Linking process variables and newsprint properties in Mazandaran Wood and paper Industries

Mohammad Hadi Moradian[1*], **Hossein Resalati**[2]

[1]Behbahan Khatam Alanbi University of Technology, Natural Resources Department, Behbahan, Iran,
[2] Gorgan Agricultural Science and Natural Resources, Department of Pulp and Paper Technology, Gogan, Iran
[*]Corresponding author: e-mail: moradian@bkatu.ac.ir

Pulp and paper industries have provided great research opportunities to control systems. The objective of this study was to investigate the relationships between 80 process variables of CMP tower and stock preparation, and 17 newsprint quality properties in Mazandaran Wood and Paper Industries (MWPI). After the preparation of two suitable data series considering the time needed for pulp to paper, the relations between process dependent and newsprint independent variables were determined using partial least squares (PLS) regression. As a result, two PLS models were developed. The first model with 4 latent vectors categorized and related CMP tower variables and the second one, through 8 latent vectors connected stock preparation variables with paper properties. PLS regression coefficients determined how much the most influencing process variables impact each paper properties.

Keywords: Newsprint, Statistical model, PLS regression, Process variables.

INTRODUCTION

Pulp and paper quality control is considered either directly or indirectly using different methods. In direct control, experimental or online data is observed by the operator to understand or guess the cause of undesirable changes of the pulp and paper properties. On the other hand, using different kinds of mathematical, statistical, neural networks and such models, is better and increasingly under development in industries. To develop statistical models, different methods and experimental designs are needed. The experiments may be very expensive and time consuming, while using historical data can be more acceptable, even the experiments may often be unnecessary[1]. The concept of statistical data mining is an overall term for using various, mainly multivariate, statistical methods and techniques for exploratory data analysis, developed to handle large data sets with many and often highly correlated variables[2–5]. Some important multivariate data mining methods used in pulp and paper researches are; principal component analysis (PCA), factor analysis (FA), partial least squares (PLS) regression, and multiple linear regression (MLR)[6–9].

PLS is a common soft modeling approach in industrial applications. PLS regression is particularly useful when we need to predict a set of dependent variables from a very large set of independent variables. The goal of PLS regression is to predict responses (Y) from regressors (X) and to describe their common structure. When Y is a vector and X is full rank, this goal could be accomplished using ordinary multiple regression. When the number of predictors is large compared to the number of observations, X is likely to be singular and the regression approach is no longer feasible because of multicollinearity. Several approaches have been developed to cope with this problem. One approach is to eliminate some predictors using stepwise methods. Another one, called principal component regression, that performs a principal component analysis (PCA) of the X matrix and then use the principal components (i.e., eigenvectors) of X as regressors on Y. By contrast, PLS regression finds components from X that are also relevant for Y. Specifically, PLS regression searches for a set of components (called latent vectors) that performs a simultaneous decomposition of X and Y with the constraint that these components explain as much as possible of the covariance between X and Y. This step generalizes PCA. It is followed by a regression step where the decomposition of X is used to predict Y[2].

Broderick et al. (1995, 1996) have comprehensively elaborated and used the PLS method. They used PLS regression to characterize the quality of a series of sulphonated refiner pulps. They successfully grouped fiber characteristics into latent vectors which were statistically related to an opposing set of latent vectors describing standard pulp handsheet properties. Also pulp fibers from 22 pilot-scale runs were examined by them for many physical and chemical characteristics including specific surface, flexibility, Klason lignin, and pentosan index. More than 75% of the variations in handsheet properties were explained by developed models[10, 11].

Grage (2004) used some statistical multivariate analysis to study on the SCA paper mill in Munksund, Sweden. He investigated the influence of paper machine, and wet end process variables on linerboard quality variables produced in that mill[12]. Nordstrom et al. (2005) continued that study but the emphasis had been on modeling the laboratory collected paper quality variables by means of the data available on-line, during production. The primary tool for prediction at non observed time instants, of variables was partial least squares (PLS). Also they obtained the relative importance of the variables used for estimation[13].

A mill-wide multiblock partial least squares model (PLS) was developed by Ortiz-Cordova et al. (2006), to identify the key parameters causing variations in MD tensile strength at an integrated TMP newsprint mill. Also the role of each processing stage and each intermediate product was characterized and the actions required to reduce the tensile strength variability were identified by them. In that study, 80 process variables were investigated in four different groups and in result, fiber length in pulp was found as the most important factor affecting tensile in machine direction, provided that the freeness remains constant. Thus, in this respect, the furnish composition would have an important effect on the tensile strength. Generally, the variations in MD tensile strength, were determined between 15 to 50 percent by different PLS models[14].

Mazandaran pulp and paper industry is the largest paper manufacturer in Iran and the largest wood-based paper manufacturer in the Middle East. It produces 175 000 tons of different papers per year. Newsprint production line with CMP pulping process suffers from some fluctuations in paper qualities. In order to recognize the most influencing variables on paper qualities and generate predictive models to enhance quality control, online and offline data were used in this study. Online data is the data stored by automated machinery in the production line, while offline data measured from laboratory experiments on sampled materials taken from the production line. CMP tower and stock preparation variables were considered including 80 process variables and nearly 7000 observations. Several data sets for different stages were generated considering the residence time needed in each stage. Determining the connections between independent process variables and the final paper and generating predictive models among large amount of data is very important. Therefore the objective of this research includes identification of the most important variables for quality control, with respect to the 17 paper quality variables in MWPI.

EXPERIMENTAL WORK

After preparing about 7000 of online and offline observations of 80 process variables (including CMP tower and stock preparation parts) and 17 newsprint quality properties, two data sets were prepared according to the residence time of the pulp in each stage of the process to be paper, through synchronization. Then partial least square (PLS) regression is used to determine the relationships between process variables and paper properties. PLS is a recent technique that predicts or analyses a set of dependent variables from a set of independent variables or predictors by extracting a set of orthogonal factors or latent variables. This method is particularly useful when we have multidimensional and collinear data from complex systems[2] as in paper industries[9]. The general underlying model of PLS is:

$$X = TP^T + E$$

$$Y = UQ^T + E$$

where: X is an n×m matrix of predictors, Y is an n×p matrix of responses; T and U are n×l matrices that are, respectively, projections of X (the *X score*, *component* or *factor* matrix) and projections of Y (the *Y scores*). P and Q are, respectively, m×l and p×l orthogonal *loading* matrices; and matrices E and F are the error terms, assumed to be independent and identically distributed random normal variables. The decompositions of X and Y are made so as to maximize the covariance of T and U.

Prior to performing the PLS regression, the frequency distributions of the variables were examined for derivation from normality, and then data centering and scaling were done. The predictors and responses were centered and scaled to have mean 0 and standard deviation 1. Centering ensures that the criterion for choosing successive factors is based on how much variation they explain, in either the predictors or the responses or both. Scaling places all predictors and responses on an equal footing relative to their variation in the data.

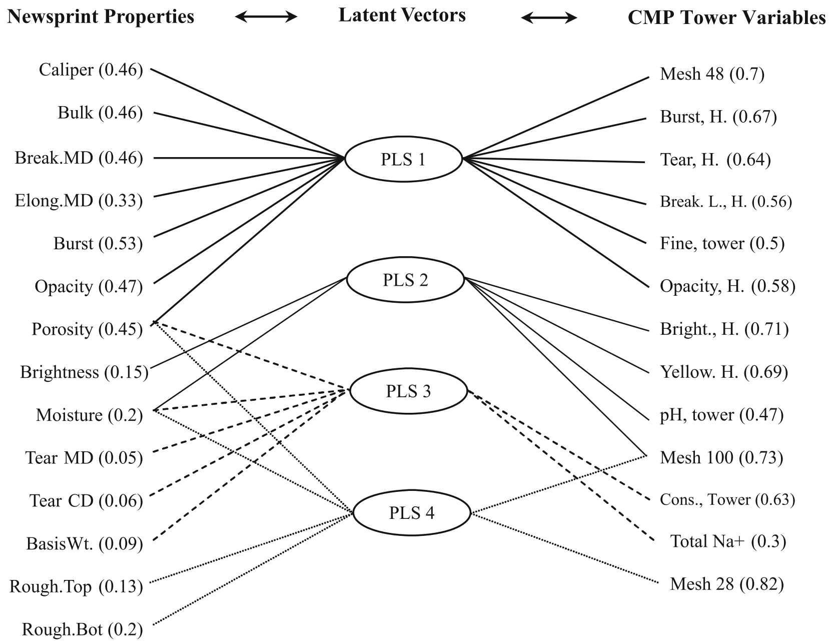
The quality of the models depends on the quality of the data, therefore obvious outliers and observations that looked suspicious, produced for example by the failure of a sensor or a machine outage, were eliminated by measuring the Euclidean distances. This allowed for a reduction of the fluctuation in the data. Also model dimension reduction was performed since not all variables are important for understanding the underlying phenomena of interest. To explore which predictors can be eliminated from the analysis, we can look at the regression coefficients for the standardized data. Predictions with small coefficients (in absolute value) make a small contribution to the response prediction. Another statistic summarizing method is the *Variable Importance for Projection* (VIP) introduced by Wold (1994) and used in this study. While the regression coefficients represent the importance of each predictor in the prediction of just the response, the VIP represents the value of each predictor in fitting the PLS model for both predictors and responses[15].

In order to choose the number of extracted factors, split-sample cross validation was done. Thus, Van der Voet's (1994) randomization based model comparison test was performed, to test models with different numbers of extracted factors against the model that minimizes the predicted residual sum of squares (PRESS)[16]. All the statistical analysis and models were developed using the SAS software.

RESULTS AND DISCUSSION

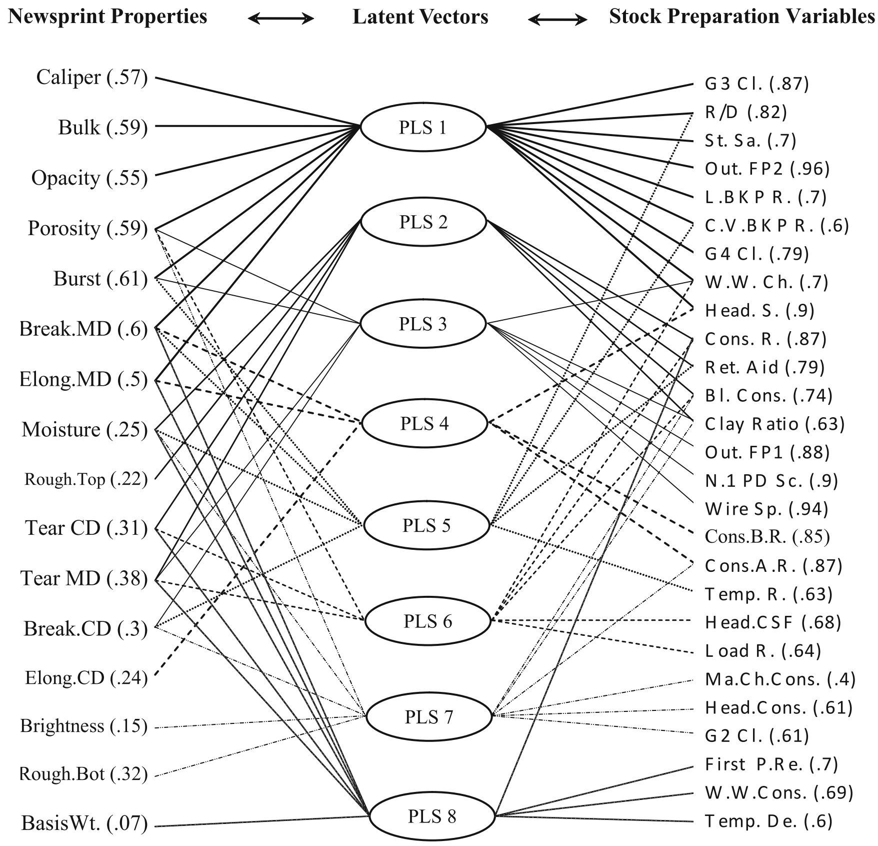
The first PLS model was developed between 16 CMP tower variables as independent and 17 newsprint properties as dependent variables using the first data set including 303 observations (more than 6000 observations were lost through synchronization). To eliminate outliers, the Euclidean distance from each point to the PLS model in both the standardized predictors and responses were measured and since 143^{rd}, 225^{th}, and 290^{th} of independent and 290^{th} of dependent variables had noticeable distances, they were omitted from the model. Then, split-sample cross validation was done for the first 10 factors. Since the predicted residual sum of squares (PRESS) for the fifth factor was the least amount (0.915), selecting five latent vectors results in the least error and more than 5 latent vectors causes over fitting. However, the *t* test analysis showed that the differences between the first, second, and third, factor and the fifth one was 99 percent significant but, the fourth and fifth factors were insignificantly different. Thus, since the model with fewer factors is preferred, four latent vectors would suffice to develop the PLS model. After developing such PLS model, the weights, loadings and regression coefficients were measured.

The network diagram in Figure 1, illustrates the relationships between CMP tower variables and newsprint properties by linking these variables through PLS latent vectors extracted from larger PLS weights. The coefficient of determination next to each variable shows variations explained by the model.



Mesh 100, 48, 28: The amount of fibers remained on the +100, +48 and, +28 mesh screens, Burst, H., Tear, H., Opacity, H., etc: the hand sheet properties made by CMP tower pulp, Cons., Tower: consistency of pulp in CMP tower, Total Na+: total sodium ions in pulp, Break.MD: breaking length in machine direction, Elong.MD: elongation resistance in machine direction, Basis Wt.: basis weight, Rough.Top: the top roughness of paper.

Figure 1. PLS Network between CMP tower variables and paper properties and variance explained through 4 latent vectors



G3 Cl.: stock pressure in the third group cleaners, R/D: rush to drug speed ratio, St. Sa.: stock save-all flow, Out. FP2: output of the second fan pump, L.BKP R.: load of long fiber refiner, C.V.BKP R.: circular valve of long fiber refiner, G4 Cl.: stock pressure in the fourth group cleaners, W.W.Ch.: white water chamber valve, Head. S.: headbox slice opening degree, Cons. R.: consistency of refiner, Ret. Aid: retention aid, Bl. Cons.: blending chest consistency, Clay Ratio: clay ratio, Out. FP1: output of the first fan pump, N.1 PD Sc.: pressure of the first screen, Wire Sp.: wire speed, Cons.B.R.: pulp consistency before refiner, Cons.A.R.: pulp consistency after refiner, Temp. R.: refiner temperature, Head. CSF: headbox freeness, Load R.: refiner load, Ma.Ch.Cons.: consistency of machine chest, Head.Cons.: headbox consistency, G2 Cl.: stock pressure in the second group cleaners, First P.Re.: first pass retention, W.W.Cons.: white water consistency, Temp. De.: temperature of decollators, Break.MD: breaking length in machine direction, Elong.MD: elongation resistance in machine direction, BasisWt.: basis weight, Rough.Top: the top roughness of paper.

Figure 2. PLS Network between stock preparation variables and paper properties and variance explained through 8 latent vectors

The first vector of the model as the most important factor, determined nearly 50 percent changes of 7 newsprint properties including; Caliper, Bulk, Breaking Length MD, Elongation MD, Burst, Opacity and Air Resistance. The amount of fiber remained on the +48 mesh screens including long fiber fractions, was the most influencing variable on the above seven newsprint properties. The fiber length in this range increases the paper strengths and decreases caliper, bulk and opacity of paper. The influence of +48 mesh screens variable is mostly on the network connectivity and strengths improvements of paper. Other important CMP tower variables in this latent vector are influenced by this variable too, since they showed high correlation to it. Controlling this variable is very important to reach the acceptable amount of above seven newsprint properties. Broderick et al. (1995, 1996) studied on modeling the handsheet and fiber properties using PLS method and could relate more than 75% of the variations in handsheet properties to changes of fiber characteristics through five latent vectors. They found +48 mesh screens, affected on the fiber elasticity and increased the paper breaking length, burst and opacity while decreased the paper bulk[10, 11]. The model variation determination for Basis Weight, Break Length CD, Tear MD, Tear CD, and Yellowness were very low. So, it can be concluded that CMP tower variables have little influence on the above newsprint properties.

Second, third and fourth vectors have explained not very much variation, but they connect other input and output variables and showed different dimensions. For instance the influence of pH on the paper brightness and moisture in PLS 2, that also reported by Ortize-Cordova et. al. (2006)**[14]**. In order to determine the influencing variables on the paper properties, and quantitatively define the direct or indirect impacts of them, it's better to look at the regression coefficients other than PLS weights, because of different influences of variables in different vectors especially in the second PLS model. The regression coefficients of the most influencing process variables of the first and second PLS models on each paper properties, are shown in Figure 3.

Using the second data set including 515 observations, the second PLS model could be developed between 56 independent and 17 dependent variables, but to reduce the model dimension, the low impact independent variables were defined and eliminated from the model using *Variable Importance for Projection* (VIP) of Wold (1994)**[15]**. In result, 17 process variables that had very small VIP (less than 0.5) were removed to model 39 remained independent variables with 17 newsprint properties. Ortiz-Cordova et al. (2006) used VIP to determine which process variables would have the strongest influence on some newsprint properties**[14]**. The Euclidean distances from each point to the PLS model were measured and since 166th and 167th observations of dependent variables had noticeable distances, they were omitted as outliers from the model. Then, split-sample cross validation for 25 factors was done and since the predicted residual sum of squares (PRESS) for the 24th factor was the least amount (0.833), selecting 24 latent vectors results in the least error and more than that causes over fitting. However, the *t* test analysis showed that the differences between first to seventh factors with 24th factor were 99 percent significant but, the 8th and 24th factors were insignificantly different. Thus, eight latent vectors were

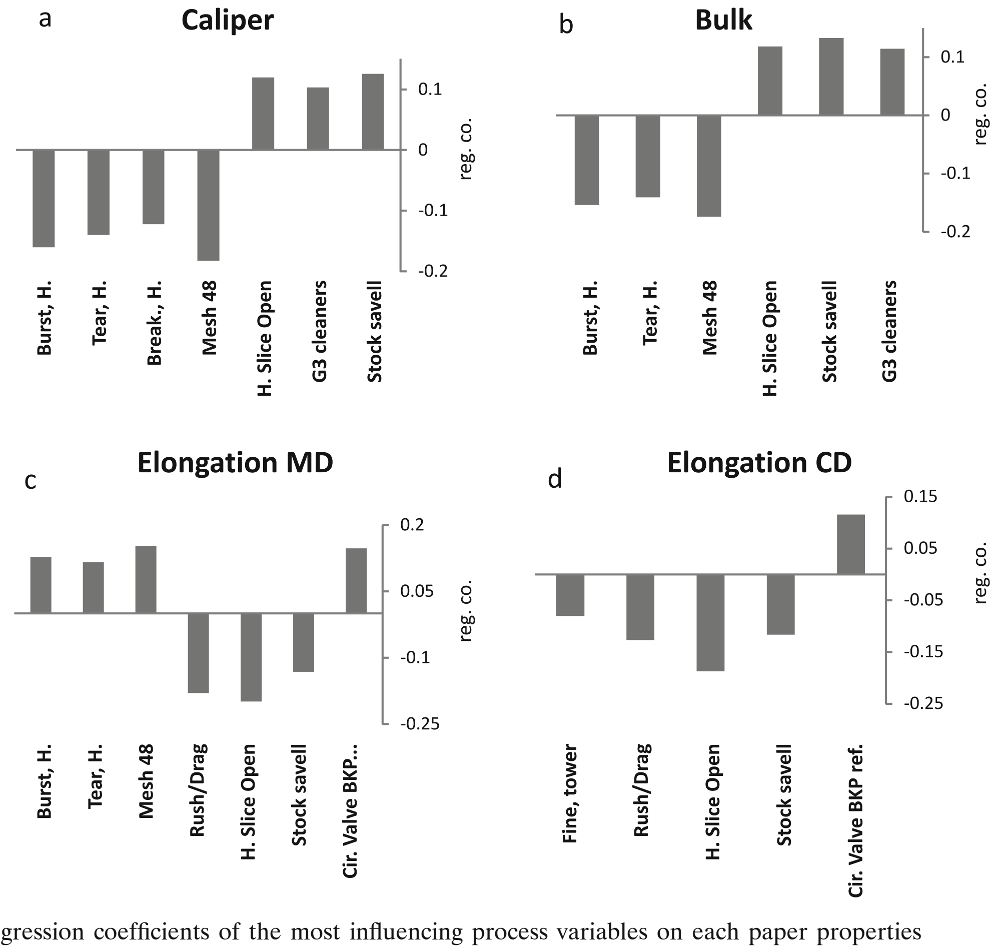


Figure 3. PLS regression coefficients of the most influencing process variables on each paper properties

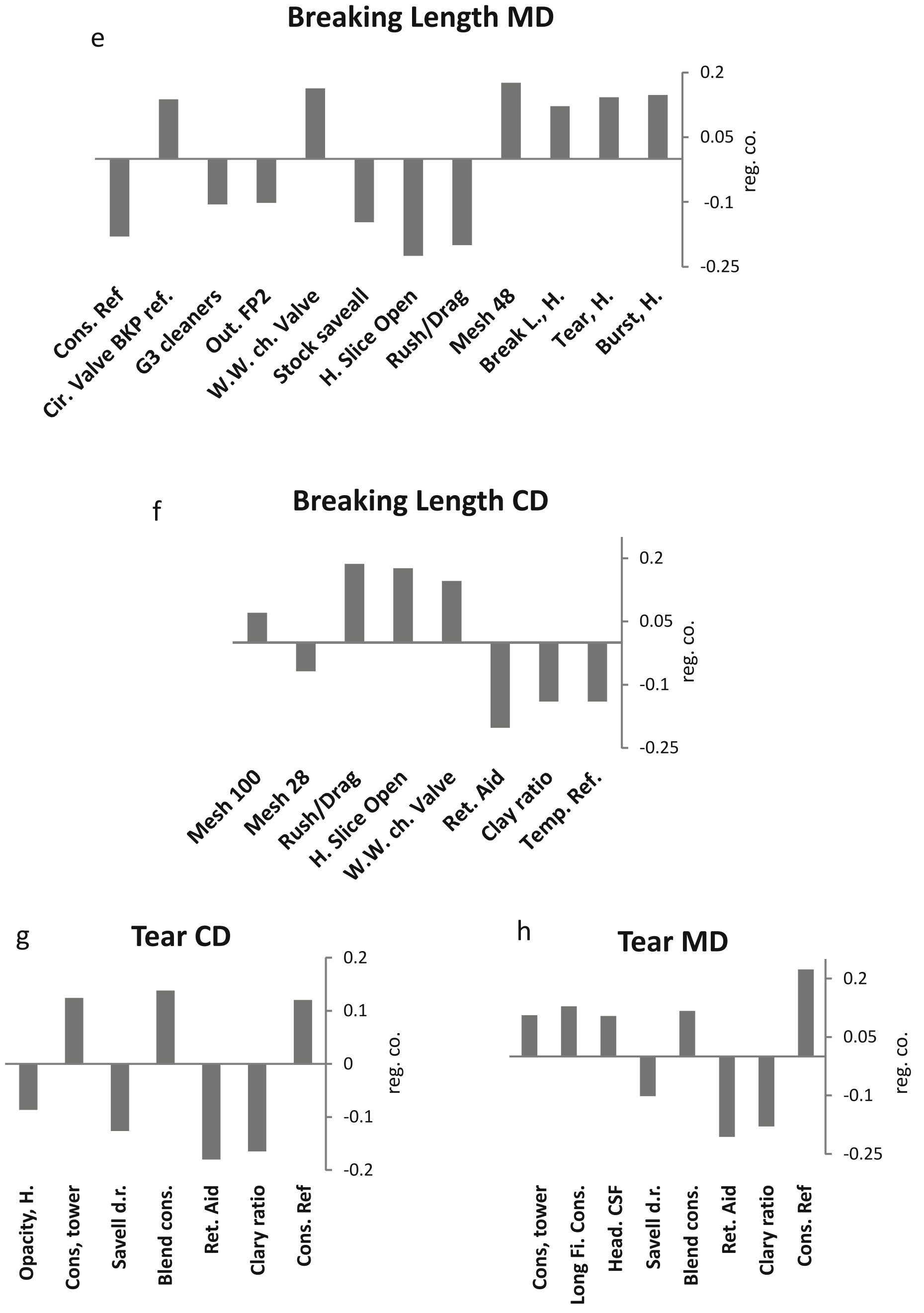


Figure 4. PLS regression coefficients of the most influencing process variables on each paper properties

selected to develop the second PLS model. These 8 vectors accounted for more than 70 percent variations of stock preparation variables and up to 60 percent changes of paper properties. On the other hand the newsprint yellowness and basis weight were very little explained by these 39 process variables. Also the newsprint breaking length and elongation resistance were defined much more in machine direction other than cross direction, just the same as what Nordstrom et al. (2005) reported[13].

Figure 2, illustrates the relationships between stock preparation variables and newsprint properties through 8 latent vectors and figure 3 quantitatively shows the direct or indirect influences of the most important process variables on 11 paper properties through PLS regression coefficients. The first latent vector includes 9 stock preparation variables that influences on 7 correlated newsprint properties that also appeared in the first PLS model. These process variables influence on the paper network connectivity and fortunately they are operational and adjustable. The second latent vector accounted for paper tear resistance (MD and CD) and moisture, explained by process variables as refiner and blending chest consistency, also retention aids and clay percentage. The sixth vector also accounted for tear resistance but they explained by white water percentage and headbox freeness. Retention aids, clay and white water increase small particles and so decrease long fiber ratio, thus they can drop paper tear resistance. Also as shown in figure 3 (part g) headbox freeness had direct impact on the tear resistance because it increases with safe fiber ratio than small fractions. Jones (1993) modeled a corrugated-medium paper machine using integrating novel system of performance attribute models, and found machine or wire speed had minor effect on most paper properties. Here in this research the wire speed had a moderate influence on PLS3 paper quality variables.

In PLS 5, rush to drag ratio has great impact on CD and MD breaking length. Figure 3 (part *e* and *f*) shows direct impact of rush/drag and headbox slice opening degree on the CD breaking length and indirect impact

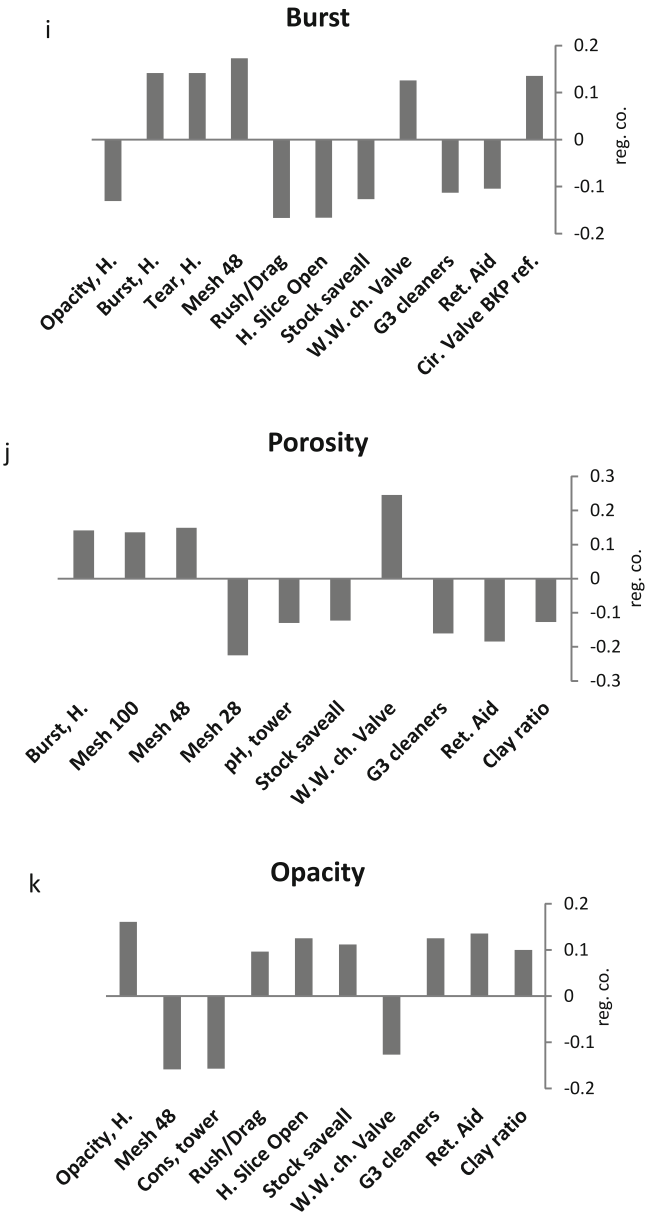


Figure 5. PLS regression coefficients of the most influencing process variables on each paper properties

on MD breaking length. As rush to drag ratio and headbox slice opening degree increase, more fibers can get along the machine direction. This increases the breaking length in cross direction and decreases the breaking length in machine direction. Therefore, using the model and adjusting these variables can optimize the breaking length in machine and cross directions. The newsprint breaking length and elongation are more explained in machine direction (MD) than cross direction (CD) by the model, just the same as Nordstrom et al. (2005) results[13]. Other latent vectors in Figure 2, however, relate different paper properties to the peer process variables and regression coefficients in Figure 3 determine how much the most influencing process variables impact on each paper properties.

The application of the PLS2 model results, for most of the stock preparation influential variables are instant and direct, since they are operational and adjustable. But in using PLS1, although the suitable level of CMP tower variables was defined, more investigation on previous variables especially refiners and furnish composition is necessary to optimize the CMP tower process variables to reach the acceptable newsprint properties in MWPI.

CONCLUSION

This study was aimed at developing PLS models that could connect some process variables to the final newsprint properties in Mazandaran Wood and Paper Industries. In result, two PLS models with 4 and 8 latent vectors were successfully developed from two prepared data sets. Also, the use of PLS model was defined as an appropriate method for identifying correlations between variables in large database of MWPI, as well as recognition of the main sources of variability in the process.

– The first vector of both models as the most important factor, determined more than 50 percent changes of 7 newsprint properties including; Caliper, Bulk, Breaking Length MD, Elongation MD, Burst, Opacity and Air Resistance.

– The amount of fiber remained on the +48 mesh screens including long fiber fractions, was the most influencing CMP tower variable on the above seven newsprint properties. The influence of +48 mesh screens variable is mostly on the network connectivity and strengths improvements of paper. Controlling this variable is very important to reach the acceptable amount of the above seven newsprint properties. Furnish composition and refiner optimization can be very important factors affecting +48 mesh screen fraction that can be the future assessment in MWPI.

– The first model variation determination for Basis Weight, Break Length CD, Tear MD, Tear CD, and Yellowness were very low. So, it can be concluded that CMP tower variables have little influence on the above newsprint properties.

– Stock pressure in the third group cleaners, rush to drag ratio, stock save-all flow, output of second fan pump, headbox slice opening degree, and white water chamber valve were defined as the most important and operational and adjustable stock preparation variables affecting on the mentioned 7 correlated newsprint properties.

– The models facilitated the systematic identification and quantification of the effect of individual process variables on the newsprint properties.

LITERATURE CITED

1. Schweiger, C.A. & Rudd, J.B. (1994). Prediction and control of paper machine using adaptive technologies in process modeling. *TAPPI J.* 77(11), 201–208.

2. Abdi, H. (2007). Partial Least Square Regression (PLS--Regression). Encyclopedia of Measurement and Statistics. Thousand Oaks, USA.

3. Bjorkstrom, A. (2007). Regression methods and their interconnections. Technical report, Stockholm University, Sweden.

4. Farshadfar, E. (2007). *Basis and Methods of Multivariate Statistics* (2end ed.). Taghbostan Press, Razi university.

5. Fridén, H. & Tano, K. (2005). Using PLS models with both controlled and uncontrolled X variables for" Waht if..." prediction. In The 9th Scandinavian Symposium on Chemometrics, Reykjavik, Iceland 2005-09-30. Ornsköldsvik: NPI.

6. Suwannarangsee, S., Bunterngsook, B., Arnthong, J., Paemanee, A., Thamchaipenet, A., Eurwilaichitr, L. & Champreda, V. (2012). Optimisation of synergistic biomass-degrading enzyme systems for efficient rice straw hydrolysis using an experimental mixture design. *Bioresource Technol.* 119, 252–261.

7. Kallioinen, M., Huuhilo, T., Reinikainen, S.P., Nuortila--Jokinen, J. & Mänttäri, M. (2006). Examination of membrane performance with multivariate methods: A case study within a pulp and paper mill filtration application. *Chemometr Intell. Lab.* 84(1), 98–105.

8. Lahtinen, K. & Kuuipalo, J. (2008). Statistical prediction model for water vapour barrier of extrusion-coated paper. *TAPPI J*. 9(2008), 8–15.

9. Mercangoz, M. & Doyle, F.J. (2006). Model-based control in the pulp and paper industry. *Control Systems*, Ieee. 26(4), 30–39. DOI: 10.1109/MCS.2006.1657874.

10. Broderick, G., Paris, J., Valade, J.L. & Wood, J. (1995). Applying latent vector analysis to pulp characterization. *PAP Puu-Pup Tim*. 77(6/7), 410–418.

11. Broderick, G., Paris, J., Valade, J.L. & Wood, J. (1996). Linking the fiber characteristics and handsheet properties of a high-yield pulp. *TAPPI J.* 79(1), 161–169.

12. Grage, H. (2004). A statistical analysis of data from the production line at the Munksund paper mill. Technical report, Lund Institute of Technology, Sweden.

13. Nordstrom, F., Lindstrom, T. & Holst, J. (2005). Statistical models for on-line monitoring quality properties. Technical report, Lund Institute of Technology.

14. Ortiz-Cordova, M.H.A., Orccotoma, J.B.J. & Begin, B.P.J. (2006). MATHEMATICAL MODELS-Analysis of paper strength variability in an integrated newsprint mill. *Pulp Pap--Canada*. 107(10), 37–43.

15. Wold, S. (1995). PLS for multivariate linear modeling. Chemometric methods in molecular design 2, 195–218.

16. Van der Voet, H. (1994). Comparing the Predictive Accuracy of Models Using a Simple Randomization Test. *Chemometr Intell. Lab.* 25, 313–323.

17. Jones, G.L. (1993). Modeling a corrugating-medium paper machine for improved edgewise compressive strength, *TAPPI J.* 76(7), 122–129.