INFLUENCING PROCESS VARIABLES AND PREDICTIVE MODELS FOR OPACITY USING REAL DATA OF MWPI

MOHAMMAD HADI MORADIAN, HOSSEIN RESALATI,[*] ANTONIO DOURADO[**] and AHMADREZA ZAHEDI TABARESTANI[***]

Behbahan Khatam Alanbia University of Technology, Behbahan 6361647189, Iran
[*]*Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resources University, PO Box 737, Sari, Mazandaran, Iran*
[**]*Department of Informatics Engineering, Coimbra University, Portugal*
[***]*Mazandaran Wood and Paper Industries, Sari 48191-74555, Iran*
✉ *Corresponding author: Mohammad Hadi Moradian, moradian@bkaut.ac.ir*

Received March 22, 2013

The impact of a number of variables involved in pulp processing on the opacity fluctuation of newsprint produced by Mazandaran Wood and Paper Industries (MWPI) from hardwood chemi-mechanical pulp was studied. Using real data from MWPI paper plant, datasets were prepared and the variables that had the greatest influence on paper opacity were found using correlation and mutual information. These included stock pressure in the third group cleaners, the amount of fibres retained on 48 mesh screen, rush to drug ratio, output of second fan pump, and head box slice opening. Then, appropriate neural network predictive models were developed and tested with a suitable dataset to better control the opacity of newsprint produced at MWPI. The models were successfully validated using new real data from the mill, demonstrating the generalization capacity of the neural network models.

Keywords: data-mining, mutual information, neural networks, newsprint, predictive models, opacity

INTRODUCTION

The positive and negative effects of some variables on opacity are known. Opacity increases with increasing grammage, filler content, dyestuff or pigment addition, while wet pressing and calendaring decreases the opacity.[1] However, there are many other factors that directly or indirectly affect the opacity, which are not easy to recognize. In addition, some variables are well controlled in any pulp and paper mill and some interact with other variables, which makes the problem even more complex. Thus, some mathematical tools are very useful to recognize and furthermore prioritize the most important variables affecting the opacity.

Many paper mills have advanced digital control and online measurement systems producing huge amounts of data stored in the mills' database systems. These databases are valuable sources with a great potential for improving process knowledge, provided that they are exploited and explored in a systematic way. The concept of statistical data-mining is an overall term for using various, mainly multivariate, statistical methods and techniques for exploratory data analysis, which are developed to handle large data sets with many and often highly correlated variables.[2] Some important multivariate data-mining methods used in pulp and paper research are: principal component analysis (PCA), factor analysis (FA), partial least squares (PLS) regression, multiple linear regression (MLR) and mutual information estimations for input selection.[3-6]

A neural network is a powerful data modelling tool and several authors have addressed its general applicability to paper industry problems.[7-17] Such models have some advantages for this type of work: there is no need to assume a starting model form; the process is able to handle interacting, nonlinear models without special considerations; models can be constructed to predict multiple outputs from a single set of inputs; the models can be inverted to enable the prediction of process inputs corresponding to a desired set of outputs.[18]

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 175000 tons of different types of paper per year. The newsprint production line involving the CMP pulping process suffers from some fluctuations in paper quality. In order to recognize the most influencing variables on paper quality and generate predictive models to enhance quality control, on-line and off-line data were used in this study.

CMP tower, stock preparation and some paper machine variables were considered, including 72 process variables and nearly 7000 observations. Several data sets for different stages were generated considering the residence time needed in each stage. This study was made using Matlab computational environment.

EXPERIMENTAL

Data set preparation

Paper production lines consist of many stages. Each stage of production should receive suitable and qualified materials as input, perform definite operation at regular time, and deliver the output to the next stage. The stages follow one another to meet the final paper qualities. On the MWPI newsprint production line, it takes about 40 hours for wood chips to convert to final paper (Figure 1).

The pulp flow diagram and the descriptive statistics of all variables considered in this study are shown in Figure 2 and Table 1, respectively. Several data sets with the corresponding time to process variables were prepared. However, the methodology is intended to be applicable to other variables and other types of paper. The data sets are as follows:

1) Pulp and handsheet test variables from CMP tower with 300 observations corresponding to opacity test records;
2) Stock preparation variables corresponding to opacity test records with 850 observations;
3) Wet-end variables corresponding to opacity test records with 750 observations;

Pulp and handsheet test, stock preparation, and wet-end variables corresponding to opacity test records with 85 observations.

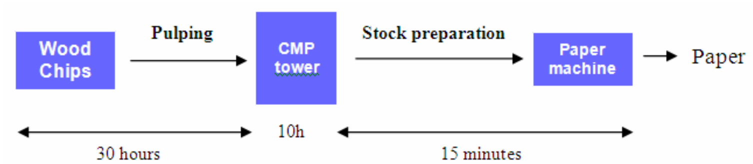
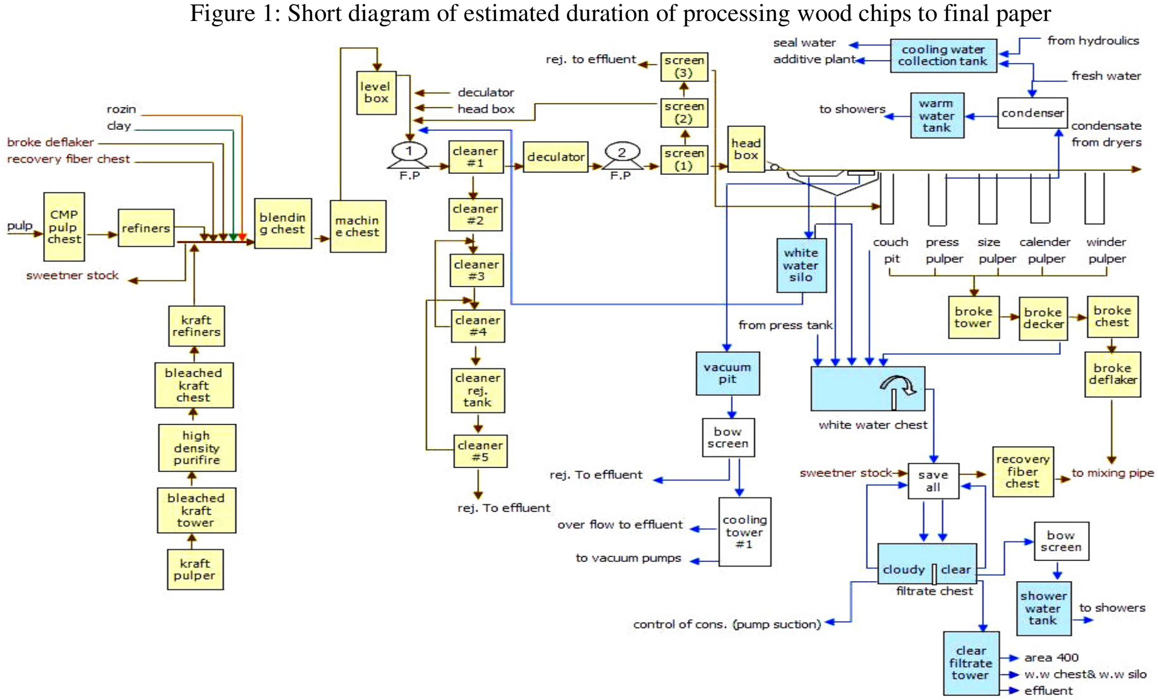


Figure 1: Short diagram of estimated duration of processing wood chips to final paper



Wood

Figure 2: Flow diagram of CMP pulping process to the wet-end part of the paper machine[28]

Table 1
Process variables and pulp and paper properties

Variables	Min	Mean	Max	Std.dev.	Variables	Min	Mean	Max	Std. dev.
pH of pulp in CMP tower	3.7	4.6	6.8	0.29	Out-Fan Pump 2	43	75	97	6.72
Cons. ″	6.2	10.7	15	1.13	pH-Screen	3.8	7.3	10	0.46
C.S.F. ″	310	408	540	23.9	No.1PD-Screen	0	0.09	0.2	0.03
Total Na^+ ″	0.15	0.36	0.85	0.08	RPM-SaveAll	501	557	599	8.85
Shive ″	0.01	0.08	0.9	0.09	W.W.Ch. Valve	12	16.4	19	0.74
Yield ″	80.6	84.7	88.9	1.1	SaveAll-Sa.L.	33	41	110	4.7
+28Mesh ″	2.1	9.5	22.1	3.35	Stock SaveAll	1602	2094	2772	110
+48Mesh ″	21.5	40	50.6	4.48	Stock W.W.Sa.	2455	24014	29980	2457
+100Mesh ″	5.3	18.8	48.9	3.92	Recovery Co.Sa.	1.8	3.48	5.65	0.32
-200 (Fine) ″	12.5	31.8	47.3	4.26	Pre.Ref.Cons.	2.8	4	4.5	0.15
Break.L. handsheet test	1.4	2.4	4.5	0.34	Pre.Ref.CSF	270	368	520	25
Tear ″	131	212	320	30.3	After Ref.Cons	2.8	4	4.7	0.16
Burst ″	35	61	200	12.1	After Ref.CSF	230	319	440	23
Brightness ″	42.5	53.1	60.2	1.35	Level Box Cons.	3.2	3.9	4.3	0.11
Yellowness ″	23.5	28.2	37.3	1.23	Level Box CSF	170	227	370	17.6
Opacity ″	82	90.1	94.2	1.16	Head Box pH	7	7.5	8.8	0.12
Cons.CMP ref.	3.2	4.1	4.9	0.13	Head B. Cons.	0.69	0.99	1.6	0.07
Cir.val.CMP ref.	15	25	40	2.75	Head B. CSF	50	79	190	11.6
Temp.CMP ref.	22	38	48	5.14	W.W. Cons.	0.2	0.42	0.65	0.05
Load refiner	180	390	625	83.8	First Pass Ret.	35	57	72	3.9
Cons.-BKP ref.	3	3.99	5.1	0.37	Long Fi. Cons.	1.8	3.7	4.5	0.26
Load 1-BKP ref.	0	220	440	184	Long Fi. CSF	260	498	630	21.4
Cir.valve BKP ref.	22	28	45	2.66	Head B. Air Pr.	0.58	0.84	0.98	0.05
Temp.-BKP ref.	24	41	49	2.33	Head B.Level	348	696	837	71
CMP ratio	75	81	92	1.77	Slice Open H.	9.99	11.43	13.77	1
BKP ratio	8	19	28	1.88	Speed-Wire	550	724	795	37.5
Broke ratio	0	61	114	23	Rush/Drug	-10.8	7	-2.5	3.87
Clay ratio	3	21	44	5.7	Basis wt. of test paper	46.5	49.3	54	0.41
Dye ratio	0	4.1	8.7	0.83	Moisture ″	7.5	8.6	13	0.34
Retention aid	0	3	7.7	1.27	Calliper ″	66	75	101	3.2
Blend Cons.	1.25	3.55	5.34	0.64	Bulk ″	1.33	1.51	2	0.07
Machine Cons.	3.14	4.05	7.01	0.07	Ash ″	.9	4.1	9.1	1.1
Broke cons.	3.02	3.97	4.8	0.27	Break L.MD ″	4.7	6.8	9	0.75
Stock valve	20	29	33	1.29	Break L.CD ″	1.2	2	4.9	0.13
Stock to F.P. flow	2892	6600	9803	577	Elongate MD ″	1	1.4	3.2	0.14
RPM-Fan Pump1	838	1081	1220	43	Elongate CD ″	2	3	4.6	0.31
Out-Fan Pump1	56	87	98	4	Burst ″	65	102	190	14.1
G2-cleaners	0.33	0.48	0.79	0.03	Tear MD ″	168	219	379	20.3
G3-cleaners	0.2	0.52	0.91	0.16	Tear CD ″	197	304	482	24.3
G4-cleaners	-0.9	-0.74	0.97	0.16	Porosity ″	7	18	65	6.8
T-Decollator	30	42	46	1.93	Rough. Top. ″	3.9	4.4	7.1	0.1
L-Decollator	425	750	912	48	Rough. Bot. ″	4	4.9	7.4	0.15
PD-Decollator	0.48	1.06	1.8	0.10	Brightness ″	46.3	50.8	60.8	1.1
Vac.Decollator	44	97	335	21	Yellowness ″	2.1	7.9	17.3	0.89
RPM-Fan Pump 2	397	846	1043	72	Opacity ″	85.3	91.3	94	1

Determination of influencing variables

Correlation analysis and mutual information are two criteria to measure dependencies between variables. Correlation analysis shows linear relations between process variables and the opacity, while mutual information shows nonlinear relations.[13] The most influencing variables on paper opacity are detected using these criteria from the mentioned data sets. The forward selection method, known as “greedy selection”, for detecting influencing variables was used. Several methods for estimating mutual information were considered, such as a numerical method proposed by Kraskov[13] and two Kernel-based methods for estimating the probability density function (FKDE and AKDE).[20]

Theoretical foundations of mutual information and its estimation

In information theory, mutual information is defined to measure the amount of information shared between two variables; so it shows the level of dependency of one variable to another.[18,19] Let X and Y be two random variables, then mutual information between them is a quantity that measures the knowledge on Y provided by X and vice versa. If X and Y are independent, then X contains no information about Y and vice versa and mutual information between them is zero.

The definition of mutual information originates from the Shannon entropy in the information theory.[20] Mutual information between two random variables X and Y is defined in equation (1) where p means probability density function:

$$I(X;Y)=\int_x\int_y p_{X,Y}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}dxdy \qquad (1)$$

For estimating mutual information we only need to estimate $p_{X,Y}(x,y)$. Most common estimation methods are Histogram and Kernel methods.[20] However, they will face the course of dimensionality when the dimension of space increases. To overcome this problem and reduce computing complexity, a recent estimator based on k-nearest neighbours' statistics is used.[13,19] It estimates mutual information between two random variables of any dimensional space.

Consider a set of N input-output pairs, $z_i=(x_i,y_i)$, $i=1,\ldots,N$, which are assumed to be i.i.d. (independent identically distributed) realizations of a random variable $Z=(X,Y)$ with density $p_{X,Y}(x,y)$ on the spaces spanned by X, Y. Input–output pairs are compared through the infinity norm (2), while any norms can be used for $\|x-x'\|$ and $\|y-y'\|$ (these spaces could be completely different):[13]

$$\|z-z'\|_\infty=\max\{\|x-x'\|,\|y-y'\|\} \qquad (2)$$

Let us consider $z_{k(i)}=(x_{k(i)},y_{k(i)})$ the k*th* nearest neighbour of z_i (with infinity norm) where k is a fixed positive integer ($k\ge 1$). We denote $\varepsilon_i/2=\|z_i-z_{k(i)}\|_\infty$ the distance from z_i to its k*th* neighbour, $\varepsilon_i^x/2=\|x_i-x_{k(i)}\|$ and $\varepsilon_i^y/2=\|y_i-y_{k(i)}\|$ the distances between the same points projected into the X and Y subspaces. Considering Equation (2), it is clear that $\varepsilon_i=\max\{\varepsilon_i^x,\varepsilon_i^y\}$.

Cubic and rectangular estimations are two versions of the estimation method. In the first version, we count the number n_i^x of points x_j whose distance from x_i is strictly less than $\varepsilon_i/2$, and similarly for y instead of x. Notice that $\varepsilon_i/2$ is a random (fluctuating) variable, and therefore n_i^x and n_i^y also fluctuate. The cubic version of estimation is then (3):

$$\hat{I}^{(1)}(X;Y)=\psi(k)-\frac{1}{N}\sum_{i=1}^{N}\left[\psi(n_i^x+1)+\psi(n_i^y+1)\right]+\psi(N) \qquad (3)$$

In the rectangular version, we replace n_i^x and n_i^y by the number of sample points with $\|x_i-x_j\|\le\varepsilon_i^x/2$ and $\|y_i-y_j\|\le\varepsilon_i^y/2$.[13] The second estimate for mutual information is then (4):

$$\hat{I}^{(2)}(X;Y)=\psi(k)-\frac{1}{k}-\frac{1}{N}\sum_{i=1}^{N}\left[\psi(n_i^x)+\psi(n_i^y)\right]+\psi(N) \qquad (4)$$

where ψ is the Digamma function (5):

$$\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}=\frac{d}{dx}\ln\Gamma(x)\,, \qquad \psi(1)\approx-0.5772156 \qquad (5)$$

where:

$$\Gamma(x)=\int_0^\infty u^{x-1}e^{-u}du \qquad (6)$$

Generally, two estimators provide very similar results. In high dimensions where ε_i tends typically to be much larger than the marginal $\varepsilon_i^{x_j}$, the second method is recommended.[13] In this paper, we use the second method to estimate mutual information. The proposed value of k = 6 is suitable for a good estimation.[21]

Algorithms for input variables selection

In this section, we propose the algorithm to find the best subset of input variables from the initial set.[22-24] The objective of the algorithm is to maximize the relevance between inputs and output. This algorithm starts with an empty set of selected input variables. The goal is to select inputs that maximize the mutual information between selected inputs and output. This ideal greedy algorithm can be described by the following procedure:

1) Initialisation: Set $L\leftarrow$ 'initial set of n inputs' $S\leftarrow$ 'empty set' $T\leftarrow$ 'output'.
2) Computation of the mutual information with the output: for each input $l\in L$ compute $I(T;l)$.
3) Choice of the first input: find the input l that maximizes $I(T;l)$; set $L\leftarrow L-\{l\}$, $S\leftarrow\{l\}$.
4) Greedy selection: repeat until desired number of input variables is selected:
   a) Computation of the joint mutual information between variables: for all variables $l\in L$; compute $I(T;l,S)$.
   b) Selection of the next input: choose the input $l\in L$ as the one that maximizes $I(T;l,S)$; set $L\leftarrow L-\{l\}$, $S\leftarrow S\cup\{l\}$.
5) Output the set S containing the selected inputs.

The stopping criterion in step 4 is to select the desired number of inputs, which can be replaced by another criterion such as a considerable decrease in $I(T;l,S)$.

Statistical and neural network modelling

Statistical models were developed using multi linear regression. The stepwise variable selection method was used to develop predictive models. Outliers were eliminated according to regression standardized residuals out of the range of 3 and -3. Also non-colinearity of models was considered having VIF coefficient less than 5.

Independent influencing process variables on paper opacity were measured by correlation and mutual information. Then these selected variables were used to investigate the possibility of generating neural network models with the least possible errors using feed forward back propagation method and tangent sigmoid activation functions. Models were then successfully validated using new real data of the mill.

RESULTS AND DISCUSSION

Influencing variables

Although most of the influencing variables show more linear relations than nonlinear, some of them have high mutual information values that were also nonlinearly dependent. Table 2 presents the most influencing CMP tower pulp variables on newsprint opacity. The average value of +28 mesh for a period of 1.5 years was about 9.5%, while it was 40% for +48 mesh and 31% for -200 mesh (fines). Accordingly, no effect, negative correlation and positive influence on the paper opacity were observed for the values of +28, +48, and -200, respectively. Small particle size fractions (values of +100 mesh plus -200 mesh) are filler-like materials with high specific surface area, they improved the scattering coefficient and minimized the show-through, and thus showed direct relation to opacity. In order to achieve the desired newsprint opacity, the fibre classifications should be controlled through efficient refining.

The values measured for +48 mesh, burst and tear indices are the most effective variables related nonlinearly to newsprint opacity, according to Kraskov.[13]

Table 2
Most influencing CMP tower pulp variables on newsprint opacity

Process variables	Correlation values	Mutual information values	Mutual information values, Forward selection method		
		Kraskov	Kraskov	FKDE	AKDE
+48Mesh	-0.54	0.17	0.17		
Opacity	0.46		0.29		
Burst	-0.46	0.1			
Tear	-0.45	0.09			
Breaking L.	-0.36		0.35		
+28 & +48Mesh	-0.36				0.59
-200Mesh	0.36				
+100Mesh	0.1			0.14	0.29

Table 3
Most influencing stock preparation variables on newsprint opacity

Process variables	Correlation values	Mutual information values	Mutual information values, Forward selection method		
		Kraskov	Kraskov	FKDE	AKDE
G3 cleaners	0.59	0.3	0.30		
Output, F.P.2	0.50				
Retention aid	0.43		0.42		
Stock SaveAll	0.42				
RPM, F.P.2	0.40				
Temp. CMP	-0.29			0.34	0.65
Blend cons.	-0.24	0.25			
PD-decollator	0.18		0.49		
Clay	-0.07			0.46	0.88
G4 cleaners	-0.06	0.23			
Stock to F.P.flow	-0.03			0.23	0.45

On the basis of forward selection and Kraskov method, +48 mesh is the first and most related variable with 0.17 mutual information value, while the sum of +48 mesh and opacity increases the Kraskov mutual information value to 0.29, and it reaches 0.35 if the effect of breaking length variable is added to those two variables. According to the forward selection and FKDE or

AKDE methods, the first nonlinearly related variable to newsprint opacity was +100 mesh and the second variable selected by the AKDE method was +48 mesh. The results of this study indicated that Kraskov's method showed more acceptable results than either FKDE or AKDE methods.

Table 4
Most influencing wet-end variables on newsprint opacity

Process variables	Correlation values	Mutual information values	Mutual information values, Forward selection method		
		Kraskov	Kraskov	FKDE	AKDE
Rush/drug	0.52	0.28	0.28		
Slice opening	0.49	0.22			
Head B.L.	0.29		0.39		
H. Air pres.	0.28	0.21	0.42	0.11	0.22
Speed wire	0.25			0.19	0.36

Table 5
Most influencing newsprint paper test variables on newsprint opacity

Process variables	Correlation values	Mutual information values	Mutual information values, Forward selection method		
		Kraskov	Kraskov	FKDE	AKDE
Burst	-0.69	0.34	0.34		
Break L. MD	-0.60	0.25			
Calliper	0.598	0.24			
Bulk	0.595				
Elongation, MD	-0.585				
Porosity	-0.57				
Brightness	0.46		0.54		
Tear, CD	-0.38				
Elongation, CD	-0.38				
Ash	0.10		0.59	0.10	0.19

Table 3 shows some highly correlated and controllable variables to newsprint opacity. The output of the third group cleaners, the output of the second fan pump, the amounts of retention aids, the stock save-all flow, and the speed of the second fan pump were the variables most correlated to opacity. Other variables presented in Table 3 were nonlinearly related to opacity. As wet-end process variables, rush/drug ratio and slice opening were strongly correlated to opacity (Table 4).

The relations between opacity and other newsprint properties are shown in Table 5. The highest correlation to opacity was observed by the burst index, while breaking length, calliper, and bulk were also highly related to opacity. These paper properties can be considered as factors affecting the opacity of newsprint produced at MWPI mill, thus the opacity can be controlled by adjusting the corresponding process variables.

Statistical models

Different statistical models were generated from the above-mentioned data sets to predict opacity. Correlation and mutual information show positive or negative relations between opacity and process variables, but it cannot be defined how much change in the process variable values is needed in order to achieve the desired improvement in newsprint opacity. These regression models are predictive tools to measure opacity when any changes occur in the independent variables. In other words, by putting the desired value of opacity in the following equations, the suitable value of different process variables can be calculated. Models 1, 2 and 3 include CMP tower, stock preparation, and wet-end variables, respectively. Models 4 and 5 include combinations of all process variables and model 6 contains all process variables plus two physical properties of final paper, namely calliper and ash. To predict the opacity from some special variables, one can select any of the suggested and tested models.

Model (1):
Calculated R-squared = 58%   Tested R-squared = 51%
Opacity = 74.936 - 0.060 (+48 Mesh) + 0.266 (Opacity) - 0.021 (Burst) - 3.554 (Total Na^+) - 0.007 (Tower CSF)

Model (2):
Calculated R-squared = 55%   Tested R-squared = 53%
Opacity = 78.892 + 2.706 (G3 cleaners) + 0.227 (Dye ratio) + 0.380 (pH Screen) + 0.197 (Retention aid) - 2.603 (PD decollator) + 0.037 (Out F.P.2) + 0.006 (Broke ratio) - 0.044 (Cir. Valve BKP ref.) + 0.014 (RPM Saveall)

Model (3):
Calculated R-squared = 55%   Tested R-squared = 45%
Opacity = 84.837+ 0.099 (Rush/drug) + 0.321 (Slice opening) + 0.004 (Speed wire)

Model (4):
Calculated R-squared = 71%   Tested R-squared = 67%
Opacity = 80.658 + 2.125 (G3 cleaners) + 0.239 (Opacity) - 0.735 (Break L.) - 1.920 (Total Na^+) - 0.486 (Blend cons.) - 0.067 (+48 Mesh) - 0.067 (Out F.P.1)+ 0.170 (Dye ratio)

Model (5):
Calculated R-squared = 70%   Tested R-squared = 68%
Opacity = 95.615 - 0.041 (+48 Mesh) + 0.117 (Opacity) – 0.008 (Tear) + 0.074 (Rush/drug) + 0.133 (Retention aid) - 1.981 (Total Na^+) - 0.153 (Brightness) - 0.102 (Yellowness)

Model (6):
Calculated R-squared = 80%   Tested R-squared = 78%
Opacity = 71.712 + 0.173 (Calliper) + 0.142 (Opacity) + 0.169 (Ash) + 0.128 (Retention aid) - 0.108 (Brightness) + 0.038 (Rush/drug) - 1.58 (Total Na^+) - 0.004 (Tear)

Neural network models

By using the variables listed in Tables 2, 3, 4 and 5 as inputs, the best neural networks with the least error and highest R-squared were developed to predict opacity according to the models 7, 8, 9, and 10 (Table 6). These models predict opacity from the most influencing process variables that were highly correlated or nonlinearly related to the opacity of newsprint produced at MWPI mill. The first model was developed from selected variables of the CMP tower pulp test (Table 2), with two hidden layers having 6 and 20 neurons, respectively (Figure 3).

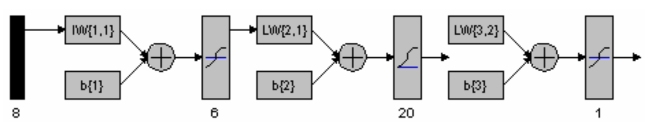


Figure 3: Neural network model for CMP tower pulp

Table 6
Specifications of neural network models

Neural network models	Number of inputs	Number of neurons		R-squared value		
		First hidden layer	Second hidden layer	Training set	Test set	Validation set
Model 7	8	6	20	63	44	55
Model 8	11	7	18	79	62	60
Model 9	5	5	15	70	58	57
Model 10	10	5	20	80	73	75

CONCLUSION

In this study, the process variables influencing newsprint opacity were defined for paper produced at Mazandaran Wood and Paper Industries (MWPI). Some of these variables are easier to control such as most of the stock preparation and wet-end variables, including stock pressure in the third group cleaners, rush to drug ratio, output of second fan pump, and head box slice opening. On the other hand, CMP tower influencing variables, such as +48 mesh screen fraction, can be controlled through the uniformity of the chip quality and optimization of chemical treatment and refining in the pulp mill.

It may be found that how much change is needed for these influencing variables to achieve the desired level of opacity by using a suitable neural network model developed in this study. Neural network models are better tools than statistical ones since they show higher R-squared values and contain variables both linearly and nonlinearly related to newsprint opacity. However, statistical models are also very useful tools, because they can easily predict any of the process variables by putting the desirable amount of opacity in regression equations. It is expected that these models would increase the production yield and reduce the consumption of materials and energy at MWPI.

ACKNOWLEDGMENTS: The authors would like to acknowledge the support provided by Mazandaran Wood and Paper Industries and Coimbra University. Also special appreciation is extended to Mr. Mohammadmahdi Rezaei Yousefi and Dr. Dinesh Kumar for their partial assistance and valuable help in programming.

REFERENCES

[1] W. E. Scott, J. C. Abbott and S. Trosset, in "Properties of Paper: An Introduction", 2nd ed., Tappi Press, 1995, pp. 106-108.
[2] M. Mercangoz and F. J. Doyle, *IEEE Contr. Syst. Mag.*, **26**, 30 (2006).
[3] K. Lahtinen and J. Kuuipalo, *Tappi J.*, **9**, 8 (2008).
[4] M. Kallioinen, T. Huuhilo, S. P. Reinikainen, J. Nuortila-Jokinen and M. Mänttäri, *Chemometr. Intell. Lab.*, **84**, 98 (2006).
[5] H. Zhang, A. F. Nitzman and T. Royappa, *Tappi J.*, **3**, 3 (2004).
[6] M. H. Moradian and H. Resalati, *Pol. J. Chem. Technol.*, **16**, 110 (2014).
[7] J. Lampinen, in *Procs. IEEE World Congress on Computational Intelligence*, *IEEE International Conference*, 27 June - 2 July, Orlando, FL, 1994, Vol. 6, pp. 3812-3815.
[8] R. Lanouette, J. Thibault and J. L. Valade, *Comput. Chem. Eng.*, **23**, 1167 (1999).
[9] H. C. Aguiar and R. M. Filho, *Comput. Chem. Eng.*, **22**, S981 (1998).
[10] T. A. Runkler, E. Gerstorfer, M. Schlang, E. Junnemann and J. Hollatz, *Control. Eng. Pract.*, **11**, 1299 (2003).
[11] C. A. Schweiger and J. B. Rudd, *Tappi J.*, **77**, 201 (1994).
[12] G. L. Jones, *Tappi J.*, **76**, 122 (1993).
[13] A. Kraskov, H. Stögbauer and P. Grassberger, *Phys. Rev. E*, **69**, 6 (2004).
[14] J. Mizuki, *Japan Tappi J.*, **63**, 930 (2009).
[15] R. Vincent, M. Rueff and C. Voillot, *Tappi J.*, **9**, 15 (2010).
[16] Z. Zainuddin, W. R. Wan Daud, O. Pauline and A. Shafie, *Bioresour. Technol.*, **102**, 10978 (2011).
[17] P. J. Garcia Nieto, J. Martinez Torres, M. Araujo Fernandez and C. Ordonez Galan, *Appl. Math. Model.*, **36**, 6137 (2012).
[18] T. M. Cover and J. A. Thomas, in "Elements of Information Theory", John Wiley & Sons, 2012.
[19] J. Yongnan, N. Reyhani, J. Hao and A. Lendasse, in *Procs. ESANN European Symposium on Artificial Neural Networks*, Bruges, Belgium, April 27-29, 2005, pp. 503-504.
[20] C. E. Shannon, *AT&T Tech. J.*, **27**, 379 (1948).
[21] F. Rossi, A. Lendasse, D. François, V. Wertz and M. Verleysen, *Chemometr. Intell. Lab.*, **80**, 215 (2006).
[22] N. Kwak and C. H. Choi, *IEEE T. Neural Network.*, **13**, 143 (2002).
[23] R. Battiti, *IEEE T. Neural Network.*, **5**, 537 (1994).
[24] A. Al-Ani and M. Deriche, in *Procs. Signal Processing and its Applications, Sixth International Symposium*, Kuala Lumpur, Malaysia, August 13-16, 2001, pp. 477-480.