to one product or another. Therefore, it was not appropriate to use it for this analysis. Only the valid data sets, which represent 79.8 % of the complete data sets were used for further analysis.

Table 4.1 The Number of Input Data Collected and Used

Product	Data Collected (case)	Complete Data		Valid Data	
		Number of case	Number of variable	Number of case	Number of variable
1. GF	70	54	17	40	13
2. GB	82	65	12	39	11
3. DP	257	208	30	192	16
Total	409*	327	59	261	40
% Total Valid Data of Complete Data				79.8	

Note : * Total data included paper machine shut down = 424 cases

The details of the data collected show that the production of all basis weights of duplex coated board (DP) (208 cases) represent more days of production than those existed for production of gypsum liner board (consisting of both GF and GB altogether) 119 cases or days of production during the period studied (Table 4.1). The percentage of the data that was used for data analysis and modeling is highest for DP and lowest for GB. This is because the production of DP is usually higher than those of GF and GB. Thus, proportion of incomplete and inappropriate data of DP is lowest, and highest for GB. Note that for DP the data collected for the different basis weight product varieties is not treated separately because the components of the inputs and outputs for the subproducts are similar. The sequence of DP subcategories in the production line is not matched with the progression of basis weight numbers. For example, it typically starts from DP 400 → DP 450 → DP 350 → DP 310 and then back to DP 400 or DP 350 → DP 400 → DP 450 → DP 310 → DP 270 → GB/GF. During the period studied, the production days of DP 450, DP 400, DP 350, DP 310, and DP 270 are 45, 36, 35, 59 and 17, respectively. This data indicates that DP 310 is most heavily produced, while DP 270 is least produced. However, there may be some data variability that could affect data analysis and modeling due to the changeover between the different basis weights.

The next step of data qualification involved consideration of the variables in order to determine if there were enough observations for the individual types of variables to make valid conclusion. For this purpose, a reasonable assumption based upon the accuracy and availability of data is that if the data for a variable appears in at least 70% of the observations then the conclusions will be valid, and therefore the variable is a valid variable. If the variable appears in less than 70% of the observations, it will provide invalid data and is therefore an invalid variable. (The issue of validity or invalidity applies, of course, only to the particular set of data collected. For other data collections, the determination of validity or invalidity could be different from the determinations made in this study.) From Table 4.2, it can be seen that after the application of the variable validity rule, the input variables are reduced from 17 variables to 13 variables for GF, from 12 variables to 11 variables for GB, and from 30 variables to 16 variables for DP.

The details for the input variables are shown in Table 4.2. For GF, four variables: A_{10}, A_5, A_{11} and A_{12} have been removed because they are invalid. For GB, only one variable: A_{10} was removed. For DP, fourteen variables: A_2, A_3, A_{10}, A_{12}, A_{13}, A_{14}, A_{15}, A_{16}, A_{17}, A_{18} (types of wastepaper), defoamer, fibran (a type of emulsifier), $CaCO_3$, and biocide are removed. The percentage of the appearances of the invalid variables that have been removed is shown in Table 4.3. If these variables had been used for building the FA model, there probably would have been no effect on solution of the model in terms of the number of significant common factors. However, their low frequency of appearance in the data could cause failure of FA model construction since the beginning. That is the primary reason for their removal before the data analysis and modeling. Note that the unit used for the input variables is in production base (per ton).

As for the output variables, the data collected shows four output variables for each of the types of paper produced. These four consist of the wastewater loads, specifically the SS, TDS, COD and BOD loads (Table 4.4). The data collected for these output variables were organized together with the input data collected on the operational day. They were put into the same cases. This means that a direct correlation could be made between the inputs into the process on a given day and the outputs, in terms of wastewater quality on the same day. The loads of wastewater were calculated by multiplying the concentration of the wastewater quality parameters by the flow of the water through the mill. The result was converted into the same unit of production base as used for the input variables. That is output variables are in units of (kg of specific wastewater load/ton of product).

Table 4.2 Total Input Variables of Industrial Paper from the Data Set Used for Building Model I

Variable	GF		GB		DP	
	Data Collected	Data Used	Data Collected	Data Used	Data Collected	Data Used
1. Water	x	x	x	x	x	x
2. Electricity	x	x	x	x	x	x
3. A_1	x	x			x	x
4. A_2	x	x	x	x		
5. A_3	x	x	x	x	x	-
6. A_4	x	x	x	x	x	x
7. Alum	x	x	x	x	x	x
8. Clay	x	x			x	x
9. Defoamer	x	x	x	x	x	-
10. Emulsifier	x	x	x	x	x	x
11. cato	x	x	x	x	x	x
12. starch	x	x	x	x	x	x
13. wet strength	x	x	x	x		
14. A_{10}	x	-	x	-	x	-
15. A_5	x	-	x	x	x	x
16. A_{11}	x	-			x	-
17. A_{12}	x	-			x	-
18. A_6					x	x
19. A_7					x	x
20. A_8					x	x
21. A_9					x	x
22. A_{13}					x	-
23. A_{14}					x	-
24. A_{15}					x	-
25. A_{16}					x	-
26. A_{17}					x	-
27. A_{18}					x	-
28. Fibran					x	-
29. $CaCO_3$					x	-
30. biocide					x	-
31. color					x	x
32. latex					x	x
33. other					x	x
Total	17	13	12	11	30	16

Note: x indicates the presence of the variable in each product, - indicates an invalid variable in each product.

Table 4.3 Percentage of Appearance of Removed Variables in Each Product Type

Input Variable	GF	GB	DP
1. A_{10}	2.5	38.5	25.5
2. A_5	5		
3. A_{11}	5		
4. A_{12}	27.5		3.7
5. A_2			9.4
6. A_3			1.0
7. defoamer			1.6
8. A_{13}			7.8
9. A_{14}			27.6
10. A_{15}			17.2
11. A_{16}			7.3
12. A_{17}			9.4
13. A_{18}			19.3
14. Fibran			24.5
15. $CaCO_3$			19.8
16. biocide			1.6

Table 4.4 Output Variables of Industrial Paper for Building Model II

Output Variables (kg/ton)	GF		GB		DP	
	Data Collected	Data Used	Data Collected	Data Used	Data Collected	Data Used
1. SS load	x	x	x	x	x	x
2. TDS load	x	x	x	x	x	x
3. COD load	x	x	x	x	x	x
4. BOD load	x	x	x	x	x	x
Total Number	54	40	65	39	208	192

4.2.2 Data Used for Model Validation:

The raw process data obtained from the industrial paper production on the same paper machine in a subsequent 7-month period was used for validation of the model (Table 4.6).

From Table 4.6, it can be seen that, through the same procedure used previously for data pre-analysis for model building, the percentages of the input variables that qualify as valid data are about 63% for GF, 50% for GB and 87% for DP. These valid data represent 67.68% of the complete data obtained for model validation. Based on the valid data, DP was still produced at the highest level. The numbers of days of production for DP 450, DP 400, DP 350, DP 310, and DP 270 are 11, 9, 9, 34, and 17, respectively. DP 310 was produced at the highest level.

As for output variables, the data for these variables were assembled in the same cases as are the input variables for the same production day.

Table 4.5 Total Numbers of Input Data for Model Validation

Product	Data Collected (case)	Number of Complete Data (case)	Number of Valid Data (case)
1. GF	48	38	24
2. GB	46	34	17
3. DP	118	92	80
Total	212	164	114
% Total Valid Data of Complete Data			67.8

Note : Total data included paper machine shut down = 214 cases

4.3 Characteristics of Original Data Matrix:

The data obtained from the data collection and the pre-analysis contain two matrices of original data for both the input and output variables for gypsum liner board. Similarly, there are five matrices from the duplex coated board data. These matrices are described more completely in the following sections.

4.3.1 Characteristics of Input Variables:

4.3.1.1 Gypsum liner board

From Table 4.7, the data profile obtained from the two original input data matrices contain 40 and 39 separate observations of the daily operation during the 14-month interval using for model building of GF and GB. The data showed the characteristics of the material input and utility consumption for gypsum liner board production (consisting of GF and GB) during that period. It was found that for GF, of the 13 variables of input used, the fibrous material, A_3 was consumed at the highest level (1450.98 kg/ton) and that the data showed a large amount of variability (SD = 2277.08). The defoamer was consumed at the lowest level (0.86 kg/ton) and this data showed only a small amount of variability (SD = 1.03) (see Table 4.7).

For GF, of the 11 variables of input used, the fibrous material, A_4 was consumed at the highest level (1197.88 kg/ton) with a medium level of data variability (SD = 154.85). The defoamer again was consumed at the lowest level (0.66 kg/ton) and shows a low data variability (SD). The data profiles of GF and GB were not substantially different, most of the input used was consumed within similar ranges.

The major difference between GF and GB is that the quality of wastepapers used (A_3 - A_5) for GB is better than the quality of some wastepapers ($A_1 - A_2$) for GF due to the difference source and use of wastepapers. Moreover, clay was not used for substitution some fibrous materials for GB production. This is because GB, the product used as back side of wall paper needs higher strength than GF that is used as front side of wall paper.

Table 4. 6 Characteristics of Input Variables for Gypsum Liner Board

Material Input Variables	GF N = 40		GB N = 39	
	Mean	SD	Mean	SD
1. Water (m^3/ton)	25	24.9	19.8	8
2. Electricity (kWhr/ton)	554.4	35.8	483.5	77.4
3. A_1 (kg/ton)	124	222.8	-	-
4. A_2 (kg/ton)	242	340.8	-	-
5. A_3 (kg/ton)	1451	2277.1	238.1	196.9
6. A_4 (kg/ton)	462.2	484.4	1197.9	154.9
7. Alum (kg/ton)	79.5	68.4	75.2	19.1
8. Clay (kg/ton)	26.1	24.8	-	-
9.Defoamer (kg/ton)	0.9	1.1	0.7	0.8
10. Emulsifier (kg/ton)	30.2	25.6	33	11
11. Cato (kg/ton)	31.5	68.7	12.5	2.5
12 Starch (kg/ton)	48.8	105.9	37.9	48.3
13. Wet strength (kg/ton)	5.6	6.7	3.2	1.8
14. A_5 (kg/ton)	-	-	252.1	99.1

4.3.1.2 Duplex coated board

As seen in the five original data matrices for the various types of DP (Table 4.7), there were 16 variables of material input and utility consumption for each of the basis weight varieties of DP. The data profiles of the input variables for all of the DPs did not differ greatly. Most of the input used was in similar ranges among the different varieties. An exception is that the use of color for DP 310 and DP 270 was higher than the analogous usage for DP 450, DP 400, and DP 350.

In addition, it is noted that the standard deviation (SD) of some variable such as alum is higher than the mean value. Because the production process is in the step of lowering and changing the use of alum for pH control. Also, the SD of some variables: color, latex, and other are higher than their mean values because sometimes there were out-of-specification of products that did not need coating application. Thus, in these cases, these coating chemicals were not used.

Table 4.7 Characteristics of Input Variables of Duplex Coated Board

Material Input Variables	DP 450		DP 400		DP 350		DP 310		DP 270	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1 Water (m^3/ton)	19.89	12.29	21 94	14.58	15.43	5.05	14.32	3.91	17.18	8.14
2.Electricity (kWhr/ton)	532.6	110.3	542.41	117.82	486.4	90.1	468.8	75.16	495.9	72.74
3. A_6 (kg/ton)	41.56	25.85	51.71	31.32	47.85	21.32	38.83	13.53	52.72	30.73
4. A_7 (kg/ton)	71.33	45.16	90.28	54.64	99.43	31.62	105.4	27.12	114.7	17.00
5. A_8 (kg/ton)	67.56	79.06	105.56	134.96	46.57	60.53	75.76	72.21	57.06	78.48
6. A_9 (kg/ton)	213.8	184.6	196.39	189.00	203.9	184.5	159.2	171.6	196.5	156.6
7. A_1 (kg/ton)	251.8	202.4	284.17	251.51	267.1	185.7	248.5	181	341.8	139.3
8. A_5 (kg/ton)	31.06	36.12	51.61	67.48	44.4	48.65	33.79	49.37	34.12	41.99
9.Clay (kg/ton)	13.50	12.79	15.86	12.17	13.93	12.94	12.31	10.67	19.06	11.28
10.Emulsifier (kg/ton)	1.94	2.80	14.42	71.48	1.79	2.92	1.93	2.13	4.04	2.59
11.Cato (kg/ton)	5.53	3.36	6.34	3.11	5.59	3.07	4.88	2.00	6.01	1.98
12.Starch (kg/ton)	20.91	13.19	18.92	10.23	17.32	10.72	15.83	8.90	15.40	7.28
13.Color (kg/ton)	69.65	22.25	82.75	38.26	87.18	16.08	102.4	104.5	94.20	20.43
14.Latex (kg/ton)	16.34	5.38	19.68	9.30	20.75	3.88	23.90	24.78	22.16	4.99
15.Other (kg.ton)	23.43	36.54	27.35	54.43	29.28	45.64	56.32	201.9	17.81	6.56
16.Alum (kg/ton)	4.17	8.10	6.83	13.58	3.57	5.78	2.85	4.45	4.22	6.66

4.3.2 Characteristics of Output Variables:

4.3.2.1 Wastewater characteristics for all industrial papers

The wastewater characteristics from the industrial papermaking were obtained from two major step of production process: 1) Stock preparation (Wastepaper plant) and 2) Paper machine.

The quality of wastewater from these step is shown in Table 4.8.

Table 4.8 Wastewater Characteristics of All Industrial Papers

Parameter	Stock Preparation	Paper Machine
1. Volume (m^3/d)	4,303	2769.9
2. pH	7	7.3
3. SS (ppm)	3,721.6	2,344
4. TDS (ppm)	1,431.9	917.3
5. COD (ppm)	3,871.7	3,222.5
6. BOD (ppm)	2,744.3	2,129.9

From Table 4.8, it was found that during the 14 month-interval, the total average of wastewater quality obtained from the step of stock preparation using for material preparation was poorer quality than the wastewater from paper machine due to the dirtiness of wastepaper that needed the cleaning in the step of stock preparation before passing through the paper machine.

4.3.2.2 Wastewater load of gypsum liner board

From Table 4.8, can be seen that based upon the same sample size of input variables in the same data matrix, the wastewater load from GB was slightly higher than the wastewater load from GF. The COD load was higher than the BOD load for both GF and GB indicating that the wastepaper components contain inorganic chemical rather than organic chemical (fibrous material). The SS load was higher than TDS for both of the products due to some fiber losses.

Table 4.9 Wastewater Load from Gypsum Liner Board Production

Wastewater Load (kg/ton)	GF N = 40		GB N = 39	
	Mean	SD	Mean	SD
1. SS	53.8	61.6	53.9	39.6
2. TDS	35.4	30.3	37.5	22.1
3. COD	78.2	79.1	79.3	41.6
4. BOD	53.9	54.2	55.4	29.6

4.3.2.3 Wastewater load of duplex coated board

From Tables 4.10, it can be seen that based on the same sample size of input variables in the same data matrices, the wastewater load for all the varieties of DP showed the same pattern of wastewater loading as did GF and GB. The COD load was highest, and the TDS load was lowest. The wastewater load for DP 270 was lowest for SS, COD and BOD since less fibrous material and chemicals were contain therein. The COD and BOD loads for DP 450 were the highest, and the SS and TDS loads for DP 400 were the highest due to their more components of fibrous material. However, the highest loads for DP 400 were close to the levels shown for DP 450.

Table 4.10 Wastewater Load of DP Production

Waste Water Load (kg/ton)	DP 450 N = 45		DP 400 N = 36		DP 350 N = 45		DP 310 N = 36		DP 270 N = 17	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. SS	51.8	23	52.6	35.5	44.7	18.6	51.8	30	42.2	12.8
2.TDS	15	7.8	17.7	13.8	13.9	13.8	12	5.6	14.4	11.8
3.COD	59.2	24	56.7	28.9	48.9	22.7	56.2	28.9	48.7	11.0
4.BOD	40.8	17.1	39.2	20.6	33.9	16.1	39.1	20.7	33.2	7.9